

TALLINN UNIVERSITY OF TECHNOLOGY

School of Science

Department of Marine Systems

**ESTIMATION OF SEA ICE EXTENT FROM
SENTINEL-1 RADAR DATA IN ESTONIAN
COASTAL SEA**

Master's thesis

Kaari Laanemäe

Supervisor: Rivo Uiboupin, PhD

Tallinn

2017

Declaration

Hereby I declare that this master's thesis, my original investigation and achievement, submitted for the master's degree at Tallinn University of Technology has not been submitted for any academic degree. All content and ideas drawn directly or indirectly from external sources are indicated as such.

Kaari Laanemäe

(Signature and date)

Supervisor: *Rivo Uiboupin*

Work meets the requirements for master's thesis.

(Signature and date)

Chairperson of the defence panel:

Accepted for defence

.....

(Name, signature, date)

TALLINNA TEHNIKAÜLIKOOL

Loodusteaduskond

Meresüsteemide instituut

**JÄÄ ULATