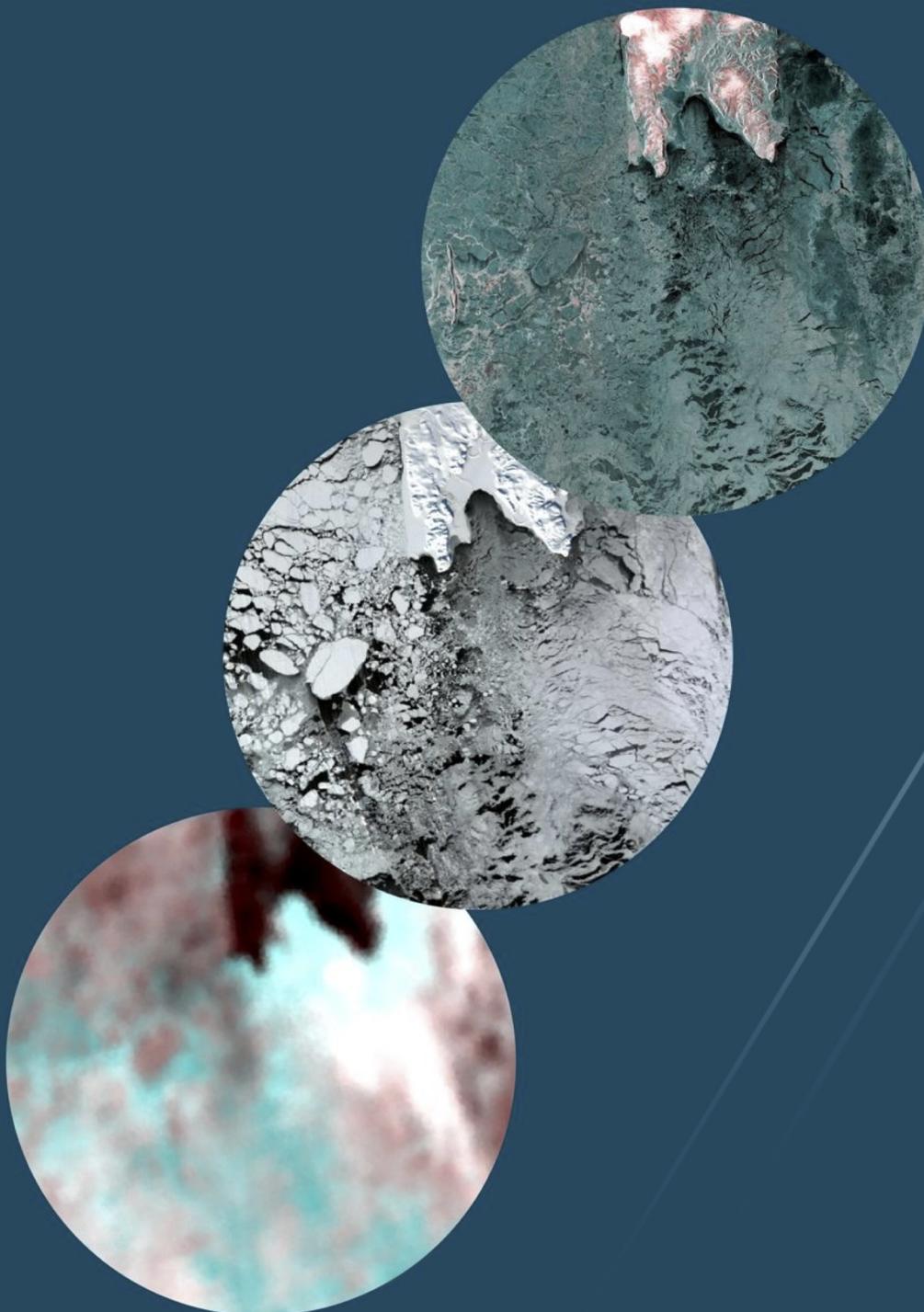UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Physics and Technology

Multimodal Integrated Remote Sensing
for Arctic Sea Ice Monitoring

Khachatrian Eduard

A dissertation for the degree of Doctor of Philosophy – March 2023

A dissertation for the degree of Doctor of Philosophy

Multimodal Integrated Remote Sensing for Arctic Sea Ice Monitoring

Eduard Khachatrian

Supervised by:

Andrea Marinoni

Wolfgang Dierking

Torbjørn Eltoft

Anthony Doulgeris

Abstract

Remote sensing data acquired from various sensors have been used for decades to monitor sea ice conditions over polar regions. Sea ice plays an essential role in the polar environment and climate. Furthermore, sea ice affects anthropogenic activities, including shipping and navigation, the oil and gas industry, fisheries, tourism, and the lifestyle of the indigenous population of the Arctic. With the continuous decline of sea ice in the Arctic the presence of human-based activities will grow. Therefore, reliable information about sea ice conditions is of primary interest to protect the Arctic and to ensure safe and effective commercial activities and polar navigation.

Currently, sea ice services produce operational ice charts manually using the knowledge of sea ice experts. However, with an increasing number of various data sources that provide different information regarding sea ice, it is important to develop automatic methods for sea ice characterization. Robust and automatic ice charting can not be achieved using only one satellite mission. It is fundamental to combine information from various remote sensors with different characteristics for more reliable sea ice monitoring and characterization. However, how do we know that all the information is actually relevant? It may be redundant, corrupted, or unnecessary for the given task, hence decreasing the performance of the algorithms from the required processing time and accuracy point of view. Therefore, it is crucial to select an optimal set of features that provides the relevant information content to enhance the efficiency and accuracy of the image interpretation and retrieval of geophysical parameters. The work in this dissertation specifically focuses on the development of such a method.

In this thesis, we employ a fully automatic, flexible, accurate, efficient, and interpretable information selection method that is based on the graph Laplacians. The proposed approach assesses relevant information on a global and local level using two metrics simultaneously and selects relevant features for different regions of an image according to their physical characteristics and observation conditions. Moreover, it is linked in a common scheme with a classification algorithm that helps to properly evaluate the performance of the information selection and provides sea ice classification maps as an output. Accordingly, in

recent studies, we investigate and evaluate the robustness and effectiveness of the proposed method for sea ice classification by testing several data combinations with various sea ice conditions. Experiments illustrate the flexibility and efficiency of the proposed scheme and clearly indicate an advantage of combining various sensors. Moreover, the results demonstrate the potential for operational sea ice monitoring that should be further thoroughly examined in future studies.

Acknowledgements

Over the last several years many people have significantly contributed to my work and dissertation. I would like to express my sincere gratitude to them in a few short but meaningful paragraphs.

I would like to start with my co-supervisors, Anthony Doulgeris and Torbjørn Eltoft. Anthony, thanks for reminding me that the main object of my research is actually sea ice. Torbjørn, thanks for the great opportunity to work on such fascinating topics and be part of CIRFA.

Many thanks to my main supervisor, Andrea Marinoni, who always filled my life with an enormous amount of joy and motivation, not only by setting unachievable deadlines but also by providing new ideas and knowledge that were crucial for my PhD project.

Most importantly, I would like to express sincere appreciation to my main scientific advisors, Wolfgang Dierking, and Saloua Chlaily, without whom this thesis would not be ever finished. Both, Wolfgang and Saloua significantly changed my way of thinking and working and eventually made me a better scientist and hopefully a better person.

Furthermore, I would like to express my gratitude to a great colleague Malin Johansson who, along with my supervisors Wolfgang Dierking and Andrea Marinoni, was a part of the evaluation committee. I am extremely grateful that I was considered worthy to become a new CIRFA member.

Finally, I would like to thank my beloved wife Varvara, my sweet family Elena, Larisa, and Matevos, and my loyal friends Julia and Nikita. Their constant love, help, and support through these years were tangible even at a distance of several thousand kilometers.

This thesis is dedicated to my beloved Grandfather ♡

Contents

List of Figures

List of Tables

List of Abbreviations

AA Average Accuracy

AARI Arctic and Antarctic Research Institute

AMSR Advanced Microwave Scanning Radiometer

ASM Angular Second Moment

CIRFA Centre for Integrated Remote Sensing and Forecasting for Arctic Operations

CIS Canadian Ice Service

DBFE Decision Boundary Feature Extraction

DP Dual Polarization

EM Electromagnetic Spectrum

ET Execution Time

FIS Fisher Score Selection

FYI First-Year Ice

FS Forward Feature Selection

GA Genetic Algorithm

GK Gaussian Kernel

GLCM Gray-Level Co-Occurrence Matrix

LiDAR Light Detection and Ranging

MI Mutual Information

MIZ Marginal Ice Zone

mRMR Minimum Redundancy and Maximum Relevance

NIR Near-Infrared

NIS Norwegian Ice Service

NISAR NASA-ISRO SAR Mission

OA Overall Accuracy

OBB Orthogonal Branch and Bound

PCA Principal Component Analysis

PFA Principal Feature Analysis

PMR Passive Microwave Radiometer

QP Quad-Polarization

RF Random Forest

RFE Recursive-Feature Elimination

RISAT-1 Radar Imaging Satellite-1

ROI Region of Interest

RSM RADARSAT Constellation Mission

SAR Synthetic Aperture Radar

SIC Sea Ice Concentration

SLIC Simple Linear Iterative Clustering

SNAP Sentinel Application Platform

SNMF Sparse Non-Negative Matrix Factorization

Spa-BS Sparse Representation Based Band Selection

SP Single Polarization

SVM Support Vector Machine

SWIR Short-Wave Infrared

TEMD Thresholded Earth's Mover Distance

TIR Thermal Infrared

UiT The Arctic University of Norway

WMO World Meteorological Organization

/1

Introduction

Over the last several decades, satellite imagery has become an important source that provides useful information about the Earth's surface. Nowadays, a region of interest (ROI) can be monitored by various sensors characterized by different modalities, i.e., acquisition techniques, using different frequency bands or wavelength channels, wide or narrow coverage, and coarse or high spatial resolution, and different revisit periods [1].

Information provided by multiple sensors reflects different aspects of the ROI, e.g., hyperspectral images reveal the material content, and light detection and ranging (LiDAR) technology provides highly accurate measurements of the vertical height of structures. Nevertheless, here we are more focused on other sensors, such as optical, synthetic aperture radar (SAR), and passive microwave radiometers (PMR). These sensors together grasp different aspects of the observed area, optical sensors provide information about the chemical and atomic composition of the surface, synthetic aperture radar responds to dielectric properties, geometry, roughness, as well as to an object's surface structure and - if the radiation penetrates into the object - to its volume structure, and passive microwave radiometers measure the emitted radiance and temperature of different surfaces. Moreover, each of these sensors has different sensing characteristics, such as units of measurement, spatial resolution, image, and geographical coordinate systems. Consequently, an accurate characterization of the Earth's surface can be achieved by combining relevant information coming from various modalities in order to obtain useful insight into different aspects of the underlying surface [2].

Although improving the understanding and characterization of the ROI sounds very tempting, the combination of the multimodal data sets raises several challenges [1, 2]. One of the challenges lies in the fact that we are dealing with heterogeneous data sets obtained from various remote sensing systems, hence combining different data domains. Furthermore, by expanding the size of a data set, we are simultaneously increasing the complexity of the records to be analyzed, especially when it is multimodal. Moreover, although different data sources provide complementary information, they can potentially also include redundant, corrupted, or simply unnecessary information for a given task. Indeed, it has been shown that increasing the number

of modalities without properly addressing an investigation of the significance and reliability of the data may even deteriorate the performance [3, 4]. This may, therefore, represent a strong limiting factor to the use of multimodal remote sensing data analysis in practical scenarios, as well as to its actual impact in operational frameworks within private and public sectors [1].

The main object of this dissertation is sea ice, which has become a hot topic in Earth observation disciplines, as it plays a vital role in the polar ecosystem, and is one of the main indicators of global climate change. Moreover, it strongly affects several anthropogenic activities in the Arctic region, such as the oil and gas industry, fisheries, shipping, tourism, and the lifestyle and welfare of the indigenous population.

Currently, there are plenty of satellite missions that observe sea ice in a wide range of the electromagnetic spectrum. A large amount of publicly available data has created a need to develop new accurate, flexible, and efficient methods capable of extracting relevant information from heterogeneous data sets [5]. Moreover, these methods in combination with data fusion can be employed to improve the performance of various algorithms in terms of accuracy and computing time. The work presented in this thesis focuses on the development and validation of the information selection method.

The information retrieval can be obtained by dimensionality reduction algorithms. Generally, they are divided into two main categories, which are feature extraction and feature selection. Each of these information retrieval techniques has its advantages and drawbacks. Feature extraction reduces the dimensionality by finding the optimal transformation matrix and projecting the original high-dimensional data onto a lower-dimensional subspace [6]. Unlike feature extraction, feature selection reduces dimensionality by selecting the most informative subset of existing records [7]. In the case of sea ice, it is crucial to preserve the original physical meaning and characteristics of the data, therefore, our choice fell to the feature selection algorithms. Among a huge variety of feature selection methods, graph-based algorithms have some unique advantages, e.g. (a) the ability to capture complex relationships/dependencies among different features, (b) the capacity to efficiently handle high-dimensional data, (c) flexibility in defining similarity measures that allow customization of the feature selection process to the specific needs, (d) capability to provide interpretable results that can help to understand the relationships among features [8]. In this work, we are employing high-dimensional remote sensing data obtained from various sensors and satellite platforms with complex relationships between features, which motivates us to choose the graph-based clustering feature selection.

The proposed method has a few crucial novelties which distinguish it from existing techniques. Among all the advantages, we would like to stress a few major contributions of this work and the proposed approach in particular: (a) it employs two similarity measures that preserve global and local particularities of the original data set; (b) it is performed on a superpixel level, i.e. homogeneous regions that share similar pixel information, therefore, it selects the most relevant features for the considered classes individually for each area of the scene; (c) it can be applied to various data combinations with different characteristics; (d) it is unsupervised, thus not requiring any prior information regarding the data.

1.1 Motivation

The work presented in this thesis is funded by the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) which is a centre for research-based innovation hosted by the Department of Physics and Technology at The Arctic University of Norway (UIT). The

main focus of CIRFA in general, and this thesis in particular is the integration and combining of remote sensing data from various sensors for improving sea ice observations.

1.1.1   Scientific Goals

The main scientific goals of this project which is a part of the CIRFA work package allocated to "Sea Ice Remote Sensing" are as follows:

Goal 1: explore approaches to properly integrate, combine and utilize various types of data with different characteristics in order to improve the knowledge regarding the observed area and sea ice characterization in particular.

Goal 2: develop new strategies to extract relevant unique information from heterogeneous multimodal data combinations in order to improve the performance of the classification algorithms in terms of accuracy and time consumption.

Goal 3: investigate the potential capabilities of the developed algorithms for operational sea ice monitoring.

1.1.2   Research Outcomes

Below is a list of main scientific publications that are specifically exploring the dissertation goals mentioned above. Furthermore, we additionally list other published journal and conference papers that were not included as a part of the dissertation, however, contain important information for this project.

1.1.2.1   Main Scientific Publications

1. E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni. "**A Multimodal Feature Selection Method for Remote Sensing Data Analysis Based on Double Graph Laplacian Diagonalization**", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Volume: **14**, 11546 – 11566, November 2021.

2. E. Khachatrian, S. Chlaily, T. Eltoft, W. Dierking, F. Dinessen, and A. Marinoni. "**Automatic Selection of Relevant Attributes for Multi-Sensor Remote Sensing Analysis: A Case Study on Sea Ice Classification**", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Volume: **14**, 9025 – 9037, July 2021.

3. E. Khachatrian, W. Dierking, S. Chlaily, T. Eltoft, F. Dinessen, N, Hughes, and A. Marinoni. "**SAR and Passive Microwave Fusion Scheme: a Test Case on Sentinel-1/AMSR-2 Data Sets for Sea Ice Classification**", *AGU Geophysical Research Letters*, Volume: **50(4)**, February 2023.

Paper I introduces the proposed method and validates the performance of the algorithm using various multimodal remote sensing data combinations for the purposes of urban and agricultural classification (*Goal 2*). In this study, we validated our method on scenes that have a proper ground

truth (not readily available for Arctic sea ice) which makes it easier to accurately evaluate the performance of the algorithm. We conducted various experiments with the algorithm in order to properly tune it and motivate the choice of choosing several important parameters, validate the robustness of the proposed approach, and compared it with existing dimensionality reduction methods.

Paper II explores the potential of applying the information selection proposed in *Paper I* for different multimodal remote sensing data sets in order to improve the sea ice characterization and classification algorithms performance (*Goal 1, Goal 2*). Moreover, we investigated features selected for various sea ice and water classes. Additionally, we explored the advantages of applying multisensor/multifrequency data combinations by comparing to single-sensor scenarios.

Paper III continues *Paper II* with further investigation on the potential of applying multimodal remote sensing data sets for the improvements of sea ice classification. In comparison to previous studies, here we apply the most commonly used sensors for sea ice observation and charting, which are SAR and PMR, in order to test the proposed method for operational sea ice monitoring (*Goal 1, Goal 2, Goal 3*).

1.1.2.2   Other Scientific Publications

1. E. Khachatrian and N. Sandalyuk. "**On the Exploitation of Multimodal Remote Sensing Data Combination for Mesoscale/Submesoscale Eddy Detection in the Marginal Ice Zone**", *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022, Art no. 3513805, doi: 10.1109/LGRS.2022.3215202.

2. E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni. "**Unsupervised Information Selection In Multimodal Sea Ice Remote Sensing**", PROCEEDINGS OF THE 25th INTERNATIONAL SYMPOSIUM ON ICE Trondheim, Norway, 23rd – 25th November 2020. The International Association for Hydro-Environment Engineering and Research (IAHR) 2020 ISBN 978-82-7598-120-0. s. 208-217.

3. E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni. "**Selecting Principal Attributes in Multimodal Remote Sensing for Sea Ice Characterization**", EUSAR 2021 : 13th European Conference on Synthetic Aperture Radar 29 March – 1 April 2021, VDE Verlag GmbH 2021 ISBN 978-3-8007-5457-1. s. 531-536.

4. E. Khachatrian, S. Chlaily, T. Eltoft, P. Gamba, and A. Marinoni. "**Unsupervised Band Selection for Hyperspectral Data Sets by Double Graph Laplacian Diagonalization**", International Geoscience and Remote Sensing Symposium (IGARSS), 2021.

1.2   Notation

For notational convenience, random scalars are denoted by lowercase letters, e.g., z. Random vectors are designated by bold lowercase letters, e.g., $\mathbf{z}$. Bold upper case letters refer to matrices, e.g., $\mathbf{A}$. $|\mathbf{A}|$ denote the determinant of the matrix $\mathbf{A}$. $\mathrm{diag}\{d_1, \ldots, d_N\}$ refers to a diagonal matrix whose diagonal elements are $d_1, \ldots, d_N$ starting from the upper left.

In this thesis, we are using two similar terms, namely "feature" and "attribute". In general, they

both are used to describe the properties or characteristics of an object or data set. Within this dissertation, we are mostly using the term "feature", e.g. feature extraction or feature selection, which is a commonly used term in fields such as classification methodologies, pattern recognition, and texture analysis. Nevertheless, in Chapter 4 dedicated to the proposed method, we are using the notation "attribute" in order to not confuse it with textural feature extraction which we are using in the same chapter. The term "attribute" is from information theory which in our case refers to directly measured quantities such as, e.g., radar intensities at different frequencies and polarizations, optical reflectance, and parameters such as textural or polarimetric features.

1.3   Thesis Outline

The rest of this thesis is organized as follows. Chapter 2 introduces the main object of this research - sea ice. Chapter 3 covers the basics of dimensionality reduction as an important part of pattern recognition. Chapter 4 gives an overview of the proposed method and its novelty. Chapter 5 demonstrates the basic principles of optical, SAR, and PMR remote sensing, while Chapter 6 displays the challenges and importance of multimodal remote sensing in particular for sea ice monitoring. The papers overview as well as the papers themselves are presented in Chapters 7-10. Chapter 11 shows the main conclusions along with future research plans and suggestions.

/ 2

Sea Ice

In the last decades, sea ice research has become a crucial topic in Earth observation disciplines. This is due to the fact that sea ice loss is one of the most visible signs of a current dramatic change in the climate. It is especially crucial in the Arctic region where sea ice extent and volume are declining rapidly [9]. Scientists from all over the world are trying to improve the understanding and knowledge of the potential sea ice decline effect on Earth's climate, ecosystems, and anthropogenic activities in the Arctic. Figure 2.1 displays the Arctic and Antarctic sea ice extent for several years including the record minimum for both poles along with the 1981 to 2010 median. It is evident that the Arctic suffers a dramatic sea ice decline, while the Antarctic is having less significant fluctuations in sea ice extent. Nevertheless, both Arctic and Antarctic are extremely sensitive to climate change and especially to anthropogenic influence in comparison to the rest of the world.

2.1 Sea Ice Importance

Sea ice plays an essential role in the global climate system, being an important indicator of any climate-related fluctuations in the atmosphere and the ocean. There are several characteristics of sea ice that affect components of the climate system. First and foremost, sea ice as a surface layer completely prevents or significantly reduces the exchange of heat, momentum, and water vapor between the atmosphere and the ocean [11]. Secondly, the reflection and absorption of incoming short-wave radiation is controlled by surface albedo, which changes dramatically with sea ice melting [12]. Naturally, sea ice has a significantly higher albedo compared to the surrounding ocean. For comparison, the albedo of the ocean is approximately 0.06, while sea ice albedo varies between 0.5 to 0.9 depending on the age, i.e., thickness and snow cover on top of it [13, 14]. Therefore, sea ice absorbs less solar energy and keeps the underlying surface cooler. Moreover, sea ice plays an important role in brine and freshwater distribution in the ocean. Sea ice melting and formation affect ocean circulation and stratification, which is indirectly
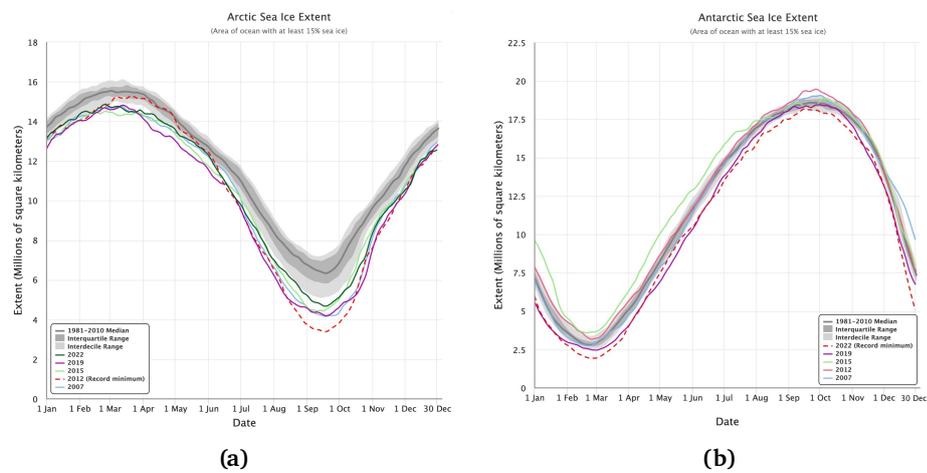
Figure 2.1: Graphs of Arctic (a) and Antarctic (b) sea ice extent for 5 previous years (2022, 2019, 2015, 2012, 2007) along with the 1981 to 2010 median shown in dark grey color. The gray areas around the median line show the interquartile and interdecile ranges of the data. Figure was generated from [10].

influencing the climate in different areas [11, 15].

Sea ice covers large areas of the Arctic Ocean and plays a major role in the marine ecosystem. Different species from large mammals to microorganisms inaccessible to the human eye depend on the presence of sea ice. While birds and mammals use sea ice as a platform for hunting, smaller organisms such as algae live either internally within the sea ice or at the ice-water interface [17, 18]. Moreover, sea ice organisms are an important food source for different pelagic animals and they are a reason for algae blooms occurring during the sea ice melt [17].

Additionally, sea ice considerably influences various human activities in the polar regions. The lifestyle and welfare of the indigenous population strongly rely on the presence of sea ice [19, 20]. Sea ice poses a great challenge to navigation in the polar seas, which is especially important given the growth of shipping activity in the Northern Sea Route. Figure 2.2 illustrates the main shipping routes in the Arctic Ocean, as well as the potential routes that can be possibly employed in near future due to significant and unprecedented sea ice decline. Therefore, the knowledge about its type, concentration, thickness, deformation, and extent is extremely important for various activities, such as marine transportation and offshore operations, and for stakeholders from the oil and gas industry, fisheries, and tourism, among others [21, 22]. Some of the ice charting groups around the world, such as the Norwegian Ice Service (NIS), Canadian Ice Service (CIS), and the Arctic and Antarctic Research Institute (AARI) are providing the daily sea ice conditions overview in the form of maps. One of the examples of such a map is illustrated in Figure 2.3. The ice chart provides information about sea ice concentration (SIC) and the location of the original SAR data that was used. It should be noted that this kind of ice chart is produced mostly manually and requires an expert's knowledge.
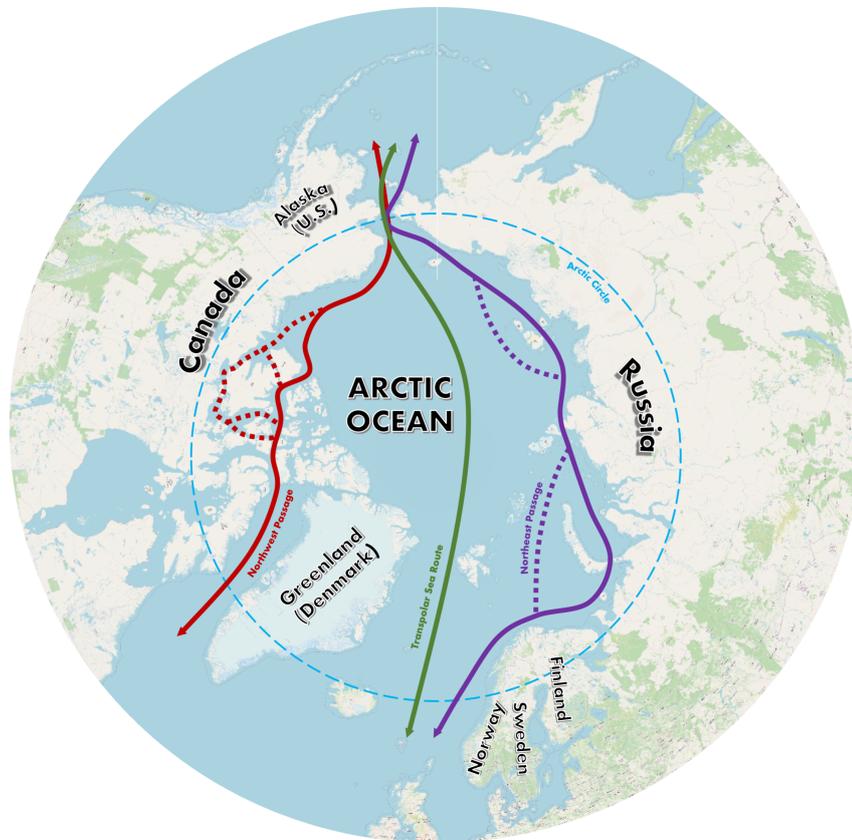
Figure 2.2: Map of the main Arctic shipping routes that become significantly more employed due to the unprecedented sea ice concentration decrease. The red line refers to Northwest Passage, the purple to Northeast Passage, and the green line to the potential Transpolar Sea Route. The light blue line corresponds to the Arctic Circle.

2.2 Sea Ice Observations

2.2.1 In-Situ Observations

In-situ observations of sea ice are crucial for understanding the processes that control its formation, growth, and decay. Many studies require extraction and examination of ice cover samples, in order to obtain the physical, chemical, and biological properties of sea ice [23]. They involve direct measurements of various properties of the ice, including its thickness, roughness, temperature, salinity, and structure.

Even though the interest in sea ice research has grown significantly over the last decades, with an increasing number of polar expeditions, in-situ observations are still very sparse and can only reveal local information about sea ice.

Overall, in-situ observations of sea ice are still critical for understanding complex processes of sea ice evolution. They provide researchers with valuable data that can further be used to
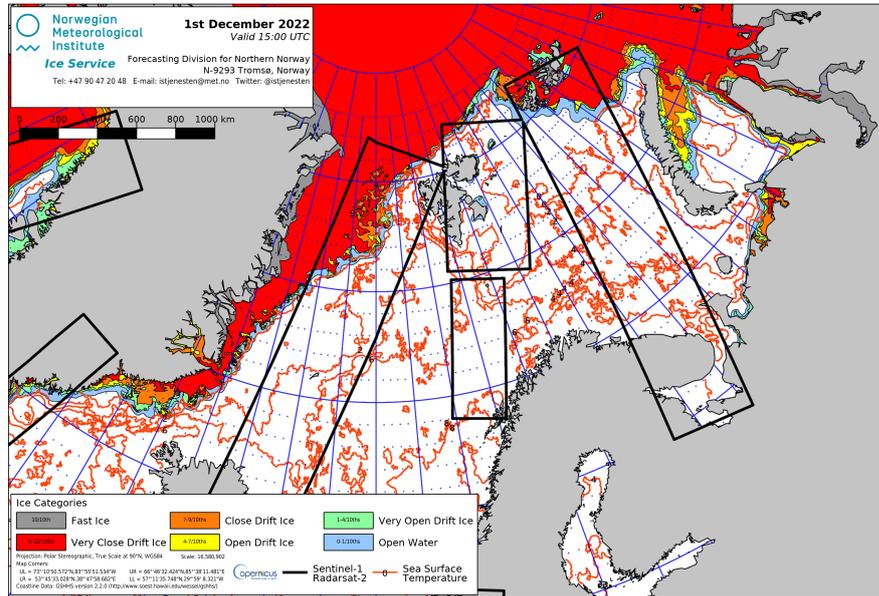
Figure 2.3: Example of an ice chart produced by Norwegian Ice Service for the 1st of December, 2022. Sea ice concentration is demonstrated by different colors, while the black rectangles show the SAR imagery frames used in the production of this ice chart. In addition to various SAR sources, the Ice Service also uses data from other sensors, such as optical and passive microwave in order to produce detailed ice charts. Such ice charts are freely available from [16].

improve various sea ice models to make more accurate predictions about atmosphere - sea ice - ocean interactions.

2.2.2   Remote Sensing Observations

Sea ice observations have a long history of more than a century. They were carried out visually from coastal stations, ships, and aircraft. Aircraft survey was the main observation method until the 1980s, however, it was spatially and temporally limited [22]. Regular sea ice monitoring over larger regions became possible in the late 1970s with the advent of satellites and over the last several decades, it became impossible to imagine operational sea ice monitoring without satellite data [22, 24]. Since then, the technology for acquiring and analyzing sea ice data has been considerably improved and extended.

The sea ice observations are carried out over a wide range of the electromagnetic spectrum, at visible, infrared, and microwave wavelengths, using both active and passive sensors. Active sensors generate their own signal and are independent of a natural radiation source, while passive sensors rely either on solar illumination, as in the case of optical sensors, or on radiation that is naturally emitted from the Earth's surface and atmosphere, in the case of passive microwave

radiometry.

Nowadays, the main sources of information about sea ice conditions and for climatological studies are data from passive microwave radiometers (PMR), and synthetic aperture radars (SAR). The latter is preferably used for strategic and tactical navigational support and for local studies requiring data at high spatial resolution. Both sensors are commonly used due to their independence from cloud and light conditions and therefore their ability to provide imaging of the Earth's surface continuously during day and night, and for almost all weather situations [25].

2.3   Sea Ice Types

Sea ice can be classified based on different criteria, such as its thickness, structure, and environmental conditions in which it forms. In the case of sea ice classification, we are more interested in ice thickness which is related to the ice age and deformation.

The ice types based on World Meteorological Organization (WMO) Sea Ice Nomenclature are the most common classification scheme for sea ice [26]. Even though the SAR and PMR data can not directly measure the sea ice thickness, except the thickness of very thin ice that can be retrieved from PMR data, the class labels for sea ice are mainly related to ranges of ice thickness (nilas, grey ice, grey-white ice), in some cases directly focussing on age (e.g. young ice, first-year ice, multi-year ice) [27].

For microwave remote sensing, it is important to understand the interaction of the radar signal with the sea ice, which depends on various radar and radiometer data as well as on the ice properties, in order to properly identify and distinguish different sea ice classes. In this thesis, we used various sensors, namely optical, SAR, and PMR that provide different information about sea ice types. It should be noted, that each of these sensors can provide complementary information to distinguish different sea ice types. Optical data helps distinguish water, nilas, and young ice, and SAR can provide information about older ice types, while from PMR data one can derive the sea ice concentration and extent. Therefore, it is extremely beneficial for visual inspection and manual identification of various sea ice and water classes to use various sensors in combination in order to grasp all the available information about the ROI.

In *Paper II* several sea ice classes were determined after the visual data inspection, namely nilas, grey ice, grey-white ice, level first-year ice (FYI), deformed FYI, thick FYI, and thin FYI. It should be noted, that we used optical and SAR high-resolution data combinations that allowed us to distinguish more sea ice types than shown in *Paper III*, where we are using SAR and PMR data that significantly differ in spatial resolution. Accordingly, in *Paper III* only 3 sea ice classes were determined after the visual inspection, namely brash ice, grey ice, and thin FYI.

3

Dimensionality Reduction

The majority of scientific contributions in this dissertation focus on the development and validation of a new information selection algorithm, therefore, this chapter introduces the basics of dimensionality reduction which is an important part of pattern recognition in general.

3.1 Dimensionality Reduction Principles

Dimensionality reduction is one of the crucial parts of any pattern recognition task. It corresponds to the methods applied to represent data with fewer features/attributes in order to discover a compact and informative representation of the original data [6, 28]. There are several reasons behind dimensionality reduction that makes it an important part of many applications. First and foremost, dimensionality reduction methods can remove information redundancies that usually exist in the set of samples obtained by the measuring devices [7]. This redundant information can significantly increase the complexity of the analysis leading to a decrease or saturation of the accuracy. Secondly, decreasing the number of data dimensions will lead to a reduction of necessary computational resources and an increase in the overall algorithm performance. Moreover, in the case of data sets with a large number of features, some models become too complex and dimensionality reduction helps avoid the problem of overfitting. Likewise, it can be used to transform non-linear data into a linearly-separable form. Additionally, it should be noted, that not all the features are actually valuable since they can be redundant, corrupted, or simply unnecessary for the given task and can potentially even decrease the method's performance. Accordingly, to improve the algorithm's performance as well as knowledge about an observed area, there is a need to develop an automatic method to select the relevant information.

Dimensionality reduction can be achieved through supervised and unsupervised methods. In recent years, unsupervised dimensionality reduction techniques have raised considerable interest, since they do not require any prior knowledge about the data, which gives them a significant

advantage over the supervised methods, especially for applications where it is challenging to obtain proper labeling [29]. Nevertheless, despite the huge variety of existing methods, generally, dimensionality reduction methods can be divided into two main approaches: feature extraction and feature selection.

3.1.1 Feature Extraction

Feature extraction is one of the dimensionality reduction approaches that reduce dimensionality by finding an optimal transformation matrix and projecting the original high-dimensional data onto a lower dimensional subspace [30, 31]. In this way, the separability of the data is increased, however, the physical interpretability of the original data is not preserved.

In order to demonstrate the idea of feature extraction methods in general we will use one of the most popular and widely applied unsupervised dimensionality reduction techniques, principal component analysis (PCA) [7, 30]. PCA converts a set of features of potentially correlated variables into a set of linearly uncorrelated variables, called principal components. It projects the original set into a lower dimensional space spanned by a subset of the principal eigenvectors of the data's covariance matrix. Thus, it reduces the size of the original set while maximizing its variance [7].

Let M be the number of all features, in our case multimodal remotely sensed images, e.g., polarization intensities, spectral channels, textural features, etc. It should be noted, that the images acquired by different sensors may have different physical units, different resolutions, and different coordinate systems. The first step of analysis consists of making the data comparable, by means of normalization, subsampling, and spatial alignment on the same coordinate system such that the images perfectly overlap. We denote by $\mathbf{x}_i \in \mathbb{R}^M$ the set of M images, where each image represents one feature associated with the i-th pixel, and by $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{M \times N}$ the matrix of features, where N is the number of pixels.

In order to remove the redundancy and reduce the dimensions of the original features, PCA generates a new representation using the eigenvectors of their covariance matrix. The sample covariance matrix of the features can be written as follows:

$$\Sigma \quad = \quad \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu}) (\mathbf{x}_i - \boldsymbol{\mu})^T \tag{3.1}$$

where $\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$, and $.^T$ denote the transpose operator. Using singular value decomposition [32], we obtain:

$$\Sigma = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^T \tag{3.2}$$

where $\mathbf{V} \in \mathbb{R}^{M \times M}$ is a unitary matrix, whose columns are the eigenvectors of Σ, while

$\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_M\}$ is a diagonal matrix, whose elements are the eigenvalues of Σ, sorted such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$.

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1K} & \cdots & v_{1M} \\ v_{21} & v_{22} & \cdots & v_{2K} & \cdots & v_{2M} \\ & & & \vdots & & \\ v_{M1} & v_{M2} & \cdots & v_{MK} & \cdots & v_{MM} \end{pmatrix}$$

K principal eigenvectors

$$(3.3)$$
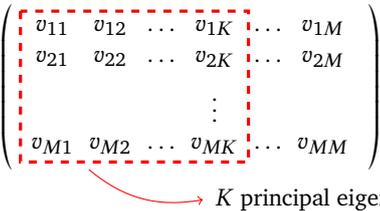
Σ depicts the second-order statistical relationships between the features. Moreover, the eigenvectors of Σ can be geometrically interpreted as the axes that best represent the directions of the data plane [33]. Their corresponding eigenvalues reflect the variability of the features along the axes. Accordingly, eigenvectors with large eigenvalues reveal most of the information about data variance [33]. Therefore, PCA uses principal eigenvectors corresponding to the K largest eigenvalues of Σ (which hold most of the variation about the features) [34], as the basis of a lower-dimensional space onto which the data set is projected. Respectively, we use the rows of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K]$, after discarding the eigenvectors corresponding to the smallest eigenvalues, as the new representations of the features (see Equation 3.3). This new representation that accounts only for the largest eigenvalues presents a powerful discrimination compared to the original features. The new lower-dimensional space is given by:

$$\mathbf{Y} \;=\; \mathbf{X}\mathbf{V}^{T} \tag{3.4}$$

where $\mathbf{Y}$ is the transformed output image, $\mathbf{X}$ is the original set of input features defined above, and $\mathbf{V} \in \mathbb{R}^{M \times K}$ is the matrix of normalized eigenvectors of the input data set covariance matrix, which is ordered by the respective eigenvalues. Depending on the data set and task, the first several components are selected as relevant representative features.

As was mentioned above, PCA is a widely used dimensionality reduction technique. It projects the original data onto a new transformed space, a direct mapping between the original features in the data set and the PCA components [33]. This strongly affects remote sensing data analysis, since it makes it difficult to achieve an interpretation of the physics behind the outcomes of the analysis, i.e., it is not possible to understand what physical factors (identified by the features originally collected in the data set) drive the information extraction process. This is a crucial drawback, especially in disciplines (e.g., environmental monitoring, climate change adaptation, and mitigation) that rely on a robust characterization of the events occurring on the Earth's surface [35, 36].

One of the first results of this PhD project that we achieved was improving the principal feature analysis (PFA) technique [37] which is a feature selection method that was based on PCA. PFA exploits the same tools as PCA to generate a new representation of the data, however, this new representation, as opposed to PCA, can be mapped back to the original domain and hence preserves the physical interpretation of the data set. In [38] we applied PFA for multimodal information selection for remote sensing of sea ice. In contrast to the classic PFA, we perform the selection in a patch-wise manner, which allows us to capture the most relevant information for smaller homogeneous regions. Moreover, we take into account the particularity of each object in the observed scene, since our approach only chooses the relevant features to its characterization. Experimental tests demonstrated the robustness and flexibility of the improved algorithm. This experience and knowledge helped us in developing a new feature selection algorithm that we subsequently proposed.

Among the methods of feature extraction, besides PCA and its variations, we may also cite, for instance, decision boundary feature extraction (DBFE) [39]. Unlike PCA, DBFE is a supervised approach that uses the training set to determine the decision boundary between classes. The eigenvectors of the decision boundary matrix determine the direction of projection of the original set of features. Thus, it provides a minimum number of transformed features that achieve the same accuracy as the original set [39].

3.1.2   Feature Selection

Feature selection is a part of different scientific fields, such as pattern recognition, machine learning, data mining, and statistical analysis. Unlike feature extraction, feature selection reduces dimensionality by selecting the most informative subset of records, preserving the characteristics of the original data without working in a different space [40]. It determines a relevant subset of the original set according to different criteria, such as information, similarity, or correlation.

3.1.2.1   Categories

There are different ways how the feature selection methods are divided, in this work we will categorize them into several main categories: ranking-based, searching-based, sparsity-based, embedding learning-based, clustering-based, and hybrid-based. The following paragraphs provide a brief description of each category:

Ranking methods measure the significance of each feature individually by sorting them in terms of relevance using different criteria [41, 42]. They can be very efficient, however, not very precise due to the fact that they do not consider the relationship between the features. Among this feature selection category, it is worth citing Fisher score selection (FIS). FIS is a supervised algorithm that uses the Fisher score as a criterion to select the most relevant features. Fisher score measures the ability of each feature to reduce the intraclass distance while increasing the interclass distance [43].

Searching methods select the optimal subset in an incremental, removal, or update manner using a search method, such as the genetic algorithm [44] or branch and bound [45]. This category of feature selection methods is more accurate compared to ranking methods since it considers the interaction between the data, however, such methods are limited by the size of the searching space. In the case of large data sets, computation time significantly increases, and the searching methods fail to achieve optimal results. Within this category, it is worth highlighting forward feature selection (FS) [46], orthogonal branch and bound (OBB) [45], and the genetic algorithm (GA) [44]. FS determines the optimal subset in an incremental fashion. The algorithm starts with a minimum number of features, and with each new step, it adds one feature that improves the accuracy until no further improvement is noticed [46]. The OBB is a backtracking feature selection algorithm, that is based on the assumption that the adopted criterion function fulfills the monotonicity condition. Hence, it guarantees finding the optimal subset while omitting many feature subset evaluations. The branching step consists of constructing the tree such that the subtree of each level is constructed by deleting one feature until the required number of features is reached. The bounding step represents the process of traversing the tree to find the optimal subset [45]. GA is an adaptive algorithm that finds the global optimum solution for an optimization problem based on the mechanics of natural genetics and biological evolution. Genetic algorithms operate on a population of

individuals to produce better approximations. In feature selection, each individual in the population represents a predictive model with genes that correspond to the total number of features in the data set. Genes are encoded as binary values that show if the feature is included or not in the subset [44].

Sparsity methods emphasize underlying structures within data by solving an optimization problem with sparsity constraints [47]. Among this category, we can mention several methods, one with sparse non-negative matrix factorization (SNMF) that uses additional constraints, such as the thresholded Earth's mover distance (TEMD) [48], and sparse representation-based band selection (Spa-BS) [49] that does not require non-negative matrix factorization.

Clustering methods divide the components of the original set of features into different groups/clusters, and from each cluster, selects a representative relevant element to compose the optimal subset [50]. It is worth noting that the approaches within this category can be further divided into three sub-categories, k-means-based [47, 50], affinity propagation-based [51], and graph-based [52]. This thesis shall focus on graph-based methods.

Embedding learning-based methods incorporate feature selection into the optimization of a specific model for various applications, such as classification, target detection, unmixing, etc [47]. Here we can mention a method that automatically eliminates redundant features by adopting logistic voting to improve the convergence speed of recursive-feature elimination (RFE) [53]. Additionally, there is an RFE-SVM method that was modified into the kernel-based feature-selection algorithm, which applied a magnitude of the support vector machine (SVM) coefficients as a ranking criterion to select the relevant features and improve the classifier training [54].

Hybrid methods integrate multiple above-mentioned operations to obtain relevant feature selection. Among the hybrid-based algorithms, we can highlight a method where the ranking and eliminating searching schemes were combined together in order to rank the features according to minimum redundancy and maximum relevance (mRMR) along with sequential backward elimination to choose the most relevant features [55].

3.1.2.2 Graph Clustering

Among the various subcategories of clustering-based feature selection methods, graph-based clustering methods play a key role. Even though the definition of clustering feature selection methods might look similar to the ranking methods, in reality, there are some significant differences. While ranking methods involve assigning a score or rank to each feature based on some metric or criterion, clustering methods group similar features together based on some similarity measure or distance metric. The intuitive goal of clustering itself is to separate the data points into several groups/clusters such that points in the same group are similar, while the points in different groups are dissimilar to each other. The graph clustering-based approaches find the relevant features by partitioning the graph into subgraphs and selecting the representative feature from each of them [56]. Consequently, the crucial difference between the above-mentioned feature selection methods lies in the fact that the ranking techniques focus on the individual features and arrange them depending on the rank or score, while the clustering-based approaches group features together and select them based on their similarity or correlation with each other.
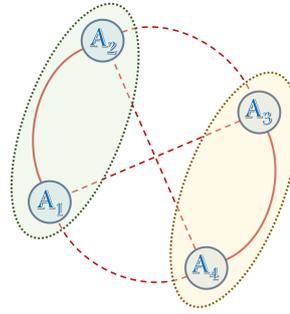
Figure 3.1: Example of the graph of four features. Blue circles represent the set of vertices/nodes, while the red lines correspond to the set of edges that connect the nodes. The dashed red line shows a weak connection, while the solid red line shows a strong connection between features, therefore A1 and A2 are grouped into one cluster, while A3 and A4 are included in another cluster.

Graphs are one of the natural ways to represent various types of data. A graph is a set of nodes or vertices, where each node corresponds to one feature, with a corresponding set of edges that connect them (see Figure 3.1). The edges may be either directed or undirected and in some cases, they can have weights. If there is not much information apart from similarities between the data points provided, a nice way of representing the data is to form a similarity graph. There are various similarity graphs that differ from the construction point of view. The main idea is to properly model the local neighborhood relationships between the data points. The most commonly used graphs are ϵ-neighborhood graph, k-nearest neighbor graph, and fully connected graph. The ϵ-neighborhood graph is usually considered as an unweighted graph that connects all points whose pairwise distances are smaller than ϵ. The k-nearest neighbor graph is a directed graph where vertexes are connected among their k-nearest neighbors. The fully connected graph simply connects all points with positive similarity with each other, and weights all the edges. In the case of a fully connected graph, we use a similarity function to reflect the local neighborhood relationships.

For the sake of clarity, we present the principles of the graph clustering feature selection first at a pixel level. The adaptation to the superpixel/patch level will be demonstrated in Section 3. In feature selection, we specifically aim to find the smallest subset of features, that preserves the structure and information content of the original set. Let $\{x_1, \ldots, x_M\}$ be a set of data points/pixels, where M is the number of features used, then $\{x_1, \ldots, x_K\}$ corresponds to a smaller subset that we are targeting to find, where K refers to the number of features to be selected. The undirected and fully connected similarity graph is introduced as follows:

$$\mathcal{G}(\mathbb{V}, \mathbb{E}) \tag{3.5}$$

where $\mathbb{V}$ denotes the set of data points from our features that are represented as nodes/vertices, while $\mathbb{E}$ corresponds to the edges that connect the nodes and reflect their similarities.

The most commonly used function to create a fully connected graph is the Gaussian similarity

function which is given by:

$$w_{i,j} = \exp\left(-\frac{(x_i - x_j)^2}{2\sigma^2}\right) \tag{3.6}$$

where $\sigma > 0$ is a parameter that controls the width of the neighborhoods [56]. A large value of σ will indicate a larger similarity even if the Euclidean distance between x_i and x_j is relatively large, whereas conversely, a lower value of σ weakens the resemblance judged from only the Euclidean distance.

In graph theory, there are a few graph representations that can be used to reflect the relationships between data points and to construct a graph Laplacian matrix, namely affinity (weighted adjacency) and adjacency matrices. However, they differ in their definitions and the information they provide. The adjacency matrix $\mathbf{W}$ is a graph representation, where the row and column indices represent the nodes, and the entries represent the absence or presence of an edge between the nodes, either 0 or 1. An affinity matrix is similar to an adjacency matrix, however, the value for a pair of points signifies how similar those points are to each other. Therefore, the affinity acts like the weights for the edges on our graph. For both, affinity and adjacency matrices, 0 corresponds to complete dissimilarity between two points, while 1 means that the points are identical. The main difference between the two matrices is that the affinity matrix reflects the similarity between data points, while the adjacency matrix provides information about the structure of the graph. Whereas the symmetrical matrix $\mathbf{W}$ represents the connections between the nodes in a graph, the degree matrix $\mathbf{D}$ is a diagonal matrix that contains information about the degrees of the nodes in the graph, thereby demonstrate how many edges connect to a particular node.

Once the graph is defined according to the operations that have been previously introduced, we can build the Laplacian matrix that plays a crucial role in graph clustering. There are various variants of graph Laplacians each of which has different properties, however, the most commonly used definitions are the unnormalized Laplacian and the normalized Laplacian [56]. The normalized graph Laplacian matrix is defined as

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \tag{3.7}$$

where $\mathbf{I}$ is the identity matrix. $\mathbf{W} = (w_{ij})$ is the adjacency matrix of the graph $\mathcal{G}$, and $\mathbf{D} = \text{diag}(\sum_{i \neq j} w_{ij})$ is a degree matrix, which counts the number of times an edge terminates at the particular vertex.

The next step is to perform the eigendecomposition of the Laplacian matrix that provides information about the graph's structure that can be used for further clustering. The corresponding eigenvectors of the Laplacian matrix form a set of orthogonal basis vectors that span the space of all possible vertex configurations. Accordingly, by projecting the vertices onto these eigenvectors, we can obtain a low-dimensional representation of the graph that preserves the graph's connectivity [56]. The eigenvectors that correspond to the smallest eigenvalues of the Laplacian matrix are especially crucial, since they capture the local structure of the graph, therefore, being the most informative. After that, the clustering algorithm can be performed on this low-dimensional representation in order to group similar vertices together. The eigendecomposition is defined as follows:

$$\mathbf{L} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \qquad\qquad (3.8)$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, ..., \lambda_M)$, is a diagonal matrix of the corresponding eigenvalues, and $\mathbf{V}$ is the matrix of eigenvectors.

The last step is clustering itself, which is performed on the eigenvectors that were built from the graph or the data set. Correspondingly, we stack the most informative first K eigenvectors of $\mathbf{L}$, $\mathbf{u}_i$ $(i = 1, \dots, K)$, into one matrix $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_K]$, where K is the number of clusters to be created. It should be noted, that K is directly connected to the number of features to be selected. There are various ways to determine the appropriate K for a clustering algorithm, such as Elbow Method, Silhouette Method, and Gap Statistic [57, 58, 59]. However, each of these methods can provide a different optimal value for K, even for the same data set. Moreover, some of these algorithms are time-consuming. Therefore, in some cases, it is better to tune the K for specific data sets. The graph partition can be done using a clustering method, which is an unsupervised division of the data points into a number of groups with similar values. There are plenty of algorithms that are used for clustering purposes, among them those that are particularly suitable for sea ice analysis [60]. In this work, we are employing a commonly used method in spectral clustering, namely k-means [7], which is a simple and fast clustering algorithm.

The k-means algorithm is one of the simplest unsupervised learning techniques, that group a given data set through a particular number of clusters that are fixed apriori. The basic idea behind the k-means algorithm is that we define K centroids, one for each cluster, respectively, that should be placed as far away as possible from each other.

Generally, k-means can be divided into 3 crucial steps. As a first step, namely **(1) cluster initialization** we randomly select the K centroids, which in our case corresponds to the number of relevant features we want to select. The second step is **(2) cluster assignment**, where we take each point that belongs to a given data set and associate it with the nearest centroid. The final step of this algorithm is **(3) cluster update**, where we re-calculate K new centroids, by averaging the locations of data points corresponding to each centroid and re-assign the centroid to that location. Additionally, we perform the steps **(2)** and **(3)** until convergence (until the end of the change in centroids position).

In conclusion of this paragraph, I would like to add that, for sea ice monitoring, it is not a very common scenario to have large data sets in terms of a number of features. Even if it is a multimodal case, like in this particular dissertation, we can potentially end up with less than a dozen of features. Although, it is possible to increase the number of features by extracting the textural or polarimetric parameters. However, even when we have only a few features, feature selection can still be a useful technique to improve the performance of the classification.

4

Proposed Scheme

The backbone of the dissertation and the most important part of *Paper I*, *Paper II*, and *Paper III* is the proposed dimensionality reduction method that is integrated into the scheme for classification. In this section, we briefly introduce the proposed information selection method, which is based on the representation of Graph Laplacians, as well as the whole scheme for data processing and classification. The scheme itself consists of three main parts, which are displayed in Figure 4.1, namely *Attribute Generation*, *Attribute Selection*, and *Classification*.

4.1 Attribute Generation

Attribute Generation is the very first step that includes *Pre-Processing* and *Feature Extraction*. In *Paper I* we used publicly available data sets that are fully pre-processed. In *Paper II* the pre-processing part was fully performed by the dissertation author, while in *Paper III* other authors were involved in the data curation part.

4.1.1 Pre-Processing

One of the main goals of this thesis is the employment of multimodal data sets which are acquired by various sensors that operate at different frequencies, hence can potentially have different characteristics, such as units of measurement, spatial resolution, and geographical systems. As a very first step, we make sure that all the acquired images are perfectly spatially compatible.

Therefore, the images should be spatially aligned such their attributes overlap; calibrated if needed; resampled to the same spatial resolution; denoised, which is especially important for SAR, where we have to consider thermal noise and speckle; and stacked together. Most of these

(1) Attribute Generation

(a) All the available images are stacked together in a data cube.

Flowchart of the proposed scheme

(3) Classification

(f) For all superpixels apply parallel classification using Random forest classifier.

(2) Attribute Selection

(b) Example of superpixel segmentation algorithm application.

(c) The attributes of the l-th superpixel are stacked in one matrix.

(d) Graph of four attributes with two similarity functions at the l-th superpixel.

(e) Example of graph clustering, similar attributes are grouped together.

(2.1) Segmentation   **(2.2) Graph Building**   **(2.3) Graph Clustering**

Figure 4.1: Flowchart of the proposed algorithm with all the crucial steps for a multi-modal sea ice classification case.

steps were performed in ESA's Sentinel Application Platform (SNAP).

4.1.2 Textural Features Extraction

One of the crucial parts of Attribute Generation is the extraction of textural features from the gray-level co-occurrence matrix (GLCM) [61, 62, 63]. This step was performed in each of the included papers, in order to increase the number of attributes and extract some additional unique information from the original data.

It is worth mentioning, that there are other features besides textural that can potentially be applied to sea ice analysis. Especially, polarimetric features should be highlighted, since they are able to provide more detailed and accurate information about the physical properties of sea ice than traditional intensity-based features. Moreover, by analyzing the polarization properties of the radar signal, it is possible to obtain crucial information about sea ice, such as its surface, volume, and dielectric properties. Even though polarimetric features were not used in this thesis, they can potentially be integrated into future studies.

In the case of sea ice, we used a directional average for $0°$, $45°$, $90°$, and $135°$ which is common practice to account for the possible rotation of sea ice floes, leads, or any other sea ice structure or roughness pattern on the ocean surface. There are also other parameters that can be potentially changed, such as distance/displacement or window size, however, it is noteworthy that minor changes in these parameters are not significantly affecting the performance. Therefore, it is more beneficial to use averaged parameters in order to reduce the number of GLCM matrices. Figure 4.2 demonstrates a few GLCM textural features extracted from one optical band and SAR polarization for eddy identification in the marginal ice zone that we used in other studies. It is evident that the textural features can potentially expand the amount of unique information that will further be evaluated by the proposed selection method in order to decide whether it is complementary or redundant.
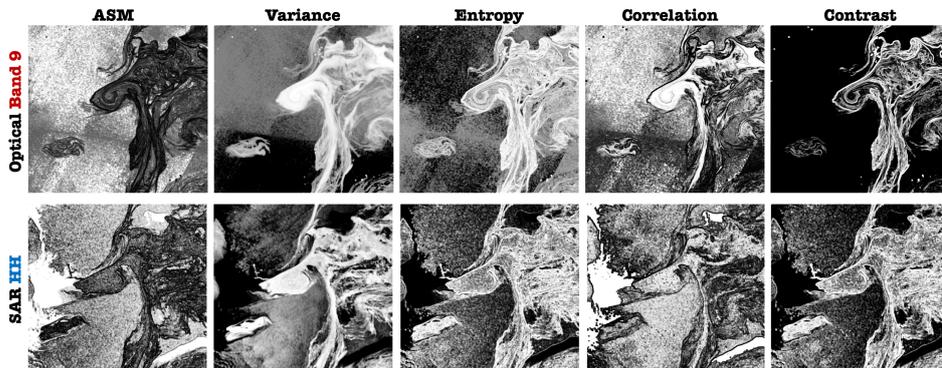
Figure 4.2: Examples of GLCM texture features (angular second moment (ASM), variance, entropy, correlation, contrast) extracted from optical Band 9 and SAR HH polarization for the marginal ice zone over the Greenland Sea, East Greenland [64].

4.2   Attribute Selection

Here we present the proposed fully unsupervised, adaptive, interpretable, and flexible information selection algorithm. In the remaining of this section, we refer to our method as GKMI, based on the two similarity measures that were used, namely the Gaussian Kernel and Mutual Information. The proposed information selection method consists of three main steps: *Segmentation*, *Graph Building*, and *Graph Clustering*.

4.2.1   Segmentation

Taking into account the large data size of the remote sensing images, performing the information selection in a pixel-wise manner is computationally expensive. Therefore, in order to alleviate the computational complexity, we implement the algorithm superpixel-wise, i.e., patch-wise. Moreover, depending on the size of superpixels, it allows us to preserve more local particularities, which is crucial since some image parts might require different types of attributes to effectively represent different classes and characteristics.

It should be noted that there are other patch-wise approaches, e.g. using regularly spaced windows. However, in comparison to fixed windows, superpixels include pixels that share similar information since they are generated using segmentation, which is a grouping of homogeneous pixels. It makes the selection more accurate and precise. Figure 4.3 illustrates an example of superpixel segmentation for the sea ice scenario.

There are various segmentation methods that can be used to divide the image into superpixels, such as Watershed [65] or Felzenszwalb [66], and Simple Linear Iterative Clustering (SLIC) [67] segmentation method. Watershed and SLIC are two commonly used algorithms in computer vision and image processing, that partition an image into small, compact regions called superpixels. Watershed treats an image as a topographical surface and partitions it based on the intensity and spatial gradients, while SLIC, on the other hand, is a clustering-based approach that partitions an image into a fixed number of superpixels based on color and spatial distance.

(a)

(b)
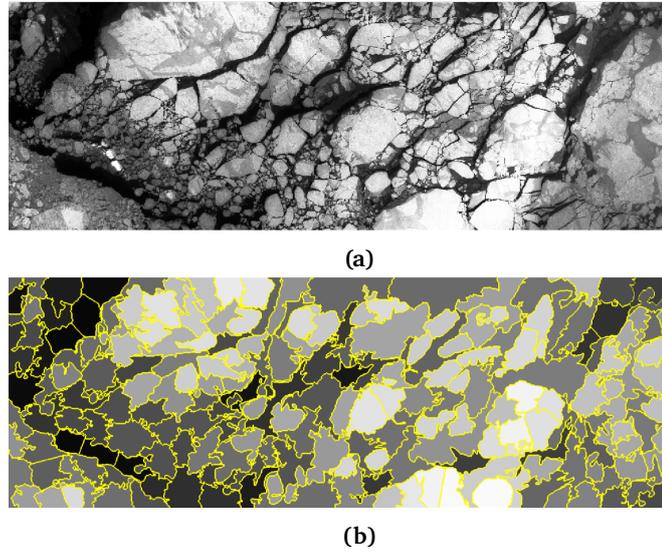
Figure 4.3: Example of superpixel segmentation for sea ice: original image (a), segmented image (b).

Furthermore, the number of superpixels is a parameter that can be changed within the algorithm depending on user preferences and applications. It is worth mentioning, that superpixel segmentation is only used as a part of the information selection step and despite other valuable benefits, the main purpose of the superpixel-wise approach is the computational optimization of the attribute selection algorithm.

4.2.2 Graph Building

To find the attributes which are best suited for sea ice classification, we apply a selection method, that relies on information theory metrics and on a representation based on graph Laplacians. The following subsections, namely *Graph Building* and *Graph Clustering* are based on the approach and equations that are fully described in Section 3.1.2.2. Therefore, the main focus will be on the crucial differences, between the classical approach and our proposed attribute selection.

First and foremost, in comparison to the classical pixel-wise approach, the proposed method is performed on a superpixel level, therefore, let L be the number of superpixels extracted using the segmentation method and M the number of initial attributes.

Furthermore, unlike existing graph-based clustering methods that are only using kernels as similarity measures (representing the mutual relations between the data points), we are also considering the information content of the original data. Therefore, the similarity is quantified using two metrics simultaneously, which allows us to capture relevant information at different scales which improves the precision of the selection. The mutual information (MI) is performed globally and provides a better estimation of the attributes-shared information, while the Gaussian kernel (GK) is applied locally and preserves the structure of the original data.

Gaussian Kernel w_{lij}^{GK} is a commonly used similarity measure that permits the preservation of

the structure of the original data set and in our method, it is performed on each superpixel separately using Equation 3.6. The difference is that instead of scalars x_i and x_j we are using the vectors $\mathbf{x}_{li}$ and $\mathbf{x}_{lj}$ consisting of all pixels corresponding to the i-th and j-th attribute of the given superpixel l respectively.

On the other hand, MI reports the shared information between two attributes [68]. The selection via information is performed image-wise in order to capture the global information of the original data. It is defined as follows:

$$
\begin{aligned}
w_{i,j}^{\mathrm{MI}} &= I\left(\mathbf{x}_i, \mathbf{x}_j\right), \quad 1 \leq i, j \leq M \\
&= \sum_{i=1}^{L} \sum_{j=1}^{L} P\left(\mathbf{x}_i, \mathbf{x}_j\right) \log \left(\frac{P\left(\mathbf{x}_i, \mathbf{x}_j\right)}{P\left(\mathbf{x}_i\right) P\left(\mathbf{x}_j\right)}\right)
\end{aligned}
\tag{4.1}
$$

where $P(\mathbf{x}_i, \mathbf{x}_j)$ is the joint density function of $\mathbf{x}_i$ and $\mathbf{x}_j$, and $P(\mathbf{x}_i)$ and $P(\mathbf{x}_j)$ are the marginals. Small values of $w_{i,j}^{\mathrm{MI}}$ exhibit independency of $\mathbf{x}_i$ and $\mathbf{x}_j$ which means that both attributes encompass different information. Conversely, high values of $w_{i,j}^{\mathrm{MI}}$ show dependency between the attributes which means that both reflect similar information.

It should be mentioned, that for the image-wise metric that reveals the information content of the original attributes, other metrics could be potentially used. For instance, the Pearson correlation coefficient measures the linear relationship between two attributes, thereby showing how similar they are [69]. In one of the studies, we applied the Pearson correlation coefficient as an intercorrelation criterion to directly evaluate the performance of the different information selection methods [70].

Accordingly, for each superpixel l, we build a multi-graph $\mathcal{G}_l(\mathbb{V}_l, \mathbb{E}_l^{GK}, \mathbb{E}^{MI})$, where $\mathbb{V}_l$ denote the set of attributes, $\mathbb{E}_l^{GK}$ and $\mathbb{E}^{MI}$ are two set of edges that connect the vertices. Consequently, the weights of the edges (strength of the connections) are given by Gaussian kernel w_{lij}^{GK}, and mutual information $w_{i,j}^{\mathrm{MI}}$. The MI is measured image-wise, considering pixels of the original images, therefore, the weights are equal for all superpixels, while the GK is calculated superpixel-wise, using only the pixels within each superpixel.

Applying the method on the superpixel level as well as exploiting two similarity measures are two main novelties of the proposed information selection algorithm that distinguish our method from the classical graph-based feature selection approaches. These novelties give us the possibility to extract both global and local information about our data which allows us to enhance the performance of the proposed method.

4.2.3 Graph Clustering

By partitioning the graph $\mathcal{G}_l$ into sub-graphs, GKMI groups similar attributes together, according to two similarity measures. Thus, by selecting a representative attribute from each sub-graph we obtain a subset that preserves the structure and the information content of the original set of attributes within the l superpixel.

The Graph Clustering part of the method applies similar steps as in Section 3.1.2.2. Accordingly, since we have two similarities, we should calculate two normalized Laplacian matrices, $\mathbf{L}^{\mathrm{GK}}$ and

$\mathbf{L}^{MI}$, that are associated with the $\mathcal{G}_l$. They are defined using Equation 3.7.

After the Laplacian matrices are calculated, we can perform joint approximate diagonalization in order to obtain joint eigenvectors and corresponding eigenvalues of the $\mathbf{L}^{GK}$ and $\mathbf{L}^{MI}$ using Equation 3.8. The common eigenspace spanned by both Laplacians enables their interaction, which might unfold the complicated structure of the graph.

The last step is performing k-means clustering on the joint eigenvectors that were built from the graphs of the data set. Accordingly, the final subset of K relevant attributes will correspond to the attributes closest to the centroids of the clusters, which is determined by comparing the distances between the centroids and each attribute. Moreover, the embedding (the matrix of eigenvectors of the Laplacian matrix associated with the eigenvalues) of the attributes increases their separability and accordingly gives better results than when performing the clustering on the original attributes [56]. Unlike the classical feature extraction approaches, attribute embedding can be mapped back to the original set since the original attributes and their new representations have the same indices. It allows for preserving the physical interpretability, which can be of crucial importance for some applications.

4.3   Classification

Dimensionality reduction in general and the proposed information selection, in particular, can be used as a pre-processing step for different kinds of applications and algorithms, such as target detection, classification, unmixing, etc. In our case, in order to quantify and evaluate the performance of the proposed method we linked it with different classification algorithms. It should be noted that the main focus of the thesis is developing the dimensionality reduction algorithm, therefore the classification methods will be only briefly described.

Throughout this thesis, we applied different supervised classification methods. Supervised methods require to have prior information about the data in order to learn specific patterns and further predict an associated discrete class. There are also unsupervised methods, also known as clustering, e.g. k-means described in Section 3.1.2.2, that do not require any preliminary knowledge about original data. They have the advantage of providing a segmentation based on the actual image information, however, it is challenging to afterward label the segments according to a given classification scheme. Moreover, another limitation is that in most clustering algorithms, the number of clusters has to be fixed before running the algorithm, which might differ from the actual number of classes on the observed scene.

4.3.1   Support Vector Machines

Support Vector Machines (SVM) is a type of classifier that is commonly used in various fields, such as vegetation [71], land cover [72], and sea ice classification [73]. It determines a set of hyperplanes that separate the data set into different classes [74].

Originally SVM is a linear classifier that searches for the optimal linear decision boundary in the training set, which is defined by maximizing the margin to the closest training samples (called support vectors) of each particular class. However, depending on the case, if the points in the data set are not linearly separable, SVM can model nonlinear decision boundaries through the use of various kernel functions [7]. Figure 4.4 shows an example of simulated training samples for

(a)   **(b)**
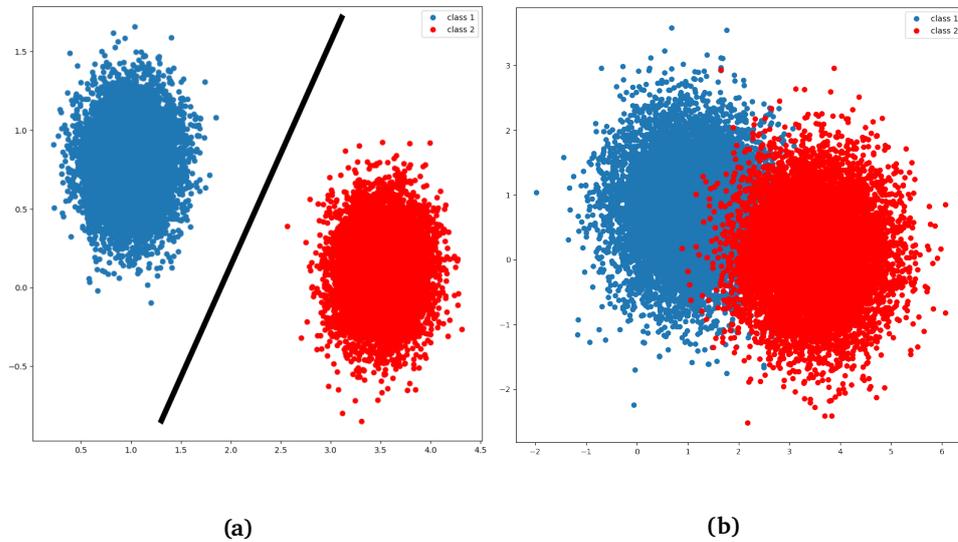
Figure 4.4: Example of simulated training samples for the two-dimensional two-class
          problem: linearly separable case (a), non-separable case (b). Various colors
          represent different classes. The black line displays the optimal decision
          boundary for a linearly separable case.

the two-dimensional two-class problem, where (A) shows the linearly separable case, while (B)
shows the non-separable case. It should be noted, that the choice of the proper kernel is a crucial
parameter that affects the performance and should be optimized for a particular task.

In *Paper I* in order to validate the performance and show the consistency and robustness of our
algorithm, we implement SVM and Random Forest (RF) since they are two of the widely applied
classifiers in remote sensing.

4.3.2   Random Forests

Random Forest is another type of classifier that is commonly used in remote sensing. RF generates
an ensemble of individual decision trees and combines their outputs to get an accurate prediction
of the class [75]. Each individual decision tree in the forest is a multistage decision system that
sequentially rejects classes along a path of nodes until the final proper path for a particular class
is created. There are various ways to split the trees depending on the splitting criteria and other
factors. The most common trees split the feature space linearly into hyper-rectangles with sides
parallel to the axes of the feature space.

RF was used in *Paper I*, *Paper II*, *Paper III*. In *Paper I* we used RF along with SVM in order to
properly demonstrate the robustness of the proposed information selection algorithm as well as
to tune the parameters within the method. Whereas on the smaller urban scenes used in *Paper I*,
both RF and SVM provided similar performance, on the large sea ice scenes employed in *Paper
II* and especially in *Paper III*, the SVM needed significantly more computer time. Therefore, in
Paper II and *Paper III* we integrated the RF classifier into the sea ice classification scheme since
it was providing less computational complexity than SVM, which is crucial if we are aiming
toward operational sea ice monitoring. Moreover, we further investigated the potential of the

proposed method for different data combinations and more sophisticated scenarios.

4.3.3   Validation Metrics

Once the supervised classifier is finally developed and properly tuned, it is crucial to evaluate the performance and estimate the quality of the outputs. The validation is performed on the data which is preliminary split into the training and test set. There are several metrics that we used throughout the thesis to quantitatively evaluate different aspects of the proposed scheme.

The *Overall Accuracy* (OA) shows the percentage of correctly classified samples, *Average Accuracy* (AA) quantifies the mean of class-specific accuracies for all classes, while *Cohen's Kappa coefficient* (k) measures the agreement between the classification and the reference data [76]. Additionally, we used the *Execution Time* (ET) to estimate the time complexity of the algorithm's performance.

While *Paper I* had the ground truth labels that were specifically intended for the classification, for the sea ice analyses of *Paper II* and *Paper III* the data sets were labeled by analysts from operational ice services and sea ice experts from institutions contributing to the articles. It should be noted that labeling different sea ice polygons and distinguishing the regions of interest (ROIs) for different sea ice types is a challenging task and always requires expert knowledge. For all the experiments of *Paper I* and *Paper II*, we randomly choose 20% of the pixels from the ROIs as a training set, while the remaining 80% of pixels were used as a test set for performance evaluation, which is a common approach in pattern recognition. However, for *Paper III* we employed the predefined and mutually exclusive ROI polygons for training and testing in order to better evaluate the differences between single-sensor and multi-sensor cases as well as the performance of the information selection algorithm in general. Since *Paper III* exploits the data acquired from sensors with significantly different spatial resolutions, we specifically tried to place the ROIs with large enough distances between them to avoid mapping upsampled PMR pixels corresponding to a single original PMR pixel into different ROIs. Furthermore, for various superpixels, different numbers and attributes are selected, that were considered in the pixel-wise classification which improves the performance of the algorithm in terms of accuracy and required computational time.

4.3.4   Summary

In order to sum up everything mentioned in the *Proposed Method* section, in this thesis, we introduce an information selection method that is combined with a classification algorithm for multimodal data sets. Among all the mentioned advantages we would like to stress the main contributions and novelties of this work and the proposed approach in particular:

- **Two Similarities**: it simultaneously employs two similarity measures that preserve global and local particularities of the original data set, which allows selecting the most relevant attributes.

- **Flexible Selection**: the method is performed in a superpixel-wise manner, therefore it selects the most relevant attributes for the considered classes across the different superpixels.

Moreover, here are some minor advantages, that are less significant in general, and have been employed in existing works, however, still worth mentioning:

- **Multimodal**: it is flexible, therefore it can be applied to various data combinations with different characteristics.

- **Unsupervised**: the part with information selection is completely application-independent, thus, it does not require any prior knowledge regarding the data sets or class labels in particular.

- **Interpretable**: the method keeps the crucial advantages of both dimensionality reduction strategies, namely attribute extraction and selection, such as preserving the physical meaning of the original data, while increasing its separability.

5

Basics of Optical, SAR, and PMR Remote Sensing

Since the beginning of the satellite era in the late 1970, space-borne remote sensing has become the main source of data for numerous applications. The measurements are acquired in a wide range of the electromagnetic spectrum, at visible, infrared, and microwave wavelengths. Remote sensing systems are most commonly divided into active and passive sensors. Active sensors generate their own energy, then record the reflection of that energy from the Earth's surface, which makes them independent from natural radiation sources, e.g. incoming sun radiation and Earth emission. Dependent on the type, passive sensors either measure reflected sunlight emitted from the sun, e.g., optical, or radiation that is naturally emitted from the Earth's surface, e.g., passive microwave radiometry [77]. Each of these sensors provides unique information since they grasp different aspects of the area of interest. Moreover, each of these sensors has some advantages and limitations. In the following subsections, some of the sensors used throughout this thesis as well as their basic principles will be briefly described.

5.1   Optical Sensors

Optical Remote Sensing uses that part of the electromagnetic spectrum (EM) that is characterized by the wavelengths from the visible to the near-infrared (NIR) or short-wave infrared (SWIR), and sometimes up to thermal infrared (TIR), as in the case of Landsat mission. Figure 5.1 displays different types of electromagnetic radiation, in particular, the tiny part related to optical remote sensing. Optical sensors detect the solar radiation reflected and emitted (in the case of TIR) from the observed surfaces and targets. Various materials reflect and absorb differently at particular wavelengths, therefore, targets can be identified by their spectral reflectance signatures in the formed remotely sensed images [78]. There are various types of optical remote sensing that are
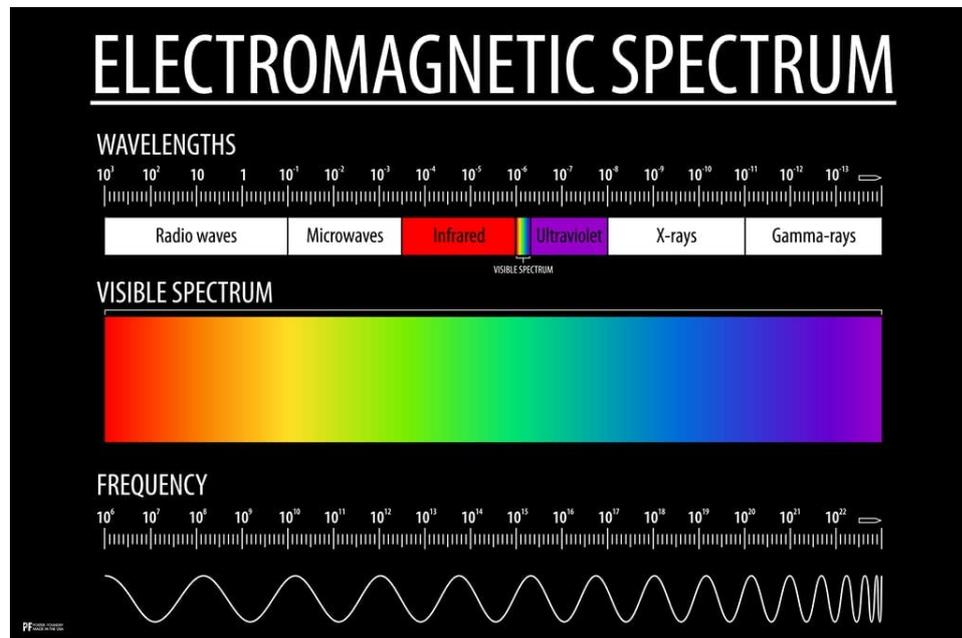
Figure 5.1: Electromagnetic spectrum with the range of various types of EM radiation. Figure from *Review of Myopia Management* [79].

characterized depending on the number of spectral bands and wavelengths used.

- **Panchromatic Imaging System.** This sensor is a single-channel detector that is sensitive to radiation within a broad wavelength range [78]. Since it is a single-channel, the output of the panchromatic sensor is a grey-scale image, therefore, the spectral information about the color of targets is lost.

- **Multispectral Imaging System.** The sensor is a multichannel detector that is operating with only a few spectral bands. Each of the provided channels is sensitive to radiation within a narrow wavelength band. The multichannel sensor provides a multilayer image that contains both the brightness and spectral information about the surface and targets that were observed.

- **Superspectral Imaging System.** A superspectral imaging sensor has more spectral channels in comparison to a multispectral sensor, usually around 10. Each band has a narrower bandwidth, therefore, enabling the finer spectral characteristics of the targets to be captured by the sensor.

- **Hyperspectral Imaging System.** A hyperspectral imaging system acquires from a few dozen to a few hundred wavelength channels with more contiguous spectral ranges. The spectral information contained in a hyperspectral image enables a more accurate characterization, identification, and separation of targets [77].

Optical remote sensing is widely used for various applications such as *agriculture* (monitoring of types, moisture, health, maturity of the crops), *vegetation* (identification of different vegetation types, health, separation from other types of land cover, forest fires), *coastal management*

(a)   (b)
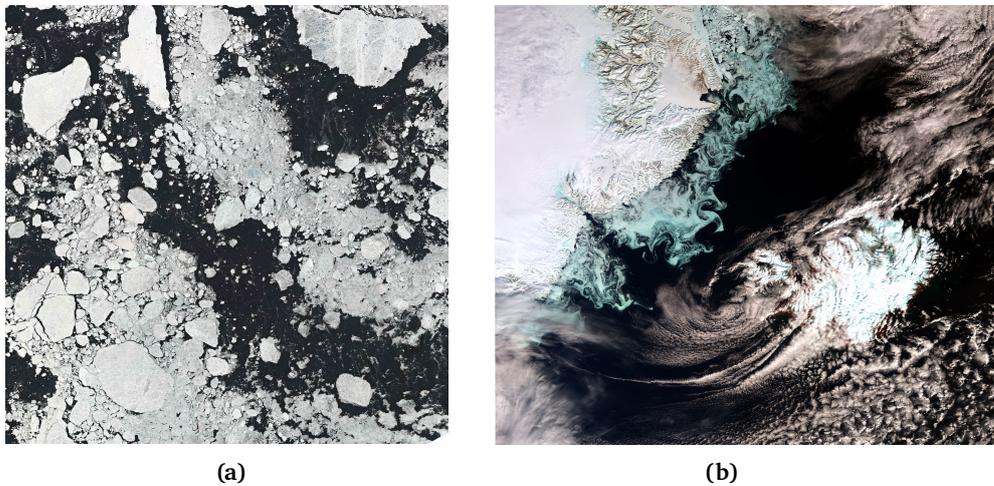
Figure 5.2: Example of natural-color composite for optical data: (Bands 4, 3, 2) of Sentinel-2 (a) with swath width 290 km and (Bands 17,6,3) of Sentinel-3 (b) with swath width 1270 km.

(monitoring of phytoplankton bloom, water pollution, bathymetry), *soil* (monitoring the desertification processes, soil erosion), *glaciers* (estimation of size, shape, melting), etc [77]. Figure 5.2 illustrates an example of optical data obtained from Sentinel-2 and Sentinel-3 sensors that have a significant difference in spatial coverage and resolution.

In *Paper I* we are employing several optical publicly available data sets for classification purposes in order to perform the evaluation of the proposed information selection method. Moreover, in *Paper II* we are additionally using images obtained from Sentinel-1 and Landsat 8 satellites for sea ice classification. It should be noted, that optical sensors are limited due to sensitivity to light and atmospheric conditions, which is well visible in Figure 5.2 (b), where a huge part of the scene is blocked due to the cloud cover. Nevertheless, it is still an important source of information about sea ice, which can be beneficial depending on the task. It can provide a lot of complementarity for visual inspection. Optical images can be crucial for labeling some sea ice types that are harder to interpret and separate using other sensors.

Table 5.1 shows several currently operating optical sensors and satellite platforms that are commonly used for Earth Observation. Each of these sensors provides a different number of spectral bands in different wavelengths. Throughout the thesis, we employed the data obtained from some of the satellites mentioned in Table 5.1, namely Sentinel-2, Landsat 8, and Sentinel-3, as well as some airborne hyperspectral scenes. It is evident from the table, how the spatial resolution differs with the change in swath width. While the high resolution of Sentinel-2 and Landsat 8 makes them an ideal source for mapping and monitoring small objects on the ground, the wide swath width of Sentinel-3 and MODIS makes them optimal for observing large-scale oceanic and atmospheric processes, such as sea surface temperature, ocean color, and surface wind speed.

Table 5.1: Examples of optical sensors that are currently operating. Resolution is indicated in meters. Swath Width is demonstrated in kilometers.

Optical Sensors				
Sensor	**MODIS**	**Landsat 8**	**Sentinel-2**	**Sentinel-3**
Bands	36	11	13	21
Launched	2011	2013	2015	2016
Country	USA	USA	Europe	Europe
Resolution	250-1000	15-100	10-60	300
Swath Width	2330	185	290	1270

5.2   Synthetic Aperture Radar

During the last decades, SAR has become one of the main sensors in remote sensing, significantly improving various applications. Moreover, it is the main data source that is used in this thesis. Therefore, the following sections will briefly describe the SAR advantages and limitations, as well as some fundamental principles that were especially important for this dissertation.

Currently, there are several SAR instruments that are orbiting the Earth. They operate at different characteristics, such as frequency bands, polarization channels, spatial resolutions, etc. Table 5.2 shows a few of the currently operating spaceborne SAR sensors and satellite platforms. Each of these sensors is obtaining valuable information that is crucial in the areas, such as polar regions, where other sensors, such as optical, are failing to provide relevant data. In comparison to passive microwave radiometers, SAR is an active sensor that generates its own signal [80]. Moreover, the typical wavelengths of SAR are perfectly able to penetrate through dense clouds, therefore, the observation is completely independent of solar illumination and weather conditions which is additionally combined with high spatial resolution [22]. This is a crucial thing since dense cloud cover and long periods of darkness prevail in the polar regions for several months of the year. However, there are a few main limitations of SAR, which are especially significant for sea ice. First of all SAR data are highly affected by thermal and speckle noise. The second and most important is that the automatic interpretation of remote sensing data, especially SAR data, is challenging and strongly relies on expert knowledge. Figure 5.3 demonstrates an example of SAR data obtained from Sentinel-1 for the coast of West Greenland.

Table 5.2: Examples of a few spaceborne SAR sensors that are currently in operation. Spatial resolution is demonstrated in meters, while the swath width is shown in kilometers. The swath width as well as the resolution may vary depending on the imaging modes, therefore, for both parameters, the values show the range from the least to the largest possible. SP refers to Single Polarization (HH or VV), DP corresponds to Dual Polarization (HH + HV or VV + VH), and QP indicates Quad-Polarization (HH, HV, VH, and VV).

Synthetic Aperture Radar				
Sensor	**Radarsat-2**	**TerraSAR-X/TanDEM-X**	**ALOS-2**	**Sentinel-1A/1B**
Frequency	C-band	X-band	L-band	C-band
Polarization	SP, DP, QP	SP, DP, QP	SP, DP, QP	SP, DP
Launched	2007	2007/2010	2013	2014/2016
Country	Canada	Germany	Japan	Europe
Swath Width	20-500	1-30	10-350	20-400
Spatial Resolution	3-100	0.25-40	3-100	5-40

<div align="center">(a)   (b)</div>
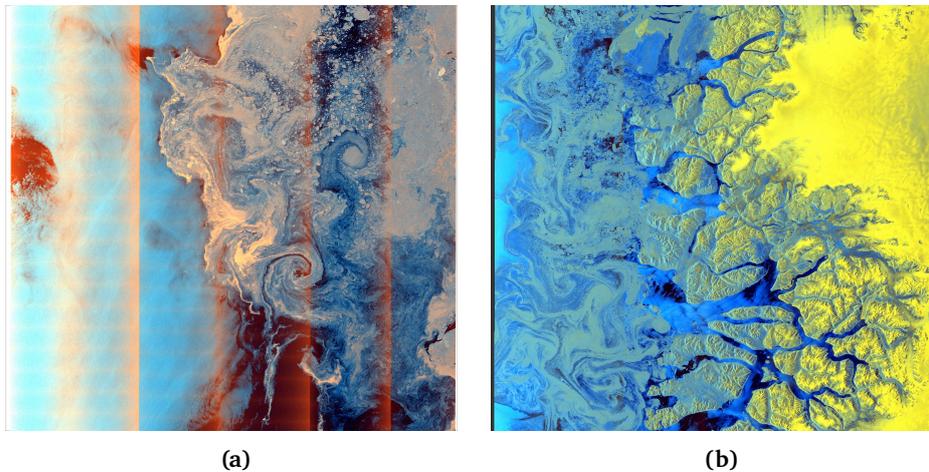
Figure 5.3: Examples of false-color composite (HV, HH, HH-HV as RGB) SAR data obtained from Sentinel-1 for the marginal ice zone along the coast of West Greenland.

5.2.1   Influence of Radar Parameters

Frequency, polarization, and incidence angle are a few important parameters for radar remote sensing that play a crucial role in the physical interaction between the electromagnetic wave and the observed surface [78].

Frequency. The choice of frequency strongly depends on the application and availability of corresponding images. In this thesis, we used images that were acquired from the most commonly used frequencies for spaceborne SAR observations, namely X-, C-, and L-band [81]. Table 5.3 lists different frequencies and wavelengths that can be used for radar remote sensing [77]. It should be noted, that the use of different frequencies can provide various complementary information about the underlying surface. Moreover, it can be especially crucial for sea ice applications since longer wavelengths can penetrate deeper into the snow cover and sea ice itself, therefore, providing a different view on specific ice properties [82]. However, it can still be challenging to acquire images at different times from different satellite platforms, especially for sea ice applications since they should be overlapped closely in space and time due to the ice drift. Nevertheless, the information provided by various sensors with different frequencies and wavelengths can significantly improve the knowledge about the ROI.

Incidence Angle. The incidence angle refers to the angle at which the sensor illuminates the target on the Earth's surface. The incidence angle increases across the swath from the near to far range and can significantly affect the resulting image intensity values, which can have an impact on the quality and interpretation of remote sensing images.

Polarization. Electromagnetic waves consist of electric and magnetic fields that are orthogonal to each other. The polarization of the signal is defined as the orientation of the electric field and can be described in terms of two orthogonal basis vectors [83]. Generally, electromagnetic waves are elliptically polarized, however, there are special cases with linear and circular polarization. SAR uses linear polarization on both the transmitter and the receiver with either horizontal (H) or vertical (V) polarization direction to the plane of wave propagation [78]. Various polarizations available from the SAR techniques have a different sensitivity to the target physical properties,

therefore providing unique additional information. Various SAR instruments can observe the surface in different polarimetric modes. There are three main modes, namely Single Polarization (HH or VV) that transmits and receives using a single channel, Dual Polarization (HH + HV or VV + VH) that utilize two polarization channels, and Quad-Polarization (HH, HV, VH, and VV) that exploits the full spectrum of linear polarization capabilities by transmitting and receiving in two polarization channels. Furthermore, several satellite missions, namely RADARSAT Constellation Mission (RSM), NASA-ISRO SAR Mission (NISAR), and Radar Imaging Satellite-1 (RISAT-1) use hybrid compact polarimetry that transmits a circularly polarized wave and receives H and V. It combines the desirable properties of Dual Polarization mode, e.g., allowing for better discrimination between different types of targets, while better balancing the power between the receive channels.

Paper II shows the performance comparison of applying sensors with different frequencies and wavelengths for better sea ice characterization and classification. Moreover, it demonstrates the advantages and limitations of each sensor when used separately and simultaneously.

Table 5.3: Frequency bands and wavelengths used for radar remote sensing. Note, that f corresponds to frequency and λ refers to wavelength. The cyan color indicates the most commonly used frequencies in Earth Observation disciplines.

Band	f **[GHz]**	λ **[cm]**
VHF	0.03-0.3	100-1000
UHF	0.3-1	30-100
L	1-2	15-30
S	2-4	7.5-15
C	4-8	3.75-7.5
X	8-12	2.5-3.75
Ku	12-18	1.67-2.5
K	18-27	1.11-1.67
Ka	27-40	0.75-1.11

5.2.2   Speckle and Noise

At full spatial resolution, SAR images reveal a grainy "salt and pepper" variation of their intensity, which is a multiplicative noise that complicates image interpretation. This effect is called speckle and it is unavoidable in SAR imaging systems. Speckle is created by constructive and destructive interference of the energy reflected from a number of separate small scatterers that are located within a single-resolution cell and resulting in a granular appearance in the image [77]. Speckle is one of the reasons why the SAR image is so challenging to interpret since sometimes it significantly degrades the image quality. There are some techniques to remove or reduce the speckle influence, however, at a cost of radiometric and spatial resolution. The most commonly used is multi-looking which involves dividing the raw SAR data into a series of sub-apertures and averaging the data from each sub-aperture to create a new image. The number of looks used for multi-looking is generally chosen depending on the desired level of image quality. The individual looks can be significantly affected by noise and other artifacts, however, averaging the multiple looks, results in reduced noise and enhanced underlying signal, but at the expense of a decreased spatial resolution. Another simple and computationally efficient way to reduce speckle is a boxcar filter that averages the pixel values within a fixed-size window. Figure 5.4 illustrates the example of speckle on Sentinel-1 SAR image before (a) and after (b) speckle reduction using a boxcar filter with 9×9 pixels.

<div style="text-align:center">(a)   (b)</div>
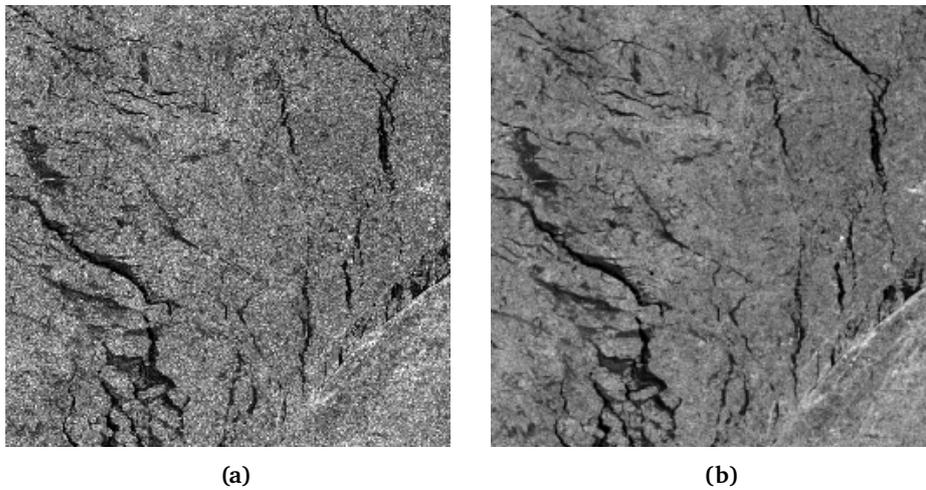
Figure 5.4: Example of speckle on Sentinel-1 SAR image: original image before reduction (a), after reduction, using boxcar filter with 9×9 window (b).

Another phenomenon that can significantly deteriorate both the manual and automatic interpretability of SAR images is thermal noise. It is a strong additive background noise that defines the sensor´s noise floor [84]. There are some techniques that reduce noise in the image in the pre-processing phase, however, sometimes they are insufficient. Therefore, there is a clear necessity for developing robust noise correction. Figure 5.5 illustrates the example of thermal noise on Sentinel-1 SAR image before (A) and after (B) correction available in the Sentinel Application Platform (SNAP). It is visually evident that most of the thermal noise was removed by the algorithm, however, there are still some noise residuals left especially in the first subswath area where the influence of noise is stronger.

<div style="text-align:center">(a)   (b)</div>
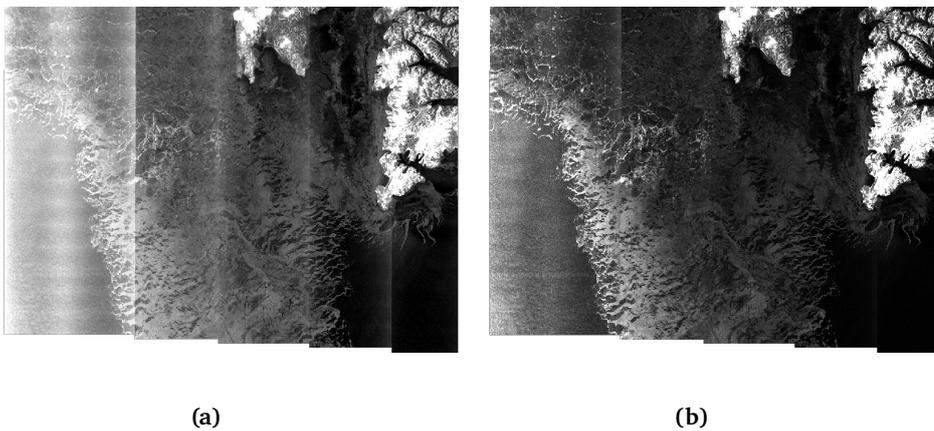
Figure 5.5: Example of thermal noise on Sentinel-1 SAR image: original image before correction (a), corrected image (b).

It is worth mentioning, that properties of the illuminated object influence the radar backscattering such that the image analysis gets more difficult, e.g. varying soil moisture or large-scale surface roughness in land applications [77]. The target properties have a strong influence on the information that can be retrieved from the radar image and their individual contributions

to the signal. Therefore, they need to be considered in the image interpretation.

5.3 Passive Microwave Radiometers

Another type of sensor that is commonly used for various applications is the Passive Microwave Radiometer (PMR). It provides important information such as sea surface characteristics, wind speed above the surface, atmospheric water vapor, cloud liquid water, and rain rate. These observations have led to significant advances in Earth observation research as well as significant improvements in monitoring and forecasting of weather and climate [85]. Moreover, they provided a most comprehensive large-scale characterization of the global sea ice cover, which is of crucial importance for the topic of this thesis [86]. Table 5.4 demonstrates a few of the currently operating PMR sensors.

(a)   (b)
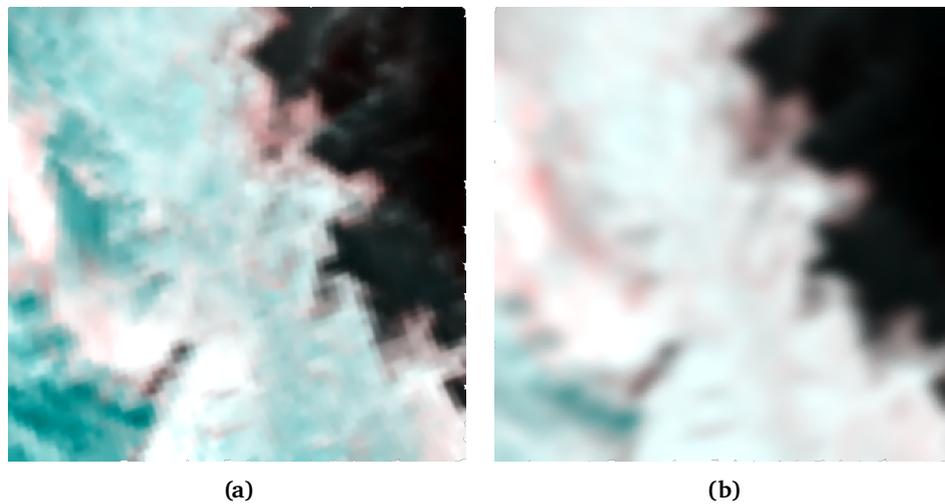
Figure 5.6: Example of false-color composite (H, V, V as RGB) PMR data obtained from AMSR-2 for different frequencies: 37 GHz (a) and 19 GHz (b). Spatial coverage is 400 km (cropped to the Sentinel-1 scene). Spatial resolution is 10 km for 37GHz and 24 km for 19GHz, respectively.

Passive Microwave Radiometers measure the emitted radiance from different surfaces, which is referred to as brightness temperature. Brightness temperature is related to the emissivity and the physical temperature of an object. The observed material can emit microwave energy differently, depending on the atomic composition and crystalline structure. Moreover, the amount of microwave energy emitted is a function of an object's dielectric constant [77]. The differences and fluctuations in microwave emissions allow for the detection and measuring of various objects, such as soil or sea ice. Figure 5.6 shows an example of a PMR image obtained from AMSR-2 for different frequencies: 37 GHz (a) and 19 GHz (b). It is evident how the resolution and the information content of the two PMR bands differ depending on the wavelength.

As mentioned above, PMR sensors are commonly used for large-scale Earth observations. These sensors operate at multiple frequencies, each of which has different spatial resolutions and sensitivity to atmospheric parameters, in particular water vapor [25]. However, even the finest spatial resolution achievable with passive microwave sensors is significantly coarser than in the

Table 5.4: Examples of passive microwave radiometers that are currently operating. Spatial resolution and swath width are shown in kilometers. The spatial resolution varies depending on the frequency, therefore, the values show the range from the least to the largest possible.

Passive Microwave Radiometers			
Sensor	**SSM/I**	**AMSR-E**	**AMSR-2**
Frequency	19.3-85.5	6.9-89.0	6.9-89.0
Channels	7	12	13
Launched	1987	2002	2012
Country	USA	USA	USA
Swath Width	1400	1450	1450
Spatial Resolution	25-50	5-50	5-50

case of other sensors used throughout this thesis, such as SAR and optical, being just over 3 kilometers. These higher spatial resolutions are achieved at higher frequencies of around 90 GHz, however, this comes with greater susceptibility to atmospheric noise. Lower frequencies reduce this, but at the expense of much reduced spatial resolutions of 10's of kilometers. The main benefit of passive microwave radiometers is that their orbit and wide swaths allow daily coverage over most of the polar regions and monitor sea ice at synoptic scales, which is crucial for the topic of this thesis, namely sea ice observations [87].

In *Paper III* we use the AMSR-2 PMR sensor along with Sentinel-1 SAR. Specifically, we test the potential of employing PMR as a complementary source for accurate sea ice classification and characterization despite its coarse resolution. Moreover, we compare the performance difference between single-case scenarios of SAR and PMR.

/6

Multimodal Remote Sensing of Sea Ice

6.1   Challenges of Multimodality

Regular sea ice monitoring over larger regions became possible in the late 1970s using image data from satellites. In general, there are several sensors that are widely used for sea ice classification, ice charting, and climatological sea ice monitoring. Nevertheless, there are other sensors that are not commonly used, however, still can be potentially complementary for specific tasks related to sea ice. Each of these sensors has some limitations and advantages. Moreover, various sensors are often operating at different frequencies and/or spatial resolutions and coverage and with different time intervals between data acquisitions over a given region. Thus, this diversity of remote sensing systems can grasp various properties and aspects of the underlying surface, in our case, sea ice, by using different physical principles. The joint use of the information provided by multiple sensors allows for gaining more knowledge regarding the area of interest as well as a better characterization of sea ice by providing complementary information [2]. Figure 6.1 demonstrates a few examples of overlapping areas for SAR and optical data obtained for several cases: Radarsat-2 and Landsat 8 (a), Sentinel-1 and Sentinel-2 (b,c), and Sentinel-1 and Landsat 8 (d). The maps display visually how different the information is provided by various sensors.

Nevertheless, the combination of multimodal data sets raises several crucial challenges, that are especially important for sea ice analysis.

- **Heterogeneity of Data:** the diverse variety of sensors that can potentially be used for sea ice characterization and monitoring are very different in terms of various characteristics, such as temporal, spatial, and radiometric resolutions, sizes, data types, and units of measurement. That results in a problem of combining the data with different domains, which

(a) (b)

(c) (d)
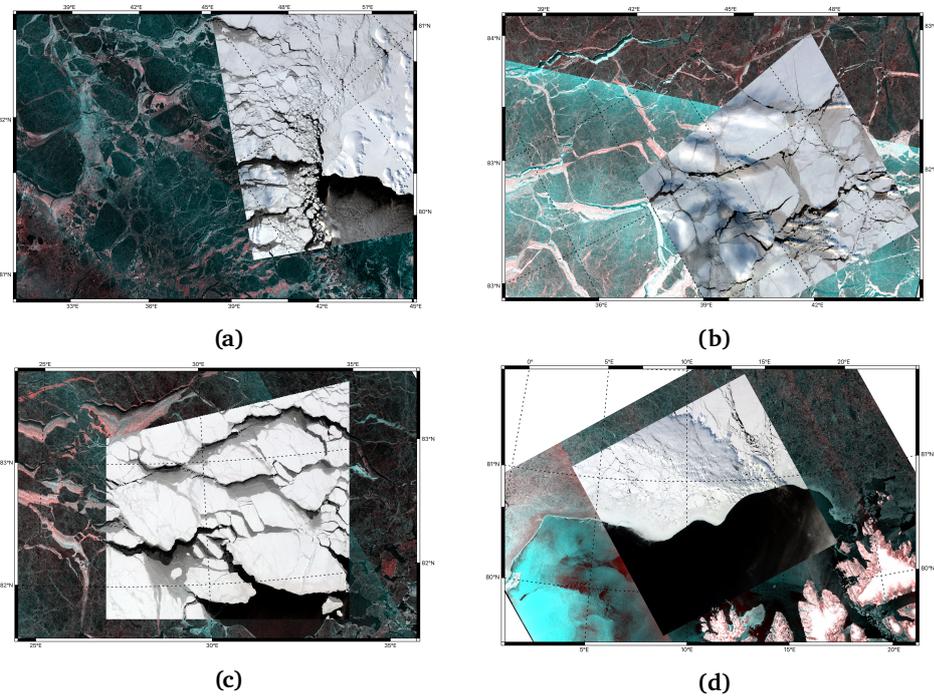
Figure 6.1: Examples of overlapping areas for SAR and optical data obtained for several cases: Radarsat-2 and Landsat 8 (a), Sentinel-1 and Sentinel-2 (b,c), and Sentinel-1 and Landsat 8 (d).

subsequently might lead to the loss of interpretability, which is crucial for sea ice analysis.

- **Complexity of Analysis:** by expanding the size of a data set, we are simultaneously increasing the complexity of the records to be analyzed. This is especially crucial for multimodal remote sensing data taking into account the large size of provided images.

- **Data Availability:** in the case of sea ice monitoring, which mostly occurs in polar regions, not all different types of satellite data are always available. Some sensors, like optical ones, are limited due to their sensitivity to weather and light conditions. Passive microwave radiometers, depending on the frequency, can be sensitive to atmospheric parameters, in particular to cloud liquid water and atmospheric water vapor. Moreover, some parts of the acquired scenes can simply be corrupted due to technical reasons, such as thermal noise in synthetic aperture radar.

- **Temporal Lag:** nevertheless, the sensors should not only be available but also overlapping in terms of time. This is especially important for sea ice applications since the majority of sea ice cover is constantly drifting. Therefore, depending on the region and application, only short-time lags, which are less than a few hours, are suitable for subsequent work.

- **Data Interpretation:** interpreting multimodal data can be a challenging task since different sensors may provide conflicting or ambiguous information. Moreover, the determination of the importance of modalities, i.e. which sensor provides the most accurate or reliable information, may be also difficult.

Moreover, there are some sensor-related challenges, that are especially crucial for sea ice monitoring, that should also be mentioned.

- **Optical Sensors:** the main limitation of such type of data is the high sensitivity to atmosphere and light conditions, which is a crucial problem since dense cloud covers and long periods of darkness prevail in the polar regions for several months of the year making it impossible to obtain any useful information. Moreover, it is quite challenging to discriminate snow-covered sea ice types using optical data.

- **Synthetic Aperture Radar:** even though SAR systems are widely used for sea ice monitoring, they have a lot of significant challenges. First and foremost, speckle salt-and-pepper noise and thermal noise can deteriorate the performance of the algorithms and the interpretability of the images. Moreover, seasonality significantly affects the dielectric characteristics of sea ice and snow cover on top of it. Therefore, the same sea ice classes might look very different in terms of intensities for the summer and winter periods, which should be taken into account when performing various sea ice applications. Another significant challenge is the automatic and manual interpretation of SAR data due to the complex relationship between radar backscatter and sea ice surface and volume properties which strongly relies on the knowledge of sea ice experts [88].

- **Passive Microwave Radiometers:** the main limitation in comparison to optical and SAR sensors is a very coarse spatial resolution. Even the finest spatial resolution achievable with passive microwave sensors is about 3 km and hence significantly coarser than in the case of SAR or optical, which means that the information about the smaller sea ice structures and formations is lost. Moreover, as well as optical sensors, PMR can be sensitive to atmospheric parameters, in particular to cloud liquid water and atmospheric water vapor depending on the used frequency [89]. The finest spatial resolution is achieved at frequencies of around 90 GHz which, however, comes with a greater susceptibility to atmospheric noise. This effect is less severe at lower frequencies at the expense of much reduced spatial resolutions at tens of kilometers.

The aforementioned challenges can either completely prevent or significantly deteriorate further analysis. Furthermore, due to these limitations, some of the considered algorithms can potentially fail to capture the data's underlying structure, i.e., not achieve an accurate and robust characterization of the physical phenomena occurring on Earth's surface. However, simply combining all the available sources without properly addressing the complementarity that they may possess is not always a good solution. Therefore, adding more sources of information does not surely lead us to the best algorithms performance since some of the available information can be redundant, corrupted, or unnecessary for the particular task [1, 2]. Moreover, increasing the number of modalities without properly addressing an investigation of their relevance and reliability may even significantly deteriorate the analysis. This, in fact, can be a strong limiting factor to the use of multimodal remote sensing data analysis for practical scenarios or operational frameworks within private and public sectors. Nevertheless, that leads us to the importance of developing and applying automatic dimensionality reduction algorithms as a pre-processing to select the relevant information from various sensors for improving the knowledge about an observed area.

6.2   Sensors Complementarity

Sea ice can be potentially monitored by various sensors and satellite techniques, such as optical, synthetic aperture radar, and passive microwave sensors. Despite the limitations mentioned above each of these sensors has some crucial advantages that will be briefly described in the following paragraphs.

- **Optical Sensors** provide images in visible and infrared channels of the electromagnetic spectrum. The most obvious benefit of such images is that they are easy to interpret visually, especially in comparison to other aforementioned sensors. Moreover, they are usually provided in a high spatial resolution, which allows identifying smaller sea ice formations and structures on the surface. Furthermore, in the case of sea ice optical data is commonly used in order to properly label and distinguish different sea ice classes as well as to separate open water from sea ice.

- **Synthetic Aperture Radar** on the other hand combines high spatial resolution and absolute independence of cloud and light conditions, which makes it one of the main sources for operational sea ice monitoring and characterization. The high spatial resolution of SAR allows identifying small-scale features of sea ice, such as leads, ridges, melt ponds, and floes. Moreover, it provides images with different polarization modes, which can provide complementary information about the characteristics of various sea ice types. Furthermore, in comparison to optical sensors, SAR is capable to discriminate snow-covered ice types (under freezing conditions) since the radar signal penetrates dry snow.

- **Passive Microwave Radiometers** are another type of sensor that is commonly used for large-scale sea ice observations. Like SAR, passive microwave radiometers can operate in almost all weather conditions, including darkness, fog, and clouds, especially at the lower frequencies. These sensors operate at multiple frequencies, each of which has different spatial resolutions and can separate water and ice and hence provide information about ice concentration and the position of the sea ice edge. At the same time, the single channels are to varying degrees affected by atmospheric parameters. Overall, even though PMR is unable to obtain the same detailed information as we have from SAR, it still can be used as a complementary source. The crucial benefit of PMR is that their wide swaths allow daily coverage over most of the ice-covered polar regions which are extremely useful for monitoring sea ice at synoptic scales [87]. Furthermore, the regular monitoring of ice-covered areas also allows to retrieve the large-scale (regional) patterns of ice drift.

- **Other Sensors** can potentially be used for various sea ice applications. Among them are laser and radar altimeters which are instruments that measure the height of the surface and allow us to estimate the sea ice thickness at an ocean scale. The principle consists of measuring the height of both sea ice floes and leads or polynyas and after that calculating the height difference between them, which allows determining the freeboard [90]. Knowing the density of ice and the water, as well as considering the snow load, we can estimate the sea ice thickness using the hydrologic equilibrium equation. Another source of information about sea ice comes from thermal sensors, which measure the thermal radiation emitted by the sea ice surface that has a different thermal signature than the surrounding water and land, so it can be further identified. Furthermore, thermal sensors are useful to separate thin and thick ice. Moreover, thermal sensors can be used during the polar night, however, only during cloud-free conditions.

(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g) **(h)** **(i)**
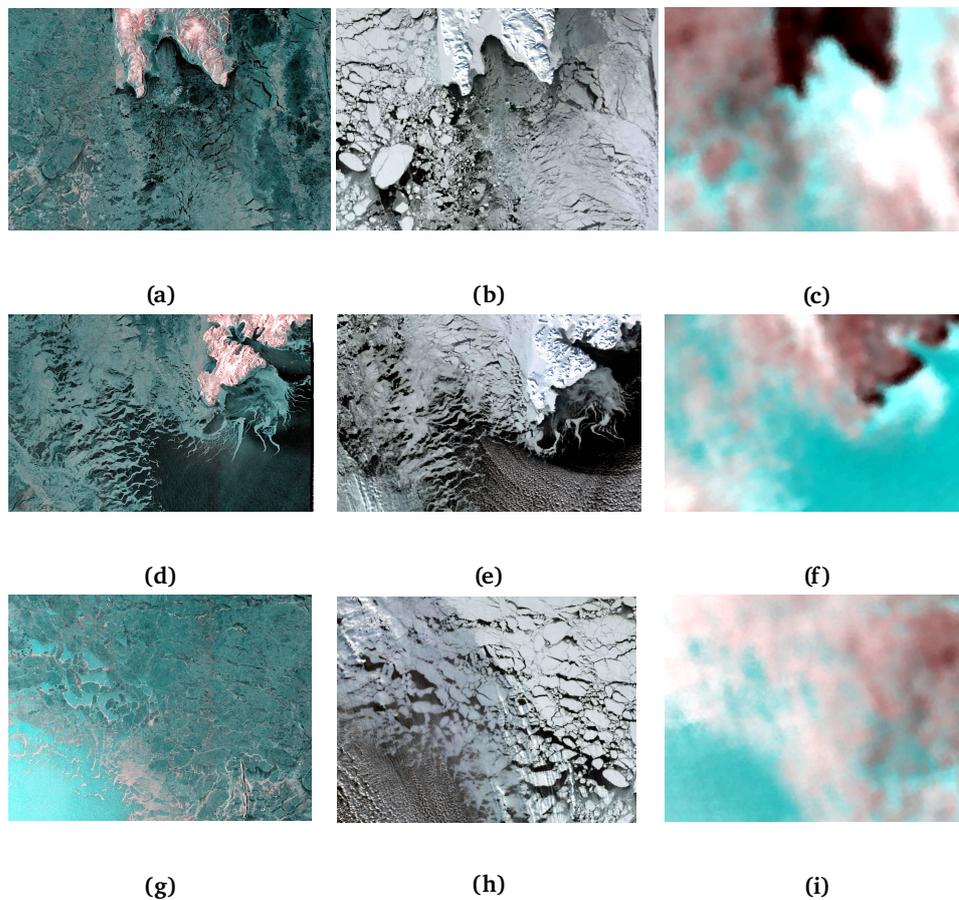
Figure 6.2: Color representation of the cropped overlapping area for Sentinel-1/Sentinel-3/AMSR2 multisensor case: SAR false-color composite (HV, HH, and HH as RGB) (a,d,g), optical natural color composite (RGB) (b,e,h), passive microwave false-color composite (H, V, and V as RGB) (c,f,i). The time gap between SAR and PMR scenes was a few minutes, while the optical scene was acquired a few hours later.

Figure 6.2 demonstrates a few examples of sensor complementarity for sea ice applications, including color representation of cropped overlapping area obtained from SAR Sentinel-1, optical Sentinel-3, and passive microwave AMSR2 (Advanced Microwave Scanning Radiometer). The scenes perfectly illustrate the limitations and advantages of each sensor. An optical sensor usually provides a higher or similar resolution, than SAR, however, in the case of Sentinel-3, the resolution is significantly coarser. It contains clouds that can significantly corrupt some parts of the image. Moreover, it is visible that the optical image is performing better in distinguishing younger ice types, such as nilas, and separating open water from the sea ice. However, it is often unable to provide enough information regarding older sea ice types because of the snow cover on the surface, while SAR provides an enormous amount of information about the older ice types due to the deeper signal penetration. However, it should be noted that the radar penetration depth is much reduced during melting, therefore, the separation of snow-covered ice types is much easier under dry freezing conditions. It is evident from the image, that SAR provides more information about mixed sea ice classes in the areas where the optical sensor fails. However,

SAR data are still harder to evaluate visually, especially due to the thermal and speckle noise, and the interpretation of images requires expert knowledge. Passive microwave images have a much coarser resolution, nevertheless, they are still providing relevant information about sea ice. Nevertheless, each sensor provides unique different information that can significantly improve the algorithm performance and our knowledge regarding the region of interest.

/7

Overview of Publications

7.1 Paper Summaries

Paper I

E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni. "**A Multimodal Feature Selection Method for Remote Sensing Data Analysis Based on Double Graph Laplacian Diagonalization**", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Volume: **14**, 11546 – 11566, November 2021.

This paper introduces a new information selection approach for multimodal remote sensing data sets (Figure 7.1). The proposed approach relies on a representation based on graph Laplacians and information theory metrics. It should be noted that this method contains several novelties that distinguish it from other dimensionality reduction algorithms, even ones that are also based on graph theory. While the existing works using graph Laplacians exploit the attributes' structure using kernels as similarity measures, we additionally consider the attribute's information content. Therefore, we simultaneously exploit two metrics, mutual information (MI), and the Gaussian kernel (GK) as similarity measures in order to capture the most relevant attributes within the data sets. Moreover, two similarity measures not only provide various information about the original data set but are also applied at different detail levels. The MI is used globally, considering the pixels within the images to ensure a better estimation of the attributes shared information. On the other hand, the GK is employed locally to preserve the particularity of homogeneous areas within the images. Accordingly, different attributes are selected for different parts of an image that might belong to different classes and measured under different conditions (i.e., different noise levels, cloud coverage, etc.).

The main motivation behind this study is to develop a method that can improve the classic graph-Laplacian-based approaches in the separation of the attributes when they are heterogeneous. We are tackling this limitation with the joint employment of the two metrics simultaneously,
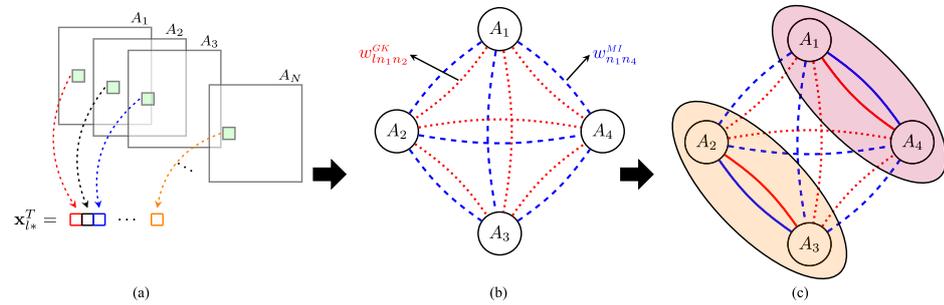
Figure 7.1: Flowchart of the multimodal information selection approach proposed in *Paper I*. (a) Attributes of the l-th pixel are stacked in one vector $\mathbf{x}_{l*}$. (b) Graph of four attributes with two similarity functions at the l-th pixel. (c) Graph of four attributes with two similarity functions at the l-th pixel. Similar attributes are grouped together.

and at different scales that ensures a better separability of the attributes and, hence, a more precise selection. Accordingly, the proposed approach guarantees high accuracy of the analysis so that the potential of multimodal remote sensing data analysis can be exploited in multiple applications.

Our new information selection method can be applied for different multimodal remote sensing data combinations. We specifically tested it on two LiDAR and hyperspectral data combinations acquired over the University of Houston campus and agricultural area in the south part of the city of Trento and two optical data combinations acquired over the cities of Berlin and Paris from Sentinel-2 and Landsat-8 satellites. It should be noted that in this study we are not using any sea ice scenes since they require proper labeling which is a challenging task. Therefore, as a first step towards the sea ice application, we tested and tuned our proposed method on publicly available urban scenes that have accurate ground truth labels. In order to evaluate the performance of the proposed method, we report how the proposed information selection algorithm can be used in order to enhance the performance of classification. Moreover, we properly tuned various important parameters and compared the proposed algorithm with other dimensionality reduction techniques. The experimental results obtained from several multimodal data sets consistently demonstrated the effectiveness and robustness of the proposed method for the processing of multimodal remote sensing data sets.

Paper II

E. Khachatrian, S. Chlaily, T. Eltoft, W. Dierking, F. Dinessen, and A. Marinoni. "**Automatic Selection of Relevant Attributes for Multi-Sensor Remote Sensing Analysis: A Case Study on Sea Ice Classification**", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Volume: **14**, 9025 – 9037, July 2021.

In this study, we provide a new scheme to obtain information on sea ice conditions by remote sensing data analysis (Figure 7.2). In fact, combining information obtained from different image modes or from various sensors (using for example different spectral ranges and spatial resolutions) is instrumental in order to achieve advanced characterization of sea ice and automatic sea ice monitoring. However, when analyzing several images with both complementary and redundant characteristics, it is necessary to select an optimal set of image attributes that provides the relevant information content to enhance the efficiency and accuracy of the system.
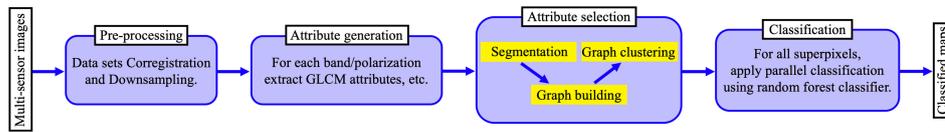
Figure 7.2: Flowchart of the proposed scheme tested in *Paper II* and improved in *Paper III* including Pre-Processing, Attribute Generation, Attribute Selection, and Classification steps.

While *Paper I* was introducing the method along with its main novelties for the multimodal data combinations, *Paper II* focuses on employing the recently developed fully automatic and efficient graph-based information selection method to capture the most relevant attributes for better separation and characterization of different sea ice classes. The main motivation of this paper is to introduce the information selection as a pre-processing step of the classification schemes for sea ice applications, show the flexibility of the method, and compare the performance of different sensors separately and simultaneously in order to distinguish their limitations and advantages. Moreover, one of the main ideas was to check if it is possible to generalize the information selected for different sea ice classes by comparing the attributes that were chosen as relevant by our method.

It should be mentioned, that the focus of this paper is on using different frequencies/wavelengths by combining optical and SAR images that do not differ significantly from the spatial resolution point of view. Therefore, we combined a multi-frequency data set from an airborne SAR measurement campaign (ICESAR), and two satellite data sets, namely Radarsat-2 and Landsat-8, and Sentinel-1 and Sentinel-2.

The results indicate the flexibility of the approach to process different combinations of data sets for different sea ice conditions, and the importance of deploying multi-sensors for the characterization of ice types, by comparing the performance of using several sensors separately and simultaneously. Experiments demonstrate that the attributes selected by our method result in high classification accuracies. Moreover, we discovered, that it is hard to generalize the attribute relevance for different sea ice classes since they not only depend on sensor types and imaging modes, but also on local environmental and sea ice conditions during the data acquisition.

Paper III

E. Khachatrian, W. Dierking, S. Chlaily, T. Eltoft, F. Dinessen, N. Hughes, and A. Marinoni. "**SAR and Passive Microwave Fusion Scheme: a Test Case on Sentinel-1/AMSR-2 Data Sets for Sea Ice Classification**", *AGU Geophysical Research Letters*, Volume: **50(4)**, February 2023.

This work was a direct and logical continuation of *Paper II*. Here we demonstrate a flexible, highly accurate, efficient, and adaptive multi-sensor fusion scheme applied to SAR and PMR data combination, i.e. two most commonly used remote sensing sensors for ice charting and sea ice monitoring (Figure 7.3). In comparison to the previous paper, in *Paper III* we combine sensors that differ significantly in spatial resolution as well as in the measured quantities, which makes it challenging to properly combine them. Nevertheless, each of these sensors provides different complementary information that can be relevant for sea ice characterization. Moreover, these sensors are not hindered by cloud and light conditions as in *Paper II*, where we use SAR and optical data combinations.

In this paper, we investigate the potential of combining SAR and PMR data for sea ice classification, which we demonstrate for a test case using a combination of data from the Sentinel-1 C-band (5.405 GHz) SAR, and the 89 GHz channel of the Advanced Microwave Scanning Radiometer 2 (AMSR-2). Accordingly, we illustrate both the advantages and limitations of applying each sensor separately and simultaneously. The results illustrate, that there is a huge potential in PMR and SAR combination, and a competent fusion of advantages of such data can lead to better classification performance from both accuracy and time consumption point of view.

(a)   (b)   (c)   (d)
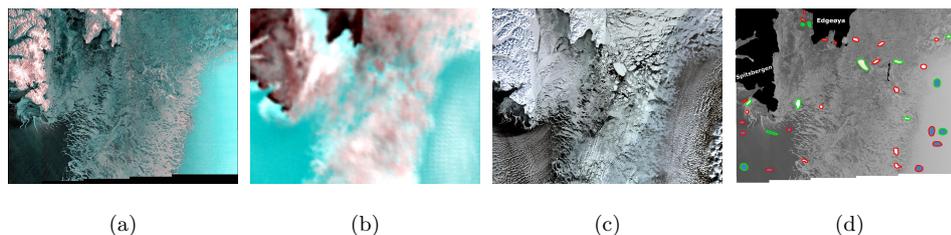
Figure 7.3: Color representation of the data set used in *Paper III*: (a) false-color composite SAR (HV, HH, and HH as RGB), (b) PMR (H, V, and V as RGB), (c) natural-color composite of an optical image from Sentinel-3 OLCI (Bands 8, 6, 4), and (d) spatial distribution of ROIs that were used for training (polygons with green boundaries) and testing (red boundaries); here the blue color refers to Open Water (OW), white color corresponds to Brash Ice (BI), cyan illustrates the young Grey Ice (GI), pink color shows the Thin First-Year Ice (TFYI), and black color indicates the landmask for Svalbard.

The experimental results of *Paper III* provide solid ground for future research that will be investigating the potential of applying the proposed scheme for operational sea ice monitoring. We are planning to evaluate the performance of the proposed fusion scheme on a larger number of Sentinel-1/AMSR2 data sets gathered over a full season and with more complex ice conditions.

7.2 Other Scientific Contributions

The following section includes other scientific contributions, such as journal and conference papers. From all the scientific contributions that are not crucial parts of the thesis, I would like to highlight a scientific paper published in IEEE Geoscience and Remote Sensing Letters. While the conference papers demonstrate the shorter versions of the journal papers on the proposed information selection method, with less description and results, this paper not only contains some of the topics that are crucial for this study, namely multimodality and sea ice but also provides a slightly different application where some of the techniques used before can be employed. The main focus of this article is to explore the potential of applying various sensors for submesoscale/mesoscale eddy monitoring in the marginal ice zone (MIZ).

Note, that various symbols correspond to different types of contributions, such as ✧ referring to journal papers, ◆ to a conference paper, and ❖ to other contributions.

As First Author

✧ E. Khachatrian and N. Sandalyuk. "**On the Exploitation of Multimodal Remote Sensing Data Combination for Mesoscale/Submesoscale Eddy Detection in the Marginal Ice Zone**", *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022, Art no. 3513805, doi: 10.1109/LGRS.2022.3215202.

◆ E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni. "**Unsupervised Information Selection In Multimodal Sea Ice Remote Sensing**", PROCEEDINGS OF THE 25th INTERNATIONAL SYMPOSIUM ON ICE Trondheim, Norway, 23rd – 25th November 2020. The International Association for Hydro-Environment Engineering and Research (IAHR) 2020 ISBN 978-82-7598-120-0. s. 208-217.

◆ E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni. "**Selecting Principal Attributes in Multimodal Remote Sensing for Sea Ice Characterization**", EUSAR 2021: 13th European Conference on Synthetic Aperture Radar 29 March – 1 April 2021, VDE Verlag GmbH 2021 ISBN 978-3-8007-5457-1. s. 531-536.

◆ E. Khachatrian, S. Chlaily, T. Eltoft, P. Gamba, and A. Marinoni. "**Unsupervised Band Selection for Hyperspectral Data Sets by Double Graph Laplacian Diagonalization**", International Geoscience and Remote Sensing Symposium (IGARSS), 2021.

❖ E. Khachatrian, S. Chlaily, and A. Marinoni. "**Automatic Information Selection in Multimodal Remote Sensing**", CIRFA Annual Conference, 2019.

❖ E. Khachatrian, S. Chlaily, T. Eltoft, W. Dierking, F. Dinessen, and A. Marinoni. "**Automatic Selection of Relevant Attributes for Multi-Sensor Remote Sensing Analysis: A Case Study on Sea Ice Classification**", CIRFA Annual Conference, 2021.

As Co-Author

◆ C. Taelman, S. Chlaily, E. Khachatrian, F. van der Sommen, and A. Marinoni. "**On the Exploitation of Heterophily in Graph-Based Multimodal Remote Sensing Data Analysis**", CDCEO 2021: 1st Workshop on Complex Data Challenges in Earth Observation, 2022.

❖ S. Chlaily, E. Khachatrian, T. Eltoft, and A. Marinoni. "**A Multisensor Information Selection for an Accurate and Efficient Sea-Ice Classification**", Arctic Frontiers, 2020.

8

Paper 1: A Multimodal Feature Selection Method for Remote Sensing Data Analysis Based on Double Graph Laplacian Diagonalization

Eduard Khachatrian, Saloua Chlaily, Torbjørn Eltoft, and Andrea Marinoni

A Multimodal Feature Selection Method for Remote Sensing Data Analysis Based on Double Graph Laplacian Diagonalization

Eduard Khachatrian ⓘ, *Student Member, IEEE*, Saloua Chlaily ⓘ, *Member, IEEE*, Torbjørn Eltoft ⓘ, *Member, IEEE*, and Andrea Marinoni ⓘ, *Senior Member, IEEE*

Abstract—**When dealing with multivariate remotely sensed records collected by multiple sensors, an accurate selection of information at the data, feature, or decision level is instrumental in improving the scenes' characterization. This will also enhance the system's efficiency and provide more details on modeling the physical phenomena occurring on the Earth's surface. In this article, we introduce a flexible and efficient method based on graph Laplacians for information selection at different levels of data fusion. The proposed approach combines data structure and information content to address the limitations of existing graph-Laplacian-based methods in dealing with heterogeneous datasets. Moreover, it adapts the selection to each homogenous area of the considered images according to their underlying properties. Experimental tests carried out on several multivariate remote sensing datasets show the consistency of the proposed approach.**

Index Terms—**Gaussian kernel (GK), graph Laplacians, multimodal remote sensing, mutual information (MI), unsupervised information selection.**

I. Introduction

OVER the past several decades, satellite imagery has become a crucial source in providing a vast amount of information about the Earth's surface. Thanks to technological advances, a region of interest (ROI) can be monitored by various sensors characterized by different acquisition techniques (*modalities*), using different spectral, temporal, or spatial resolutions [1]. The information provided by multiple sensors grasps different aspects of the area of interest. For instance, hyperspectral images might reveal the material content of the observed region, while synthetic aperture radar (SAR) complements the capabilities of optical imaging by reporting the topographic (interferometry) and surface roughness information, and the

light detection and ranging (LiDAR) technology provides highly accurate measurements of the vertical height of structures. Accordingly, robust characterization of the Earth's surface can be achieved by combining data coming from different modalities to obtain useful insight into various aspects of the underlying surface [2].

The combination of multimodal datasets raises several challenges [1], [2]. These challenges are limited not only to dealing with the heterogeneity of the multimodal images in terms of temporal, spatial, and radiometric resolutions, sizes, and data types [2], but also to accurately selecting the relevant information that maximizes the benefits of the multimodal analysis. By expanding the size of a dataset, we are simultaneously increasing the complexity of the records to be analyzed, especially when it is multimodal. Hence, the considered algorithms might fail to capture the data's underlying structure, i.e., not achieving an accurate and robust characterization of the physical phenomena occurring on Earth's surface. Indeed, it has been shown that increasing the number of modalities without properly addressing an investigation of the significance and reliability of the data may deteriorate the analysis [3], [4]. This may, therefore, represent a strong limiting factor to the use of multimodal remote sensing data analysis in practical scenarios, as well as to its actual impact in operational frameworks within private and public sectors [1].

In fact, not all information provided by several sensors is valuable; it can be redundant, corrupted, or unnecessary for the given task [1], [2]. Therefore, to get the most use of a multimodal dataset, it is crucial to select only relevant information. In this way, it is expected that the accuracy of the analysis will increase, whereas the time complexity will be reduced. Consequently, to improve the knowledge about an observed area, there is a need to develop an automatic method to select the relevant information from various sensors [1]–[3].

Dimensionality reduction has been proven as an effective tool to tackle these issues in remote sensing data analysis [2], [3], [5]–[7]. Feature extraction and feature selection methods are able to strongly enhance the quality of understanding and assessment of physical–chemical phenomena on the ground, especially when data collected by means of homogeneous acquisition techniques (i.e., sensors with similar properties of the sensing devices) are analyzed. Nonetheless, traditional methods

The authors are with the Department of Physics and Technology, University of Tromsø—The Arctic University of Norway, N-9037 Tromsø, Norway (e-mail: eduard.khachatrian@uit.no; saloua.chlaily@uit.no; torbjorn.eltoft@uit.no; andrea.marinoni@uit.no).

for dimensionality reduction might fail in capturing the details of elements, materials, and dynamic phenomena on Earth's surface when multimodal datasets are explored [2], [3].

We note that the term *feature* is commonly used in different fields such as classification methodologies, pattern recognition, and texture analysis. However, in our work, to prevent confusion with textural features, we introduce the notation *attribute* from information theory, which refers to directly measured quantities as, e.g., optical/hyperspectral/LiDAR reflectance across the electromagnetic spectrum, and additional parameters such as textural features.

In the case of multimodal datasets, which reside on a nonlinear manifold, graphs are the appropriate representation of the data. The graph is composed of the multimodal attributes as nodes, and their similarities will give the weights to their connecting edges. The dimensionality reduction is carried out by determining similar nodes and picking a representative attribute from each group. The graph partition reveals the pattern of the attributes; as such, the chosen attributes preserve the structure of the graph. The problem of graph partitioning or clustering to group similar nodes is nondeterministic polynomial-time hard (NP-hard), but it can be approximated via several techniques, such as spectral clustering (SC) [8]. In fact, the graph structure can be understood and analyzed via the Laplacian of the adjacency matrix that summarizes the nodes' similarities. In particular, the eigenvectors of the Laplacian matrix, associated with the lowest eigenvalues, reveal the structure information of the graph [9]. However, in the case of graphs of heterogeneous degrees, where the nodes interact differently, the graph's eigenvalues scatter across the spectrum. Accordingly, it will be hard to distinguish the lowest eigenvalues and determine the informative eigenvectors, which will undermine the attributes selection's pertinence and efficacy [10]–[12].

In this article, we introduce an approach to information selection in multimodal remote sensing datasets that relies on a representation based on graph Laplacians. While the existing works using graph Laplacians exploit the attributes' structure using kernels as similarity measures, we additionally consider the attribute's information content. As such, we jointly exploit mutual information (MI) and the Gaussian kernel (GK) similarity metrics to capture the most relevant attributes within the records. The two similarity measures are applied at different detail levels. The MI is used globally, considering all the pixels within the images to ensure a better estimation of the attributes' shared information. On the other hand, the GK is employed locally to preserve the particularity of homogeneous areas within the images. Accordingly, different attributes are selected for different parts of an image that might belong to different classes or be measured under different conditions (i.e., different noise levels, clouds coverage, etc.).

The main motivation of this work is, thus, the limitation of classic graph-Laplacian-based approaches at separating the attributes when they are heterogeneous, as it has been shown in [10]–[12]. Nevertheless, the joint employment of the MI and the GK at different scales ensures a better separability of the attributes and, hence, a more precise selection. Accordingly, the proposed approach guarantees high accuracy of the analysis and

reduces the computational complexity so that the potential of multimodal remote sensing data analysis can be exploited in multiple applications.

The rest of this article is organized as follows. Section II reports a brief summary of the main methods for information selection in remote sensing data analysis and the main contribution of the proposed approach. Section III provides details of the proposed architecture. Section IV presents an experimental validation of the proposed method. Finally, Section V concludes this article.

For notational convenience, random scalars are denoted by lowercase letters, e.g., z. Random vectors are designated by bold lowercase letters, e.g., $\mathbf{z}$. Bold uppercase letters refer to matrices, e.g., $\mathbf{A}$. $|\mathbf{A}|$ and $\text{Tr}(\mathbf{A})$ denote the determinant and trace of the matrix $\mathbf{A}$, respectively. $\text{diag}\{d_1, \ldots, d_N\}$ refers to a diagonal matrix whose diagonal elements are $d_1, \ldots, d_N$ starting from the upper left. The $\text{ddiag}(\mathbf{A})$ operator is set to zero the off-diagonal entries of $\mathbf{A}$.

II. Background and Motivation

A. Existing Work

In order to select the most informative subset of attributes and discard the irrelevant ones, it is possible to use several dimensionality reduction methods. Generally, dimensionality reduction methods can be separated into two main approaches: attribute extraction and attribute selection [13], [14].

1) *Attribute extraction* reduces the dimensionality by projecting the original data into a lower dimensional space [15], [16]. As such, the separability of the data is increased but at the expense of physical interpretability, which is essential in remote sensing analysis. Among the methods of attribute extraction, we may cite, for instance, principal component analysis (PCA) [17] and decision boundary feature extraction (DBFE) [5]. PCA converts a set of attributes of potentially correlated variables into a set of linearly uncorrelated variables, called principal components. It projects the original set into a lower dimensional space spanned by the principal eigenvectors of data's covariance matrix. Thus, it reduces the size of the original set while preserving its variance [17]. DBFE is a supervised approach that uses the training set to determine the decision boundary between classes. The eigenvectors of the decision boundary matrix determine the direction of projection of the original set of attributes. As such, it provides a minimum number of transformed attributes that achieve the same accuracy as the original set [5].

2) *Attribute selection* reduces dimensionality by selecting the most informative subset of records preserving the characteristics of the original data without working on a different space [18]. Attribute selection determines a subset of the original set that is more relevant according to some criteria, such as information, similarity, or correlation. The methods for attribute selection can be divided into three categories: ranking, searching, and clustering.

a) *Ranking methods* sort the attributes with respect to a given criterion and select the most significant elements [6], [19]. They are very efficient, but they might not be very precise because they do not consider the relationships among the attributes. Among this family of attribute selection methods, we can cite Fisher score for attribute selection (FIS). FIS is a supervised approach that selects the subset of attributes with a large Fisher score. The Fisher score measures the ability of each attribute to reduce the intraclass distance while increasing the interclass distance [20].

b) *Searching methods* select the optimal subset in an incremental, removal, or update manner using a search method, such as a genetic algorithm (GA) [21] or branch and bound [22]. This class of attribute selection algorithms is more accurate than ranking methods since it considers the interaction between the data. However, such methods are limited by the size of the searching space. In the case of large datasets, computation time significantly increases, and the searching methods fail to achieve optimal results. Here, we can highlight forward attribute selection (FS) [23], orthogonal branch and bound (OBB) [22], and GA [21]. FS determines the optimal subset in an incremental fashion. The algorithm starts with a minimum number of attributes, and with each new step, it adds one attribute that improves the accuracy until no further improvement is noticed [23]. The OBB is a backtracking attribute selection algorithm. It is based on the assumption that the adopted criterion function fulfills the monotonicity condition. Hence, it guarantees to find the optimal subset while omitting many attribute subset evaluations. The branching step consists of constructing the tree such that the subtree of each level is constructed by deleting one attribute until the required number of attributes is reached. The bounding step represents the process of traversing the tree to find the optimal subset [22]. The GA is an adaptive algorithm that finds the global optimum solution for an optimization problem based on the mechanics of natural genetics and biological evolution. GAs operate on a population of individuals to produce better approximations. In attribute selection, each individual in the population represents a predictive model with genes that correspond to the total number of attributes in the dataset. Genes are encoded as binary values that show if the attribute is included or not in the subset [21].

c) *Clustering methods* divide the components of the original set into different groups, and from each cluster, a representative element is selected to compose the optimal subset [24]. The approaches within this category can be further divided into three subcategories: k-means-based [24], [25], affinity propagation-based [26], and graph-based [27].

Among the various subcategories of clustering methods, graph-based clustering methods play a key role. The graph-clustering-based approaches find the relevant attributes by partitioning the graph into subgraphs (clusters) and selecting the representative attribute from each of them [27]. In this representation, the nodes would correspond to data points, while the edge between two nodes is weighted by their similarity. It is important to note that data representation through graphs has attractive characteristics since it enables grasping the local and global properties efficiently. This effect is obtained by the intrinsic ability of graph representation to naturally address local neighborhoods, paths, and global connectivity in its definition [28]–[30]. In this sense, a graph can enhance the characterization of complex manifolds, giving graph-based methods a key role in investigating realistic datasets. Moreover, it can help in reducing the computational complexity of data investigation [27], [30], [31].

When performing dimensionality reduction on graph structures, two main approaches can be addressed. On one side, graph-based clustering algorithms might work on similarities among the nodes according to specific criteria and metrics derived on the attributes associated with each vertex in the graph [27]. Methods belonging to this category (i.e., methods addressing *vertex similarity*) attempt to capture the global geometry of the overall dataset by constructing graphs based on measures of global connectivity of the ensuing graph. The intuition behind this algorithm is that random perturbations of the points in a high-dimensional space will induce changes in a nonhomogeneous fashion in different parts of the graph inducing the given dataset to show minimal global distance. Thus, depending on how globally important certain edges of the graph are, the algorithms working on vertex similarity will aim to capture the globally important edges in the perturbed ensemble [29], [31], [32].

A popular way to ensure such global connectivity addressing vertex similarity is through the minimum spanning tree (MST) approach [27], [33], [34]. The main step of the MST approach consists of determining the MST of the graph, which connects its vertices without cycles and with the minimum total edge weight. MST identifies the graph's cluster by removing the inconsistent edges according to a certain criteria [34]. MST-based approaches can, thus, capture the geometry of nonhomogeneously sampled data points in a high-dimensional space since the MST contains not only local but also global features of the dataset [32], [33], [35].

Another way of performing dimensionality reduction on graphs relies on identifying clusters to fulfill a specific target condition, i.e., a *fitness criterion* [27]. Several forms of fitness criteria have been proposed in technical literature, typically as a function of the density of the clusters to be detected and/or the amount of edges in the graph necessary to reach the maximum value for cliques in the induced subgraphs [27], [36], [37]. In this respect, community detection (CD) algorithms and methods based on dominant set (DS) search play a key role.

Let us consider CD schemes [27], [36], [38], [39]. In general, CD algorithms depend on the definition of the resolution parameter that leads to multiscale CD. Specifically, for small values of this resolution parameter, the number of detected communities is large, and the communities capture the graph's local information. As the resolution parameter becomes larger,

there are fewer communities, and the communities are able to capture the global features of the graph [38], [39]. For instance, Markov stability is a quality measure for CD, which adopts a dynamical perspective to unfold relevant structures in the graph at all scales as revealed by a diffusion process [29], [38], [39].

On the other hand, within fitness-criteria-based graph clustering approaches, DS clustering generalizes the problem of finding a maximal clique to edge-weighted graphs [27], [28], [37]. At each iteration, a DS is extracted, and its subsets of nodes are removed from the graph (this is called the peeling-off strategy). The process iterates on the remaining nodes until all are assigned to a cluster. Hence, the DS approach determines the clusters sequentially using a relative measure that quantifies the clusters' homogeneity [9], [37], [40], [41].

Unlike the ranking and searching algorithms, clustering methods guarantee the nonredundancy of the selected attributes. In this way, the subset of selected attributes is more representative of the original set. Hence, the performance of the remote sensing analysis will be enhanced. Thus, clustering methods, as well as searching methods, are quite accurate. However, graph-based clustering approaches, in particular, are more advantageous than searching methods for their pertinence in dealing with nonconvex datasets. Computational complexity can vary depending on the clustering algorithm that is used and the size of the dataset.

It is worth noting that methods based on deep learning, such as autoencoders, can be used for attribute extraction [42]. By using a training set of data, autoencoders learn a mapping that preserves the structure, from the original data space to a lower dimensional space. Many variants of autoencoders have been proposed for attribute selection as well as to tackle the issue of interpretability loss. Xu *et al.* [43] select the subset of attributes that contributed the most to the output, while Tomar *et al.* [44] backpropagate the network through more probable links, to name a few. The main drawback of approaches based on deep learning is their heavy dependence on the density of the training set. The training dataset should be rich in quality and size to reflect on the structure of the underlying manifold, especially if it has a complex structure. However, due to the difficulty of procuring such dense training sets, such methods can be hardly employed to obtain accurate and reliable results. Moreover, the aforesaid frameworks are not flexible in dealing with heterogeneous datasets. All this adds up to the complexity of implementation. We would like to emphasize that, in this study, we are comparing the proposed method only with unsupervised dimensionality reduction approaches, while neural-network-based approaches are either supervised or semi-supervised; therefore, we are not using any of these approaches since they require a training set [42]–[44].

The methods described above can be classified as supervised if they require labeled data, or unsupervised, otherwise. However, unsupervised methods are more convenient since acquiring labeled data, which in most cases involves the implication of an expert, is costly and time consuming. Indeed, in contrast to other research fields, providing very accurate labels is challenging in the case of remote sensing, for instance, when dealing with complex scenes or when considering modalities that are difficult to interpret, such as SAR images of sea ice in polar areas.

B. Related Work

In this study, we propose an information selection method based on the graph Laplacian. Since this approach has been widely employed for multimodal analysis in remote sensing, it is worth to mention several works based on the graph Laplacian and generally on segmentation of multimodal datasets. The graph Laplacian is a matrix representation of the graph that reflects its properties [9], [12], [45]. In particular, the eigenvectors of the Laplacian constitute a low-dimensional embedding of the nodes (that represent attributes), which increases their separability by revealing their hidden pattern. As opposed to attribute extraction approaches, this embedding can be mapped back, preserving the attributes' physical interpretability. As such, it combines the advantages of attribute selection and attribute extraction methods [17].

The graph Laplacian has been widely applied for multimodal analysis in remote sensing. For instance, we might cite manifold alignment applications that aim to determine a common latent space where multimodal datasets have a unified representation and become comparable [46]. In [47], Tuia *et al.* propose a semi-supervised framework for a manifold alignment that avoids geometric comparisons between modalities since it only compares their labels while preserving each domain's geometry via domain-specific graph Laplacians. A successful outcome of this approach relies on the quality of labels that should be similar among the datasets and representative of their connections. Hong *et al.* [48] propose to consider unlabeled information additionally to labeled samples. In particular, their approach exploits labeled samples from the overlapped area of hyperspectral and multispectral modalities and pseudo labels given only by the multispectral modality. The pseudo labels are updated using a data-driven Laplacian matrix learned on the latent subspaces of both modalities. As opposed to [47], the approach in [48] requires the datasets to be coregistered and overlapped. Furthermore, some deep learning framework attempts to increase the capability of information blending between multimodalities using different strategies, such as multiscale fusion, bidirectional symmetrical mechanism, and highly dense connectivity, have been proposed [49]. Moreover, while, generally, the joint modality representation is used, some methods are building the disjunct subnetworks in order to learn the discriminative features independently for each modality and integrate them with various structured constraints, which can be measured by similarity, correlation, or sequentiality, onto the resulting encoder layers [50]. In addition, some GAN-induced models have also been investigated [51], [52]. Among these methods, it is worth to mention the strategy proposed in [51], where the robustness of the features is increased by eliminating the effects of the adversarial noises. Moreover, the algorithm in [52] models the adversarial perturbation into end-to-end multimodal networks to obtain large-scale semantic segmentation.

Another application is multimodal segmentation, specifically by combining LiDAR and hyperspectral datasets. In [53], Iyer *et al.* proposed an approach based on SC for multimodal segmentation. To combine information from multimodal datasets, the similarity between the pixels is given by the minimum of all

similarities considering different modalities. As such, two classes are similar if and only if they are similar in all modalities. The eigenvectors of the fused graph are then used for segmentation in a semi-supervised manner using the MBO algorithm [53]. Xia *et al.* [54] also propose to combine hyperspectral and LiDAR features in a semi-supervised manner. Their approach exploits both labeled and unlabeled samples to optimally fuse both modalities' spectral, elevation, and spatial features. Hong *et al.* [55] as well as the aforementioned authors further extended their model to a semi-supervised version by learning a graph structure for the alignment of labeled and unlabeled samples. In the case of multimodal datasets that involve the use of information-rich data, which accompanied by high storage and computational costs, it seems relevant to train the model employing only a limited part of the multimodal dataset. Thus, training in more compact and varied cross-modal representations facilitates predicting larger scale semantic segmentation results [55].

For graph building, the GK, also called the heat kernel function or radial basis function (RBF), is typically used to assess the graph's nodes' similarity. In the case of heterogeneous datasets, GK might be a valid choice. However, GK will not be able to reveal the structure of data from different domains [48]. This limitation can be circumvented by comparing the heterogeneous datasets' labels as in [47], assuming that they include similar classes, or by learning the graph from the dataset as in [48]. Both approaches heavily rely on the quality and density of the labels.

C. Contributions

With this in mind, we developed a method for flexible attribute selection based on graph Laplacian representation induced by metrics computed at global and local scales across the given multimodal datasets. When analyzing multimodal data, classical spectral methods are struggling to perform on such highly heterogeneous datasets [10]–[12]. Therefore, in this work, we are suggesting adding another criterion to weight a graph edges in order to solve the limitations of the classic SC approaches. Unlike the commonly used functions to weigh the edges of a graph, such as GK, MI can assess nodes' similarity from different domains since it only compares their probability density functions (PDFs). MI measures the statistical dependence between two random variables. It is defined as the Kullback–Leibler divergence of their joint PDF and the product of marginals. Instead of only exploiting the MI to assess the similarities of multimodal attributes, we propose to combine it with the GK. The GK will compensate for the incapacity of MI to capture the local structure of the attributes. Several works employ two similarity measures for information selection [13], [56]–[58]. In contrast to those methods, we exploit both measures simultaneously and not sequentially. In this way, the results will not be biased by the order in which the measures were applied, i.e., both criteria are equally important, and hence, the selection will be more precise.

Accordingly, while the existing works using graph Laplacians only rely on the attributes' structural similarity using kernels, we also consider the attribute's information content. We jointly employ GK and MI to identify the most relevant attributes within the records. Bearing in mind the variability of the Earth's surface properties, the attributes' relevance will vary among the different classes within the remotely sensed images.

Correspondingly, the second major contribution is that the two similarity measures are applied at different detail levels to preserve more information about original data. The MI is applied globally, i.e., image-wise, so to provide a better estimation of the attributes' shared information. On the other hand, the GK is performed locally, i.e., patch-wise, in order to preserve the structure and particularity of homogeneous areas within the images. This allows us to increase the flexibility and accuracy of information selection since different relevant attributes are selected for various homogeneous areas.

Thereby, the proposed approach guarantees high accuracy of the analysis and reduces the computational complexity so that the potential of multimodal remote sensing data analysis can be exploited in multiple applications. The different experimental tests conducted on several multimodal datasets illustrate the ability of such an approach in revealing the complex pattern of the heterogeneous attributes that ensures a more precise selection than the existing works.

It is worth noting that, as opposed to [53] and [54], our approach employs the Graph Laplacian for attribute selection and not to extract new attributes. As such, we preserve the physical interpretability of the attributes that might be exploited, for instance, in understanding the contribution of each modality in the underlying analysis. Moreover, given the difficulty in acquiring dense and rich labels in remote sensing, and to avoid the imprecision of the selection in case of uncertain labels, our approach is applied in an unsupervised manner.

In order to sum up everything mentioned above, in this article, we introduce an unsupervised, flexible, interpretable, and accurate method for information selection that is applied for multimodal datasets. Among all the mentioned advantages, we would like to stress the main contributions and novelties of this work and proposed approach in particular.

1) *Two Similarities*: It simultaneously employs two similarity measures that preserve global and local particularities of the original dataset, which subsequently allows selecting the most relevant attributes.
2) *Flexible Selection*: The method is performed patch-wise; therefore, it selects the most relevant attributes for the considered classes across the different areas of the ROI.

Additionally, here are some minor advantages, which are less significant, and have been employed in existing works, nevertheless still worth mentioning.

1) *Multimodal*: It is flexible; therefore, it can be applied to various data combinations with different characteristics.
2) *Unsupervised*: The method is completely application independent; thus, it does not require any prior knowledge regarding the datasets or class labels in particular.
3) *Interpretable*: The method keeps the crucial advantages of both dimensionality reduction strategies, namely, attribute extraction and selection, such as preserving the physical meaning of the original data, while increasing its separability.
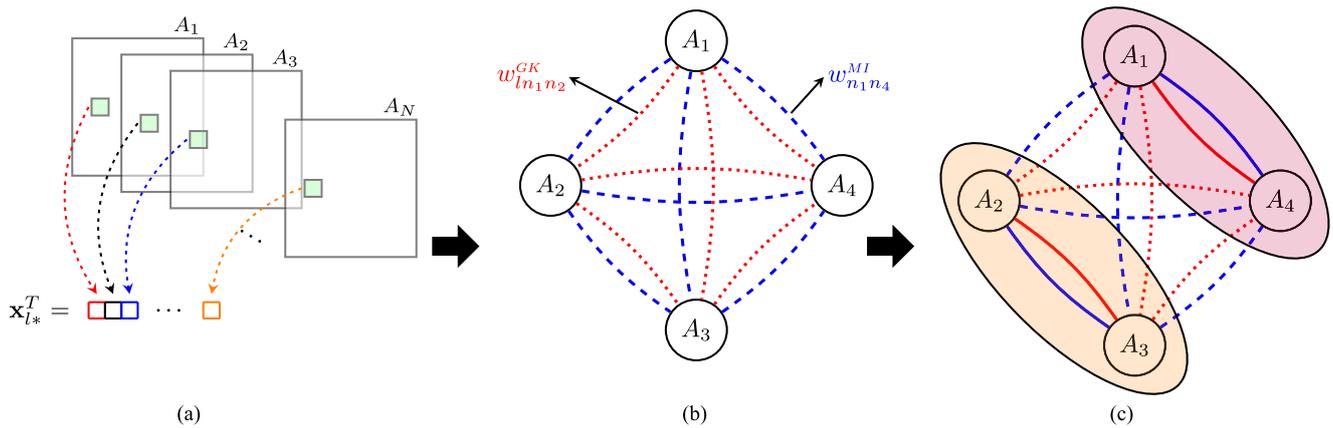
Fig. 1. Flowchart of the multimodal information selection approach proposed in this work. (a) Attributes of the l-th pixel are stacked in one vector $\mathbf{x}_{l*}$. (b) Graph of four attributes with two similarity functions at the l-th pixel. (c) Graph of four attributes with two similarity functions at the l-th pixel. Similar attributes are grouped together.

III. METHODS

This section reports the detailed description of the main steps of the proposed information selection method (see Fig. 1).

A. Attribute Generation

The very first step is attribute generation. We assume that the images are spatially aligned such that their attributes overlap. Let M be the number of available images, including bands and polarizations, and L be the number of pixels in each image. We assume that N attributes (images, textural features, etc.) could be associated with each pixel across the whole dataset, and we stack them all in $\mathbf{X} = (x_{ln}) \in \mathbb{R}^{L \times N}$ [see Fig. 1(a)]. We denote the n-th column of $\mathbf{X}$, which corresponds to the n-th attribute by $\mathbf{x}_{*n}$, so it is possible to write $\mathbf{X} = [\mathbf{x}_{*1}, \ldots, \mathbf{x}_{*N}]$. Analogously, we denote the l-th row of $\mathbf{X}$, which details the values of attributes at the l-th pixel, by $\mathbf{x}_{l*}$; hence, $\mathbf{X} = [\mathbf{x}_{1*}^T, \ldots, \mathbf{x}_{L*}^T]^T$.

B. Graph Building

For the sake of clarity, we present our approach first at a pixel level. The adaptation to the superpixel/patch level will be detailed in Section III-D. We aim to find, for a given pixel l, the smallest subset of attributes, $\{x_{l1}, \ldots, x_{lK}\}$, that preserves the structure and information content of the original set. To perform such selection, we apply the graph theory [9] since graphs are a natural way to represent various types of data.

In the proposed method, the set of N attributes will constitute the vertices of an undirected fully connected graph $\mathcal{G}_l(\mathbb{V}_l, \mathbb{E}_l^{\text{GK}}, \mathbb{E}^{\text{MI}})$, where $\mathbb{V}_l$ denote the set of attributes ($\mathbb{V}_l = \{A_1, \ldots, A_N\}$, A_n refers to the n-th attribute the values of which are given by $\mathbf{x}_{*n}$), $\mathbb{E}_l^{\text{GK}}$ and $\mathbb{E}^{\text{MI}}$ are two set of edges that connect the nodes ($\mathbb{E}_l^{\text{GK}}, \mathbb{E}^{\text{MI}} = \{(A_i, A_j), A_i, A_j \subset \mathbb{V}\}$). The weights of the edges are defined by two similarities, GK and MI, to increase the accuracy of analysis (see Fig. 1(b) for an example of four attributes at the l-th pixel).

It is worth noting that two vertices are connected by two edges. The weight of the first edge, between attributes x_{ln_1} and x_{ln_2},

is determined using the GK

$$w_{ln_1n_2}^{\text{GK}} = \exp\left(-\frac{(x_{ln_1} - x_{ln_2})^2}{2\sigma}\right), \qquad 1 \le n_1, n_2 \le N \tag{1}$$

where σ controls the width of the neighborhood in the graph. The width of the neighborhood, i.e., the number of connected vertices, increases with σ. In this work, we set σ to 1 by default, since it produces a more accurate result; however, this parameter does not affect the performance significantly.

A large value of $w_{ln_1n_2}^{\text{GK}}$ implies that the attributes x_{ln_1} and x_{ln_2} are very similar, and hence, it will be sufficient to only consider one of them to obtain accurate characterization of the dataset. Conversely, small values of $w_{ln_1n_2}^{\text{GK}}$ mean that the attributes are different and, therefore, likely to carry different information, so that they must be both considered for the analysis.

The weight of the second edge, between attributes $\mathbf{x}_{*n_1}$ and $\mathbf{x}_{*n_2}$, is defined using MI, as follows:

$$w_{n_1,n_2}^{\text{MI}} = I(\mathbf{x}_{*n_1}, \mathbf{x}_{*n_2}), \qquad 1 \le n_1, n_2 \le N$$

$$= \sum_{i=1}^{L} \sum_{j=1}^{L} P(x_{in_1}, x_{jn_2}) \log\left(\frac{P(x_{in_1}, x_{jn_2})}{P(x_{in_1})P(x_{jn_2})}\right) \tag{2}$$

where $\mathbf{x}_{*n}$ is a vector of measures corresponding to the n-th column of matrix $\mathbf{X}$, i.e., n-th attribute. $P(\mathbf{x}_i, \mathbf{x}_j)$ is the joint density function of $\mathbf{x}_i$ and $\mathbf{x}_j$, and $P(\mathbf{x}_i)$ and $P(\mathbf{x}_j)$ are the marginals. MI quantifies the shared information between two random variables [14]. Accordingly, large values of $I(\mathbf{x}_{*n_1}, \mathbf{x}_{*n_2})$ imply redundancy in information. Conversely, low values of $I(\mathbf{x}_{*n_1}, \mathbf{x}_{*n_2})$ imply synergy (novelty).

The similarity measure based on the GK represents the structure of the attribute set. In our method, it is applied at local level (i.e., on pixels or segments) in order to preserve the local particularities of the original data. On the other hand, MI reports the information content of the attribute set by discarding redundant ones. The selection via information is performed image-wise to capture the global information of the observed region. Thus, we

extract both global and local information about our data in order to enhance the performance of the proposed method.

C. Graph Clustering

Once the graph is defined according to the operations that have been previously introduced, we perform the partition of the graph using a procedure inspired by the SC approach [9] so as to identify and select the most relevant attributes in the dataset. In order to understand the main steps of this strategy, let us suppose that we only use the GK as a similarity measure as is the case in classic SC. The partition is performed by grouping the vertices of the graph into subgraphs so that two vertices of the same subgraph have strong connections (weights), while two vertices from different subgraphs have weak connections. Such a problem can be formalized using the normalized cut criterion [9], which can be defined as follows:

$$\sum_{k=1}^{K} \frac{\sum_{i \in V_{lk}} \sum_{j \in V_l \backslash V_{lk}} w_{lij}^{\mathrm{GK}}}{\sum_{i \in V_{lk}} \sum_{j \in V_{lk}} w_{lij}^{\mathrm{GK}}} \tag{3}$$

where w_{lij}^{GK} is the weight of the edge defined by the GK, and $V_{l1}, \ldots, V_{lK}$ are the K partitions of the graphs, i.e., $\bigcup_k V_{lk} = V_l$. It is also worth recalling that K identifies the number of relevant attributes that are meant to be selected out of the original records. The normalization in (3) ensures that the clusters are large enough to avoid clusters of single vertices. The criterion in (3) is then minimized over the K graph partitions to select the K most relevant attributes in the original dataset.

The aforesaid optimization of the normalized cut criterion is NP-hard and, hence, very cumbersome to efficiently address. To enhance the partition procedure, Shi and Malik proposed to replace the normalized cut minimization with an approximated problem [59]

$$\min_{\mathbf{H}} \operatorname{Tr}\left(\mathbf{H}^T \mathbf{L}_l^{\mathrm{GK}} \mathbf{H}\right) \qquad \text{subject to} \quad \mathbf{H}^T \mathbf{H} = \mathbf{I} \tag{4}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{R}^{N \times K}$, and $\mathbf{h}_k$ denotes the indicator vector of the i-th subgraph. $\mathbf{L}_l^{\mathrm{GK}}$ is the so-called symmetric normalized Laplacian matrix based on the GK, and it is defined as follows:

$$\mathbf{L}_l^{\mathrm{GK}} = \mathbf{I} - \mathbf{D}_l^{\mathrm{GK}^{-1/2}} \mathbf{W}_l^{\mathrm{GK}} \mathbf{D}_l^{\mathrm{GK}^{-1/2}} \tag{5}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{W}_l^{\mathrm{GK}} = (w_{lij}^{\mathrm{GK}})$ is the adjacency matrix, and $\mathbf{D}_l^{\mathrm{GK}} = \operatorname{diag}(\sum_{i \neq j} w_{lij}^{\mathrm{GK}})$ is the degree matrix. The n-th element of the graph indicator $\mathbf{h}_k$ can be constrained to assume a nonnull value of $(\sum_{i,j \in V_{lk}} w_{i,j}^{\mathrm{GK}})^{-\frac{1}{2}}$ if and only if the n-th node of the graph belongs to the k-th subgraph.

It is worth noting that such a discrete constraint leads to an NP-hard problem that can be relaxed by allowing the solutions to be in $\mathbb{R}$ [59]. In this case, according to the Rayleigh–Ritz theorem, the solution of (4) is given by the first K eigenvectors of $\mathbf{L}_l^{\mathrm{GK}}$ [9]. In fact, the multiplicity of the null eigenvalue of the Laplacian matrix equals the number of the connected components in the graph, and their corresponding eigenvectors are indicators of different subgraphs [9]. Moreover, since the discrete constraint

on the indicators was discarded, a clustering of the rows of $\mathbf{H}$ is required to refine the results [9], [59]. Indeed, the n-th row of $\mathbf{H}$ corresponds to the n-th attribute. As such, the same results of the clustering on the rows of $\mathbf{H}$ apply to the attributes. Moreover, the subset of relevant attributes is constituted by picking, from each cluster, the closest attribute to the centroid. Accordingly, the rows of $\mathbf{H}$ can be considered a revertible low-dimensional embedding of the attributes.

At this point, it is worth recalling that the graph representation of the datasets we aim to analyze is associated with a fully connected graph. In this case, the graph is one connected component. Hence, there will be one null eigenvalue of the Laplacian matrix corresponding to a constant eigenvector [9]. As a consequence, the graph indicators are given by the eigenvectors related to the next lowest eigenvalues. Therefore, graph clustering success relies on the identifiability of these informative eigenvalues related to the graph indicators. As such, they need to be isolated from other eigenvalues [9], [12].

The isolation of the eigenvalues is directly associated with the clusters' separability, which is more plausible in homogeneous graphs, where similar interactions occur among the nodes. On the other hand, the attributes of multimodal datasets are heterogeneous, and they interact differently. In this case, however, it has been shown that the classic graph clustering will fail at separating the clusters [10]–[12]. To tackle this issue, we propose considering the MI in addition to the GK. Incorporating the MI will reflect different relationships between the attributes from the GK. This new variability will help isolate the informative eigenvalues and increases the clusters' separability, which will translate into a precise attributes selection.

Now, if we consider the MI in addition to the GK, we would like to partition the graph such that the vertices of the same subgraph have strong connections via both links, while the vertices from different subgraphs have one or two weak connections, either GK or MI [see Fig. 1(c)]. An approximation of this problem can be written as follows:

$$\begin{cases} \min_{\mathbf{H}} \operatorname{Tr}\left(\mathbf{H}^T \mathbf{L}_l^{\mathrm{GK}} \mathbf{H}\right) \\ \\ \min_{\mathbf{H}} \operatorname{Tr}\left(\mathbf{H}^T \mathbf{L}^{\mathrm{MI}} \mathbf{H}\right) \end{cases} \quad \text{subject to} \quad \mathbf{H}^T \mathbf{H} = \mathbf{I} \tag{6}$$

where $\mathbf{L}^{\mathrm{MI}}$ denotes the Laplacian matrix based on MI

$$\mathbf{L}^{\mathrm{MI}} = \mathbf{I} - \mathbf{D}^{\mathrm{MI}^{-1/2}} \mathbf{W}^{\mathrm{MI}} \mathbf{D}^{\mathrm{MI}^{-1/2}} \tag{7}$$

where the corresponding adjacency matrix and degree matrix are defined as $\mathbf{W}^{\mathrm{MI}} = (w_{ij}^{\mathrm{MI}})$, and $\mathbf{D}^{\mathrm{MI}} = \operatorname{diag}(\sum_{i \neq j} w_{ij}^{\mathrm{MI}})$, respectively. The solution of (6) is given by the common eigenspace of $\mathbf{L}_l^{\mathrm{GK}}$ and $\mathbf{L}^{\mathrm{MI}}$, i.e., their joint eigenvectors. The common eigenspace spanned by both Laplacians enables their interaction, which might unfold complicated structure of the graph. The joint eigenvectors of the graph Laplacians, $\mathbf{L}_l^{\mathrm{GK}}$ and $\mathbf{L}^{\mathrm{MI}}$, are defined so that the following equations hold:

$$\mathbf{L}_l^{\mathrm{GK}} = \mathbf{V}_l \boldsymbol{\Lambda}_l^{\mathrm{GK}} \mathbf{V}_l^T \tag{8}$$

$$\mathbf{L}^{\mathrm{MI}} = \mathbf{V}_l \boldsymbol{\Lambda}_l^{\mathrm{MI}} \mathbf{V}_l^T \tag{9}$$

Algorithm 1: SC Algorithm for Local Pixel/Superpixel-wise Selection.

Input:
- Attributes of the l-th pixel—$\{x_{l1}, \ldots, x_{lN}\}$/$l$-th superpixel—$\{\mathbf{x}_{l1}, \ldots, \mathbf{x}_{lN}\}$
- Number of selected attributes—$K < N$

Output: Subset of N Attributes

1) Compute the adjacency matrices $\mathbf{W}^{\mathrm{MI}}$ using (2) and $\mathbf{W}_l^{\mathrm{GK}}$ using (1) for pixel-wise selection and (11) for superpixel-wise selection.
2) Compute the degree matrices $\mathbf{D}_l^{\mathrm{GK}}$ and $\mathbf{D}^{\mathrm{MI}}$
3) Construct the Laplacians $\mathbf{L}_l^{\mathrm{GK}}$ and $\mathbf{L}^{\mathrm{MI}}$ as in (5) and (7), respectively.
4) Compute the first K smallest joint eigenvectors of $\mathbf{L}_l^{\mathrm{GK}}$ and $\mathbf{L}^{\mathrm{MI}}$, $\mathbf{v}_{l1}, \ldots, \mathbf{v}_{lK}$.
5) Form $\mathbf{V}_l = [\mathbf{v}_{l1}, \ldots, \mathbf{v}_{lK}] \in \mathbb{R}^{L \times K}$.
6) Normalize the rows of $\mathbf{V}_l$ to 1.
7) Cluster the rows of $\mathbf{V}_l$ into K clusters using K-means
8) Assign $\mathbf{r}_{li}$ to the same cluster as the i-th row of $\mathbf{V}_l$
9) Return, for each cluster, the closest attributes to the centroid.

where $\mathbf{V}_l = [\mathbf{v}_{l1}, \ldots, \mathbf{v}_{lN}]$ is the matrix of eigenvectors, and $\mathbf{\Lambda}_l^{\mathrm{GK}} = \mathrm{diag}(\lambda_{l1}^{\mathrm{GK}}, \ldots, \lambda_{lN}^{\mathrm{GK}})$ and $\mathbf{\Lambda}_l^{\mathrm{MI}} = \mathrm{diag}(\lambda_{l1}^{\mathrm{MI}}, \ldots, \lambda_{lN}^{\mathrm{MI}})$ are diagonal matrices of the corresponding GK- and MI-based eigenvalues, respectively.

In general, a joint diagonalization (JD) exists if and only if $\mathbf{L}_l^{\mathrm{GK}}$ and $\mathbf{L}_l^{\mathrm{MI}}$ commute in multiplication [60], which is not always valid in practice. Thus, $\mathbf{V}_l$ is determined using approximate JD algorithms [61] instead, which minimize a criterion of diagonality of $\mathbf{V}_l^T \mathbf{L}_l^{\mathrm{GK}} \mathbf{V}_l$ and $\mathbf{V}_l^T \mathbf{L}^{\mathrm{MI}} \mathbf{V}_l$. Different diagonalization constraints and distances can be used leading to a multitude of algorithms. In this work, we perform the JD using the Quasi-Newton algorithm [61], which minimizes the log-likelihood measure introduced by Pham and Cardoso [62], i.e.,

$$\mathcal{L}(\mathbf{V}) = \log \frac{\left| \mathrm{ddiag}\left(\mathbf{V}_l^T \mathbf{L}_l^{\mathrm{GK}} \mathbf{V}_l\right) \right|}{\left| \mathbf{V}_l^T \mathbf{L}_l^{\mathrm{GK}} \mathbf{V}_l \right|} + \log \frac{\left| \mathrm{ddiag}\left(\mathbf{V}_l^T \mathbf{L}^{\mathrm{MI}} \mathbf{V}_l\right) \right|}{\left| \mathbf{V}_l^T \mathbf{L}^{\mathrm{MI}} \mathbf{V}_l \right|}. \tag{10}$$

Once the original set of attributes is embedded into a lower dimensional manifold using the joint null eigenvectors of the Laplacian matrices, a classical clustering method, such as K-means, is then applied to partition the embedding, i.e., to cluster the rows of the matrix $\mathbf{H} = [\mathbf{v}_{l1}, \ldots, \mathbf{v}_{lK}]$ into $K < N$ clusters. This new representation enhances the efficiency of standard clustering methods by increasing the separability of data, mainly if it is nonlinearly separable. Moreover, it eliminates the sensitivity to initialization of such methods. Finally, the centroids of the clusters will form the set of selected attributes. It should be noted that the number of selected attributes K is not determined automatically in this work. Algorithm 1 reports the main steps of the proposed information selection method inspired by Ng *et al.* [63].

Superpixel Approach
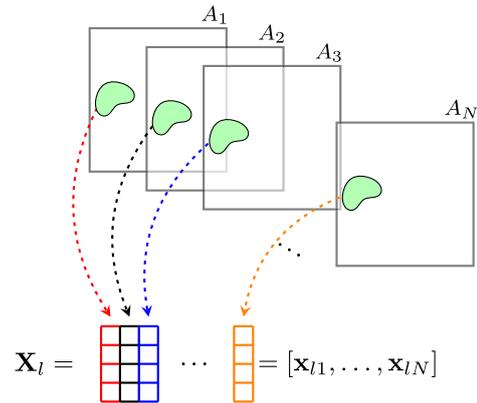
Fig. 2. Attributes of the l-th superpixel are stacked in one matrix $\mathbf{X}_l \in \mathbb{R}^{L_l \times N}$. L_l denotes the number of pixels in the l-th superpixel.

D. Superpixel Approach

Considering the large size of the remote sensing images, performing the selection at each pixel is computationally expensive [64]. To alleviate the computational complexity while preserving each pixel's local particularity, we propose to implement selection on a superpixel-level instead, i.e., patch-wise (see Fig. 2). As opposed to other patch-wise approaches, e.g., windowing, superpixels include pixels that share similar information since they are generated using segmentation (i.e., the grouping of homogeneous pixels) [65]. As such, the selection is more precise since it is particular to the properties of homogeneous pixels.

The first step of a superpixel selection consists of segmenting the image into homogeneous areas. This step can be achieved using segmentation methods such as Watershed [66], [67] or simple linear iterative clustering [68]. In our work, we use Watershed superpixel segmentation.

In the superpixel approach, similar steps as in Algorithm 1 applies except for the calculation of the GK adjacency matrix. In the case of the superpixel-based definition of the graph to be used for attribute selection as previously mentioned in this section, the elements of the adjacency matrix $\mathbf{W}_l^{\mathrm{GK}}$ are calculated using all the pixels within the l-th superpixel, i.e.,

$$w_{l n_1 n_2}^{\mathrm{GK}} = \exp\left(-\frac{||\mathbf{x}_{ln_1} - \mathbf{x}_{ln_2}||^2}{2\sigma}\right), \quad 1 \le n_1, n_2 \le N \tag{11}$$

where $||.||$ denotes the Frobenius norm. The graph is then explored and the eigenanalysis is performed according to the steps detailed in the previous subsection and summarized in Algorithm 1 in order to identify and select the K most relevant attributes in the dataset.

IV. ANALYSIS AND EXPERIMENTAL RESULTS

The following section reports the experimental analysis and performance evaluation of the proposed method, as well as comparison results with existing methods using several multimodal

datasets. In the remaining of this section, we refer to our method as GKMI–Gaussian kernel and mutual information.

Attribute selection can be applied as a preprocessing step of several remote sensing applications, e.g., target detection, classification, unmixing, etc. However, for the validation of our method, we only consider the improvement of classification accuracy.

The segmentation step used as part of GKMI might produce superpixels that include a different number of classes, and the classes may differ from one superpixel to another. To tackle the heterogeneity of the superpixels, we classify them separately using parallel classification. Accordingly, we employ L classifiers for the L superpixels that constitute the image $\mathbf{X}$. To train the classifiers, we use the same training set $\mathbb{T} = \{\mathbf{t}_1, \mathbf{t}_2, \dots \mathbf{t}_\theta\}$ that constitutes a certain percentage of the original dataset, where $\mathbf{t}_i \in \mathbb{R}^N$ is the i-th pixel in the training set. However, the attributes of the training set are adapted to each superpixel. As such, for a given superpixel $\mathcal{S}_l$, only a subset of the elements of $\mathbf{t}_i$ is considered. The indices of these elements are the indices of the attributes selected for $\mathcal{S}_l$.

Various classifiers can potentially produce different accuracy results on the same dataset. To validate the performance and show the consistency and robustness of our algorithm, we implement two of the widely applied classifiers in remote sensing: support vector machine (SVM), and random forest (RF) [69], [70].

SVM is a classification method that determines a set of hyperplanes that separate the dataset into different classes [71]. To perform a nonlinear classification, we choose the RBF as a kernel. The optimal parameters c and γ of the RBF kernel are determined by parameter tuning.

RF generates an ensemble of individual decision trees and combines their outputs to get an accurate prediction of the class [72]. In other words, RF is a classifier consisting of a collection of tree-structured classifiers.

Both classifiers are supervised methods that strongly rely on an analyst to define the classes for subsequent classification. To quantitatively estimate the classification result, we use the overall accuracy (OA) index, average accuracy (AA) index, and Cohen's kappa statistic (Kappa). OA shows the percentage of correctly classified samples, AA quantifies the mean of class-specific accuracies for all classes, while Kappa measures the agreement between the classification and the reference data [73].

This section is divided into four subsections, which aim to display the capacity of the proposed method according to the following organization.

1) Section IV-A introduces the datasets that were investigated in this work.
2) Section IV-B investigates the algorithm's sensitivity to the number of selected attributes, the size of superpixels, and the size of the training sample.
3) Section IV-C reports the relevance of using two similarity functions and the pertinence of a superpixel selection versus pixel-wise and image-wise selection.
4) Section IV-D shows the validation of GKMI performance and comparison with different information selection methods on the considered multimodal datasets.

TABLE I
GLCM FEATURES

Features	Definition
Contrast	$\sum_{i,j=0}^{Q-1} g_{i,j} \, (i-j)$
Dissimilarity	$\sum_{i,j=0}^{Q-1} g_{i,j} \, |i-j|$
Homogeneity	$\sum_{i,j=0}^{Q-1} \frac{g_{i,j}}{1+(i-j)^2}$
ASM	$\sum_{i,j=0}^{Q-1} g_{i,j}^2$
Energy	$\sqrt{ASM}$
Correlation	$\sum_{i,j=0}^{Q-1} g_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \right]$

$g_{i,j}$ denotes the (i,j) element of the GLCM matrix $\mathbf{G}$. Q is the number of gray levels used, and $\mu = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} i \, g_{i,j}$ and $\sigma^2 = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} (i - \mu) \, g_{i,j}^2$ are, respectively, the GLCM mean and variance. ASM refers to the angular second momentum.

TABLE II
TYPES AND NUMBER OF ATTRIBUTES FOR THE CONSIDERED DATASETS

Dataset	Hyperspectral		Optical			LiDAR		N
	Original	GLCM	S2	L8	GLCM	Original	GLCM	
Berlin	–	–	10	18	168	–	–	196
Paris	–	–	10	18	168	–	–	196
Trento	63	–	–	–	–	2	12	77
Houston	144	–	–	–	–	1	6	151

S2 and L8 refer to Sentinel-2 and Landsat-8, respectively. N denotes the total number of attributes for each dataset. It should be noted that the GLCM attributes, listed in Table I, are generated for each band.

A. Dataset Description

To evaluate the performance of the proposed GKMI method for attribute selection, we consider different multimodal datasets obtained from various satellite platforms. In this work, we only consider data and feature levels of multimodal data fusion, although the GKMI method can also be applied at the decision level.

To increase the number of attributes and extract some additional information from the original data, along with the bands of optical and LiDAR datasets, we use textural features, while for hyperspectral datasets, we only use existing bands. To extract textural features, we use the gray-level co-occurrence matrix (GLCM) [74]–[76]. Table I illustrates the extracted features as well as their mathematical definitions.

Table II summarizes the number and types of the attributes considered in this article, and Table III reports the list of ground truth labels for each dataset. A detailed description of the datasets is presented as follows.

1) Berlin/Paris: The datasets were acquired over the cities of Berlin and Paris, and both consist of images obtained from two optical sensors: Sentinel-2 and Landsat-8. The datasets were obtained from the 2017 IEEE GRSS Data Fusion Contest [77].

Both datasets (Sentinel-2 and Landsat-8) were resampled at 100-m resolution. Berlin and Paris test sites were prelabeled for the subsequent classification and include 12 ground truth labels corresponding to various built-up (anthropogenic constructions)

TABLE III
GROUND TRUTH LABELS FOR ALL THE DATASETS USED IN THIS ARTICLE

	Berlin	Paris	Trento	Houston
ω_1	Compact midrise	Compact high-rise	Buildings	Grass healthy
ω_2	Open high-rise	Compact midrise	Wood	Grass stressed
ω_3	Open midrise	Open high-rise	Apple trees	Grass synthetic
ω_4	Open low-rise	Open midrise	Roads	Tree
ω_5	Large low-rise	Open low-rise	Vineyard	Soil
ω_6	Sparsely built	Large low-rise	Ground	Water
ω_7	Dense trees	Sparsely built		Residential
ω_8	Scattered trees	Dense trees		Commercial
ω_9	Bush and scrub	Scattered trees		Road
ω_{10}	Low plants	Low plants		Highway
ω_{11}	Bare soil or sand	Bare rock or paved		Railway
ω_{12}	Water	Water		Parking lot 1
ω_{13}				Parking lot 2
ω_{14}				Tennis court
ω_{15}				Running track

(a)

(b)

Fig. 5. False-color composite representation of Trento (a) hyperspectral and (b) LiDAR datasets.

(a)

(b)
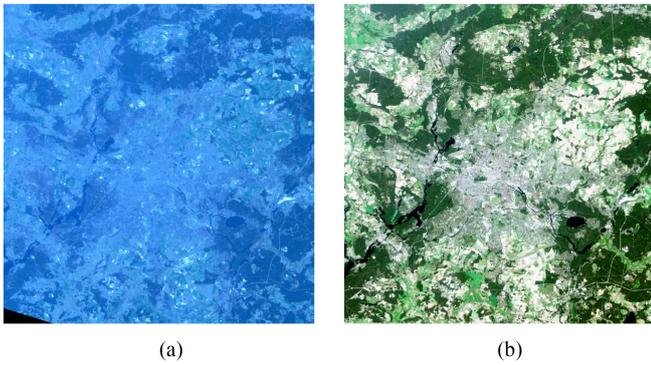
Fig. 3. Overlapping area of the Berlin dataset. (a) Landsat-8 and (b) Sentinel-2 natural color composite images.

(a)

(b)
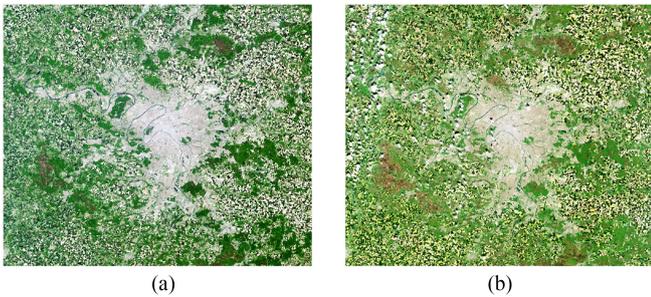
Fig. 4. Overlapping area of the Paris dataset. (a) Landsat-8 and (b) Sentinel-2 natural color composite images.

and land cover types. Sentinel-2 dataset contains ten bands in the visible, near-infrared, and short-wave infrared part of the spectrum. Landsat-8 contains nine bands in visible, short, and long infrared wavelengths (according to the notation in Section III-A, $M = 28$, 1×Sentinel-2 dataset + 2×Landsat-8 datasets). Moreover, from each band, we extract six textural features (see Table I). Therefore, the final datasets that were used contain $N = 196$ extracted attributes.

Figs. 3 and 4 show the overlapping area of the two datasets. The overlapping test size area for Berlin example is of 666×643 pixels, and for Paris, it is of 988×1160 pixels.

2) Trento: This dataset was acquired on an agricultural area in the south part of the city of Trento, Italy. It consists of LiDAR and hyperspectral data. Hyperspectral data were acquired by the

AISA Eagle sensor with a 1-m spatial resolution and includes 63 bands ranging from 0.40 to 0.99 μm, where the spectral resolution is 9.2 nm. The LiDAR data were acquired by the Optech ALTM 3100EA sensor. The available ground truth labels consist of six classes.

The Trento dataset contains 63 hyperspectral bands and two LiDAR bands ($M = 65$). Additionally, we extracted six textural features for each of the available LiDAR bands (see Table I). The final dataset that was used contains $N = 77$ attributes with an overlapping test size area of 600×166 pixels. Fig. 5 illustrates the false-color composite representation of the Trento dataset for both sensors.

3) Houston: The last dataset consisted of LiDAR and hyperspectral data acquired over the University of Houston campus and the neighboring urban area and was distributed for the 2013 IEEE GRSS Data Fusion Contest [78]. Hyperspectral data were acquired from the Compact Airborne Spectrographic Imager with a 2.5-m spatial resolution. The hyperspectral dataset includes 144 spectral bands ranging from 0.38 to 1.05 μm. The available ground truth labels consisted of 15 classes.

The Houston dataset contains hyperspectral data (144 bands) and 1×LiDAR data (including one band and six textural features). The final dataset that was used consisted of $N = 151$ attributes with an overlapping test size area of 1905×349 pixels. Fig. 6 demonstrates the Houston test site for both sensors.

B. Parameter Sensitivity Analysis

Several parameters may affect the performance of GKMI, mainly the number of selected attributes, the size of superpixels, and the size of the training set. In the following, we tune one parameter at a time to understand how it influences our proposed approach.

1) Number of Attributes: Fig. 7 illustrates the overall accuracies for the proposed GKMI attribute selection over a different number of selected attributes for all datasets used in this work. The blue curve identifies the OA results obtained on the Berlin dataset, the red line indicates the OA results obtained on the Paris dataset, the green line refers to the Trento dataset, whereas the black line shows the results for the Houston dataset. The stars

Fig. 6. False-color composite representation of Houston (a) hyperspectral and (b) LiDAR datasets.

Fig. 7. Overall accuracies of GKMI as a function of different numbers of selected attributes for Berlin (blue dotted line), Paris (red dashed line), Trento (green solid line), and Houston (black dash-dotted line) test sites using the SVM classifier.

indicate the point where the accuracy reaches its maximum. It can be seen from Fig. 7 that the OA curves rise sharply until the number of attributes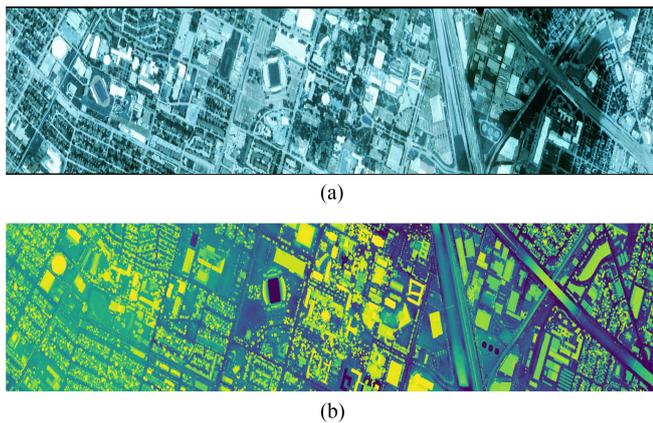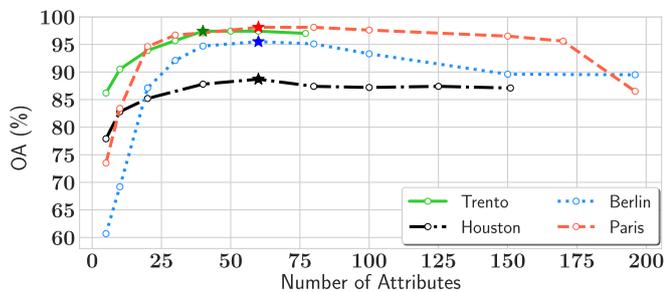 chosen reaches 40 for the Berlin dataset and 30 for the Paris dataset. After that point, OA curves keep stable high till 75 and start to decrease. The Trento and Houston curves grow abruptly until the number of attributes chosen reaches 20 for both datasets.

All the curves, in general, have a similar pattern that indicates that a large number of selected attributes do not necessarily lead to the best classification result. The number of selected attributes reaches some particular point where additional attributes can hardly provide any extra information for subsequent classification. Depending on the original data, additional attributes may even reduce the accuracy of classification. This result shows the relevance of our method, since using the total number of attributes leads to lower accuracy. Actually, the maximum efficiency is reached using less than half of the attributes.

2) Size of Superpixels: The size of superpixels is another parameter that may impact the performance of our method. Since the same set of attributes is assigned to the pixels of the same superpixel, we expect that too large or too small superpixels may deteriorate the results. Large superpixels may include several homogeneous regions; hence, the selected subset may not be representative of all pixels. On the other hand, small superpixels
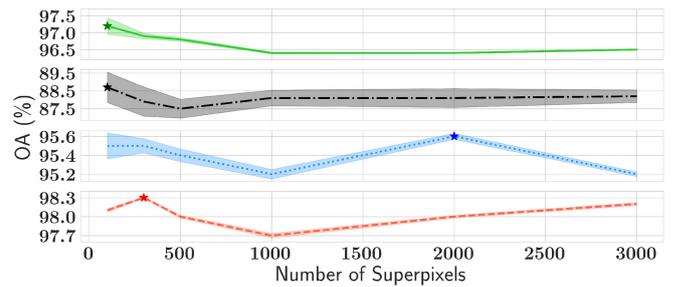
Fig. 8. Mean (curves) and variance (faded surfaces) of overall accuracies of GKMI as a function of different numbers of superpixels for Berlin (blue), Paris (red), Trento (green), and Houston (black) test sites using the SVM classifier. The same color legend as in Fig. 7 applies here. Note that the figure consists of four different subfigures with different scales on the vertical axes.

may not contain the whole homogeneous area, and various attributes can be chosen for the same region.

It is important to recall that the superpixel selection is used only during the attribute selection process, while classification is performed for each pixel separately. In other terms, let us assume that the i-th superpixel $\mathcal{S}_i$ consists of P_i pixels $\{\mathbf{x}_{p*}\}_{p=1,\ldots,P_i}$. The pixels belonging to $\mathcal{S}_i$ cannot be associated with any other superpixel, i.e., $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset \, \forall (i,j) \in \{1,\ldots,L\}^2, i \neq j$, where L is the total number of the considered superpixels in the dataset. Then, the attribute selection procedure in Section III selects for all the pixels in $\mathcal{S}_i$ a subset of K attributes Ω_i. The p-th pixel in $\mathcal{S}_i$ is, hence, classified independently from the others by taking into account only the attributes in Ω_i.

To investigate the impact of this parameter on GKMI, we illustrate in Fig. 8 the OA of the proposed method, over a different number of superpixels, for all datasets. The blue line shows the OA result for Berlin, the red line for Paris, the black line for Houston, and the green line for Trento. The stars show the point with maximum OA. The faded area displays the variance of the overall accuracies for different sizes of superpixels. The number of superpixels is representative of the size of superpixels in the dataset, i.e., the higher the number of superpixels, the smaller the size of the superpixels.

From Fig. 8, we can observe that the curves for each dataset are quite stable, and there are no significant fluctuations, which means that the size of the superpixels has a minor impact on the classification accuracy. Moreover, from the curves, it is possible to appreciate that the variance of the overall accuracies (faded area) is decreasing with the size of the superpixels for the Trento, Houston, and Berlin datasets. This indicates that for these particular examples, increasing the number of superpixels leads to a more stable results, while for the Paris dataset, there are no significant fluctuations in variance throughout the curve.

The OA displayed in Fig. 8 is the result of the attribute selection process as a function of the number of superpixels (L in the previous discussion). Therefore, it is possible to state that Fig. 8 shows how robust the proposed method is with respect to the L parameter. In fact, although the pixels of a given superpixel have the same set of chosen attributes, they might belong to different classes. Thus, the proposed approach is able to combine the benefits provided by the superpixel grouping
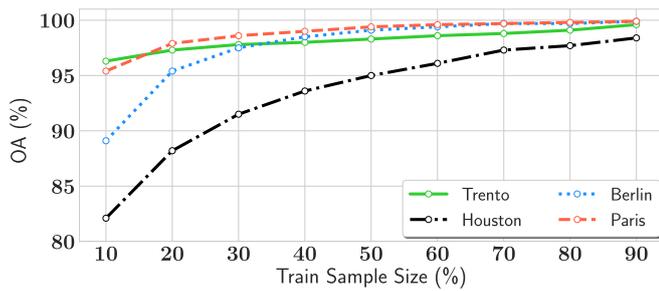
Fig. 9. Overall accuracies of GKMI as a function of different sizes of training samples for Berlin, Paris, Trento, and Houston test sites using the SVM classifier. The same color legend as in Fig. 7 applies here.

and graph clustering while avoiding biasing the results in the classification step.

3) Size of Training Sample: Another parameter that is affecting the performance is the size of the training sample. Fig. 9 demonstrates the OA of the proposed method over a different size of training samples. It is quite evident from Fig. 9 that the increase in accuracy is directly proportional to the increase in the sample size: this behavior is verified for all considered datasets.

In all other experiments of this study, we are using 20% of ground truth labels as a training sample as commonly used in practice.

C. Performance Analysis

Let us now investigate the impact of the chosen similarity measures, their weight, as well as the relevance of the superpixel analysis compared to pixel-wise and image-wise selection.

1) Kernel Comparison: Let us start by discussing the metric to be used to estimate the similarity representing the structure of the attributes' set in Section III. In this respect, it is worth noting that the choice of the function to model the similarity among attributes is a critical task in remote sensing data analysis [79], [80, ch. 9].

In fact, defining the kernel to be employed to quantify the structure of the data (and thus defining similarity between pairs of samples) is crucial to obtain a reliable understanding of the relevance of the attributes and their actual role in the characterization of the interactions among the records [79]. Furthermore, it is important to recall that a proper choice of the kernel to quantify the similarity among attributes can provide a consistent and well-founded theoretical framework for developing nonlinear techniques. Moreover, kernel functions are used in practice to unfold the complicated structure of a dataset, thus enabling the ability to deal with a low number of (potentially high dimensional) training samples, the investigation of heterogeneous records, as well as considering multiple noise sources [79], [81].

On the other hand, it is also true that the definition of the proper kernel for the aforesaid purpose might be particularly cumbersome, especially when the design of architectures for data analysis that is intended to be unsupervised, versatile, and flexible is targeted [79], [81]. Indeed, it is possible to state that the definition of a kernel mapping function that would accurately quantify the similarity among samples represents a bottleneck

TABLE IV
DEFINITIONS OF THE DIFFERENT KERNELS BETWEEN FIRST AND SECOND
ATTRIBUTES IN THE l-TH SUPERPIXEL

	Definition	Parameters
Euclidean distance (ED)	$\|\mathbf{x}_{ln_1} - \mathbf{x}_{ln_2}\|$	–
Linear kernel (LK)	$\mathbf{x}_{ln_1}^T \mathbf{x}_{ln_2} + c$	c: Optional constant
Polynomial kernel (PK)	$\left(s\mathbf{x}_{ln_1}^T \mathbf{x}_{ln_2} + c \right)^d$	c: Optional constant d: Polynomial degree s: Slope
Gaussian kernel (GK)	$\exp\left(-\frac{\|\mathbf{x}_{ln_1} - \mathbf{x}_{ln_2}\|^2}{2\sigma} \right)$	σ: scale parameter

TABLE V
PERFORMANCE COMPARISON AMONG DIFFERENT SIMILARITY MEASURES,
DEFINED IN TABLE IV, USING THE SVM CLASSIFIER

Method	Berlin		Paris		Trento		Houston	
	K	OA (%)	K	OA (%)	K	OA (%)	K	OA (%)
ED	100	93.8	80	97.7	60	97.2	100	87.3
LK	100	88.0	100	96.3	60	97.1	100	87.6
PK	100	90.2	100	96.2	60	97.2	125	86.8
GK	80	**94.9**	80	**97.9**	60	**97.6**	80	**87.9**

K refers to the optimal number of selected attributes for which the OA is obtained. In this experiment, according to the notation in Table IV, $c = 1$, $d = 3$, and $s = 1/N$.

for any kernel-based analysis approach. At the same time, it is worth remembering that not all kernel similarity functions are allowed. Specifically, valid kernels must fulfill Mercer's theorem, i.e., being positive-definite similarity matrices. This property is fundamental when no *a priori* knowledge on the interclass and intraclass statistical distributions is available [79], [81]. As a result, the kernel functions that are most commonly employed in this context are using Euclidean distance (ED) and linear, polynomial, and Gaussian functions (i.e., Linear kernel (LK), Polynomial kernel (PK), and GK in Table IV) as similarity measures.

Thus, to assess the relevance of the choice we proposed in Section III, we compare the performance of the SC in (4) when using ED, LK, PK, and GK to define the weights of the graph structure. In Table V, we represent the maximum OA achieved by the attributes selected using the different kernels. It can be seen that ED, PK, and LK show a slightly lower accuracy than GK. Moreover, the GK always achieves the highest OA with a fewer number of attributes. Compared to ED, LK, and PK, the GK is able to unfold the finer structure of the attributes since it is highly nonlinear. This strengthens the assumptions we have used in designing the data analysis steps in the proposed multimodal feature selection method. This result is consistent with the proven ability of GK to be more flexible in characterizing the data structure in complex systems, especially when an investigation of large-scale and heterogeneous datasets is conducted [45].

2) Significance of Similarity Functions: It is also worth to investigate the relevance of using MI together with the GK similarity to build the graph representing the structure of the dataset. According to the assumptions we have detailed in Section III, the proposed method assumes that both the GK and the MI are

necessary to obtain a solid characterization of the data structure to be analyzed. Indeed, the graph representation plays a key role in describing the interactions among attributes [9], [30]. Therefore, exploring the impact of the chosen similarity metrics in the definition of the graph induced by the considered dataset is crucial to understand what role the quantities used to describe the attributes' relevance can play in different applications, as well as to estimate the reliability of the proposed approach in operational use.

Let us then investigate the impact of GK and MI in the selection process outlined in Section III. In particular, to this aim, the weight of MI and GK metrics could be, in principle, unevenly distributed. Specifically, we can rewrite the function in (10) as follows:

$$\mathcal{L}(\mathbf{V}) = \left(\alpha \log \frac{\left| \mathrm{ddiag}\left(\mathbf{V}_l^T \mathbf{L}_l^{\mathrm{GK}} \mathbf{V}_l \right) \right|}{\left| \mathbf{V}_l^T \mathbf{L}_l^{\mathrm{GK}} \mathbf{V}_l \right|} \right.$$
$$\left. + (1-\alpha) \log \frac{\left| \mathrm{ddiag}\left(\mathbf{V}_l^T \mathbf{L}^{\mathrm{MI}} \mathbf{V}_l \right) \right|}{\left| \mathbf{V}_l^T \mathbf{L}^{\mathrm{MI}} \mathbf{V}_l \right|} \right). \qquad (12)$$

In other terms, the parameter α is used to change the weight (i.e., importance) of the similarity metrics employed in the selection process, i.e., high values of α give more weight to the GK, while low values of α give more weight to the MI. Particularly, only MI is considered when $\alpha = 0$, and only the GK is utilized when $\alpha = 1$.

2) Eigenvalue analysis: At this point, we study the impact of the two similarity metrics on classification performance by investigating the spectrum of the eigenvalues for different values of α. In fact, as previously mentioned in Section III-C, the eigenvectors of the Laplacian matrices used to describe the graph connectivity induced by the given dataset are directly linked to the solution of the feature selection process itself. Indeed, it is worth recalling that the key idea of graph clustering based on Laplacian matrices is that the indicators of data (attributes) classes are given by the eigenvectors of the Laplacian corresponding to the lowest eigenvalues [63]. Furthermore, heterogeneity in the graph node degrees would translate in spreading the eigenvalues of the Laplacian matrix across the spectrum [12]. This means that in the case of complex datasets (i.e., datasets where it is not possible to draw linear hyperplanes in the attribute space to perform graph clustering and therefore dimensionality reduction), it is not possible to associate the informative eigenvectors with the smallest eigenvalues anymore. Actually, by losing the isolation of informative eigenvalues, the associated eigenvectors tend to merge with the eigenvectors associated with close-by (noninformative) eigenvalues [12]. Hence, an effective dimensionality reduction can be performed only when it is possible to identify the smallest eigenvalues and clearly separate them by the eigenvalues with higher amplitude. On the contrary, when the spectrum of the eigenvalues is generally flat, then it is possible to expect that the dimensionality reduction process would not be able to achieve reliable and robust results in terms of informativity maximization [12], [63].

With this in mind, we computed the eigenvalues of the Laplacian matrix resulting from setting α to several values in [0,1].

We then considered their spectra to understand how easy it would be to identify and discriminate the smallest eigenvalues from their total set. In this respect, Fig. 10 shows the eigenvalues of the Laplacian matrices L^{GK} (case $\alpha = 1$ in (12)—yellow solid line) and L^{MI} (case $\alpha = 0$ in (12)—red dotted line) defined in (5) and (7), respectively, as well as their common eigenvalues used by GKMI (case $\alpha = 0.5$ in (12)—blue dashed curve) obtained on Paris and Trento datasets.

We notice that the curves corresponding to the eigenvalues of $\mathbf{L}^{\mathrm{GK}}$ and $\mathbf{L}^{\mathrm{MI}}$ are essentially flat [see, for instance, the enlarged section of the graph on the GK eigenvalues' trend in Fig. 10(a)]. The amplitude of the eigenvalues varies in both cases in the order of 10^{-12}), showing a low separability of the data since it is hard to isolate the eigenvalues related to class indicators [12]. On the other hand, when using both similarities according to the proposed method in Section III, it is possible to appreciate that the variability of the eigenvalues' amplitude is more pronounced in terms of several orders of magnitude. Therefore, it is a lot easier to identify the smallest eigenvalues and separate them from the total set of eigenvalues, leading to a more accurate identification of the relevant attributes in the dataset.

Fig. 10, thus, demonstrates how the heterogeneity of multimodal attributes makes their structure so complex in the attribute space such that the classic SC fails to reveal it. It is indeed worth noting that this result is compliant with recent findings in technical literature, where it has been shown that SC fails at detecting the classes of a graph with heterogeneous degrees [10]–[12], as it is the case in this work.

2) Impact of α on OA: The aforesaid results are confirmed when exploring the classification accuracy obtained when dimensionality reduction is performed for different values of α. Specifically, Fig. 11 shows the gain in the OA of GKMI compared to SC [i.e., $\alpha = 1$ in (12)] as a function of the parameter α. We notice that a negative gain (loss) is only achieved for $\alpha = 0$, implying that MI shows lower performance than the GK and that the exploitation of both measures always improves the OA. The maximum accuracy is achieved when both GK and MI are employed to define the graph [i.e., $\alpha \neq \{0, 1\}$ in (12)], specifically for $\alpha = 0.7$ for the Berlin dataset and $\alpha = 0.5$ for other datasets. According to the trends in Fig. 11, $\alpha = 0.5$ seems the best design choice to achieve high accuracy performance while guaranteeing wide applicability of the system to data with different properties.

2) Impact of α on K: To further demonstrate the pertinence of using two similarities, we compare the results obtained when using one similarity at a time and when used together. Table VI demonstrates the maximum OA of SVM classification for Berlin, Paris, Trento, and Houston datasets, when the selection is performed using only GK, using only MI, and with both similarities (GKMI). In contrast to Fig. 11, where a fixed number of chosen attributes are used, Table VI shows the optimal number for which the maximum OA is achieved. The different approaches show almost similar performance. However, GKMI reaches the maximum accuracy with less number of attributes for each dataset. It achieves an OA of 95.5% for Berlin and 98.1% for Paris, for less than a third of the original attribute set. Moreover, it achieves an OA of 88.7% for Houston, with less
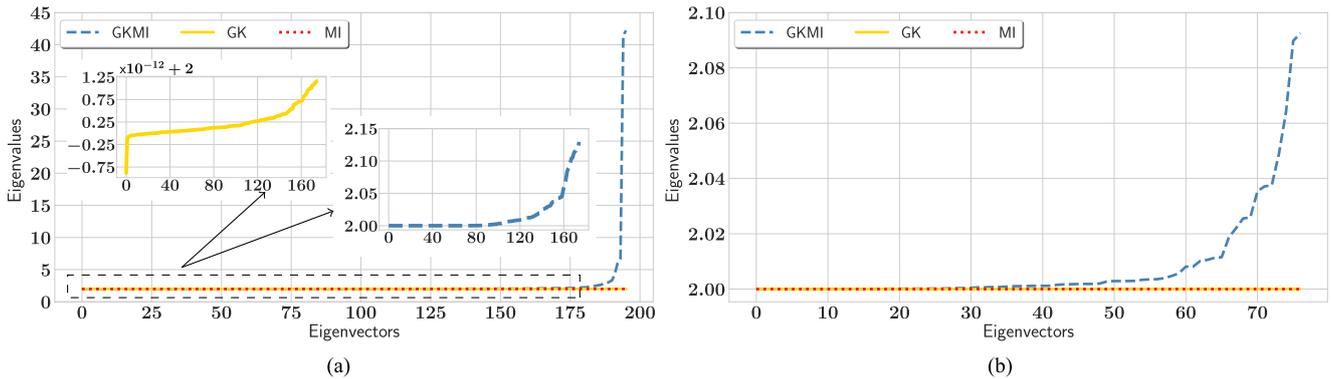
(a)   (b)

Fig. 10.   Amplitude of the eigenvalues of different Laplacian matrices for (a) Paris and (b) Trento datasets as a function of the number of eigenvectors (directly linked to the number of clusters that can be drawn in graph partitioning [9], [12]). The amplitude curves associated with eigenvalues of $\mathbf{L}^{\text{GK}}$ [as for (5)—case $\alpha = 1$ in (12)] and the eigenvalues of $\mathbf{L}^{\text{MI}}$ [as for (7)—case $\alpha = 0$ in (12)] are plotted in yellow solid line and red dotted line, respectively. The amplitude of the eigenvalues obtained via JD [proposed approach in Section III—case $\alpha = 0.5$ in (12)] is displayed in dashed blue line.
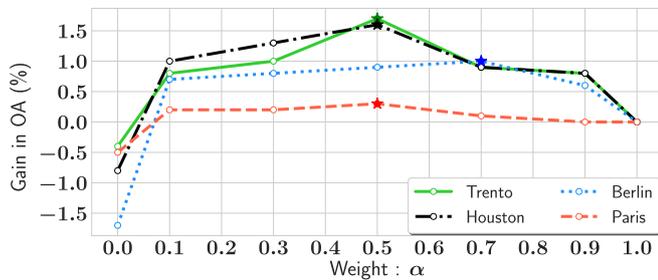

Fig. 11.   Gain in the overall accuracies of GKMI compared to SC over a different values of α for Berlin (blue), Paris (red), Trento (green), and Houston (black) test sites using the SVM classifier. The same color legend as in Fig. 7 applies here.

TABLE VI
PERFORMANCE COMPARISON OF SINGLE AND JOINT SIMILARITY MEASURES,
USING THE SVM CLASSIFIER

Method	Berlin		Paris		Trento		Houston	
	K	OA (%)	K	OA (%)	K	OA (%)	K	OA (%)
MI	80	93.7	80	97.5	60	96.8	80	86.6
GK/Spectral clustering	80	94.9	80	97.9	60	**97.6**	80	87.9
GKMI	**60**	**95.5**	60	**98.1**	40	97.4	**60**	**88.7**

K refers to the optimal number of selected attributes for which the maximum OA is obtained.

than half of the dataset. For the Trento dataset, the OA reaches the highest number 97.4% with 40 attributes, almost half of the dataset. Accordingly, we can conclude that GKMI ensures a more precise selection.

3) Fusion of Similarity Functions: For a given superpixel l, GKMI generates a graph with two edges that summarize the attributes' similarities via the GK and via MI. The two edges are then combined in a nonlinear manner by extracting the joint eigenspace of their corresponding Laplacians, $\mathbf{L}_l^{\text{GK}}$ and $\mathbf{L}^{\text{MI}}$. The fusion of the edges can be performed differently. One of the easiest approaches is by taking their mean

$$\mathbf{W}_l^{\text{mean}} = \frac{1}{2}\left(\mathbf{W}_l^{\text{GK}} + \mathbf{W}^{\text{MI}}\right). \qquad (13)$$

In this case, the indicators of the subgraphs are given by the first K eigenvectors of the Laplacian matrix

$$\mathbf{L}_l^{\text{mean}} = \mathbf{I} - \mathbf{D}_l^{\text{mean}^{-1/2}}\mathbf{W}_l^{\text{mean}}\mathbf{D}_l^{\text{mean}^{-1/2}} \qquad (14)$$

where $\mathbf{D}_l^{\text{mean}} = \text{diag}(\sum_{i \neq j} w_{lij}^{\text{mean}})$.

Another approach of graph fusion was proposed by Iyer *et al.* [53]. Their approach assumes that two nodes are similar if and only if they are similar via both similarity functions. As such, they define the weight of the combined edge as the maximum of both edges normalized

$$\mathbf{W}_l^{\text{max}} = \max\left(\frac{\mathbf{W}_l^{\text{GK}}}{\text{std}(\mathbf{W}_l^{\text{GK}})}, \frac{\mathbf{W}^{\text{MI}}}{\text{std}(\mathbf{W}^{\text{MI}})}\right) \qquad (15)$$

where $\text{std}(\mathbf{A})$ denotes the standard deviation of the elements of the matrix $\mathbf{A}$. In this case, as for the "mean" approach, the indicators of the subgraphs are given by the first K eigenvectors of the Laplacian matrix

$$\mathbf{L}_l^{\text{max}} = \mathbf{I} - \mathbf{D}_l^{\text{max}^{-1/2}}\mathbf{W}_l^{\text{max}}\mathbf{D}_l^{\text{max}^{-1/2}} \qquad (16)$$

where $\mathbf{D}_l^{\text{max}} = \text{diag}(\sum_{i \neq j} w_{lij}^{\text{max}})$.

To evaluate the performance of the different approaches, we compare their spectra to assess their ability in separating the different classes of the heterogeneous attributes.

Fig. 12 shows the eigenvalues of the Laplacian matrices obtained by the *"mean"* approach, $\mathbf{L}_l^{\text{mean}}$, the *"max"* approach, $\mathbf{L}_l^{\text{max}}$, and by the *"joint"* decomposition used by the GKMI approach. The flatness of the curves corresponding to *"mean"* and *"max"* demonstrates the incapacity of these approaches in emphasizing the informative eigenvalues, corresponding to the eigenvectors indicators of the attributes' clusters, since they are inseparable from the total set of eigenvalues. On the other hand, when using GKMI, the informative eigenvalues are well isolated. This outcome demonstrates the effectiveness of our approach.

In fact, the JD of the Laplacian matrices corresponding to the GK and MI enables their interaction, revealing their nonlinear connections and, hence, the hidden structure of the heterogeneous attributes. By connection between the similarity functions, we mean the connection of the graph structures that each
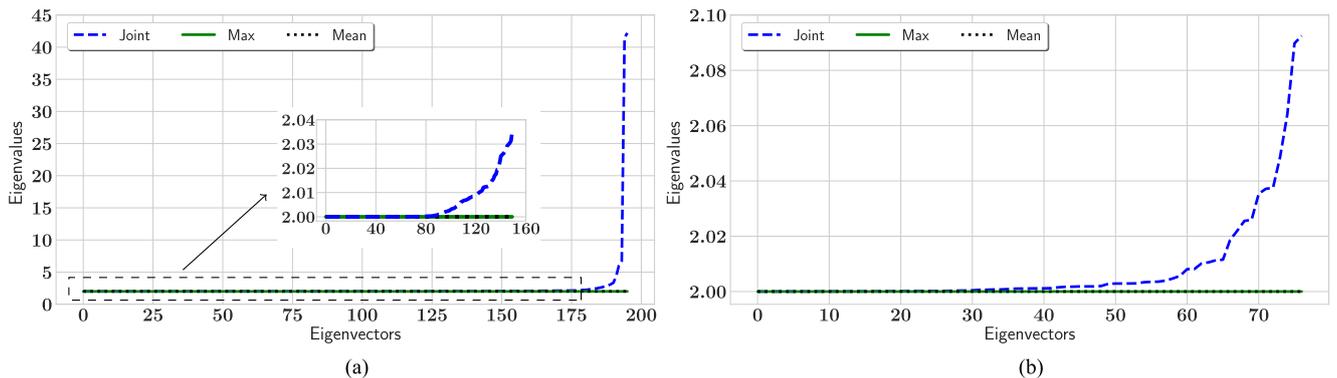
Fig. 12.   Amplitude of the eigenvalues of the different approaches to combine the similarity functions for (a) Paris and (b) Trento datasets as a function of the number of eigenvectors (directly linked to the number of clusters that can be drawn in graph partitioning [9], [12]). The amplitude curves associated with eigenvalues of $\mathbf{L}^{\text{mean}}$ (14) and the eigenvalues of $\mathbf{L}^{\text{max}}$ [proposed in [53] and defined in (16)] are plotted in black dotted line and green solid line, respectively. The amplitude of the eigenvalues obtained via JD [proposed approach in Section III in (10)] is displayed in dashed blue line.

TABLE VII
PERFORMANCE COMPARISON OF PIXEL, SUPERPIXEL, AND IMAGE-WISE GKMI APPROACHES FOR EACH DATASET

Dataset	K	Approach	OA (%)		ET (sec)
			SVM	RF	
Berlin	40	Pixel-Wise	19.9	24.2	44320
		Image-Wise	94.2	91.6	394
		Superpixel	**94.7**	**93.5**	**246**
	N = 196	No selection	89.5	93.3	465
Paris	40	Pixel-Wise	38.2	51.4	52880
		Image-Wise	96.8	94.6	889
		Superpixel	**97.1**	**95.3**	**267**
	N = 196	No selection	86.5	95.4	758
Trento	40	Pixel-Wise	55.2	66.6	3900
		Image-Wise	96.4	96.4	**23**
		Superpixel	**97.4**	97.3	64
	N = 77	No selection	97.0	**97.7**	26
Houston	40	Pixel-Wise	15.7	32.3	35310
		Image-Wise	84.3	81.0	296
		Superpixel	**87.8**	**85.8**	81
	N = 151	No selection	87.1	85.5	**58**

For the superpixel part of the calculations, 100 superpixels were used.

similarity represents. Conversely, the mean and max approaches assume linear and simple links between the similarity functions, which fail to identify and characterize their nonlinear links, and hence do not exploit their full potential.

4) Levels of Spatial Detail: The proposed GKMI method can be applied at different fusion levels. Nonetheless, its versatility allows us to investigate its application at different spatial detail levels, as GKMI can run at an image, superpixel, and pixel levels. Each of these approaches produces a different result in terms of classification accuracy and time complexity. Therefore, it is interesting to investigate how this design choice might affect the final outcome of the attribute selection procedure.

Table VII shows the OA and execution time (ET) for image, pixel, and superpixel GKMI on all datasets used in this

work. It is clear from these results that the superpixel method produces higher accuracy outcomes and outperforms pixel and image-wise approaches for different classifiers, in terms of both accuracy and computational complexity. Furthermore, the ET can be further enhanced by applying parallel computing on the different superpixels.

The superpixel procedure accounts for the particularity of each superpixel, in contrast to the image-wise, and selects the same attributes for homogeneous regions, as opposed to pixel-wise. These two reasons make the superpixel approach more accurate and effective.

Let us now investigate in more detail the GKMI approach performed at the superpixel level. In this case, the adjacency matrix using the GK is measured using all pixels of a given superpixel, as shown in (11). However, in view of the fact that the superpixels in our analysis are formed by grouping homogeneous pixels, we can improve our analysis' time complexity by performing the selection by considering each attribute's mean over all pixels or by picking a representative pixel randomly. As such, for a given superpixels with L pixels, the input of Algorithm 1 for attribute selection is a set of scalars instead of vectors given by the mean of the attributes $\{\frac{1}{L}\sum_l \mathbf{x}_{l1}, \ldots, \frac{1}{L}\sum_l \mathbf{x}_{lN}\}$ or by the attributes of the l-th randomly picked pixel $\{x_{l1}, \ldots, x_{lN}\}$. The algorithm's output, i.e., the subset of relevant attributes, will then be applied to all pixels within the superpixel. Table VIII shows the comparison of these approaches for Berlin, Paris, Trento, and Houston datasets. The results show that by randomly picking a representative pixel, the time complexity reduces without significantly affecting the OA.

In order to strengthen the idea and motivation behind the employment of the information selection on a superpixel level, we additionally analyzed the attributes that were selected by the proposed method for each class of the Trento dataset. As was mentioned earlier, the Trento dataset consists of 77 attributes (63 hyperspectral bands ranging from 402.89 to 989.09 nm, and 14 LiDAR + GLCM textural features) and has six ground truth classes, including Apple trees, Vineyard, Wood, Roads, Ground, and Buildings. Accordingly, Fig. 13 illustrates the chord diagrams that represent selected attributes for five different superpixels that fall into the area of the ground truth labels for each class of the Trento dataset. The vertices show 77 available

	Berlin			Paris			Trento			Houston		
	OA			OA			OA			OA		
	μ	σ^2	ET	μ	σ^2	ET	μ	σ^2	ET	μ	σ^2	ET
All pixels	95.5	0.11	359	**98.1**	0.03	336	**97.4**	0.29	**68**	**88.7**	0.71	88
Mean of the pixels	**95.7**	0.06	311	98.0	0.02	279	95.9	0.52	**68**	87.6	0.61	73
One pixel randomly picked	95.5	0.19	**309**	98.0	0.05	**277**	96.3	0.56	**68**	87.8	0.64	**70**

μ and σ^2 refer, respectively, to the mean and variance of the OA obtained over the 100 superpixels used. ET is presented in seconds.
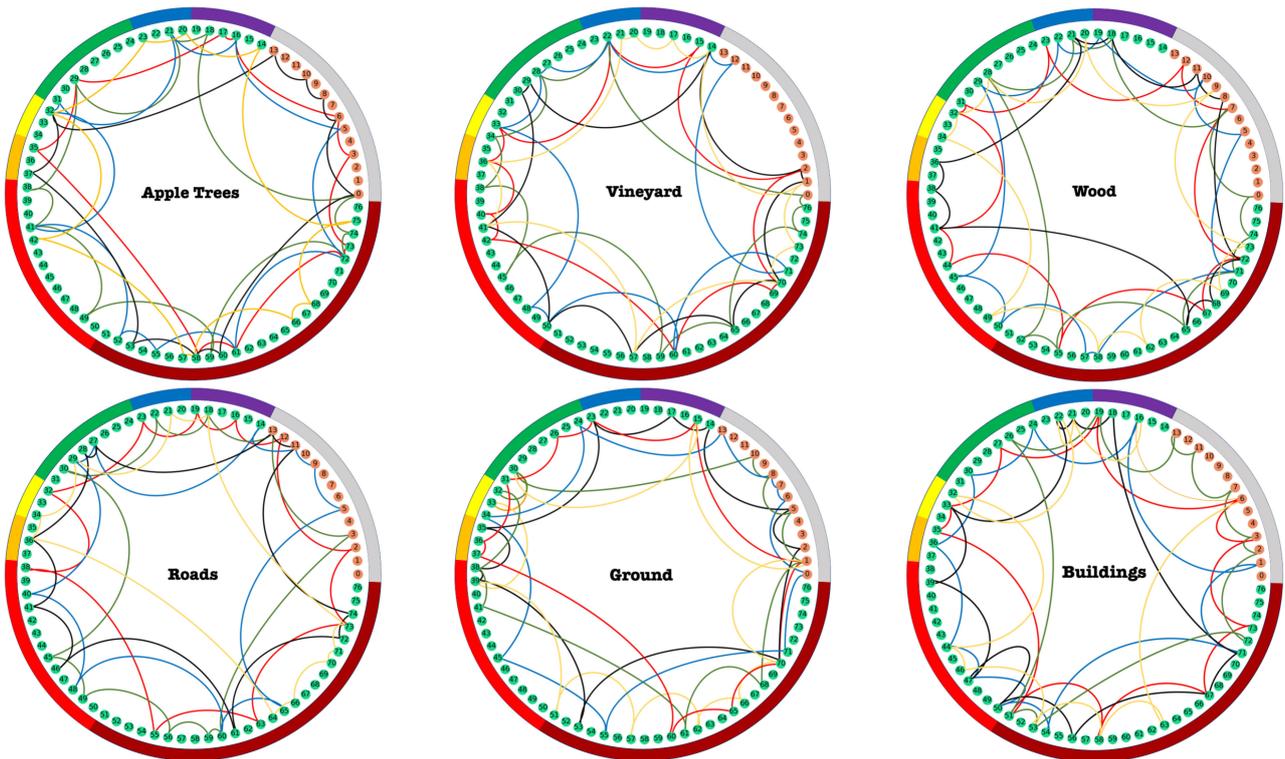
Fig. 13. Chord diagrams of selected attributes for different classes of the Trento dataset.

attributes, such as green—hyperspectral, and brown—LiDAR, while the edges illustrate the attributes that were selected by the GKMI method (color of the connections represent different homogeneous areas, i.e., superpixels). The outer circle depicts different macroscopic intervals of the spectral channels from the visible (violet 380–450 nm, blue 450–495 nm, green 495–570 nm, yellow 570–590 nm, orange 590–620 nm, and red 620–750 nm, according to the visible wavelength color representations) to the near-infrared (dark red 750–1300 nm) range with respect to the attribute numbers. The gray color represents the LiDAR attributes. Hence, the chord diagrams show that even for the same class, relevant attributes can vary and can be grouped differently. It means that if the various image parts represent the same class, they still might be observed under different technical or environmental conditions. Therefore, it is crucial to select the relevant attributes for separate zones of an image in order to reflect their particularity. The aforementioned results show the

flexibility and adaptivity of the proposed information selection scheme.

5) Selected Attributes: Additionally, in order to further investigate the effectiveness of the proposed approach, we analyzed the attribute selection method with datasets that includes corrupted attributes. Accordingly, to each dataset, we added a various number of corrupted attributes, which were randomly generated by Gaussian noise with different mean $\mu = [0.1, \ldots, 1]$ and standard deviation $\sigma = [0.1, \ldots, 1]$.

Fig. 14 shows the graph of occurrences of corrupted attributes for each dataset among a different number of noisy attributes added to the original datasets. Red color refers to attributes selected by SC, while blue color demonstrates the proposed method. It can be clearly seen from the curves that there is no clear superiority of any method for Trento and Houston datasets. For a different number of noisy attributes, each of the methods shows almost equal performance, with a

Fig. 14.   Occurrences of different number of noisy attributes for SC and proposed method (GKMI).

TABLE IX
MEAN μ AND VARIANCE σ^2 OF THE INTERCORRELATION BETWEEN THE SELECTED BANDS USING SC AND GKMI, OVER ALL SUPERPIXELS

Method	Berlin		Paris		Trento		Houston	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
SC	0.04	0.05	0.09	0.02	0.14	0.12	0.54	0.08
GKMI	**0.01**	0.05	**0.02**	0.02	**0.03**	0.04	**0.48**	0.01

slight advantage to one side or another. Nevertheless, for Berlin and Paris datasets, the predominance of the GKMI method becomes clearly visible. Moreover, for some parts, the percentage of selected noisy attributes using the GKMI method is several times less than using SC. Therefore, this result additionally strengthens the idea of applying two similarity metrics simultaneously.
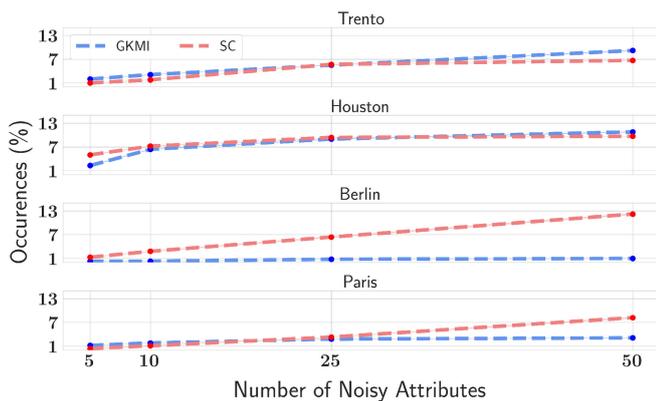
6) Correlation Sensitivity: As an alternative metric to assess the ability of the proposed method to effective dimensionality reduction, information dependence can be taken into account. This metric is one of the commonly used criteria for feature selection, especially for hyperspectral bands that are highly correlated. Specifically, computing the Pearson correlation coefficient provides insight into the strength of a linear association between two variables. Basically, a Pearson product moment correlation attempts to draw a line of best fit through the data of two variables [82, ch. 4]. The Pearson correlation coefficient indicates how far away all these data points are to this line of best fit (i.e., how well the data points fit this new model/line of best fit).

To determine the strength of association based on the Pearson correlation coefficient, it is possible to rely on the amplitude of the outcome. Specifically, if the linear relationship among attributes increases, then the Pearson correlation coefficient would increase as well. Therefore, the ability of a dimensionality reduction algorithm to identify the most informative subset of features in the dataset should show up in terms of low values of Pearson correlation coefficients [82].

In our case, it is particularly important to assess the necessity of including the MI Laplacian in the dimensionality reduction process in order to improve the ability to select informative attributes in the given dataset. To demonstrate the relevance of using MI, in Table IX, we show the intercorrelation between the selected attributes using the classic SC in (5) and GKMI. Table IX reports the mean correlation and variance among all superpixels. It is evident from the results that the incorporation of MI significantly decreases the correlation of selected attributes as opposed to the SC that only utilizes the GK.

Hence, given the previous observations, employing MI in the JD procedure as in (6) appears as a key step in order to enhance

the selection of relevant attributes delivered by a system based on a classic SC, especially when multimodal datasets are taken into account. Therefore, this is compliant with the results we have shown and commented on previously in this section and confirms from a statistical point of view the findings we have achieved when addressing the eigenanalysis of the selection capacity.

D. Method Comparison

In order to validate the proposed attribute selection method, we compare the achieved results with nine other dimensionality reduction algorithms:

1) *one ranking approach:* FIS;
2) *two attribute extraction methods:* PCA and DBFE;
3) *three searching strategies:* FS, OBB, and GA;
4) *three graph clustering approaches:* MST clustering, DS, and CD.

The aforementioned methods were described in detail earlier in Section II. It should be emphasized that, in this work, we do not compare our method with neural-network-based approaches since they require a training set.

Tables X–XIII report the performance comparison of the GKMI method with existing methods over various multimodal datasets and using two classifiers. It is evident from the tables that graph-based approaches outperform all the classical methods for feature selection in technical literature in terms of OA, AA, and Kappa since the latter methods are not flexible enough to process the multimodal datasets. However, GKMI ensures higher accuracies over all the considered datasets with the least number of attributes since it is performed on the superpixel level. Hence, it is possible to conclude that the proposed method was finding the best descriptive attributes for each homogeneous superpixel. Moreover, it is worth noting that the two similarity measures that are employed in the GKMI scheme are apparently able to ensure a more robust definition of the connections among vertices in the graphs associated with the considered datasets. This effect allows a better characterization of the subgraphs associated with the relevant attributes. Taking into account the observations drawn previously in this work (especially when considering the parameter sensitivity analysis and the trend of the eigenvalues in Fig. 10), these results further highlight the ability of GKMI to provide robust and reliable performance in selecting the most relevant attributes under diverse analysis conditions.

TABLE X
PERFORMANCE COMPARISON AMONG DIFFERENT DIMENSIONALITY REDUCTION METHODS AND DIFFERENT CLASSIFIERS FOR THE TRENTO DATASET

| Method | K | RF | | | | | | SVM | | | | | |
| | | OA (%) | | Kappa | AA (%) | | ET (sec) | OA (%) | | Kappa | AA (%) | | ET (sec) |
		μ	σ^2		μ	σ^2		μ	σ^2		μ	σ^2	
PCA	26	88.1	2.1	81.2	87.8	1.9	108	88.7	1.6	88.9	88.9	1.8	**12.1**
DBFE	25	86.3	2	86.3	86.1	1.8	110.2	82.9	1.5	83.8	84.1	1.9	13.4
FIS	25	84.2	1.8	84	83.2	1.6	109.6	85.1	1.4	84.6	85.7	1.3	12.6
FS	28	84	1.2	84	83.8	1.3	110.1	82.1	1.2	81.2	82.2	1	12.8
OBB	24	90.6	0.8	89.2	90.8	0.7	124.3	90.2	0.6	87.2	90	0.7	18.8
GA	23	90.4	0.7	89.4	90.3	0.6	123.2	90.3	0.55	89.2	90.5	0.58	17.9
MST	42	95.7	0.23	94.2	95.6	0.22	**13.1**	95.6	0.07	94.1	95.8	0.05	13.5
DS	38	88.7	0.04	84.9	88.2	0.08	14.9	86.7	0.03	82.2	85.9	0.05	24.5
CD	52	96.2	0.11	94.9	96.1	0.16	16.1	96.3	0.02	95.1	96.1	0.02	14.2
GKMI	**20**	95.0	0.33	93.4	94.4	0.73	139	93.9	0.28	91.8	93.2	0.38	41
	40	**97.3**	0.14	**96.4**	**96.8**	0.28	153	**97.4**	0.29	**96.2**	**96.6**	0.56	68

TABLE XI
PERFORMANCE COMPARISON AMONG DIFFERENT DIMENSIONALITY REDUCTION METHODS AND DIFFERENT CLASSIFIERS FOR THE HOUSTON DATASET

| Method | K | RF | | | | | | SVM | | | | | |
| | | OA (%) | | Kappa | AA (%) | | ET (sec) | OA (%) | | Kappa | AA (%) | | ET (sec) |
		μ	σ^2		μ	σ^2		μ	σ^2		μ	σ^2	
PCA	29	79.6	1.2	79.7	79.8	1.1	188	83.2	1.1	81.0	83.6	1.3	98
DBFE	31	78.9	1.3	77.6	78.8	1.09	196	79.4	1.12	78.6	79.5	1.07	99
FIS	30	80.3	1.02	80.0	80.4	1.04	188.8	80.6	1.08	80.2	80.6	1.05	102.4
FS	30	77.5	1.3	77.1	77.5	1.21	190	77.2	1.15	76.8	77.3	1.18	100.8
OBB	26	78.5	1.11	78.4	78.7	1.09	209	77.2	1.13	75.9	77.3	1.08	116
GA	23	80.2	1.03	80.1	80.2	1.04	200.1	78.6	1.05	77.3	78.7	1.02	111
MST	40	84.9	1.68	84.0	85.6	1.71	53.5	85.7	0.72	84.7	86.3	0.65	43
DS	30	71.1	1.09	69.7	72.3	1.47	**45**	77.5	1.05	75.9	78.1	1.01	**31**
CD	64	85.0	1.98	83.9	85.6	1.79	52	85.6	0.78	84.8	86.7	0.59	47
GKMI	**20**	82.6	1.67	81.3	83.0	1.56	150	85.2	0.88	84.2	85.5	0.97	69
	60	**86.5**	1.41	**85.5**	**86.6**	1.42	161	**88.7**	0.71	**87.9**	**89.1**	0.58	88

TABLE XII
PERFORMANCE COMPARISON AMONG DIFFERENT DIMENSIONALITY REDUCTION METHODS AND DIFFERENT CLASSIFIERS FOR THE BERLIN DATASET

| Method | K | RF | | | | | | SVM | | | | | |
| | | OA (%) | | Kappa | AA (%) | | ET (sec) | OA (%) | | Kappa | AA (%) | | ET (sec) |
		μ	σ^2		μ	σ^2		μ	σ^2		μ	σ^2	
PCA	47	86.2	0.13	85.3	86.3	0.13	312	85.7	0.11	84.9	85.8	0.12	319
DBFE	48	86.5	0.11	86.3	86.4	0.12	313.6	85.4	0.09	85.2	85.4	0.1	320.3
FIS	47	86.6	0.11	86.5	86.8	0.12	315	86.3	0.11	85.8	86.3	0.1	323
FS	49	83.6	0.1	82.4	83.6	0.08	316.2	82.6	0.09	82.1	82.8	0.11	321
OBB	45	85.9	0.08	85.7	85.9	0.07	323	85.8	0.07	84.7	85.9	0.06	330
GA	45	86.4	0.07	86.1	86.5	0.05	320	85.7	0.04	85.2	85.8	0.05	327
MST	41	88.4	0.12	86.6	88.3	0.51	61	91.1	0.23	89.8	89.6	0.39	184
DS	68	90.1	0.12	88.6	92.8	0.08	56.8	69.6	2.51	63.6	94.6	0.02	196
CD	61	89.3	0.13	87.6	89.3	0.24	**47**	92.4	0.16	91.2	91.2	0.36	**94**
GKMI	**40**	93.5	0.22	92.6	94.9	0.17	242	94.7	0.14	93.2	93.6	0.28	243
	60	**94.2**	0.18	**93.4**	**95.5**	0.11	253	**95.5**	0.11	**94.8**	**95.2**	0.20	359

TABLE XIII
PERFORMANCE COMPARISON AMONG DIFFERENT DIMENSIONALITY REDUCTION METHODS AND DIFFERENT CLASSIFIERS FOR THE PARIS DATASET

Method	K	RF						SVM					
		OA (%)		Kappa	AA (%)		ET (sec)	OA (%)		Kappa	AA (%)		ET (sec)
		μ	σ^2		μ	σ^2		μ	σ^2		μ	σ^2	
PCA	36	88.7	0.06	87.7	88.8	0.05	331	88.4	0.08	87.9	88.5	0.07	279
DBFE	35	88.4	0.07	88.3	88.6	0.07	333	86.2	0.08	85.3	86.3	0.06	281
FIS	35	90.6	0.05	89.2	90.6	0.06	332	89.8	0.04	88.9	89.9	0.06	281
FS	37	86.3	0.06	85.6	86.3	0.05	332	86.2	0.05	85.9	86.3	0.07	283
OBB	36	87.8	0.04	87.4	87.9	0.03	344	87.5	0.03	87.1	87.6	0.02	296
GA	33	90.3	0.03	88.6	90.4	0.02	342	90.8	0.04	89.2	90.9	0.03	290
MST	41	93.8	0.06	91.8	88.6	9.83	**110**	82.4	0.15	75.3	91.0	18.03	257
DS	67	91.7	0.12	89.0	93.0	0.25	108	96.8	0.08	95.8	95.7	0.13	283
CD	83	94.0	0.02	92.0	88.4	0.98	121	82.0	0.11	74.7	91.1	0.13	443
GKMI	**30**	94.5	0.09	92.7	94.6	0.36	282	96.4	0.17	95.2	92.7	1.43	**248**
	60	**95.9**	0.07	**94.7**	**95.7**	0.22	299	**98.1**	0.03	**97.5**	**96.8**	0.46	336

V. CONCLUSION

A new unsupervised attribute selection method based on two different similarity measures has been proposed for multimodal remote sensing data. The main merits of the method are as follows.

1) *Unsupervision:* The method is application independent; therefore, it is implemented without any prior information about class labels.
2) *Flexibility:* It can be applied to datasets obtained from various sensors with different characteristics.
3) *Accuracy:* It employs two similarities that account for global and local particularities of the original dataset, which, in turn, allows selecting the most relevant attributes.
4) *Versatility:* The method is performed on a superpixel level; therefore, it selects the best descriptive attributes for each homogeneous superpixel.
5) *Interpretability:* The method retains the advantages of both attribute extraction and selection methods (preserves the physical meaning of the data and increases the separability).

The experimental results obtained from several multimodal datasets consistently demonstrated the effectiveness and robustness of the proposed method for the processing of the multimodal remote sensing datasets.

This article introduces the GKMI attribute selection method with all its crucial steps and relevant novelties. Future work directions will be focused on adding the automatic selection of the number of attributes for each superpixel, so that the multimodal data analysis can be adapted to the different conditions of the records that can be acquired on large-scale scenarios, and on developing an adaptive classifier that can deal with superpixels of heterogeneous sizes and attributes.

ACKNOWLEDGMENT

REFERENCES

[1] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[2] M. D. Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[3] N. Longbotham et al., "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009-2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.

[4] S. Chlaily, P. Amblard, O. Michel, and C. Jutten, "Impact of noise correlation on multimodality," in *Proc. 24th Eur. Signal Process. Conf.*, Aug. 2016, pp. 195–199.

[5] S. B. Serpico, M. D'Inca, F. Melgani, and G. Moser, "Comparison of feature reduction techniques for classification of hyperspectral remote sensing data," *Proc. SPIE*, vol. 4885, pp. 347–358, 2003.

[6] S. Georganos et al., "Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application," *GISci. Remote Sens.*, vol. 55, no. 2, pp. 221–242, 2018.

[7] J. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[8] C. Bichot and P. Siarry, *Graph Partitioning*. Hoboken, NJ, USA: Wiley/ISTE, 2013.

[9] U. Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, pp. 395–416, Dec. 2007.

[10] L. Gulikers, M. Lelarge, and L. Massoulié, "A spectral method for community detection in moderately sparse degree-corrected stochastic block models," *Adv. Appl. Probab.*, vol. 49, no. 3, pp. 686–721, 2017.

[11] H. T. Ali and R. Couillet, "Improved spectral community detection in large heterogeneous networks," *J. Mach. Learn. Res.*, vol. 18, no. 225, pp. 1–49, 2018.

[12] L. Dall'Amico, R. Couillet, and N. Tremblay, "Optimal laplacian regularization for sparse spectral community detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 3237–3241.

[13] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 907–948, 2020.

[14] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 175–186, 2014.

[15] M. A. Hossain, M. Pickering, and X. Jia, "Unsupervised feature extraction based on a mutual information measure for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 1720–1723.

[16] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.

[17] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Orlando, FL, USA: Academic, 2008.

[18] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images," *Pattern Recognit.*, vol. 51, pp. 295–309, 2016.

[19] Y. Zhou, R. Zhang, S. Wang, and F. Wang, "Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy," *Sensors*, vol. 18, 2018, Art. no. 2013.

[20] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," in *Proc. 27th Conf. Uncertainty Artif. Intell.*, 2012, pp. 266–273.

[21] S. Sivakumar and C. Chandrasekar, "Feature selection using genetic algorithm with mutual information," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 2871–2874, 2014.

[22] P. Somol, P. Pudil, F. J. Ferri, and J. Kittler, "Fast branch & bound algorithm in feature selection," in *Proc. 4th World Multiconf. Syst., Cybern., Inform.*, Orlando, FL, USA, 2000, vol. 7, pp. 646–651.

[23] M. Fauvel, C. Dechesne, A. Zullo, and F. Ferraty, "Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2824–2831, Jun. 2015.

[24] M. Ahmad, D. Ulhaq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *Int. J. Eng. Technol.*, vol. 3, pp. 606–614, Dec. 2011.

[25] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.

[26] J. Feng, L. Jiao, T. Sun, H. Liu, and X. Zhang, "Multiple kernel learning based on discriminative kernel clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6516–6530, Nov. 2016.

[27] S. Schaeffer, "Graph clustering," *Comput. Sci. Rev.*, vol. 1, pp. 27–64, 2007.

[28] M. T. Altuncu, E. Mayer, S. N. Yaliraki, and M. Barahona, "From free text to clusters of content in health records: An unsupervised graph partitioning approach," *Appl. Netw. Sci.*, vol. 4, 2019, Art. no. 2.

[29] M. Schaub, J.-C. Delvenne, R. Lambiotte, and M. Barahona, "Multiscale dynamical embeddings of complex networks," *Phys. Rev. E*, vol. 99, 2019, Art. no. 062308.

[30] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.

[31] T. Berry and T. Sauer, "Consistent manifold representation for topological data analysis," *Found. Data Sci.*, vol. 1, no. 1, pp. 1–38, 2019.

[32] M. Carreira-Perpiñán and R. Zemel, "Proximity graphs for clustering and manifold learning," in *Proc. 17th Int. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2004, pp. 225–232.

[33] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[34] O. Grygorash, Y. Zhou, and Z. Jorgensen, "Minimum spanning tree based clustering algorithms," in *Proc. 18th IEEE Int. Conf. Tools Artif. Intell.*, 2006, pp. 73–81.

[35] M. Beguerisse-Díaz, B. Vangelov, and M. Barahona, "Finding role communities in directed networks using role-based similarity, Markov stability and the relaxed minimum spanning tree," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Austin, TX, USA, 2013, pp. 937–940.

[36] R. Liu, S. Feng, R. Shi, and W. Guo, "Weighted graph clustering for community detection of large social networks," *Procedia Comput. Sci.*, vol. 31, pp. 85–94, 2014.

[37] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 167–172, Jan. 2007.

[38] R. Lambiotte, J. Delvenne, and M. Barahona, "Random walks, Markov processes and the multiscale modular organization of complex networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 1, no. 2, pp. 76–90, Jul.–Dec. 2014.

[39] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, "Stability of graph communities across time scales," *Proc. Nat. Acad. Sci.*, vol. 107, no. 29, pp. 12755–12760, 2010.

[40] R. Tripodi, S. Vascon, and M. Pelillo, "Context aware nonnegative matrix factorization clustering," in *Proc. IEEE Int. Conf. Pattern Recognit.*, 2016, pp. 1719–1724.

[41] S. Vascon, M. Cristani, M. Pelillo, and V. Murino, "Using dominant sets for k-NN prototype selection," in *Proc. Int. Conf. Image Anal. Process.*, 2013, pp. 131–140.

[42] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[43] X. Xu, H. Gu, Y. Wang, J. Wang, and P. Qin, "Autoencoder based feature selection method for classification of anticancer drug response," *Front. Genetics*, vol. 10, 2019, Art. no. 233.

[44] D. Tomar, Y. Prasad, M. K. Thakur, and K. K. Biswas, "Feature selection using autoencoders," in *Proc. Int. Conf. Mach. Learn. Data Sci.*, Dec. 2017, pp. 56–60.

[45] R. Couillet and M. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *J. Multivariate Anal.*, vol. 131, pp. 99–120, 2014.

[46] J. Hu, D. Hong, Y. Wang, and X. Zhu, "A comparative review of manifold learning techniques for hyperspectral and polarimetric SAR image fusion," *Remote Sens.*, vol. 11, no. 6, pp. 1–28, 2019.

[47] D. Tuia, M. Volpi, M. Trolliet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7708–7720, Dec. 2014.

[48] D. Hong, J. Kang, N. Yokoya, and J. Chanussot, "Graph-induced aligned learning on subspaces for hyperspectral and multispectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4407–4418, May 2021.

[49] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[50] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[51] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal GANs: Toward crossmodal hyperspectral-multispectral image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5103–5113, Jun. 2021.

[52] D. Hong, N. Yokoya, G.-S. Xia, J. Chanussot, and X. X. Zhu, "X-modalNet: A semi-supervised deep cross-modal network for classification of remote sensing data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 12–23, 2020.

[53] G. Iyer, J. Chanussot, and A. L. Bertozzi, "A graph-based approach for data fusion and segmentation of multimodal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4419–4429, May 2021.

[54] J. Xia, W. Liao, and P. Du, "Hyperspectral and Lidar classification with semisupervised graph fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 666–670, Apr. 2020.

[55] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LEMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019.

[56] J. Dy and C. Brodley, "Feature selection for unsupervised learning," *J. Mach. Learn. Res.*, vol. 5, pp. 845–889, Aug. 2004.

[57] S. Doan and S. Horiguchi, "An efficient feature selection using multi-criteria in text categorization for Naïve Bayes classifier," *WSEAS Trans. Inf. Sci. Appl.*, vol. 2, no. 2, 2005, Art. no. 34.

[58] L. Rokach, B. Chizi, and O. Maimon, "Feature selection by combining multiple methods," in *Advances in Web Intelligence and Data Mining*, vol. 23. New York, NY, USA: Springer, 2006, pp. 295–304.

[59] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1997, pp. 731–737.

[60] G. Strang, *Linear Algebra and its Applications*. Belmont, CA, USA: Thomson, Brooks/Cole, 2006.

[61] P. Ablin, J. Cardoso, and A. Gramfort, "Beyond Pham's algorithm for joint diagonalization," 2018, *arXiv:1811.11433*.

[62] Dinh-Tuan Pham and J. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.

[63] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2001, pp. 849–856.

[64] A. Plaza and C.-I. Chang, *High Performance Computing in Remote Sensing*. New York, NY, USA: CRC Press, 2007.

[65] Y. Liu, Q. Ren, J. Geng, M. Ding, and J. Li, "Efficient patch-wise semantic segmentation for large-scale remote sensing images," *Sensors (Switzerland)*, vol. 18, no. 10, pp. 1–16, 2018.

[66] S. Beucher, "The watershed transformation applied to image segmentation," in *Proc. 10th Pfefferkorn Conf. Signal Image Process. Microsc. Microanal.*, 1992, pp. 299–314.

[67] P. Neubert and P. Protzel, "Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms," in *Proc. Int. Conf. Pattern Recognit.*, 2014, pp. 996–1001.

[68] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[69] A. Zafari, R. Zurita-Milla, and E. Izquierdo-Verdiguier, "Evaluating the performance of a random forest kernel for land cover classification," *Remote Sens.*, vol. 11, no. 5, 2019, Art. no. 575.

[70] J. Xia, N. Falco, J. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1601–1609, Apr. 2017.

[71] S. R. Gunn, "Support vector machines for classification and regression," 1998.

[72] A. Liaw and M. Wiener, "Classification and regression by random forest," *Forest*, vol. 23, pp. 18–23, Nov. 2001.

[73] P. Bharatkar and R. Patel, "Approach to accuracy assessment tor RS image classification techniques," *Int. J. Sci. Eng. Res.*, vol. 4, no. 12, pp. 79–86, 2013.

[74] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[75] U. Kandaswamy, D. A. Adjeroh, and M. C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2075–2083, Sep. 2005.

[76] F. Albregtsen, "Statistical texture measures computed from gray level coocurrence," *Boundary 2*, vol. 3, no. 1, p. 45, 1974.

[77] *2017 IEEE GRSS data fusion contest*, 2017. [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/2017-ieee-grss-data-fusion-contest-2/.

[78] *2013 IEEE GRSS data fusion contest*, 2013. [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-contest/.

[79] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. New York, NY, USA: Wiley, 2009.

[80] S. Prasad, L. Bruce, and J. Chanussot, *Optical Remote Sensing, Augmented Vision and Reality*. Berlin, Germany: Springer, 2011.

[81] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, 2012.

[82] R. Rousseau, L. Egghe, and R. Guns, Eds., *Becoming Metric-Wise* (Chandos Information Professional Series). Amsterdam, The Netherlands: Elsevier, 2018.

Eduard Khachatrian (Student Member, IEEE) received a double M.Sc. degrees in polar and marine sciences from the Faculty of Mathematics, Informatics, and Natural Sciences, University of Hamburg, Hamburg, Germany, and the Institute of Earth Sciences, Saint Petersburg State University, Saint Petersburg, Russia, in 2017. He is currently working toward the Ph.D. degree with the Center of Integrated Remote Sensing and Forecasting for Arctic Operations, University of Tromsø—The Arctic University of Norway, Tromsø, Norway.

From 2017 to 2018, he was a Junior Scientist with the Nansen International Environmental and Remote Sensing Centre, Bergen, Norway. His research interests include multimodal data analysis, image processing, and remote sensing of polar areas.

Saloua Chlaily (Member IEEE) received the M.Sc. degree in electronics engineering from the École Nationale Supérieure d'Électrotechnique, d'Électronique, d'Informatique, d'Hydraulique et des Télécommunications, Toulouse, France, and the M.Sc. degree in electrical engineering from the Hassania School of Public Works, Casablanca, Morocco, both in 2013, and the Ph.D. degree in signal, image, speech, and telecommunication from Grenoble Alpes University, Grenoble, France, in 2018.

She is currently a Researcher with University of Tromsø—The Arctic University of Norway (UiT). She conducts her research with the Center of Integrated Remote Sensing and Forecasting for Arctic Operations, UiT. Her research interests include multimodal data analysis, signal processing, and image processing.

Dr. Chlaily is a Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Torbjørn Eltoft (Member, IEEE) received the M.Sc. and Ph.D. degrees from the University of Tromsø, Tromsø, Norway, in 1981 and 1984, respectively.

In 1988, he joined the Department of Physics and Technology, University of Tromsø—The Arctic University of Norway (UiT), Tromsø, where he is currently a Professor and the Director of the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations, a center for research-based Innovation awarded by the Norwegian Research Council in 2014, whose objective is to develop knowledge and remote sensing technology for arctic applications. From 2013 to 2015, he was the Head of the Department of Physics and Technology, UiT. He was an International Researcher with the University of California, Irvine, CA, USA, from 1992 to 1993 and then from 1997 to 1998, and with the University of California, San Diego, CA, from 2004 to 2005. He has a significant publication record in the areas of signal processing and remote sensing. His research interests include multidimensional signal and image analysis, statistical modeling, neural networks, and machine learning, with emphasis on applications in multichannel synthetic aperture radar remote sensing and multisensor remote sensing for the Arctic.

Dr. Eltoft was an Associate Editor for *Pattern Recognition* from 2005 to 2011 and a Guest Editor for *Remote Sensing*'s Special Issue for the PolInSAR 2017 Conference. He was the co-recipient of the Outstanding Paper Award in Neural Networks awarded by the IEEE Neural Networks Council in 2000, Honorable Mention for the 2003 *Pattern Recognition* Journal Best Paper Award, and the 2017 UiT Award for Research and Development from the UiT.

Andrea Marinoni (Senior Member, IEEE) received the B.S., M.Sc. (*cum laude*), and Ph.D. degrees in electronic engineering from the University of Pavia, Pavia, Italy, in 2005, 2007 and 2011, respectively.

He is currently an Associate Professor with the Earth Observation Group, Centre for Integrated Remote Sensing and Forecasting for Arctic Operations, Department of Physics and Technology, University of Tromsø—The Arctic University of Norway, Tromsø, Norway, and a Visiting Academic Fellow with the Department of Engineering, University of Cambridge, Cambridge, U.K. From 2013 to 2018, he was a Research Fellow with the Telecommunications and Remote Sensing Laboratory, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy. In 2009, he has been a Visiting Researcher with the Communications Systems Laboratory, Department of Electrical Engineering, University of California, Los Angeles, CA, USA. In 2011, he was the recipient of the two-year "Applied research grant," sponsored by the Region of Lombardy, Italy, and STMicroelectronics N.V. In 2017, he was the recipient of the INROAD grant, sponsored by the University of Pavia and Fondazione Cariplo, Italy, for supporting excellence in design of European Research Council proposal. In 2018, he was the recipient of the "Progetto professionalità Ivano Becchi" grant funded by the Fondazione Banco del Monte di Lombardia, Italy, and sponsored by the University of Pavia and the NASA Jet Propulsion Laboratory, Pasadena, CA, for supporting the development of advanced methods of air pollution analysis by remote sensing data investigation. He was the recipient of Asgard Research Program and Asgard Recherche+ Program grants funded by the Institut Français de Norvége, Oslo, Norway, in 2019 and 2020, respectively, for supporting the development of scientific collaborations between French and Norwegian research institutes. From 2015 to 2017, he was a Visiting Researcher at the Earth and Planetary Image Facility, Ben Gurion University of the Negev, Be'er Sheva, Israel; the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China; the School of Computer Science, Fudan University, Shanghai, China; the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China; and the Instituto de Telecomunicações, Instituto Superior Tecnico, Universidade de Lisboa, Lisbon, Portugal. In 2020 and 2021, he was a Visiting Professor with the Department of Electrical, Computer and Biomedical Engineering, University of Pavia. His main research interests include efficient information extraction from multimodal remote sensing, nonlinear signal processing applied to large-scale heterogeneous records, Earth observation interpretation and Big Data mining, and analysis and management for human–environment interaction assessment.

Dr. Marinoni is the Founder and Current Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Norway Chapter. He is also an Ambassador for IEEE Region 8 Humanitarian activities, and a research contact point for the Norwegian Artificial Intelligence Research Consortium. He serves as a Topical Associate Editor of machine learning for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Guest Editor of three special issues on Multimodal Remote Sensing and Sustainable Development for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He is the Leader of the GR4S Committee of the IEEE GRSS, coordinating the organization of schools and workshops sponsored by the IEEE GRSS worldwide.

/9

Paper 2: Automatic Selection of Relevant Attributes for Multi-Sensor Remote Sensing Analysis: A Case Study on Sea Ice Classification

Eduard Khachatrian, Saloua Chlaily, Torbjørn Eltoft, Wolfgang Dierking, Frode Dinessen, and Andrea Marinoni

Automatic Selection of Relevant Attributes for Multi-Sensor Remote Sensing Analysis: A Case Study on Sea Ice Classification

Eduard Khachatrian ⓘ, *Student Member, IEEE*, Saloua Chlaily ⓘ, *Member, IEEE*, Torbjørn Eltoft ⓘ, *Member, IEEE*, Wolfgang Dierking ⓘ, Frode Dinessen, and Andrea Marinoni ⓘ, *Senior Member, IEEE*

Abstract—It is of considerable benefit to combine information obtained from different satellite sensors to achieve advanced and improved characterization of sea ice conditions. However, it is also true that not all the information is relevant. It may be redundant, corrupted, or unnecessary for the given task, hence decreasing the performance of the algorithms. Therefore, it is crucial to select an optimal set of image attributes which provides the relevant information content to enhance the efficiency and accuracy of the image interpretation and retrieval of geophysical parameters. Comprehensive studies have been focused on the analysis of relevant features for sea ice analysis obtained from different sensors, especially synthetic aperture radar. However, the outcomes of these studies are mostly data and application-dependent and can, therefore, rarely be generalized. In this article, we employ a feature selection method based on graph Laplacians, which is fully automatic and easy to implement. The proposed approach assesses relevant information on a global and local level using two metrics and selects relevant features for different regions of an image according to their physical characteristics and observation conditions. In the recent study, we investigate the effectiveness of this approach for sea ice classification, using different multi-sensor data combinations. Experiments show the advantage of applying multi-sensor data sets and demonstrate that the attributes selected by our method result in high classification accuracies. We demonstrate that our approach automatically considers varying technical, sensor-specific, environmental, and sea ice conditions by employing flexible and adaptive feature selection method as a pre-processing step.

Index Terms—Graph Laplacians, multi-sensor remote sensing, sea ice, unsupervised information selection.

I. INTRODUCTION

IN THE last decades, sea ice research has become a focus of Earth observation, especially in the Arctic region where sea ice extent and volume are declining rapidly [1]. Sea ice plays an essential role in the polar ecosystem [2]. It is one of the main indicators of global climate change, and it also considerably affects the indigenous population's lifestyle and welfare [2]. Moreover, sea ice poses a great challenge to navigation in the polar seas; therefore, the knowledge about its type, concentration, thickness, deformation, and extent is extremely important for various activities, such as marine transportation and offshore operations, and for stakeholders from the oil and gas industry, fisheries, and tourism, among others.

Sea ice observations have a long history of more than a century. They were carried out visually from coastal stations, ships, and aircraft [3], while they were spatially and temporally limited. Regular sea ice monitoring over larger regions became possible in the late 1970s using image data from satellites [3]. Since then, the technologies for acquiring and analyzing sea ice data have been considerably improved and extended.

Optical sensors provide information about sea ice in visible and infrared channels. The main advantage of this kind of information is an easier visual interpretability. However, optical sensors are limited to cloud-free and favorable light conditions, which causes a significant problem since dense cloud covers and long periods of darkness prevail in the polar regions for several months of the year. Therefore, spaceborne synthetic aperture radar (SAR), which combines high spatial resolution and independence of cloud and light conditions, is the main source from which detailed maps of sea ice conditions are produced. However, the automatic interpretation of remote sensing data, especially SAR data, is challenging and strongly relies on expert's knowledge. Passive microwave radiometers are another type of sensor that can be used for sea ice observations. However, in comparison to the aforementioned techniques, it has a significantly coarser spatial resolution and is, therefore, preferably used for global or large-scale observations [3]. The increasing amount of available satellite data together with more and more activities in sea ice covered waters requires a greater effort for supplementing the production of ice charts by employing fully automated methods of information selection and image analysis [3], [4].

Multi-sensor remote sensing refers to the use of different sensors usually operating at different frequencies and/or spatial resolutions and coverage and with different time intervals between data acquisitions over a given region. Therefore, diverse remote sensing systems grasp various properties of sea ice by using different physical principles. Thus, combining the information from multiple sensors allows better characterization of sea ice [5]. Nevertheless, although different data sources provide complementary information, they can potentially also include redundant, corrupted, or simply unnecessary information for a given task. Accordingly, combining these data can significantly deteriorate the performance by decreasing the accuracy and increasing the computation time and structural complexity of the algorithm. Thus, the selection of relevant information from the original data set is an essential step of multi-source data fusion that enables reliable and efficient performance [6], [7].

We note that the term "feature" is commonly used in fields such as classification methodologies, pattern recognition, and texture analysis. In our study, however, we use the notation "attribute" from information theory which refers to directly measured quantities such as, e.g., radar intensities at different frequencies and polarizations, optical reflectance, and parameters such as textural or polarimetric features. The pertinence of different attributes for sea ice characterization has been investigated for both SAR [4], [7]–[14] and optical imagery [15], [16]. In some studies, e.g., the entropy was found to be well suited for separating sea ice types [13]. In other studies, however, the same parameter was found to be less relevant for sea ice classification [8], [11] and less useful for detection of leads in the ice [9], [12]. In fact, it has been shown that the relevance of single attributes varies with the observation conditions such as incidence angle and season [7], [9]. Accordingly, the results of such analyses cannot be generalized and emphasize the need for an automatic attribute selection algorithm.

The selection of relevant information can be achieved using dimensionality reduction methods that are generally divided into two main categories: attribute extraction and attribute selection [17]. Extraction methods generate a prominent set of attributes by projecting the original set into a lower-dimensional space. Among such methods, we can highlight the principal component analysis (PCA) [18]. Selection algorithms search for the most relevant elements, according to the given criteria, within the original set, without applying any transformation. One example of this category is the forward selection (FS) [19] that determines a subset of attributes incrementally, by adding at each step the attribute that leads to the largest improvement of the information content. The transformation applied by the attribute extraction approaches improves the separation of different data classes but at the expense of reducing the physical interpretability, which may be a disadvantage for remote sensing data analyses. Unlike the attribute extraction approach, attribute selection preserves the data's physical interpretability by working in the same space. The approaches mentioned above can be classified as supervised [20], if they require training data, or unsupervised [21]. In the case of remote sensing data, especially when dealing with complex scenes or considering modalities (i.e., various sensors characterized by different acquisition techniques) that are difficult to interpret such as sea ice SAR images, providing accurate labels is challenging even with the assistance of an expert. Therefore, the unsupervised attribute selection method is the right choice in the case of multi-sensor remote sensing sea ice analysis.

In this article, we use a recently developed graph-based method (referred to as GKMI) [22] that relies on information theory metrics to capture the most relevant attributes for different sea ice classes. The acronym GKMI recalls the Gaussian kernel (GK) and the mutual information (MI) used as similarity measures within this approach. GKMI combines the advantages of both attribute extraction and selection approaches since it generates a new representation of the attributes that can be reassigned to the original ones. As such, it increases their separability (even if they are non-linearly separable) while preserving their physical interpretability. Moreover, it selects relevant attributes for separate zones of an image that might belong to different ice classes and/or are measured under different conditions (e.g., different radar incidence angles, varying sun elevation angles). Accordingly, it enhances the classifier's ability to separate different sea ice classes even in challenging scenarios.

The considered method was paired with a supervised classification approach [23]. Regions that appear homogeneous in a single image and can also be identified in images acquired from different sensors are labeled as a certain ice type with the help of sea ice experts. In our study, we focus on data that were acquired under freezing conditions. Hence, the resulting attribute selections cannot be generalized for the melting season or freeze–melt cycles. However, the GKMI method can be employed in the same manner to other than freezing conditions.

The rest of this article is organized as follows. Section II describes the data sets used in this study. Section III provides details of the proposed architecture. Section IV presents an experimental validation of the proposed method. Finally, the discussion and conclusion are presented in Section V.

II. Data Sets

This section describes the different data sets, consisting of SAR and optical images that we used in our analysis. To increase the validity range of our conclusions, we consider several data sets of different spatial resolutions representing various sea ice conditions: a multi-frequency data set from an airborne SAR measurement campaign (ICESAR), two multi-sensor data sets combining Radarsat-2 and Landsat-8 imagery, and Sentinel-1 and Sentinel-2 imagery. The data sets were labeled based on direct field observations (ICESAR) or by analysts from operational ice services and sea ice experts from institutions contributing to this article.

Table I provides general information about the data sets, and Fig. 1 illustrates their geographical location. SAR images were acquired in linear polarization, and the notation "HV" means that a horizontally polarized wave was transmitted and the backscattered signal was received at vertical polarization.

A more detailed description of each data set is presented in the following subsections.

TABLE I
CHARACTERISTICS OF REMOTELY SENSED DATA SETS USED IN THIS STUDY. INTENSITIES ARE GIVEN AS BACKSCATTERING COEFFICIENT SIGMA-NOUGHT

| Name	Date	Region	Measured Quantities	Pixel Size [m]	Coverage [km]
ICESAR	March 20, 2007	Fram Strait	L-band : HV, HH, VV Intensities C-band : VH, VV Intensities	1.5	5.3×1.5
Sentinel-1/Sentinel-2	April 08, 2018	North-East of Svalbard	SAR : HH and HV Intensities Optical : 13 bands from $0.443\,\mu m$ - $2.19\,\mu m$	60	87.5×109.8
Radarsat-2/Landsat-8	April 25, 2018	North-East of Svalbard	SAR : HH and HV Intensities Optical : 11 bands from $0.433\,\mu m$ - $12.5\,\mu m$	100	102.5×35.5
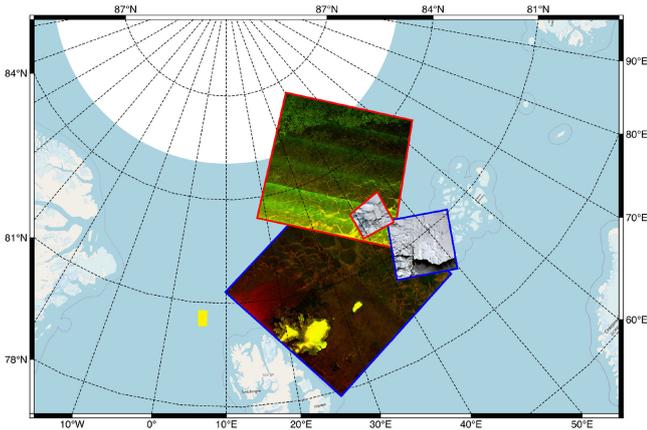
Fig. 1. Location of the data sets used in this study. Boundary colors refer to a specific data set: Sentinel-1/Sentinel-2 (Red), Radarsat-2/Landsat-8 (Blue). Given the small area of coverage of the ICESAR data set, an approximation of its location is given by a yellow square. It should be noted that only the overlapped area of the multi-sensor data sets, mainly Sentinel-1/Sentinel-2 and Radarsat-2/Landsat-8, are used.

Fig. 2. False-color composite representation of ICESAR data set. (a) C-band (VH, VV, and VV as RGB). (b) L-band (HV, HH, and VV as RGB).

A. ICESAR

From the ICESAR measurement campaign, we used images acquired by the airborne SAR of the German Aerospace Center (DLR) and an optical scanner operated on an aircraft of the Alfred Wegener Institute (AWI). The SAR data set was recorded at C-band (dual-polarization, VH, and VV) and L-band (quad-polarization, HV, HH, and VV). The time difference between C-band and L-band measurements varies approximately between 10 and 30 min. A more detailed description can be found in [4] and [24]. The ICESAR data set includes six classes which are open water, nilas, grey ice, grey-white ice, level first-year ice (FYI) and deformed FYI according to the WMO sea ice nomenclature [25]. Fig. 2 shows the false-color composite of one scene at C- and L-bands.

B. Sentinel-1/Sentinel-2

The Sentinel-1/Sentinel-2 multi-sensor data set consists of SAR and optical data obtained from Sentinel-1 and Sentinel-2 satellites. The time difference between SAR and optical data was less than an hour. Sentinel-1 imagery was acquired in extra-wide (EW) swath mode at dual polarization (HH and HV), which is commonly used for sea ice monitoring. From Sentinel-2, we used 13 bands in the visible, near-infrared, and shortwave infrared part of the spectrum. Sentinel-1 data have a pixel size
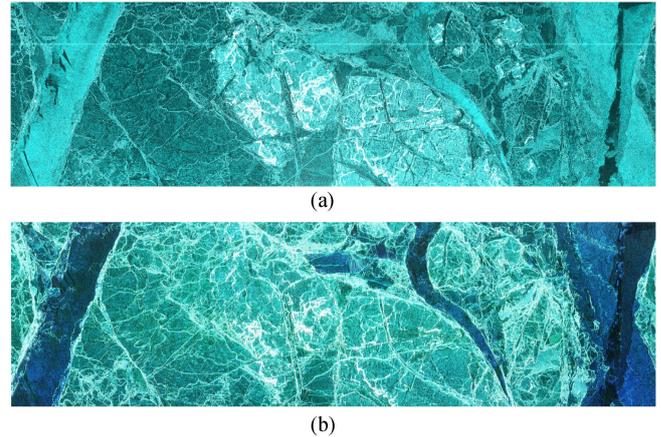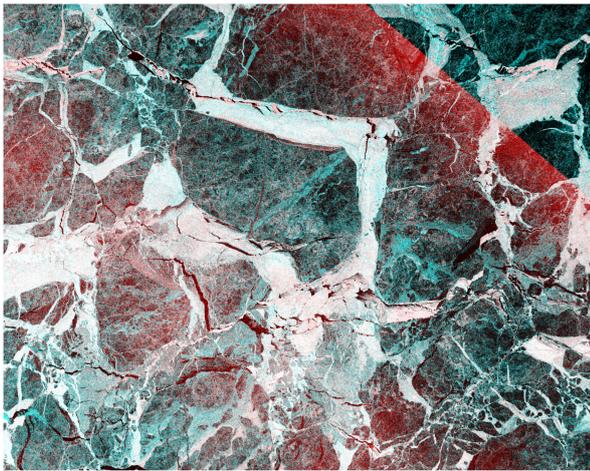
of 40 m, while the pixel size of Sentinel-2 varies from 10 to 60 m depending on the spectral band. Both Sentinel-1 and Sentinel-2 data sets were downsampled to the same pixel size of 60 m. The data set includes several classes, such as thick FYI, thin FYI, grey ice, grey-white ice and open water. Fig. 3 shows the SAR false-color composite and optical natural color composite images of the Sentinel-1/Sentinel-2 data set. In Fig. 3(b), clouds can be recognized, which is quite common for optical data.

C. Radarsat-2/Landsat-8

The Radarsat-2/Landsat-8 consists of optical and SAR data acquired from Landsat-8 and Radarsat-2, respectively. The time difference between SAR and optical data was less than an hour. The SAR images were acquired at HH and HV polarizations. From Landsat-8, we have images for 11 different spectral bands. The Radarsat-2 ScanSAR Wide A image product has a pixel size of 50 m, and the pixel size of Landsat-8 data varies in interval from 15 to 100 m. Both SAR and optical imagery were downsampled to the same pixel size (100 m) by the nearest neighbor resampling method and projected onto the same coordinate system. The data set includes several sea ice types (thick FYI, thin FYI, young ice, and nilas) and open water. Fig. 4 shows the SAR false-color composite image along with the optical natural color composite image of the Radarsat-2/Landsat-8 overlapped area. In contrast to the Sentinel-2 image of Fig. 3, the optical scene of the Radarsat-2/Landsat-8 data set is not affected by clouds [Fig. 4(b)].
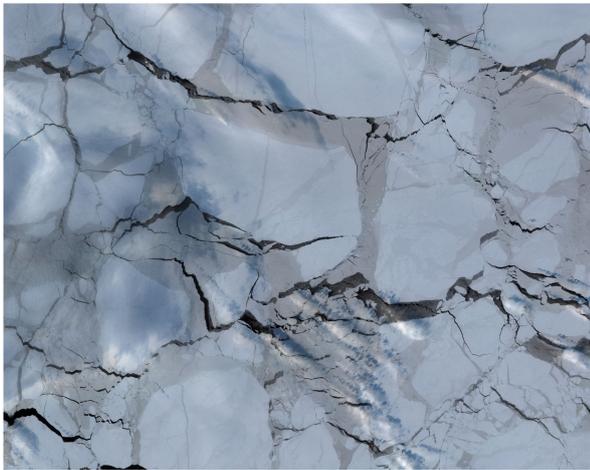
Fig. 3. Color representation of Sentinel-1/Sentinel-2 data set. (a) SAR false-color composite (HV, HH, and HH as RGB). (b) Optical natural color composite (RGB).

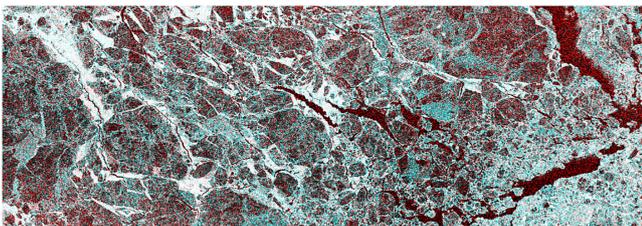Fig. 4. Color representation of Radarsat-2/Landsat-8 data set. (a) SAR false-color composite (HV, HH, and HH as RGB). (b) Optical natural color composite (RGB).

III. METHODS

Fig. 5 represents the flowchart of the proposed approach that consists of pre-processing (collocation and downsampling of the original data sets), the generation and combined extraction and selection of attributes, and, finally, the classification.

In the following sections, random scalars are denoted by lower case letters, e.g., z. Random vectors are designated by bold lower case letters, e.g., $\mathbf{z}$. Bold upper case letters refer to matrices, e.g., $\mathbf{A}$. $|\mathbf{A}|$ denotes the determinant of the matrix $\mathbf{A}$. $\mathrm{diag}\{d_1, \ldots, d_N\}$ refers to a diagonal matrix whose diagonal elements are $d_1, \ldots, d_N$ starting from upper left. The $\mathrm{ddiag}(\mathbf{A})$ operator set to zero the off-diagonal entries of $\mathbf{A}$.

A. Pre-processing

The images acquired by different sensors can have different characteristics, such as units of measurements, spatial resolution, image, and geographical coordinate systems. The first step of our analysis consists of making the data compatible by means of subsampling, alignment on the same coordinate system, collocating, and, finally, extracting the overlapping area.

We note that we did not use any additional filters or correction schemes usually applied to optical data, such as atmospheric correction, cloud masking or to SAR images, such as thermal noise removal, incidence angle compensation.

During the ICESAR campaign, the radar images at L- and C-bands were acquired with a time difference between 10 and 30 min, while for Sentinel-1/Sentinel-2 and Radarsat-2/Landsat-8, the time difference between multi-sensor data was less than 1 h. Nevertheless, after a detailed analysis of the images, especially focusing on individual ice formations and areas of open water, we can assume that there was no significant drift effect that needs to be considered.

B. Attributes Extraction

In addition to the original radar intensities or optical reflectances, we consider the textural attributes in this study. It is noteworthy that GKMI is independent of the data type; hence, it can also be applied to other attributes such as polarimetric features.

For each image layer (frequency/wavelength band and/or polarization), we extract the textural features using the gray-level co-occurrence matrix (GLCM) [26], [27]. We use the directional average for $0°$, $45°$, $90°$, and $135°$ which is common practice to account for the possible rotation of sea ice floes, leads, or any other sea ice structure or roughness pattern on the ocean surface. Moreover, it reduces the number of GLCM matrices. We also considered the average of the distance/displacement parameter that we varied from 1 to 5 (corresponding to the half of the window's size that was set to 11×11). It is noteworthy that minor changes of window size (± 5) do not significantly affect the classification accuracy. Table II illustrates the extracted features as well as their mathematical definitions.
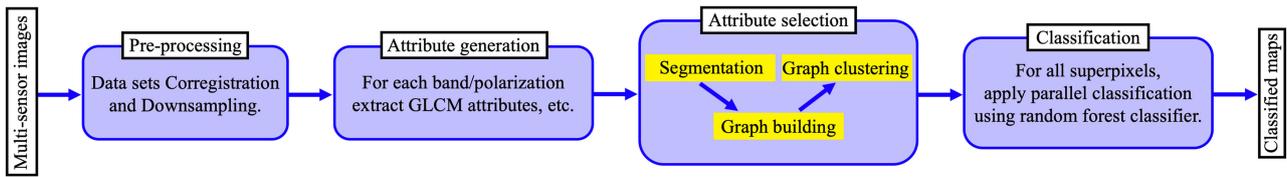
Fig. 5. Flowchart of the proposed approach.

TABLE II
MATHEMATICAL DEFINITION OF GLCM FEATURES

Texture Features	Definition	Texture Features	Definition
Angular Second Moment/ Energy (ASM)	$\sum_{i,j=0}^{Q-1} g_{i,j}^2$	**Maximum Correlation Coefficient (MCC)**	$\sqrt{\lambda}$
Correlation (CORR)	$\sum_{i,j=0}^{Q-1} g_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \right]$	**Variance (VAR)**	$\sum_{i,j=0}^{Q-1} g_{i,j}(i-\mu)^2$
Contrast (CST)	$\sum_{i,j=0}^{Q-1} g_{i,j} \, |i-j|^2)$	**Sum Average (SAVE)**	$\sum_{n=2}^{2Q} g_{x+y}(n)$
Sum Variance (SVAR)	$\sum_{n=2}^{2Q} \left(n - \sum_{n=2}^{2Q} g_{x+y}(n)^2 \right) g_{x+y}(n)$	**Sum Entropy (SENT)**	$-\sum_{n=2}^{2Q} g_{x+y}(n) \log \left(g_{x+y}(n) \right)$
Entropy (ENT)	$\sum_{i,j=0}^{Q-1} g_{i,j} \log \left(g_{i,j} \right)$	**Difference Variance (DVAR)**	$\sum_{n=0}^{Q-1} \left(n - \mu_{x-y} \right)^2 g_{x-y}(n)$
Difference Entropy (DENT)	$-\sum_{n=0}^{Q-1} g_{x-y}(n) \log \left(g_{x-y}(n) \right)$	**Information Correlation (IC)**	$-\frac{\sum_{i,j}^{Q-1} g_{i,j} \log \frac{g_{i,j}}{g_x(i) g_y(j)}}{\max \left(H_x, H_y \right)}$
Inverse Different Moment/ Homogeneity (IDM)	$\sum_{i,j=0}^{Q-1} \frac{g_{i,j}}{1+(i-j)^2}$	**Coefficient of Variation (CV)**	$\frac{\sqrt{\sum_{i,j=0}^{Q-1} g_{i,j}(i-\mu)^2}}{\sum_{i,j=0}^{Q-1} i g_{i,j}}$

Note: $g_{i,j}$ denotes the elements of the GLCM matrix $\mathbf{G}$. Q is the number of gray levels used. $g_x = \sum_i g_{i,j}$, $g_y = \sum_j g_{i,j}$, $g_{x+y}(n) = \sum_{i+j=n} g_{i,j}$, $g_{x-y}(n) = \sum_{|i-j|=n} g_{i,j}$. $\mu_{x-y} = \sum_{n=0}^{Q-1} n g_{x-y}(n)$. $H_x = -\sum_i g_x(i) \log(g_x(i))$, and $H_y = -\sum_i g_y(i) \log(g_y(i))$. λ is the second largest eigenvalue of $\mathbf{A}$, where $\mathbf{A}(i,j) = \sum_k \frac{g_{i,k} g_{j,k}}{g_x(i) g_y(k)}$.

C. Attributes Selection

In this section, we briefly describe the GKMI method that we employ to find relevant attributes [22]. This approach consists of three steps: segmentation, graph building, and graph clustering.

Segmentation: GKMI finds different attributes for separate zones of an image (superpixels) to reflect their particularity. In fact, different parts of an image might represent separate entities and may require different types of attributes to be well represented. Moreover, even if the various parts represent the same entities, they might be observed under different conditions (different noise levels, light conditions, incidence angles, etc.). As such, the first step of GKMI consists of determining the superpixels using the Watershed segmentation method [28], [29].

Graph building: Let L be the number of superpixels extracted using the segmentation method and N the number of initial attributes (see Table III).

In order to select K relevant attributes, among the initial N attributes, GKMI employs two similarity measures, GK and MI. GK permits to preserve the structure of the original set and is defined as follows:

$$w_{i,j}^{\text{GK}} = \exp \left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2} \right) \tag{1}$$

where $||.||$ is the Frobenius norm, $\mathbf{x}_i$ and $\mathbf{x}_j$ are the vectors corresponding to the i-th and j-th attribute, respectively, and $\sigma > 0$ is a parameter that controls the measure of similarity

of the attributes [30]. A large value of σ will indicate a larger similarity even if the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is relatively large, whereas, conversely, a lower value of σ weakens the resemblance judged from only the Euclidean distance. Accordingly, σ might be interpreted as a scale factor that controls the strength of the similarity measure between attributes. In this work, we set σ to the default value 1. It is worth noting that we tried other values of σ, but the performance of the analysis did not change significantly (less than 1% of accuracy).

MI quantifies the shared information between two attributes and is defined as follows [31]:

$$w_{i,j}^{\text{MI}} = D_{\text{KL}} \left(P(\mathbf{x}_i, \mathbf{x}_j) || P(\mathbf{x}_i) P(\mathbf{x}_j) \right) \tag{2}$$

where $D_{\text{KL}}(.||.)$ is the Kullback–Leibler divergence, $P(\mathbf{x}_i, \mathbf{x}_j)$ is the joint density function of $\mathbf{x}_i$ and $\mathbf{x}_j$, and $P(\mathbf{x}_i)$ and $P(\mathbf{x}_j)$ are the marginals. Small values of $w_{i,j}^{\text{MI}}$ exhibits independency of $\mathbf{x}_i$ and $\mathbf{x}_j$ which means that both attributes encompass different information. Conversely, high values of $w_{i,j}^{\text{MI}}$ show dependency between the attributes, which means that both reflect similar information.

For each superpixel l, we build a multigraph $\mathcal{G}_l(V, E)$ with N vertices corresponding to the N attributes connected by two edges. The weights of the edges (strength of the connections) are given by GK (1) and MI (2). The MI is measured image-wise, considering all pixels in the image, while the GK is calculated superpixel-wise, using only the pixels within each superpixel.

Graph clustering: By partitioning the graph $\mathcal{G}_l$ into subgraphs, GKMI groups similar attributes together, according to MI and GK measures. As such by selecting a representative attribute from each subgraph, we obtain a subset that preserves the structure and the information content of the original set of attributes within the l superpixel.

Two Laplacian matrices, $\mathbf{L}^{GK}$ and $\mathbf{L}^{MI}$, are associated with the graph $\mathcal{G}_l$, and are defined as follows:

$$\mathbf{L}^{GK} = \mathbf{I} - \mathbf{D}^{GK^{-1/2}} \mathbf{W}^{GK} \mathbf{D}^{GK^{-1/2}} \qquad (3)$$

$$\mathbf{L}^{MI} = \mathbf{I} - \mathbf{D}^{MI^{-1/2}} \mathbf{W}^{MI} \mathbf{D}^{MI^{-1/2}} \qquad (4)$$

where $\mathbf{I}$ is the identity matrix. $\mathbf{W}^{GK} = (w_{ij}^{GK})_{ij}$ and $\mathbf{W}^{MI} = (w_{ij}^{MI})_{ij}$ are the adjacency matrices of the graph $\mathcal{G}_l$, and $\mathbf{D}^{GK} = \mathrm{diag}(\sum_{i \neq j} w_{lij}^{GK})$ and $\mathbf{D}^{MI} = \mathrm{diag}(\sum_{i \neq j} w_{ij}^{MI})$ are their corresponding degree matrices, respectively.

The number of subgraphs within the graph $\mathcal{G}_l$ is equal to the multiplicity of the null eigenvalues of $\mathbf{L}^{GK}$ and $\mathbf{L}^{MI}$. Moreover, each eigenvector associated with a null eigenvalue is an indicator of a subgraph [30]. Accordingly, the partition of the graph $\mathcal{G}_l$ into subgraphs can be performed by embedding the attributes into a new manifold spanned by the joint null eigenvectors (eigenvectors associated with the null eigenvalues) of the Laplacian matrices. The joint eigenvectors of $\mathbf{L}^{GK}$ and $\mathbf{L}^{MI}$ are obtained by a joint approximate diagonalization [32]

$$\mathbf{L}^{GK} = \mathbf{V} \mathbf{\Lambda}^{GK} \mathbf{V}^T \qquad (5)$$

$$\mathbf{L}^{MI} = \mathbf{V} \mathbf{\Lambda}^{MI} \mathbf{V}^T \qquad (6)$$

where $\mathbf{V} = (v_{i,j})_{0 \leq i,j \leq N}$ is the matrix of eigenvectors, and $\mathbf{\Lambda}^{GK} = \mathrm{diag}(\lambda_1^{GK}, \ldots, \lambda_N^{GK})$ and $\mathbf{\Lambda}^{MI} = \mathrm{diag}(\lambda_1^{MI}, \ldots, \lambda_N^{MI})$ are diagonal matrices of the corresponding eigenvalues. The K first eigenvectors corresponding to the smallest eigenvalues will define a new representation of the attributes, i.e., the vector $\mathbf{u}_k = [v_{k,1}, \ldots, v_{k,N}]^T$ is the new representation of the k-th attribute $\mathbf{x}_k$, as follows:

$\mathbf{u}_1$ is the representative of the 1st attribute

K eigenvectors associated with the K smallest eigenvalues

The final step consists of performing a clustering on the new attributes $\mathbf{u}_1, \ldots, \mathbf{u}_K$. In our case, we perform the clustering using K-means [18]. The subset of K relevant attributes correspond to the closest attributes to the centroids of the clusters. It is worth noting that the embedding of the attributes increases their separability and accordingly gives better results than when performing the clustering on the original attributes [30]. Moreover, unlike the feature extraction approaches, the attributes embedding can be mapped back to the original set (since the original attributes and their new representations have the same indices), hence preserving their physical interpretability.

TABLE III
INITIAL ATTRIBUTES CONSIST OF THE ORIGINAL BANDS/POLARIZATIONS AND 14 TEXTURAL FEATURES (TABLE II)

Data set	Original	Derived	N
ICESAR	$2 \times$ C-band $+ 3 \times$ L-band	Textural	75
Sentinel-1/Sentinel-2	$2 \times$ SAR $+ 13 \times$ optical	Textural	225
Radarsat-2/Landsat-8	$2 \times$ SAR $+ 11 \times$ optical	Textural	195

TABLE IV
CLASSES DETERMINED BY SEA ICE EXPERTS FROM VISUAL INSPECTION, AND NUMBER OF TRAINING SAMPLES AND REGIONS OF INTEREST FOR ALL THE DATA SETS USED IN THIS ARTICLE

	Class Name	Class Description	Training Samples	ROI
ICESAR	OW	Open Water	2398	3
	NI	Nilas	12946	8
	GI	Grey Ice	2342	5
	GWI	Grey-White Ice	10640	9
	LFY	Level FYI	14233	8
	DFY	Deformed FYI	6356	12
S-1/S-2	OW	Open Water	911	3
	GI	Grey Ice	9906	3
	GWI	Grey-White Ice	6205	3
	TNFY	Thin FYI	5054	3
	TKFY	Thick FYI	8009	3
R-2/L-8	OW	Open Water	401	3
	NI	Nilas	774	6
	YI	Young Ice	889	5
	TNFY	Thin FYI	746	4
	TKFY	Thick FYI	1381	6

Note: Training samples refers to the number of pixels available for classification performance evaluation.

D. Classification

It is noteworthy that the main focus of this article is the application of the GKMI method to determine an optimal set of attributes, which we here demonstrate for the separation of ice types. We perform the sea ice classification using the random forest method, which is a widely applied classifier in remote sensing [23].

Table IV illustrates the sea ice and water classes for all the data sets used in this article as well as the number of regions of interest (ROIs) and available training samples that have been applied for creating the training and test data sets to evaluate the classification performance. Additionally, Figs. 6(a), 7(a), and 8(a) show the distribution of ROIs that were used for the classification of all the data sets.

In all the experiments, we randomly choose 20% of the samples from each label as a training set, while the remaining 80% of samples are used as a test set for performance evaluation.

IV. EXPERIMENTS

In this section, we study the relevance of various attributes, described in the previous section, for the characterization of different sea ice types employing GKMI. To quantitatively evaluate the result of sea ice classification, we use two measures: the overall accuracy (OA) index and Cohen's Kappa coefficient (k). The OA shows the percentage of correctly classified samples,
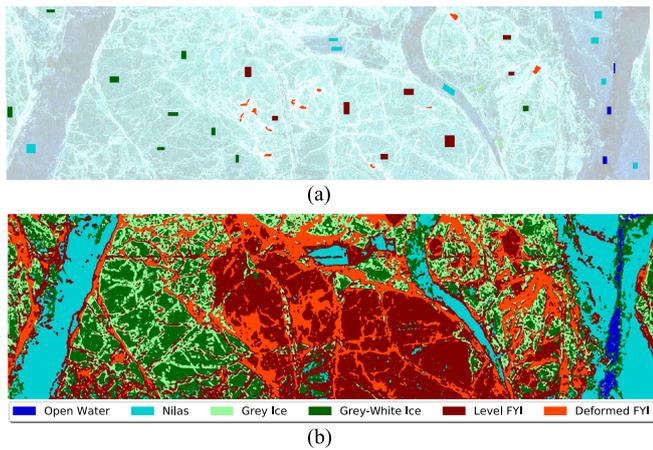
Fig. 6. (a) ROI and (b) classified map for ICESAR data set.

while Kappa measures the agreement between the classification and the reference data [33].

In the following, using the ICESAR, Radarsat-2/Landsat-8, and Sentinel-1/Sentinel-2 data sets, we run different analyses to demonstrate the relevance of automatic feature selection in Section IV-A, the importance of the combination of several imaging modes in Section IV-B, and the data dependency of the relevant attributes in Section IV-C. Finally, in Section IV-D, we compare GKMI to the commonly automatic approaches for attributes selection, PCA and FS.

A. Performance Analysis

Before conducting our analysis of the relevant attributes for the characterization of different ice types, we evaluate the performance of GKMI.

Fig. 9 illustrates the overall accuracies of the GKMI attribute selection method dependent on the number of selected attributes for the ICESAR, Sentinel-1/Sentinel-2, and Radarsat-2/Landsat-8 data sets. The red line indicates the accuracies obtained with the RF classifier for ICESAR, the blue line refers to Sentinel-1/Sentinel-2, and the black line illustrates the Radarsat-2/Landsat-8 data set. The stars show the points where the accuracies reach their peak. It can be clearly seen that all curves rise sharply until the number of attributes reaches 20 for ICESAR and Sentinel-1/Sentinel-2 and 30 for Radarsat-2/Landsat-8, whereupon the accuracies become stable and high for Sentinel-1/Sentinel-2 and Radarsat-2/Landsat-8 and slightly decreasing for ICESAR. Moreover, the maximum accuracy was reached with almost half of the attributes for ICESAR, less than one-fifth for Radarsat-2/Landsat-8, and less than one-seventh of the attributes for the Sentinel-1/Sentinel-2 data set. Table V shows the OA and Kappa coefficient (k) obtained for the optimal number of selected attributes K, and the ones obtained for the total number of attributes N. After reaching the point of the highest accuracy, the inclusion of additional attributes in the classification does not provide any further information which could improve the classification performance for Sentinel-1/Sentinel-2 and Radarsat-2/Landsat-8 data sets. In case of the ICESAR data set, more attributes slightly decline the classifier performance.
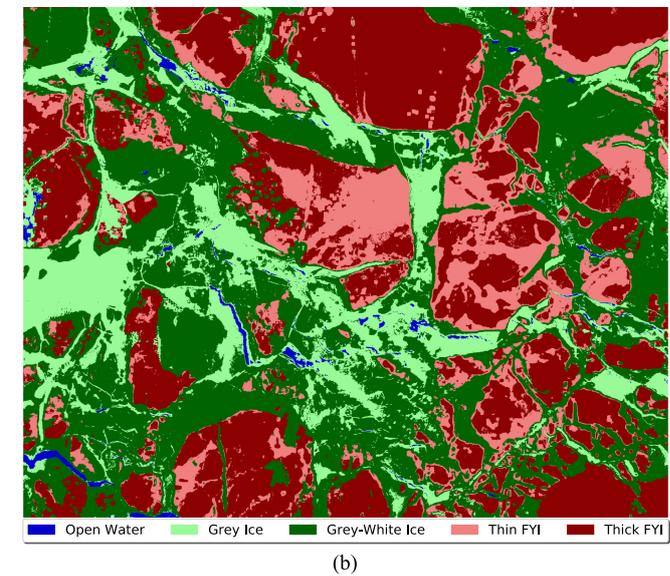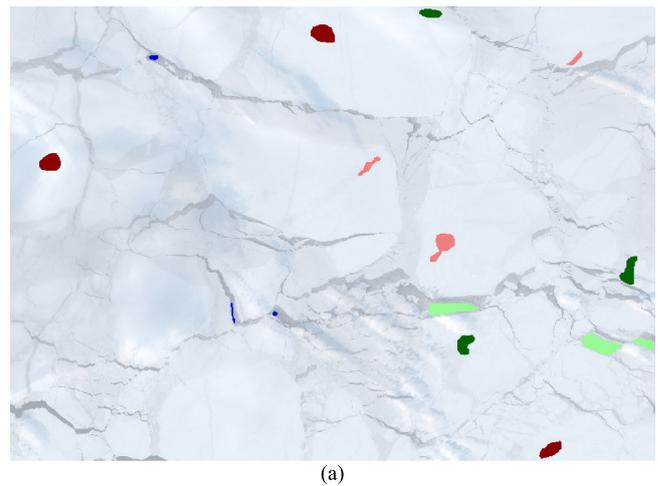

(a)

(b)

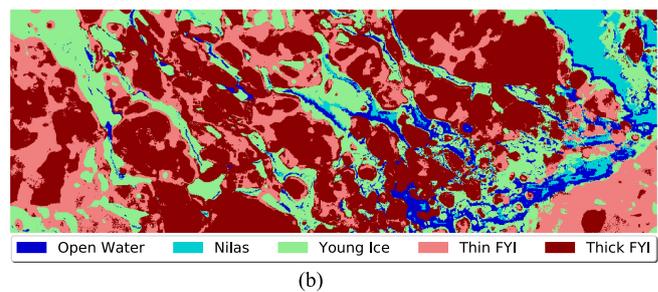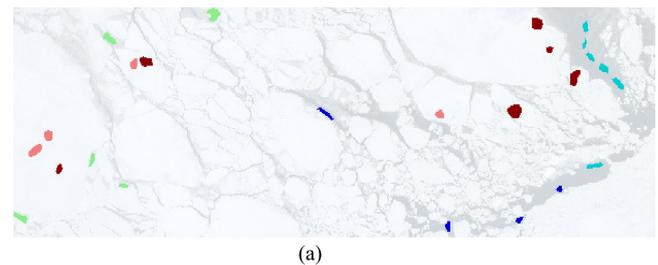Fig. 7. (a) ROI and (b) classified map for Sentinel-1/Sentinel-2 data set.

(a)

(b)

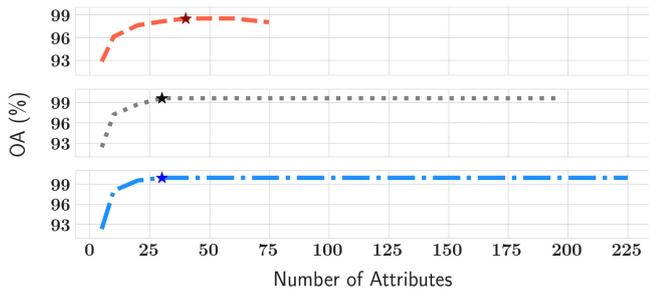Fig. 8. (a) ROI and (b) classified map for Radarsat-2/Landsat-8 data set.

Fig. 9. Overall accuracies of GKMI as a function of the number of selected attributes for ICESAR, Sentinel-1/Sentinel-2, and Radarsat-2/Landsat-8 data sets.

TABLE V
CLASSIFICATION PERFORMANCE OBTAINED OVER DIFFERENT DATA SETS

	K	OA	k	N	OA	k
ICESAR	40	**98.5**	**98.0**	75	98.0	97.5
Radarsat-2/Landsat-8	30	**99.6**	99.3	195	**99.6**	99.5
Sentinel-1/Sentinel-2	30	**99.9**	99.8	225	**99.9**	99.9

Note: K refers to the optimal number of selected attributes, N shows the total number of attributes, k refers to the Kappa coefficient, and OA (%) corresponds to the overall accuracy. The maximum values of the OA and k are shown in bold. Note that the table shows the classification performance for both optimal and total number of attributes.

TABLE VI
CLASSIFICATION PERFORMANCE OBTAINED USING C-BAND, L-BAND, AND
MULTI-FREQUENCY ATTRIBUTES FOR ICESAR DATA SET

L-band			**C-band**			**Multi-frequency**		
K	OA	k	K	OA	k	K	OA	k
40	93.2	91.2	30	90.1	87.2	40	**98.5**	**98.0**
20	92.4	90.2	20	89.6	86.5	20	**97.6**	**96.9**

These results clearly demonstrate the prominence of information selection.

Figs. 6(b), 7(b), and 8(b) show the classified maps for multi-frequency and multi-sensor data sets used in this work. Since the ICESAR data were acquired with a higher spatial resolution, narrow structures such as ice ridges, cracks, and small leads are easier to identify in the classification maps.

B. Multi-sensor vs. Single-Sensor

To demonstrate the advantage of combining data obtained from various sensors, we compare the performance of the classification algorithm when only using the attributes of each imaging mode individually and when using their attributes combined.

1) ICESAR: Table VI shows the OA and Kappa coefficient k obtained for K, the optimal number of selected attributes among the attributes extracted for L-band, C-band, and multi-frequency data set (i.e., when we are using them together). Note that the optimal number is different for each case. To better show the difference in performance, we also added results obtained for a fixed amount of selected attributes $K = 20$. It is evident that the joint use of both data sets increases the accuracy significantly. The OA of the multi-frequency data set reaches its peak with 40 attributes and is equal to 98.5%, while the L-band achieves 93.2% and C-band 90.1%. In fact, even with only 20 attributes,

TABLE VII
NORMALIZED CONFUSION MATRIX FOR CLASSIFICATION RESULTS OBTAINED
WITH ICESAR DATA SET

		Actual class					
		OW	NI	GI	GWI	LFY	DFY
Assigned class	OW	0.722	0	0.065	0.03	0	0
	NI	0	0.978	0	0.001	0.011	0
	GI	0.117	0	0.664	0.111	0	0
	GWI	0.161	0	0.271	0.854	0	0
	LFY	0	0.022	0	0.004	0.893	0.021
	DFY	0	0	0	0	0.096	0.979

(a)

		Actual class					
		OW	NI	GI	GWI	LFY	DFY
Assigned class	OW	0.961	0.007	0.009	0	0	0
	NI	0.008	0.974	0.042	0.019	0	0
	GI	0.026	0.001	0.792	0.05	0.011	0
	GWI	0.005	0.018	0.139	0.867	0.065	0
	LFY	0	0	0.018	0.064	0.924	0
	DFY	0	0	0	0	0	1

(b)

		Actual class					
		OW	NI	GI	GWI	LFY	DFY
Assigned class	OW	0.964	0	0.005	0	0	0
	NI	0	0.999	0	0	0.003	0
	GI	0.021	0	0.926	0.035	0	0
	GWI	0.015	0	0.069	0.962	0	0
	LFY	0	0.001	0	0.003	0.997	0
	DFY	0	0	0	0	0	1

(c)

Note: OW denotes the open water, NI the nilas, GI the grey ice, GWI the grey-white ice, LFY the level FYI, and DFY the deformed FYI class. (a) C-band attributes. (b) L-band attributes. (c) Multi-frequency attributes.

the multi-frequency data set achieves higher performance than the maximum achievable performance using a single band with its optimal number of attributes. Our result indicates a slightly better accuracy at L-band than at C-band. Since the penetration depth into the ice is larger at L-band, deformation structures are easier to recognize in the corresponding SAR images. In cases for which deformation structures comprise an important criterion for classification (like for our data set here), better accuracies are often achieved at L-band.

Furthermore, we analyze the interclass performance for the single-frequency and the multi-frequency approaches by calculating the normalized confusion matrices depicted in Table VII(a)–(c). The results show that the NI (nilas) and DFY (deformed FYI) were the easiest to distinguish for single-frequency and multi-frequency data sets, compared to OW (open water), GWI (grey-white ice), and GI (grey ice). Besides, we notice that GI has the lowest accuracy among all classes for both L-band and C-band data. In fact, GI is highly misclassified as GWI, as can be visible on the confusion matrices. This misclassification

TABLE VIII
CLASSIFICATION PERFORMANCE OBTAINED USING SAR, OPTICAL, AND
MULTI-SENSOR ATTRIBUTES FOR SENTINEL-1/SENTINEL-2

	SAR			Optical			Multi-sensor		
N	K	OA	k	K	OA	k	K	OA	k
225	30	98.2	97.6	100	99.8	99.7	30	**99.9**	**99.8**
43	20	97.8	97.1	13	99.4	99.1	10	**99.7**	**99.6**

results from the significant overlapping of GI and GWI signatures because of their high similarities in many radar signature attributes. Moreover, the L-band data set shows higher accuracy than the C-band data set for all classes. However, the highest accuracies were always achieved by the multi-frequency data set. The prominence of the multi-frequency approach is more apparent for the class GI (grey ice). In fact, by combining L-band and C-band, the accuracy of their most challenging class GI has significantly improved.

To show the flexibility and robustness of the attribute selection method and multi-sensor approach, we also perform a classification comparison of SAR and optical attributes for Sentinel-1/Sentinel-2 and Radarsat-2/Landsat-8. The optical data that were used as complementary source for the manual classification by experts could not be geometrically registered to the radar images because the time gap between acquisitions was too large to compensate for sea ice deformations. The Radarsat-2/Landsat-8 and Sentinel-1/Sentinel-2 data reveal a clear dominance of the optical attributes over radar attributes. Since this is dependent on the special measurement conditions (e.g., sensor characteristics and ice properties, for optical sensors sun elevation), it is difficult to generalize this result, and, correspondingly, we expect varying priorities of single attributes. Nevertheless, the dominance in the number of attributes does not always lead to better accuracy or more information content, as shown in Fig. 9. We emphasize that one of the main ideas of this study is to show that the GKMI method can be easily adapted to any sensor and image combination and for any environmental conditions.

2) Sentinel-1/Sentinel-2: Table VIII illustrates the classification performance of SAR, optical, and multi-sensor attributes for Sentinel-1/Sentinel-2 data set. This analysis was conducted using two initial sets of attributes. The first set consists of the original bands and polarizations of optical and SAR images in addition to their corresponding textural features. Conversely, in the second set, optical images' textural features were omitted (in this scenario, the data set consists of 43 attributes: 13 optical and 30 SAR). The second set was considered to appropriately evaluate each sensor's contribution. The results show that using only optical attributes, one reaches the same high accuracy as multi-sensor attributes for Sentinel-1/Sentinel-2 data set. However, when using multi-sensor attributes, accuracy is achieved with less number of attributes.

Additionally, we also extracted confusion matrices, like we did for the ICESAR data set, to better analyze the interclass performance of the single-sensor and multi-sensor approaches. Table IX(a)–(c) demonstrates the results for the Sentinel-1/Sentinel-2 data set. We notice that the optical data set achieves a better separation for all classes compared to the SAR data set. However, there is no significant accuracy superiority in any of the

TABLE IX
NORMALIZED CONFUSION MATRIX FOR CLASSIFICATION RESULTS OBTAINED
WITH THE SENTINEL-1/SENTINEL-2 DATA SET

		Actual class				
		OW	GI	GWI	TNFY	TKFY
Assigned class	OW	0.977	0	0	0	0.013
	GI	0	0.994	0.008	0	0
	GWI	0	0.006	0.959	0.032	0
	TNFY	0.014	0	0.033	0.968	0.006
	TKFY	0.009	0	0	0	0.981

(a)

		Actual class				
		OW	GI	GWI	TNFY	TKFY
Assigned class	OW	0.993	0	0	0	0
	GI	0.007	0.998	0.005	0	0
	GWI	0	0.002	0.995	0	0
	TNFY	0	0	0	1	0
	TKFY	0	0	0	0	1

(b)

		Actual class				
		OW	GI	GWI	TNFY	TKFY
Assigned class	OW	1	0	0	0	0
	GI	0	1	0.003	0	0
	GWI	0	0	0.997	0	0
	TNFY	0	0	0	1	0.001
	TKFY	0	0	0	0	0.999

(c)

Note: OW denotes the open water, GI the grey ice, GWI the grey-white ice, TNFY the thin FYI, and TKFY the thick ice class. (a) SAR attributes. (b) Optical attributes. (c) Multi-sensor attributes.

TABLE X
CLASSIFICATION PERFORMANCE OBTAINED USING SAR, OPTICAL, AND
MULTI-SENSOR ATTRIBUTES FOR RADARSAT-2/LANDSAT-8 DATA SET

	SAR			Optical			Multi-sensor		
N	K	OA	k	K	OA	k	K	OA	k
195	20	91.0	88.3	120	98.5	98.1	30	**99.6**	**99.3**
41	20	91.0	88.3	11	95.3	93.9	10	**98.6**	**98.2**

classes. Moreover, the multi-sensor approach results in higher accuracies than the single-sensor.

3) Radarsat-2/Landsat-8: Now, we evaluate the difference in performance between optical and SAR sensors for Radarsat-2/Landsat-8 data set. As for the Sentinel-1/Sentinel-2 data set, we consider two sets of initial attributes. In the second set, the textural features of the optical bands were excluded to have a comparable number of attributes for each sensor. Accordingly, for this scenario, the Radarsat-2/Landsat-8 data set includes 41 attributes (11 optical and 30 SAR). Table X illustrates the performance comparison using SAR, optical, and multi-sensor attributes. Unlike the Sentinel-1/Sentinel-2 data

TABLE XI
NORMALIZED CONFUSION MATRIX FOR CLASSIFICATION RESULTS OBTAINED
WITH RADARSAT-2/LANDSAT-8 DATA SET

		Actual class				
		OW	NI	YI	TNFY	TKFY
Assigned class	OW	0.857	0.004	0	0	0.01
	NI	0.031	0.848	0	0.005	0.147
	YI	0	0	1	0	0
	TNFY	0.017	0.016	0	0.993	0.001
	TKFY	0.095	0.133	0	0.002	0.842

(a)

		Actual class				
		OW	NI	YI	TNFY	TKFY
Assigned class	OW	0.974	0.024	0.015	0	0
	NI	0.016	0.976	0.007	0	0
	YI	0.01	0	0.978	0	0
	TNFY	0	0	0	0.998	0.001
	TKFY	0	0	0	0.002	0.999

(b)

		Actual class				
		OW	NI	YI	TNFY	TKFY
Assigned class	OW	0.985	0.013	0	0	0
	NI	0.015	0.987	0	0	0
	YI	0	0	1	0	0
	TNFY	0	0	0	1	0
	TKFY	0	0	0	0	1

(c)

Note: OW denotes the open water, NI the nilas, YI the young ice, TNFY the thin FYI, and TKFY the thick ice class. (a) SAR attributes. (b) Optical attributes. (c) Multi-sensor attributes.

set, where the single-sensor performance for both sensors was high, the Radarsat-2/Landsat-8 single-sensors show less accurate performance than multi-sensor attributes.

Table XI(a)–(c) shows the confusion matrices for Radarsat-2/Landsat-8. The results reveal that the YI (young ice) was the easiest to distinguish using SAR, while TNFY (thin FYI) and TKFY (thick FYI) showed high classification accuracy using optical data. Unlike optical, the SAR part shows significantly lower accuracies for OW (open water), NI (nilas), and TKFY (thick FYI). However, the multi-sensor attributes combine the advantages of both sensors and ensure the highest separability for all classes.

It should be noted that even though, for this particular example, optical attributes result in high accuracies, they cannot serve as basic source for operational monitoring. This is due to several familiar factors, such as limitations due to light and cloud conditions that influence the separability of rough and smooth level ice, and snow cover that prevents recognition of sea ice types.

The results obtained for the investigated data sets reflect different situations: the dominance of one frequency or one sensor for ice type separation (ICESAR and Sentinel-1/Sentinel-2), and the complementarity of sensors (Radarsat-2/Landsat-8). In all cases, the combination of data from different sensors guarantees better performance. In particular, for certain ice conditions and combinations of available data, the multi-sensor approach can increase the classification accuracy significantly if compared to the single-sensor case (in Table VII(c), e.g., the ICESAR data).

C. Relevant Attributes

Because the selection by GKMI is performed in superpixel space, different attributes are selected for different classes. The selected attributes can also differ between the superpixels of one ice-class due to the influence of several factors, such as incidence angle, range-dependent noise, light and cloud conditions, etc. However, we can still estimate the relevant attributes for a class by identifying its frequently selected attributes. To this end, we identify all superpixels belonging to a class and extract their chosen attributes.

Fig. 10 shows the histograms of the five most selected attributes for each class for the different data sets.

Fig. 10(a) shows the relevant attributes for each class of ICESAR data set. The histograms show that, in general, C-band and L-band attributes were often selected equally. However, for some classes, there is a clear predominance of L-band (open water, level FYI, and grey-white ice) or C-band attributes (nilas). Moreover, several attributes, such as information correlation and inverse different moment derived from L-band VV polarization (L VV IC and L VV IDM), were selected for several classes of the ICESAR data set.

Fig. 10(b) demonstrates the relevant attributes selected for the different classes of Sentinel-1/Sentinel-2 data set. Optical attributes are predominant for this data set, especially for thick and thin FYI. Moreover, the intensity of band B1 (B1 INTST) was frequently selected for three classes (open water, grey-white ice, and thick FYI).

Fig. 10(c) illustrates the relevant attributes for the classes of the Radarsat-2/Landsat-8 data set. Optical attributes were selected more frequently than SAR attributes; however, unlike the Sentinel-1/Sentinel-2 data set, all the classes have a mixed set of selected attributes. The most predominantly selected attribute for Radarsat-2/Landsat-8 was the maximum correlation coefficient derived from HV polarization (HV MCC), that was chosen for open water, nilas, young ice, and thick FYI.

From the results mentioned above, it is possible to conclude that it is hard to find any generalization among the attributes and corresponding sea ice types. Moreover, for a given combination of data types, selected attributes vary even between the superpixels that belong to the same ice class. This indicates that the selection procedure is sensitive to the technical setup (e.g., incidence and sun elevation angle) and slight variations of the appearance of a given ice class (e.g., varying degrees of deformation and ice concentration). Therefore, it is crucial to have a flexible attribute selection method that allows selecting the relevant attributes for different data combinations and sea ice classes even under various conditions to enhance the accuracy of classification.
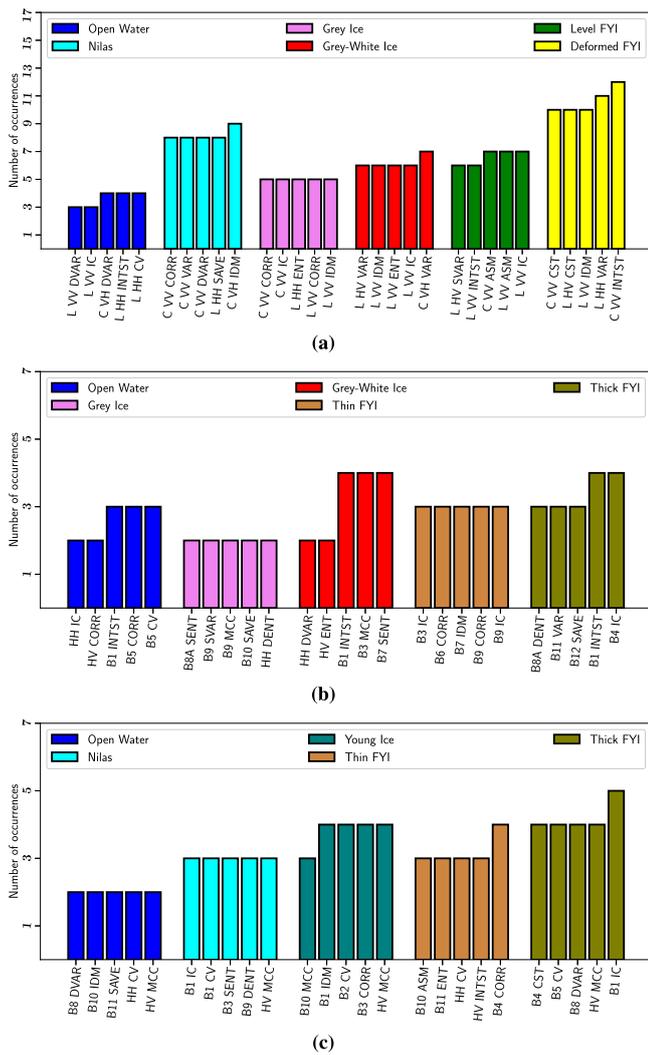
Fig. 10. Five most frequently selected attributes for each class of (a) ICESAR, (b) Sentinel-1/Sentinel-2, and (c) Radarsat-2/Landsat-8. Note that the height of the bars (number of attributes' occurrences) is different for the various classes because the number of used ROI is different for each class. INTST refers to intensity. For radar data, intensity refers to the backscattering coefficient.

TABLE XII
PERFORMANCE COMPARISON AMONG DIFFERENT METHODS FOR ATTRIBUTE
SELECTION, USING RF CLASSIFIER

	ICESAR			Sentinel-1/Sentinel-2			Radarsat-2/Landsat-8		
Method	N	OA (%)	k	N	OA (%)	k	N	OA (%)	k
FS	40	98.0	97.5	30	99.8	99.7	30	97.7	97.0
PCA	40	97.7	97.0	30	99.9	99.8	30	98.9	98.0
GKMI	40	**98.5**	**98.0**	30	99.9	99.8	30	**99.6**	**99.3**

D. Comparison of Methods

To evaluate the GKMI method's performance, we compared its results to the two mostly used automatic attributes selection methods, mainly PCA and FS. PCA is an unsupervised feature extraction approach, while FS is a supervised attribute selection. Table XII shows the accuracy and kappa for the different methods and data sets used in this article. All methods show

a high classification accuracy; however, GKMI gives slightly higher accuracy for all data sets. These results validate our approach's efficiency since it offers better performance without any supervision as for FS, and while preserving the physical interpretability of attributes as opposed to PCA.

We emphasize that in this study, the major focus was on attribute selection for maximizing the information content of satellite and airborne images as a step toward improved sea ice classification. The results of attribute selection can also be applied for an assessment of dominant scattering mechanisms or for optimizing the retrieval of ice parameters such as roughness or thickness.

In comparison to other attribute selection algorithms, particularly to the ones that were used in this study, GKMI is substantially different, which is an advantage for automated classification. Hereby, the optimal subset of attributes using the GKMI method is determined according to two different criteria, which are structure (GK) and information content (MI). These criteria are applied simultaneously and account for the global and local particularities of the original data set. Moreover, the method is performed on a superpixel level, which, in combination with the use of two similarity measures explained in Section III-C, allows us to capture relevant information at different scales. Additionally, GKMI preserves the advantages of both attribute extraction and selection approaches, namely, increases the class separability and, at the same time, preserves the physical meaning of the original data.

V. CONCLUSION

In this article, we employed GKMI, a recently developed approach for attribute selection, and applied it to the task of sea ice classification using different multi-sensor data combinations. The optimal set of selected attributes, by GKMI, is not fixed for the whole image or a sea ice class but is adaptive for each region of the observed scene according to the technical and environmental conditions (such as noise level, incidence angle, cloudiness, etc.)

Furthermore, our results show the ability of GKMI to process different combinations of data sets and the importance of deploying multi-sensors for the characterization of ice types by comparing the performance of using several sensors separately and simultaneously. Since the major idea of our approach is to introduce a highly flexible and adaptive scheme, it is not necessary to recommend fixed sets of attributes for classification, which depend on sensor types, imaging modes, and environmental and ice conditions during data acquisitions. As such, the fact that optical information is more accurate in the considered data sets is not of general validity due to the limitations of optical sensors in cloudy and dark areas.

Finally, two issues still need to be addressed. For the first, we did not yet investigate in detail how our results are affected by the observation conditions, such as radar incidence angle or sun elevation angle. This is part of future investigations. Another item concerns the availability of the investigated sensor combinations for operational mapping. Since our results clearly demonstrate the advantage of multi-sensor classification (here focussing on

GKMI), space agencies and operational services should develop strategies that make the joint use of data from different sensor types possible. For example, it has to be considered that optical data can only be used under favorable light and cloud conditions (hence requiring a "only-add-if-usable" strategy). Moreover, the combination of different radar frequencies for operational services is recently under discussion, e.g., the Copernicus ROSE-L mission (see [34]). For joint data use, it is important to consider the drift of sea ice, which requires that different sensors acquire data with smallest possible time gaps over a given region. For the combination of C-band and L-band SAR, corresponding acquisition scenarios are under investigation in the ROSE-L project.

ACKNOWLEDGMENT

REFERENCES

[1] M. R. England, L. M. Polvani, and L. Sun, "Robust arctic warming caused by projected antarctic sea ice loss," *Environ. Res. Lett.*, vol. 15, Sep. 2020, Art. no. 104005.

[2] S. Funder *et al.*, "History of sea ice in the arctic," *Quaternary Sci. Rev.*, vol. 29, no. 15/16, pp. 1757–1778, 2010.

[3] S. Sandven, O. M. Johannessen, and K. Kloster, "Sea ice monitoring by remote sensing," in *Encyclopedia of Analytical Chemistry*. Hoboken, NJ, USA: Wiley, 2006, pp. 1–43.

[4] J. Lohse, A. P. Doulgeris, and W. Dierking, "An optimal decision-tree design strategy and its application to sea ice classification from SAR imagery," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1574.

[5] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[6] N. Longbotham *et al.*, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.

[7] T. Geldsetzer, M. Arkett, T. Zagon, F. Charbonneau, J. J. Yackel, and R. K. Scharien, "All-season compact-polarimetry C-band SAR observations of sea ice," *Can. J. Remote Sens.*, vol. 41, no. 5, pp. 485–504, 2015.

[8] M. E. Shokr, "Evaluation of second-order texture parameters for sea ice classification from radar images," *J. Geophys. Res.: Oceans*, vol. 96, no. C6, pp. 10625–10640, 1991.

[9] D. Barber and E. LeDrew, "SAR sea ice discrimination using texture statistics : A multivariate approach," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 4, pp. 385–395, 1991.

[10] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, Mar. 1999.

[11] D. A. Clausi and M. E. Jernigan, "A fast method to determine co-occurrence texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 1, pp. 298–300, Jan. 1998.

[12] D. Murashkin, G. Spreen, M. Huntemann, and W. Dierking, "Method for detection of leads from Sentinel-1 SAR images," *Ann. Glaciol.*, vol. 59, no. 76pt2, pp. 124–136, 2018.

[13] J.-W. Park, A. A. Korosov, M. Babiker, J.-S. Won, M. W. Hansen, and H.-C. Kim, "Classification of sea ice types in Sentinel-1 synthetic aperture radar images," *Cryosphere*, vol. 14, no. 8, pp. 2629–2645, 2020.

[14] M.-A. Moen, "Analysis and interpretation of c-band polarimetric SAR signatures of sea ice," Ph.D. thesis, UiT Arctic Univ. Norway, Tromsø, Norway, 2015.

[15] N. C. Wright and C. M. Polashenski, "Open-source algorithm for detecting sea ice surface features in high-resolution optical imagery," *Cryosphere-Discussions*, vol. 12, no. 4, pp. 1307–1329, 2018.

[16] X. Miao, H. Xie, S. F. Ackley, D. K. Perovich, and C. Ke, "Object-based detection of arctic sea ice and melt ponds using high spatial resolution aerial photographs," *Cold Regions Sci. Technol.*, vol. 119, pp. 211–222, 2015.

[17] B. Ghojogh *et al.*, "Feature selection and feature extraction in pattern analysis: A literature review," 2019, *arXiv:1905.02845*.

[18] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Orlando, FL, USA: Academic Press, Inc., 2008.

[19] M. Fauvel, C. Dechesne, A. Zullo, and F. Ferraty, "Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2824–2831, Jun. 2015.

[20] S. H. Huang, "Supervised feature selection: A tutorial," *Artif. Intell. Res.*, vol. 4, pp. 22–37, 2015.

[21] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artif. Intell. Rev.*, vol. 53, pp. 907–948, 2020.

[22] E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni, "A multimodal feature selection method for remote sensing data analysis based on double graph Laplacian diagonalization," to be published.

[23] A. Liaw and M. Wiener, "Classification and regression by random forest," *Forest*, vol. 23, pp. 18–23, Nov. 2001.

[24] W. Dierking, "Technical assistance for the deployment of airborne SAR and geophysical measurements during the ICESAR2007; *final report - part 2: Sea ice*," 2008.

[25] A. V. Bushuyev, "Sea-ice nomenclature," WMO /OMM /BMO - No. 259. Russian Federation, 1970.

[26] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[27] U. Kandaswamy, D. A. Adjeroh, and M. C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2075–2083, Sep. 2005.

[28] S. Beucher, "The watershed transformation applied to image segmentation," in *Proc. 10th Pfefferkorn Conf. Signal Image Process. Microsc. Microanalysis*, 1992, pp. 299–314.

[29] P. Neubert and P. Protzel, "Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 996–1001.

[30] U. Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, pp. 395–416, Dec. 2007.

[31] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 175–186, 2014.

[32] P. Ablin, J. Cardoso, and A. Gramfort, "Beyond Pham's algorithm for joint diagonalization," 2018, *arXiv:1811.11433*.

[33] P. Bharatkar and R. Patel, "Approach to accuracy assessment tor RS image classification techniques," *Int. J. Sci. Eng. Res.*, vol. 4, no. 12, pp. 79–86, 2013.

[34] M. Davidson *et al.*, "Copernicus L-band SAR mission requirements document," Eur. Space Agency, Paris, France, ESA-EOPSM-CLIS-MRD-3371, Tech. Rep., Oct. 2019.

Eduard Khachatrian (Student Member, IEEE) received the dual M.Sc. degrees in polar and marine sciences from the Faculty of Mathematics, Informatics, and Natural Sciences, Hamburg University, Hamburg, Germany, and the Institute of Earth Science, Saint Petersburg State University, Saint Petersburg, Russia, in 2017.

From 2017 to 2018, he was a Junior Scientist with Nansen International Environmental and Remote Sensing Centre, Bergen, Norway. He is currently a Ph.D. Fellow with the Center of Integrated Remote Sensing and Forecasting for Arctic Operations, UiT The Arctic Univesity of Norway, Tromsø, Norway. His research interests include multimodal data analysis, image processing, and remote sensing of polar areas.

Saloua Chlaily (Member, IEEE) received the M.Sc. degree in electronics engineering from the National Higher School of Electrical Engineering, Electronics, Informatics, Hydraulics and Telecommunications, Toulouse, France, in 2013, the M.Sc. degree in electrical engineering from the Hassania School of Public Works, Casablanca, Morocco, in 2013, and the Ph.D. degree in signal, image, speech, and telecommunication from Grenoble Alpes University, Grenoble, France, in 2018.

She is currently a Researcher with the UiT The Arctic Univesity of Norway, Tromsø, Norway, where she is conducting her research with the Center of Integrated Remote Sensing and Forecasting for Arctic Operations. Her research interests include multimodal data analysis, signal processing, and image processing.

Dr. Chlaily is currently a Reviewer for the IEEE TRANSACTIONS ON GEO-SCIENCE AND REMOTE SENSING, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, and *IEEE Geoscience and Remote Sensing Letters*.

Torbjørn Eltoft (Member, IEEE) received the M.Sc. and Ph.D. degrees from the University of Tromsø, Norway, in 1981 and 1984, respectively.

In 1988, he joined the Department of Physics and Technology, UiT The Arctic University of Norway (UiT), Tromsø, Norway, where he is currently a Professor and the Director of the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations, a centre for research-based Innovation awarded by the Norwegian Research Council in 2014, whose objective is to develop knowledge and remote sensing technology for arctic applications. From 2013 to 2015, he was the Head of the Department of Physics and Technology, UiT. He was an International Researcher with the University of California, Irvine, CA, USA, from 1992 to 1993 and then from 1997 to 1998, and with the University of California, San Diego, CA, USA, from 2004 to 2005. He has a significant publication record in the area of signal processing and remote sensing. His research interests include multidimensional signal and image analysis, statistical modeling, neural networks, and machine learning, with emphasis on applications in multichannel synthetic aperture radar remote sensing and multi-sensor remote sensing for the Arctic.

Dr. Eltoft was an Associate Editor for the Elsevier journal *Pattern Recognition* from 2005 to 2011 and was the Guest Editor for *Remote Sensing*'s Special Issue for the PolInSAR 2017 Conference. He was the co-recipient of the Outstanding Paper Award in Neural Networks awarded by IEEE Neural Networks Council in 2000, Honourable Mention for the 2003 Pattern Recognition Journal Best Paper Award, and the 2017 UiT Award for Research and Development from UiT The Arctic University of Norway.

Wolfgang Dierking received the M.Sc. degree in geophysics from the University of Hamburg, Hamburg, Germany, in 1985, and the Ph.D. degree in physics (ocean remote sensing) from the University of Bremen, Bremen, Germany, in 1989.

Since 2002, he has been with the Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute (AWI), Bremerhaven, Germany, and since 2016, he has been a Professor II with the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations, UiT The Arctic Univesity of Norway, Tromsø, Norway, where he has been leading a group working with remote sensing of sea ice and icebergs. Since 2015, he has been a Lecturer of ocean remote sensing with the University of Applied Sciences, Bremerhaven, Germany. From 2007 to 2019, he was the Head of the research group Earth Observation Systems, AWI. The group dealt with remote sensing of sea ice, ice sheets, and permafrost regions. In 2007, he was a Visiting Professor with the Chalmers University of Technology, Gothenburg, Sweden. From 1996 to 2002, he was an Associate Research Professor with the Danish Centre for Remote Sensing, Technical University of Denmark, Kongens Lyngby, Denmark. His research focuses on the remote sensing of the polar regions.

Prof. Dierking was a member of different advisory groups of the European Space Agency (ESA), which include Sentinel-1 Mission Advisory Group from 2011 to 2015, SAR Advisory Group from 2003 to 2010, and Cat-1 Advisory Group from 2003 to 2015. In 2019, he became a member of Copernicus L-Band SAR Mission Advisory Group, and in 2021 a member of the Sentinel-1 Next Generation Mission Advisory Group. He participated in several field expeditions to the Arctic, Antarctic, and the Baltic Sea. Since 2008, he has been the Principal Investigator with Dragon Cooperation Programme between the Ministry of Science and Technology of China and ESA with focus on sea ice remote sensing.

Frode Dinessen received the M.Sc. degree in applied physics from UiT The Arctic University of Norway, Tromsø, Norway, in 1996.

He is currently a Research Scientist with the Division for Remote Sensing and Data Management, Norwegian Meteorological Institute, Tromsø, Norway. From 1997 to 2001, he was the Leader of the Norwegian Ice Service, Forecasting Division for the Northern Norway, Norwegian Meteorological Institute. The Ice Service is responsible for delivering regular updated sea ice analysis for the public users. From 2001 to 2007, he was a Software Engineer with Kongsberg Spacetec, Tromsø, Norway, developing satellite-based applications. In 2007, he joined the Research Department, Norwegian Meteorological Institute. He is currently the Deputy of the Sea Ice Thematic Assembling Center, providing operational sea ice products to the EU Copernicus Marine Environment Monitoring Service. His research interests include development of operational multi-sensor sea ice concentration products by combining satellite data from synthetic aperture radar, passive microwave radiometers, and national data hub for sentinel data.

Andrea Marinoni (Senior Member, IEEE) received the B.S., M.Sc. (cum laude), and Ph.D. degrees in electronic engineering from the University of Pavia, Pavia, Italy, in 2005, 2007, and 2011, respectively.

He is currently an Associate Professor with the Earth Observation Group, Centre for Integrated Remote Sensing and Forecasting for Arctic Operations, Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway, and a Visiting Academic Fellow with the Department of Engineering, University of Cambridge, Cambridge, U.K. From 2013 to 2018, he was a Research Fellow with the Telecommunications and Remote Sensing Lab, Department of Electrical, Computer and Biomedical Engineering, University of Pavia. In 2009, he was a Visiting Researcher with the Communications Systems Lab, Department of Electrical Engineering, University of California - Los Angeles, Los Angeles, CA, USA. From 2015 to 2017, he was a Visiting Researcher with the Earth and Planetary Image Facility, Ben-Gurion University of the Negev, Be'er Sheva, Israel; the School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China; the School of Computer Science, Fudan University, Shanghai, China; the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China; Instituto de Telecomunicacções, Instituto Superior Tecnico, Lisbon, Portugal; and Universidade de Lisboa, Lisbon, Portugal. In 2020 and 2021, he was a Visiting Professor with the Department of Electrical, Computer and Biomedical Engineering, University of Pavia. His main research interests include efficient information extraction from multimodal remote sensing, nonlinear signal processing applied to large-scale heterogeneous records, earth observation interpretation, big data mining, and analysis and management for human–environment interaction assessment.

Dr. Marinoni was the recipient of the two-year Applied Research Grant, sponsored by the region of Lombardy, Italy, and STMicroelectronics N.V., the INROAD Grant, sponsored by the University of Pavia and Fondazione Cariplo, Milan, Italy, for supporting excellence in design of ERC proposal, the Progetto professionalitá Ivano Becchi Grant, funded by Fondazione Banco del Monte di Lombardia, Pavia, Italy, and sponsored by the University of Pavia and NASA Jet Propulsion Laboratory, Pasadena, CA, USA, for supporting the development of advanced methods of air pollution analysis by remote sensing data investigation, and the Åsgard Research Programme Grants funded by Institut Franccais de Norvége, Oslo, Norway, in 2019 and 2020, respectively, for supporting the development of scientific collaborations between French and Norwegian research institutes. He is the Founder and the chair of the IEEE GRSS Norway Chapter. He is also an Ambassador of the IEEE Region 8 Humanitarian Activities and a Research Contact Point for the Norwegian Artificial Intelligence Research Consortium. He was a Topical Associate Editor of machine learning for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the Guest Editor of three special issues on multimodal remote sensing and sustainable development for the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. He is the Leader of the GR4S Committee within IEEE GRSS, coordinating the organization of schools and workshops sponsored by IEEE GRSS worldwide.

/10
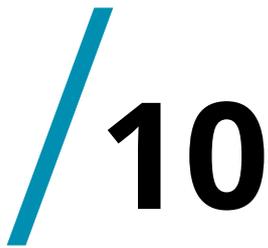
Paper 3: SAR and Passive Microwave Fusion Scheme: a Test Case on Sentinel-1/AMSR-2 Data Sets for Sea Ice Classification

Eduard Khachatrian, Wolfgang Dierking, Saloua Chlaily, Torbjørn Eltoft, Frode Dinessen, Nick Hughes, and Andrea Marinoni

AGU ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters®

Key Points:
- We propose an adaptive scheme for the fusion of passive microwave radiometers and synthetic aperture radar data for sea ice classification
- We demonstrate the flexibility and efficiency of the proposed scheme on a test case by evaluating the performance of various data scenarios
- We illustrate the advantages and limitations of applying the data of each sensor simultaneously and separately

Correspondence to:
E. Khachatrian,
eduard.khachatrian@uit.no

SAR and Passive Microwave Fusion Scheme: A Test Case on Sentinel-1/AMSR-2 for Sea Ice Classification

Eduard Khachatrian[1] , Wolfgang Dierking[1,2] , Saloua Chlaily[1] , Torbjørn Eltoft[1], Frode Dinessen[3], Nick Hughes[3] , and Andrea Marinoni[1]

[1]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway, [2]Helmholtz Center for Polar and Marine Research, Alfred Wegener Institute, Bremerhaven, Germany, [3]Norwegian Meteorological Institute, Tromsø, Norway

Abstract The most common source of information about sea ice conditions is remote sensing data, especially images obtained from synthetic aperture radar (SAR) and passive microwave radiometers (PMR). Here we introduce an adaptive fusion scheme based on Graph Laplacians that allows us to retrieve the most relevant information from satellite images. In a first test case, we explore the potential of sea ice classification employing SAR and PMR separately and simultaneously, in order to evaluate the complementarity of both sensors and to assess the result of a combined use. Our test case illustrates the flexibility and efficiency of the proposed scheme and indicates an advantage of combining AMSR-2 89 GHz and Sentinel-1 data for sea ice mapping.

Plain Language Summary The Earth's land and ocean surface is monitored from space using different sensors mounted on various satellite platforms. Each type of sensor has its advantages and limitations. Combining data from different sensors can potentially solve ambiguities in information retrievals associated with the use of only a single sensor. Here, we apply a multi-sensor fusion scheme that can be used for various data combinations in order to extract relevant information. The main goal of this work is to explore the potential of simultaneously applying two sensors for sea ice mapping and monitoring, namely synthetic aperture radar and passive microwave radiometers, in order to improve the separation of sea ice types.

1. Introduction

In the new millennium, sea ice research has become an important topic due to an unprecedented sea ice decline in the Arctic (England et al., 2020). One reason is that sea ice plays an essential role in the polar ecosystem (Funder et al., 2010). Moreover, the knowledge about sea ice conditions is crucial for polar navigation, offshore operations, weather forecasting, and climate research (Sandven et al., 2006). The main sources of information about sea ice conditions and climatological studies are data from passive microwave radiometers (PMR), and synthetic aperture radars (SAR). The latter is preferably used for tactical navigational support and for local studies requiring data at high spatial resolution. Both sensors are commonly used due to their independence of cloud and light conditions and therefore their ability to provide imaging of the Earth's surface continuously during day and night, and for almost all weather situations (Wang et al., 2016).

Spaceborne SAR provides high spatial resolution images and is one of the main sources from which detailed maps of sea ice conditions are produced for navigation in sea ice or at the ice margin (Karvonen, 2014). SAR is an active microwave sensor, which can achieve spatial resolutions ranging from about 1 m to 100 m (Johannessen et al., 2007). Image products of lower resolutions provide swath widths up to 500 km. An increased spatial resolution reduces the achievable swath width down to a few kilometers. The interpretation of SAR data is challenging due to the complex relationship between radar backscatter and sea ice surface and volume properties and strongly relies on the knowledge of sea ice experts (Zakhvatkina et al., 2019).

Passive microwave radiometers are another type of sensor and are commonly used for large-scale sea ice observations. These sensors operate at multiple frequencies, each of which has different spatial resolutions and sensitivity to atmospheric parameters, in particular to cloud liquid water and atmospheric water vapor (Spreen et al., 2008). However, even the finest spatial resolution achievable with passive microwave sensors is about 3 km and hence significantly coarser than in the case of SAR. The finest spatial resolution is achieved at frequencies of around 90 GHz which, however, comes with a greater susceptibility to atmospheric noise. This effect is less severe at
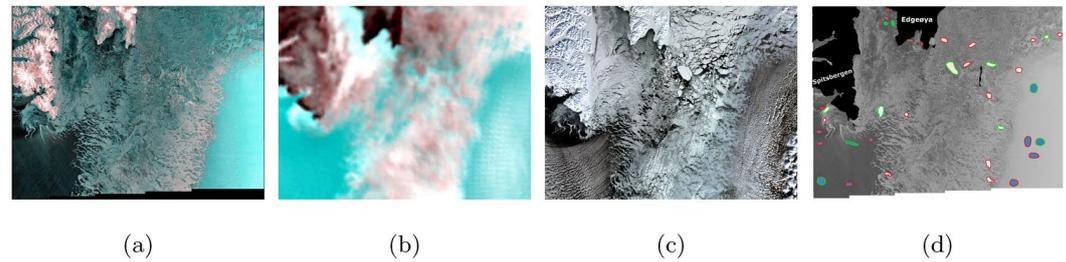
(a)   (b)   (c)   (d)

Figure 1. Color representation of the data set: (a) false-color composite SAR (HV, HH, and HH as RGB), (b) passive microwave radiometer (H, V, and V as RGB), (c) natural-color composite of an optical image from Sentinel-3 Ocean and Land Color Instrument (Bands 8, 6, 4), and (d) spatial distribution of regions of interest that were used for training (polygons with green boundaries) and testing (red boundaries); here the blue color refers to Open Water, white color corresponds to Brash Ice, cyan illustrates the young Grey Ice, pink color shows the Thin First-Year Ice, and black color indicates the landmask for Svalbard.

lower frequencies at the expense of much reduced spatial resolutions at tens of kilometers. The main benefit of PMR is that their wide swaths allow a daily coverage over most of the ice-covered polar regions which is extremely useful for monitoring sea ice at synoptic scales (Heinrichs et al., 2006).

In the past, some attempts were undertaken to combine PMR and SAR data, to improve sea ice concentration estimation that is normally performed with only PMR (Beaven et al., 1996; Karvonen, 2014; Wang et al., 2016). In our study, we focus on exploring the potential of improving SAR-based sea ice classification by additionally using PMR imagery.

We apply a data fusion scheme that includes the graph-based information selection method which relies on information theory metrics in conjunction with a supervised classification approach (Liaw & Wiener, 2001). The flexibility of the proposed scheme is well-suited for efficiently analyzing different data combinations that can be beneficial for sea ice monitoring. In this short letter, we investigate the potential of combining SAR and PMR data for sea ice classification, which we demonstrate for a test case using a combination of data from the Sentinel-1 C-band (5.405 GHz) SAR, and the 89 GHz channel of the Advanced Microwave Scanning Radiometer 2 (AMSR-2).

The rest of this paper is organized as follows. Section 2 describes the data sets used in this study. Section 3 provides details of the proposed data fusion scheme. Section 4 presents the experimental validation of the proposed method. Finally, the discussion and conclusions are presented in Section 5.

2. Data Set

The following section describes the Sentinel-1/AMSR-2 data set. The Sentinel-1 image was acquired in extra-wide swath mode in dual-polarization (HH and HV) at 40 m spatial resolution, which is commonly used for mapping sea ice. From AMSR-2 we use the brightness temperature (BT) of the 89 GHz channel at horizontal (H) and vertical (V) polarization which has a footprint of 3×5 km. The data set was acquired over the southern part of Svalbard on 17 March 2021. The Sentinel-1 data were corrected for thermal noise and calibrated to sigma-naught in dB using the ESA Sentinel-1 Toolbox. Both data sets were collocated and the AMSR-2 data were upsampled to the Sentinel-1 pixel-size which means that adjacent pixels can be completely correlated. Figure 1 shows the false-color composites of (a) SAR, and (b) PMR images, as well as (c) a natural-color composite of an optical image acquired by the Ocean and Land Color Instrument on Sentinel-3. The time gap between SAR and PMR scenes was a few minutes, while the optical scene was acquired a few hours later. In particular over the Open Water (OW) area on the right side, indications of cloud liquid water and atmospheric water vapor are visible, which corresponds to the occurrence of clouds over the same area in the optical image.

The Sentinel-3 optical data were used in addition to the SAR and PMR scenes for identifying various sea ice classes, especially the Grey Ice (GI) in the Storfjorden. One main motivation to add PMR data for sea ice mapping is based on occasional difficulties to separate OW and ice in SAR imagery. Despite its sensitivity to atmospheric parameters, we selected the 89 GHz channel because of its higher spatial resolution compared to the lower-frequency PMR bands. Using the three different data sources together was extremely beneficial for visual

Table 1
Classes Determined by Sea Ice Experts From the Visual Inspection, Along With a Number of Training Samples and Regions of Interest Used for Performance Evaluation

Name	Description	Training samples	ROIs All	ROIs Train	ROIs Test
OW	Open water	313,751	12	4	8
BI	Brash ice	145,112	9	3	6
GI	Grey ice	41,502	6	2	4
TFYI	Thin first-year ice	221,092	9	3	6

Note. The number of training samples refers to the number of pixels in the ROIs specifically used for training.

inspection and manual identification of various sea ice and water classes. Supported by the input of an expert from the Norwegian Meteorological Institute (MET Norway) we could distinguish the following sea ice types: brash ice (BI), GI, and thin first-year ice (TFYI), as well as OW areas. The spatial distribution of the regions of interest (ROIs) that were used for the classification is illustrated in Figure 1d, while the classes, number of training samples, and ROIs used are shown in Table 1. We note that the ROIs are placed with large enough distances between them to avoid mapping upsampled PMR pixels corresponding to a single original PMR pixel into different ROIs. Nevertheless, the degree of correlation between pixels is considerably higher for the PMR data than for the SAR data.

3. Methods

In the following subsections, we briefly describe the main steps of the proposed fusion scheme, namely pre-processing including the collocation and upsampling of the original data sets, as well as extraction of textural features for SAR data, the selection of relevant attributes, and finally the parallel classification. A more detailed technical description of the information retrieval method can be found in Khachatrian et al. (2021).

3.1. Pre-Processing

The Sentinel-1 and AMSR-2 images differ in their areal coverage, spatial resolution, and the coordinate system used for presenting the data. Therefore as a first step, we make the data comparable by means of collocating, upsampling the PMR data to the SAR resolution, and extracting the overlapping area. In addition to the original radar intensities, we extract 10 texture features available in the ESA's Sentinel Application Platform (SNAP) for each SAR polarization using the Gray-Level Co-Occurrence Matrix (GLCM) (Haralick et al., 1973; Kandaswamy et al., 2005) for 0°, 45°, 90°, and 135°. The results are averaged which is common practice to account for the possible rotation of different sea ice or ocean surface structures.

3.2. Attribute Selection

In this section, we briefly describe the information selection method (referred to as GKMI, i.e., Gaussian kernel and mutual information) that is part of the proposed scheme which we employ to select relevant attributes (Khachatrian et al., 2021). This approach consists of three main steps: segmentation, graph building, and graph clustering. It should be noted that even though the whole processing scheme proposed in this study is supervised due to the selected classification method, the information selection step is unsupervised.

3.2.1. Segmentation

In order to preserve the particularity of distinct areas in the observed Sentinel-1 and AMSR-2 scenes and to optimize the algorithm in terms of execution time (ET), we implemented the attribute selection on superpixels, that is, groups of neighboring pixels showing homogeneous characteristics throughout the considered multivariate data set. It should be noted that there are other patch-wise approaches, for example, using regularly spaced windows. However, fixed windows often still cover variations of characteristics, which is avoided in the superpixel approach. Moreover, superpixels allow employing the algorithm on a local scale, which is crucial since some image parts might require different types of attributes to effectively represent different ice types and characteristics. In the SAR images, the decrease of the local incidence from near- to far-range has to be considered as well. An image can be split into superpixels using different segmentation methods, such as Watershed (Beucher, 1992) or Felzenszwalb (Felzenszwalb & Huttenlocher, 2004). In this work, we determine the superpixels using the Simple Linear Iterative Clustering segmentation method (Neubert & Protzel, 2014). Furthermore, the number of superpixels, as well as their size, are parameters that can be changed within the algorithm depending on user preferences and applications. The superpixel segmentation is only used as a part of the information selection step.

3.2.2. Graph Building

To find the attributes which are best suited for sea ice classification, we apply a selection method, that relies on information theory metrics and on a representation based on graph Laplacians. Unlike existing graph-based

Table 2
Classification Performance

Data set	N	OA (%)	AA (%)	k (%)	ET (min)
Single-sensor					
SAR	2	75.6	68.3	64.1	98.1
SAR & GLCM	22	81.6	74.5	73.3	63.1
PMR	2	75.2	64.8	63.3	25.2
Multi-sensor without selection					
SAR & PMR	4	79.6	78.0	69.9	29.1
CMD	24	90.2	90.4	85.6	42.8
Multi-sensor with selection					
CMD	5	78.2	72.1	67.9	23.5
CMD	10	91.3	90.6	87.1	22.3
CMD	15	91.1	91.5	86.6	22.6
CMD	Auto	**93.2**	**93.8**	**90.1**	**21.3**

Note. N shows the number of selected/used attributes, k refers to the Kappa coefficient, AA (%) to the average accuracy, OA (%) to the overall accuracy, and ET is execution time. CMD denotes combined multi-sensor data that combines all the available sources, namely SAR, GLCM textural features, and PMR. The performance evaluation was implemented using an Intel Core i7 CPU at 2,6 GHz with 32 GB RAM. The best performance values of the OA, AA, ET, and k are shown in bold.

clustering methods that are only using kernels as similarity measures (representing the mutual relations between the data points), we are also considering the information content of the original data. Therefore, the similarity is quantified using two metrics simultaneously, which allows us to capture relevant information at different scales which improves the precision of the selection. The mutual information (MI) is performed globally and provides a better estimation of the attributes-shared information (Vergara & Estévez, 2014), while the Gaussian kernel (GK) is applied locally and preserves the structure of the original data (Luxburg, 2007).

3.2.3. Graph Clustering

Once the graph is defined according to the operations that have been previously introduced, we perform the partition of the graph using a procedure inspired by the spectral clustering approach (Luxburg, 2007) in order to identify and select the most relevant attributes in the data set. The GKMI method forms groups of similar attributes, according to MI and GK metrics, and selects the most relevant from each grouping. Thus we obtain subsets of attributes that preserve the structure and the information content within a particular superpixel. The grouping is performed using k-means, which is a simple and commonly used clustering algorithm (Theodoridis & Koutroumbas, 2008).

3.2.4. Adaptive Selection

As was mentioned above we assume that different homogeneous parts of the image, that is, superpixels, might have different relevant subsets of attributes. Not only the attributes might differ between superpixels, but also their number, depending on the specific ice characteristics. Therefore, we include an adaptive attribute selection based on the kneedle method (Satopaa et al., 2011) in the proposed scheme, which allows us to automatically determine the relevant number of attributes. The kneedle algorithm finds the maxima of the curve to detect a beneficial point, or a "knee." In our case, this means determining the optimal number of attributes after which adding any additional attributes will be redundant.

3.3. Classification

The final step of the proposed scheme is the application of the PMR and SAR data in combination for sea ice classification. The classification is carried out using the Random Forest method, which is a widely applied classifier in remote sensing (Liaw & Wiener, 2001). Furthermore, for various superpixels, different numbers and attributes are selected. We consider this in the pixel-wise classification which improves the performance of the algorithm. In our experiments, we use pixels from the predefined ROIs for training and testing the algorithm. Moreover, employing different ROIs avoids overfitting.

4. Experiments

In this section, we evaluate the performance of the proposed scheme for sea ice mapping and characterization of different sea ice types. In order to quantitatively evaluate the result of sea ice classification, we apply several metrics: the Overall Accuracy (OA) index, Average Accuracy (AA), Cohen's Kappa coefficient (k), and ET. The OA shows the percentage of correctly classified samples, AA quantifies the mean of class-specific accuracies for all classes, while Kappa measures the agreement between the classification and the reference data (Bharatkar & Patel, 2013).

4.1. Performance Analysis

In order to properly evaluate the SAR and PMR fusion scheme for sea ice classification, we conducted different experiments by varying the input data combination and assessing the performance of each combination (Table 2).

Table 2 shows the OA, AA, ET, and Kappa coefficient (k) obtained for the various cases: single- and multi-sensor scenarios without adaptive attribute selection, and with the proposed scheme using different numbers of selected

attributes. The classification accuracies are lower if only a single sensor is used, while execution times are higher in comparison to multi-sensor combinations. It is especially evident in the case of SAR intensity (Table 2: $N = 2$ for the two polarizations) that the ET is significantly larger than for other data combinations with more attributes ($N > 2$). Because of limited complementary information, the algorithm struggles to properly assign each point to a particular class which results in lower accuracy and higher ET. Using single-sensor scenarios with a low number of considered attributes, which here are the SAR intensity or the PMR BT at two polarizations, results in a lower classification performance and reveals the lowest accuracy among all the experiments. Due to the lower resolution, PMR misses narrow variations of ice structures and types, which causes lower accuracies if the classification scheme is more detailed. In the case of SAR, low accuracies are most likely due to the influence of speckle and thermal noise. Furthermore, SAR intensity contrasts do not differ strongly between the GI and TFYI in the selected scene, which makes their separation more difficult. The best classification result of the single-sensor cases is found for the combination of intensity and GLCM textural features. Nevertheless, the multi-sensor case outperforms the single-sensor cases in terms of accuracy and computation time. The best performance is achieved for the multi-sensor case with integrated information selection. We tested this on two main scenarios, with a fixed number of attributes and adaptive/automatic selection. For all the cases with selection, except for a fixed ($N = 5$), the performance scores are higher and the ET is almost halved in comparison to the whole set of attributes without the selection. Moreover, the highest accuracy scores and the lowest ET were achieved with the proposed fusion scheme with an automatic determination of the number of attributes. These results strongly support the relevance of automatic information selection and parallel classification which are the crucial parts of the proposed multi-sensor fusion scheme.

Figure 2 displays the classification results for different data scenarios. Furthermore, the red contours indicate several challenging areas for classification that was improved considerably when using SAR and PMR in combination, together with variations of the proposed scheme as shown in Table 2. The classified maps on the first row of Figures 2a, 2b, and 2c show the single-sensor scenarios. The presented images clearly demonstrate the difference in spatial resolution between PMR and SAR sensors. Because of its coarser resolution, a smoother classification map is obtained from PMR. However, it is not possible to identify sea ice structures and variations of sea ice types that are smaller than the PMR resolution cell. Furthermore, there are areas of misclassification in the OW on the right side of Figure 2a, which are caused by cloud liquid water and water vapor in the atmosphere. The original SAR image is affected by speckle and thermal noise which is transferred into the calculation of texture parameters and the final classification. Furthermore, GI is very often misclassified in the SAR data. Nevertheless, the classified map obtained from SAR considering GLCM textural features illustrates a significant improvement, especially for the GI and OW classes that were misclassified in the intensity-only case.

The classified maps on the second row of Figures 2d, 2e, and 2f illustrate the multi-sensor examples with and without the proposed attribute selection. Each of these maps provides a more accurate classification compared to the single-sensor examples. The combination of SAR and PMR shows, for example, an improvement compared to the single PMR case because the potential influence of cloud liquid water and water vapor in the atmosphere is slightly reduced. On the other hand, the classified maps based on a combination of SAR and PMR preserve the smoothness of the PMR and the ability of the SAR image to identify smaller sea ice and water surface details, such as BI in the marginal ice zone. The classified map obtained with the combined data set without and with the proposed attribute selection depicts more of the GI in Storfjorden between Spitsbergen and Edgeøya in comparison to the PMR and SAR combination without texture. The performance evaluation along with visual inspection of the classified maps shows the advantages of combining various sources in order to obtain unique information regarding the area of interest. Moreover, even though the PMR is mostly used for coarse-resolution wide-coverage products, such as in the case of sea ice concentration, our study demonstrates the usefulness of integrating the PMR for sea ice type classification.

5. Conclusions

In this test case study, we demonstrate the application of a new, flexible, adaptive, highly accurate, and efficient multi-sensor fusion scheme on specific data combinations obtained from SAR and PMR, the most commonly used sensors for sea ice classification, ice charting, and climatological sea ice monitoring. We presented sea ice classification results separately for each sensor. The results indicate that the addition of PMR can improve the SAR-based classification in certain cases. The suggested fusion scheme reveals a better classification performance and needs less computation time compared to other investigated methods.
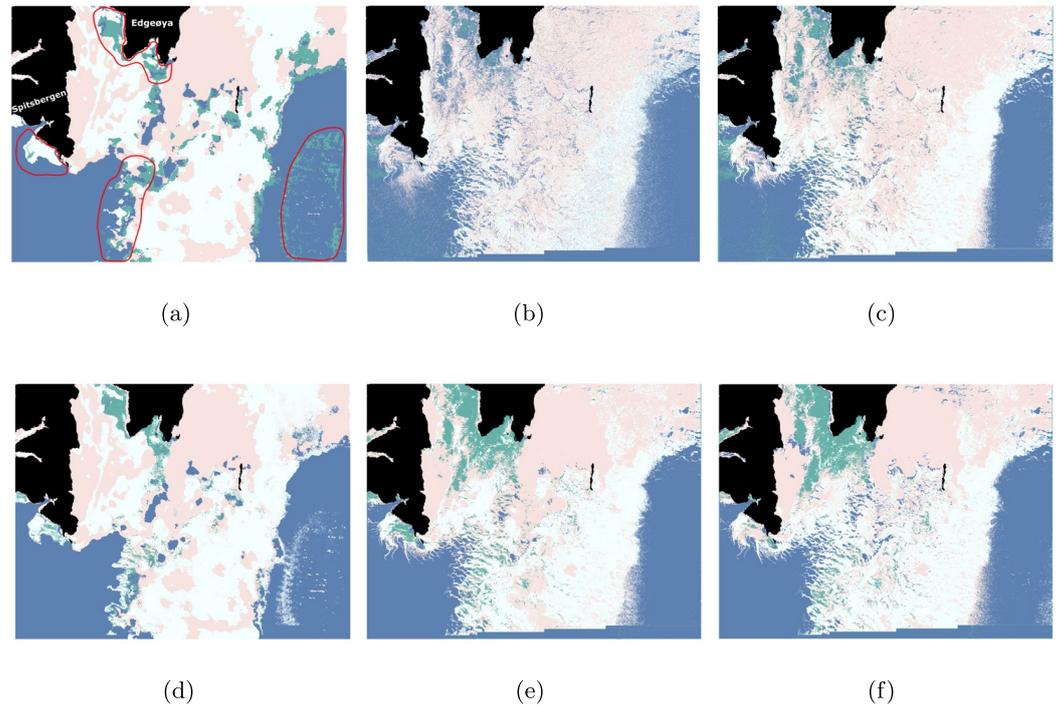
Figure 2. Classified maps using the RF method for different data sets: (a) only passive microwave radiometer (PMR) brightness temperature (BT), (b) only synthetic aperture radar (SAR) intensity, (c) SAR intensity plus Gray-Level Co-Occurrence Matrix (GLCM) textural features, (d) combined SAR intensity and PMR BT, (e) combined SAR intensity plus GLCM textural features, and PMR BT, and (f) adaptive attribute selection applied on the combination of SAR intensity plus GLCM textural features and PMR BT. The single- and multi-sensor cases in (a)–(e) were used without selection. The classes and corresponding colors are similar to Figure 1d. Red contours indicate several challenging areas for single-sensor classification. The classification was improved with the fusion of different sensors and the employment of the proposed scheme.

Because of the promising results, we plan to extend our study, which here is limited to one test case, to different ice conditions. The test case data were acquired in cold and dry winter conditions, therefore the PMR provided an image that is only marginally affected by cloud liquid water and water vapor in the atmosphere. It is also known that sea ice classification with SAR and PMR is more difficult in the melting season. Thus, future work should additionally assess the seasonal robustness of the approach and its applicability to different times of the year. We are preparing an extended data set with a larger number of Sentinel-1/AMSR-2 data acquisitions over a full season and with more complex ice conditions. In addition, we consider to investigate the usefulness of lower PMR frequency channels and derived ice concentration, considering the sensitivity of the 89 GHz channel to atmospheric parameters.

Data Availability Statement

Sentinel-1 SAR and AMSR-2 PMR data sets are publicly available through *Copernicus Open Access Hub* (https://scihub.copernicus.eu/dhus/#/home) and *GCOM-W1 Data Providing Service* (https://gportal.jaxa.jp/gpr/).

References

Beaven, S., Gogineni, S., & Carsey, F. (1996). Fusion of satellite active and passive microwave data for sea ice type concentration estimates. *IEEE Transactions on Geoscience and Remote Sensing*, *34*(5), 1172–1183. https://doi.org/10.1109/36.536534

Beucher, S. (1992). The watershed transformation applied to image segmentation. In *Proceedings of the 10th Pfefferkorn conference on signal and image processing in microscopy and microanalysis* (pp. 299–314). Retrieved from http://folk.uib.no/eha070/mat262/papers/Beucher.pdf

Bharatkar, P., & Patel, R. (2013). Approach to accuracy assessment tor RS image classification techniques. *International Journal of Scientific Engineering and Research*, *4*(12), 79–86. Retrieved from https://www.ijser.org/researchpaper/Approach-to-Accuracy-Assessment-tor-RS-Image-Classification-Techniques.pdf

England, M. R., Polvani, L. M., & Sun, L. (2020). Robust arctic warming caused by projected Antarctic sea ice loss. *Environmental Research Letters*, *15*(10), 104005. https://doi.org/10.1088/1748-9326/abaada

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, *59*(2), 167–181. https://doi.org/10.1023/B:VISI.0000022288.19776.77

Funder, S., St. John, K., Jennings, A. E., Darby, D. A., Cronin, T. M., Polyak, L., et al. (2010). History of sea ice in the Arctic. *Quaternary Science Reviews*, *29*(15–16), 1757–1778. https://doi.org/10.1016/j.quascirev.2010.02.010

Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(6), 610–621. https://doi.org/10.1109/tsmc.1973.4309314

Heinrichs, J., Cavalieri, D., & Markus, T. (2006). Assessment of the AMSR-E sea ice-concentration product at the ice edge using RADARSAT-1 and MODIS imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(11), 3070–3080. https://doi.org/10.1109/TGRS.2006.880622

Johannessen, O., Alexandrov, V., Frolov, I., Bobylev, L., Sandven, S., Pettersson, L., et al. (2007). Remote sensing of sea ice in the northern sea route: Studies and applications. https://doi.org/10.1007/978-3-540-48840-8

Kandaswamy, U., Adjeroh, D. A., & Lee, M. C. (2005). Efficient texture analysis of SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(9), 2075–2083. https://doi.org/10.1109/TGRS.2005.852768

Karvonen, J. (2014). A sea ice concentration estimation algorithm utilizing radiometer and SAR data. *The Cryosphere*, 8(5), 1639–1650. https://doi.org/10.5194/tc-8-1639-2014

Khachatrian, E., Chlaily, S., Eltoft, T., Dierking, W., Dinessen, F., & Marinoni, A. (2021). Automatic selection of relevant attributes for multi-sensor remote sensing analysis: A case study on sea ice classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 9025–9037. https://doi.org/10.1109/JSTARS.2021.3099398

Liaw, A., & Wiener, M. (2001). Classification and regression by randomForest. *Forest*, *23*.

Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416. https://doi.org/10.1007/s11222-007-9033-z

Neubert, P., & Protzel, P. (2014). Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms. In *Proceedings—International conference on pattern recognition* (pp. 996–1001). https://doi.org/10.1109/ICPR.2014.181

Sandven, S., Johannessen, O. M., & Kloster, K. (2006). Sea ice monitoring by remote sensing. *Encyclopedia of Analytical Chemistry, 1993*, 1–43. https://doi.org/10.1002/9780470027318.a2320

Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops* (pp. 166–171). https://doi.org/10.1109/ICDCSW.2011.20

Spreen, G., Kaleschke, L., & Heygster, G. (2008). Sea ice remote sensing using AMSR-E 89-GHz channels. *Journal of Geophysical Research*, *113*(C2), C02S03. https://doi.org/10.1029/2005JC003384

Theodoridis, S., & Koutroumbas, K. (2008). *Pattern recognition* (4th ed.). Academic Press, Inc.

Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing & Applications*, *24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0

Wang, L., Scott, K. A., & Clausi, D. A. (2016). Improved sea ice concentration estimation through fusing classified SAR imagery and AMSR-E data. *Canadian Journal of Remote Sensing*, *42*(1), 41–52. https://doi.org/10.1080/07038992.2016.1152547

Zakhvatkina, N., Smirnov, V., & Bychkova, I. (2019). Satellite SAR data-based sea ice classification: An overview. *Geosciences*, *9*(4), 152. https://doi.org/10.3390/geosciences9040152

11

Conclusions

11.1   General Conclusions

The main focus of this dissertation is to develop an information selection algorithm that preserves global and local information about the original data by the joint employment of two metrics simultaneously and at different scales. This ensures a better separability of the attributes and, hence, a more precise selection which leads to a better algorithms performance, in our case for sea ice classification. The focus of each paper was on improving different aspects of the proposed method, starting with tuning the main parameters (*Paper I*), exploring the potential for various data combinations for sea ice classification (*Paper II*, *Paper III*), and investigating the possibility for operational sea ice monitoring and automatic selection of attributes (*Paper III*).

In *Paper I* we introduce the method and its main novelties, tune and motivate the employment of several parameters and metrics, and compare the proposed method with existing techniques using publicly available urban scenes that have ground truth labels. The experimental results obtained from several multimodal data sets consistently demonstrated the effectiveness and robustness of the proposed method. Moreover, they reveal the crucial advantages of the proposed method, namely (a) it employs two similarity metrics that account for global and local particularities of the original data set, which, in turn, allows the selection of the most relevant attributes; (b) it is performed on a superpixel level, therefore, it selects the best descriptive attributes for local image areas; (c) it is application independent, therefore, it does not require any prior information about class labels; (d) it can be applied to data sets obtained from various sensors with different characteristics.

In *Paper II* as a first step towards sea ice application, we tested the proposed scheme on several data combinations, namely multi-frequency L and C-band airborne data and multi-sensor satellite data. The airborne data provided a better separability of ice types because of high spatial resolution. Moreover, L and C-band grasp different information about the sea ice, due to the difference in radar penetration depth. Additionally, we used satellite data combinations which are used in operational ice charting but also in studies of the ocean - sea ice - atmosphere interac-

tions. However, they usually suffer from ambiguities when only using single sensors. Therefore, in both airborne and satellite cases we specifically tested the performance of using single-source and multi-source scenarios. In all cases, the combination of data from various sources guarantees better algorithm performance. Results show the ability of the method to process different data combinations and the importance of employing multi-sensor or multi-frequency data for the characterization of sea ice types by comparing the performance of using several sensors separately and simultaneously.

As a first step towards operational sea ice monitoring in *Paper III*, we investigate the potential of combining SAR and PMR data for sea ice classification, which we demonstrate for a test case using a combination of data from Sentinel-1 and AMSR-2. We illustrate both the advantages and limitations of applying each sensor separately and simultaneously and found a clear advantage of the PMR-SAR combination for the given test case. Because of the restriction to one test case, the method should be further investigated. The first results clearly indicate that the addition of PMR can improve the SAR-based classification in certain cases.

11.2   **Future Work**

The developed algorithm has been tested on various data combinations and sea ice scenarios. The results clearly show the potential to improve the sea ice classification. However, there are still some experiments that should be done to properly integrate the proposed scheme for operational sea ice monitoring.

In *Paper I* we set some future work directions that were focused on adding the automatic selection of the number of attributes for each superpixel so that the multimodal data analysis can be adapted to the different conditions of the records that can be acquired in large-scale scenarios, and on developing an adaptive classifier that can deal with superpixels of heterogeneous sizes and attributes. Accordingly, we fulfilled this goal by integrating the automatic selection of the number of attributes in *Paper III*.

In *Paper II* we set a few more future goals. First and foremost, we should investigate in detail how our results are affected by the observation conditions, such as radar incidence angle or sun elevation angle. Another item concerns the availability of the investigated sensor combinations for operational sea ice mapping. It has to be considered that not all the sensors may be available. Moreover, even though optical data provide a better classification performance for several cases considered in *Paper II*, it is not of general validity due to the significant limitations of optical sensors and their ability to work only under favorable light and cloud conditions. Furthermore, in addition to the limitations mentioned above, optical sensors are unable to see through snow, hence cannot separate snow-covered sea ice classes. Thereby, the information provided by optical sensors is limited and cannot always provide a better classification performance. Therefore, we should employ an "only-add-if-usable" strategy. The combination of different radar frequencies (C-band and L-band) for operational services is recently under investigation in the ROSE-L project [91].

In *Paper III*, we set a goal: due to the promising results of one test case in *Paper III*, we plan to investigate the operational capabilities of our method on extended data set with a larger number of Sentinel-1/AMSR-2 data acquisitions over a full season and with more complex ice conditions. It will additionally allow us to assess the seasonal robustness of the approach and its applicability to different times of the year. In addition, we consider investigating the use of lower frequency channels and derived ice concentration, considering the sensitivity of the 89 GHz channel to

atmospheric parameters. Even though PMR is actively used mainly to separate water and ice and obtain information regarding sea ice concentration and the position of the sea ice edge, we think it is highly valuable to consider more cases in which PMR can serve as a complementary source for SAR-based sea ice classification. While SAR provides more information about older sea ice types due to the radar penetration depth, PMR can help characterize the marginal ice zone and distinguish younger ice types. Nevertheless, it should be further investigated with a larger data set, in order to better reveal the limitations and advantages of both sensors depending on various sea ice and weather conditions, especially for PMR that can be significantly affected by cloud liquid water and water vapor in the atmosphere.

Bibliography

[1] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[2] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[3] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Q. Du, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 331–342, Feb 2012.

[4] S. Chlaily, P. Amblard, O. Michel, and C. Jutten, "Impact of noise correlation on multimodality," in *2016 24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 195–199.

[5] Y. Mohan and A. P. Singh, "Challenges and opportunities in big data: A review," vol. 10, pp. 36–46, 12 2022.

[6] "Front matter," in *Anomaly Detection and Complex Event Processing over IoT Data Streams*, P. Schneider and F. Xhafa, Eds. Academic Press, 2022, pp. i–iii. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012823818900002X

[7] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th ed. Orlando, FL, USA: Academic Press, Inc., 2008.

[8] C. Gakii, P. O. Mireji, and R. Rimiru, "Graph based feature selection for reduction of dimensionality in next-generation rna sequencing datasets," *Algorithms*, vol. 15, no. 1, 2022. [Online]. Available: https://www.mdpi.com/1999-4893/15/1/21

[9] S. Funder, K. St. John, A. E. Jennings, D. A. Darby, T. M. Cronin, L. Polyak, J. T. Andrews, J. White, A. Dyke, J. Brigham-Grette, E. Wolff, M. Serreze, M. O'Regan, J. Savelle, R. Alley, J. Fitzpatrick, M. Holland, and G. Miller, "History of sea ice in the Arctic," *Quaternary Science Reviews*, vol. 29, no. 15-16, pp. 1757–1778, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.quascirev.2010.02.010

[10] National Snow and Ice Data Center (NSIDC), "Arctic sea ice news analysis," https://nsidc.org/arcticseaicenews/charctic-interactive-sea-ice-graph/, 2023.

[11] T. Vihma, "Effects of Arctic Sea Ice Decline on Weather and Climate: A Review," *Surveys in Geophysics*, vol. 35, no. 5, pp. 1175–1214, 2014. [Online]. Available: https://doi.org/10.1007/s10712-014-9284-0

[12] A. Gardner and M. Sharp, "A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization," *Journal of Geophysical Research*, vol. 115, 03 2010.

[13] R. Lindsay, "Arctic sea-ice albedo derived from rgps-based ice-thickness estimates," *Annals of Glaciology*, vol. 33, p. 225–229, 2001.

[14] R. E. Brandt, S. G. Warren, A. P. Worby, and T. C. Grenfell, "Surface albedo of the antarctic sea ice zone," *Journal of Climate*, vol. 18, no. 17, pp. 3606 – 3622, 2005. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/18/17/jcli3489.1.xml

[15] K. Aagaard and E. Carmack, "The role of sea ice and other fresh water in the arctic circulation," *Journal of Geophysical Research*, vol. 94, pp. 14,485–14,498, 10 1989.

[16] Norwegian Ice Service, https://cryo.met.no/, "Norwegian ice service," https://cryo.met.no/, 2022.

[17] M. Spindler, "Notes on the biology of sea ice in the Arctic and Antarctic," *Polar Biology*, vol. 14, no. 5, pp. 319–324, 1994. [Online]. Available: https://doi.org/10.1007/BF00238447

[18] M. Macias-Fauria and E. Post, "Effects of sea ice on arctic biota: An emerging crisis discipline," *Biology Letters*, vol. 14, p. 20170702, 03 2018.

[19] H. Eicken, B. Bluhm, R. Collins, R. Gradinger, C. Haas, M. Ingham, A. Mahoney, M. Nicolaus, and D. Perovich, *Field techniques in sea-ice research*, 01 2014.

[20] H. P. Huntington, S. Gearheard, L. K. Holm, G. Noongwook, M. Opie, and J. Sanguya, *Sea ice is our beautiful garden: indigenous perspectives on sea ice in the Arctic*. John Wiley & Sons, Ltd, 2017, ch. 25, pp. 583–599. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118778371.ch25

[21] Y. Aksenov, E. E. Popova, A. Yool, A. G. Nurser, T. D. Williams, L. Bertino, and J. Bergh, "On the future navigability of arctic sea routes: high-resolution projections of the arctic ocean and sea ice," *Marine Policy*, vol. 75, pp. 300–317, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0308597X16000038

[22] S. Sandven, O. M. Johannessen, and K. Kloster, "Sea Ice Monitoring by Remote Sensing," *Encyclopedia of Analytical Chemistry*, vol. 1993, no. March 1993, pp. 1–43, 2006.

[23] H. Eicken, R. Gradinger, M. Salganek, K. Shirasawa, D. Perovich, and M. Leppäranta, *Field Techniques for Sea-Ice Research*, 01 2010.

[24] L.-A. Breivik, T. Carrieres, S. Eastwood, A. Fleming, F. Girard-Ardhuin, J. Karvonen, R. Kwok, W. Meier, M. Mäkynen, L. Pedersen, S. Sandven, M. Simila, and R. Tonboe, "Remote sensing of sea ice," 01 2009.

[25] L. Wang, K. A. Scott, and D. A. Clausi, "Improved sea ice concentration estimation through fusing classified sar imagery and amsr-e data," *Canadian Journal of Remote Sensing*, vol. 42, no. 1, pp. 41–52, 2016. [Online]. Available: https://doi.org/10.1080/07038992.2016.1152547

[26] WMO, *WMO Sea-ice Nomenclature: Terminology, Codes, Illustrated Glossary and Symbols*, ser. WMO sea-ice nomenclature. Secretariat of the World Meteorological Organization, 1970, no. . 4. [Online]. Available: https://books.google.rs/books?id=BRVUmwEACAAJ

[27] R. Kwok, S. Nghiem, S. Yueh, and D. Huynh, "Retrieval of thin ice thickness from multifrequency polarimetric sar data," *Remote Sensing of Environment*, vol. 51, no. 3, pp. 361–374, 1995. [Online]. Available: https://www.sciencedirect.com/science/article/pii/003442579400017H

[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[29] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artificial Intelligence Review*, 2019. [Online]. Available: https://doi.org/10.1007/s10462-019-09682-y

[30] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4581–4593, 2018.

[31] M. A. Hossain, M. Pickering, and X. Jia, "Unsupervised feature extraction based on a mutual information measure for hyperspectral image classification," *International Geoscience and Remote Sensing Symposium (IGARSS)*, no. October 2015, pp. 1720–1723, 2011.

[32] P. Olver and C. Shakiban, *Applied Linear Algebra*. Prentice Hall, 2006. [Online]. Available: https://books.google.no/books?id=D2tyQgAACAAJ

[33] I. Jolliffe and Springer-Verlag, *Principal Component Analysis*, ser. Springer Series in Statistics. Springer, 2002. [Online]. Available: https://books.google.no/books?id=_olByCrhjwIC

[34] C. Syms, "Principal Components Analysis," *Encyclopedia of Ecology, Five-Volume Set*, pp. 2940–2949, 2008.

[35] C. Persello, J. D. Wegner, R. Hänsch, D. Tuia, P. Ghamisi, M. Koeva, and G. Camps-Valls, "Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 2, pp. 172–200, 2022.

[36] D. Tuia, R. Roscher, J. D. Wegner, N. Jacobs, X. Zhu, and G. Camps-Valls, "Toward a collective agenda on ai for earth science data analysis," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 2, pp. 88–104, 2021.

[37] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian, "Feature selection using principal feature analysis," in *Proceedings of the 15th ACM International Conference on Multimedia*, ser. MM '07. New York, NY, USA: ACM, 2007, pp. 301–304. [Online]. Available: http://doi.acm.org/10.1145/1291233.1291297

[38] E. Khachatrian, S. Chlaily, T. Eltoft, and A. Marinoni, "Selecting principal attributes in multimodal remote sensing for sea ice characterization," in *EUSAR 2021; 13th European Conference on Synthetic Aperture Radar*, 2021, pp. 1–6.

[39] S. B. Serpico, M. D'Inca, F. Melgani, and G. Moser, "Comparison of feature reduction techniques for classification of hyperspectral remote sensing data," *Image and Signal Processing for Remote Sensing VIII*, vol. 4885, no. March, pp. 347–358, 2003.

[40] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images," *Pattern Recognition*, vol. 51, pp. 295–309, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2015.08.018

[41] Y. Zhou, R. Zhang, S. Wang, and F. Wang, "Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy," *Sensors*, vol. 18, p. 2013, 06 2018.

[42] S. Georganos, T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, S. Kalogirou, and E. Wolff, "Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application," *GIScience & Remote Sensing*, vol. 55, no. 2, pp. 221–242, 2018.

[43] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," *CoRR*, vol. abs/1202.3725, 2012. [Online]. Available: http://arxiv.org/abs/1202.3725

[44] S. Sivakumar and C. Chandrasekar, "Feature Selection Using Genetic Algorithm with Mutual Information," vol. 5, no. 3, pp. 2871–2874, 2014.

[45] P. Somol, P. Pudil, F. J. Ferri, and J. Kittler, "Fast branch & bound algorithm in feature selection," *Proc. SCI 2000 Conf., Orlando, FL*, vol. IIV, pp. 646–651, 2000.

[46] M. Fauvel, C. Dechesne, A. Zullo, and F. Ferraty, "Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2824–2831, 2015.

[47] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 118–139, 2019.

[48] W. Sun, W. Li, J. Li, and Y. M. Lai, "Band selection using sparse nonnegative matrix factorization with the thresholded earth's mover distance for hyperspectral imagery classification," *Earth Sci. Informat.*, vol. 8, no. 4, pp. 907–918, 2015.

[49] S. Li and H. Qi, "Sparse representation based band selection for hyperspectral images," in *Proc. 18th IEEE Int. Conf. Image Processing (ICIP)*, Brussels, Belgium, Sept. 2011.

[50] M. Ahmad, D. Ulhaq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using k-means clustering," *International Journal of Engineering and Technology*, vol. 3, pp. 606–614, 12 2011.

[51] J. Feng, L. Jiao, T. Sun, H. Liu, and X. Zhang, "Multiple kernel learning based on discriminative kernel clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6516–6530, 2016.

[52] S. Schaeffer, "Graph clustering," *Computer Science review*, vol. 1, pp. 27–64, 2007.

[53] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 674–677, 2007.

[54] B.-C. Kuo, H.-H. Ho, C.-H. Li, C.-C. Hung, and J.-S. Taur, "A kernel-based feature selection method for svm with rbf kernel for hyperspectral image classification," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, pp. 317–326, 01 2014.

[55] Y. Jiang and C. Li, "Mrmr-based feature selection for classification of cotton foreign matter using hyperspectral imaging," *Computers and Electronics in Agriculture*, vol. 119, pp. 191–200, 11 2015.

[56] U. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec. 2007. [Online]. Available: http://dx.doi.org/10.1007/s11222-007-9033-z

[57] D. Marutho, S. Hendra Handaka, E. Wijaya, and Muljono, "The determination of cluster number at k-mean using elbow method and purity evaluation on headline news," in *2018 International Seminar on Application for Technology of Information and Communication*, 2018, pp. 533–538.

[58] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257

[59] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00293

[60] A. Li, P. B. Dammert, G. Smith, and J. Askne, "Fuzzy c-means clustering algorithm for classification of sea ice and land cover from SAR images," in *Image Processing, Signal Processing, and Synthetic Aperture Radar for Remote Sensing*, J. Desachy and S. Tajbakhsh, Eds., vol. 3217, International Society for Optics and Photonics.   SPIE, 1997, pp. 86 – 97. [Online]. Available: https://doi.org/10.1117/12.295637

[61] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, 1973.

[62] U. Kandaswamy, D. A. Adjeroh, and M. C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2075–2083, 2005.

[63] Albregtsen Fritz, "Statistical Texture Measures Computed from Gray Level Coocurrence," *Boundary 2*, vol. 3, no. 1, p. 45, 1974. [Online]. Available: http://www.jstor.org/stable/302397?origin=crossref{%}0Ahttps://www.jstor.org/stable/302397?origin=crossref

[64] E. Khachatrian and N. Sandalyuk, "On the exploitation of multimodal remote sensing data combination for mesoscale/submesoscale eddy detection in the marginal ice zone," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[65] S. Beucher, "The Watershed Transformation Applied to Image Segmentation," *Proceedings of the 10th Pfefferkorn Conference on Signal and Image Processing in Microscopy and Microanalysis*, pp. 299–314, 1992. [Online]. Available: http://folk.uib.no/eha070/mat262/papers/Beucher.pdf

[66] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004. [Online]. Available: https://doi.org/10.1023/B:VISI.0000022288.19776.77

[67] P. Neubert and P. Protzel, "Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms," *Proceedings - International Conference on Pattern Recognition*, pp. 996–1001, 2014.

[68] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing and Applications*, vol. 24, no. 1, pp. 175–186, 2014.

[69] K. Pearson, "Notes on the history of correlation," *Biometrika*, vol. 13, no. 1, pp. 25–45, 1920. [Online]. Available: http://www.jstor.org/stable/2331722

[70] E. Khachatrian, S. Chlaily, T. Eltoft, P. Gamba, and A. Marinoni, "Unsupervised band selection for hyperspectral datasets by double graph laplacian diagonalization," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 4007–4010.

[71] A. Braun, U. Weidner, and S. Hinz, "Support vector machines for vegetation classification - a revision Über den einsatz von support-vektor-maschinen (svm) für vegetationsklassifizierung," *Photogrammetrie - Fernerkundung - Geoinformation*, vol. 2010, pp. 273–281, 08 2010.

[72] D. Rudrapal and M. Subhedar, "Land cover classification using support vector machine," *International Journal of Engineering Research and*, vol. V4, 09 2015.

[73] N. Y. Zakhvatkina, V. Y. Alexandrov, O. M. Johannessen, S. Sandven, and I. Y. Frolov, "Classification of sea ice types in envisat synthetic aperture radar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2587–2600, 2013.

[74] S. R. Gunn, "Support vector machines for classification and regression," 1998.

[75] A. Liaw and M. Wiener, "Classification and regression by randomforest," *Forest*, vol. 23, 11 2001.

[76] P. Bharatkar and R. Patel, "Approach to Accuracy Assessment tor RS Image Classification Techniques," *International Journal of Scientific & Engineering Research*, vol. 4, no. 12, pp. 79–86, 2013. [Online]. Available: https://www.ijser.org/researchpaper/Approach-to-Accuracy-Assessment-tor-RS-Image-Classification-Techniques.pdf

[77] J. B. Campbell, *Introduction to remote sensing / James B. Campbell*, 5th ed. Guilford Press New York, 2011.

[78] C. Elachi and J. van Zyl, *Introduction to the Physics and Techniques of Remote Sensing*, ser. Wiley Series in Remote Sensing and Image Processing. Wiley, 2006. [Online]. Available: https://books.google.no/books?id=I-Wws1EMdpwC

[79] X. N. Liu, "Review of myopia management: Could violet light be the cure for myopia?" https://reviewofmm.com/could-violet-light-be-the-cure-for-myopia/, 2020.

[80] J. A. Richards, "Remote sensing with imaging radar," 2009.

[81] IEEE, "Ieee standard letter designations for radar-frequency bands," *IEEE Std 521-2019 (Revision of IEEE Std 521-2002)*, pp. 1–15, 2020.

[82] W. Dierking, "Sea ice monitoring by synthetic aperture radar," *Oceanography*, vol. 26, pp. 100–111, 06 2013.

[83] J. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, ser. Optical Science and Engineering. CRC Press, 2017. [Online]. Available: https://books.google.rs/books?id=1nAvp2HW_gwC

[84] F. D. Carsey, "Microwave Remote Sensing of Sea Ice," *Washington DC American Geophysical Union Geophysical Monograph Series*, vol. 68, Jan. 1992.

[85] C. Gentemann, F. Wentz, M. Brewer, K. Hilburn, and D. Smith, "Passive microwave remote sensing of the ocean: An overview," *Oceanography from Space: Revisited*, 02 2010.

[86] J. C. Comiso, D. J. Cavalieri, C. L. Parkinson, and P. Gloersen, "Passive microwave algorithms for sea ice concentration: A comparison of two techniques," *Remote Sensing of Environment*, vol. 60, no. 3, pp. 357–384, 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425796002209

[87] J. Heinrichs, D. Cavalieri, and T. Markus, "Assessment of the amsr-e sea ice-concentration product at the ice edge using radarsat-1 and modis imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3070–3080, 2006.

[88] N. Zakhvatkina, V. Smirnov, and I. Bychkova, "Satellite sar data-based sea ice classification: An overview," *Geosciences*, vol. 9, no. 4, 2019. [Online]. Available: https://www.mdpi.com/2076-3263/9/4/152

[89] G. Spreen, L. Kaleschke, and G. Heygster, "Sea ice remote sensing using amsr-e 89-ghz channels," *Journal of Geophysical Research: Oceans*, vol. 113, no. C2, 2008. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JC003384

[90] J. C. Landy, J. Bouffard, C. Wilson, S. Rynders, Y. Aksenov, and M. Tsamados, "Improved arctic sea ice freeboard retrieval from satellite altimetry using optimized sea surface decorrelation scales," *Journal of Geophysical Research: Oceans*, vol. 126, no. 12, p. e2021JC017466, 2021, e2021JC017466 2021JC017466. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JC017466

[91] M. Davidson, M. Chini, W. Dierking, S. Djavidnia, J. Haarpaintner, G. Hajduch, G. Laurin, M. Lavalle, C. López-Martínez, T. Nagler, N. Pierdicca, and B. Su, "Copernicus l-band sar mission requirements document," 10 2019.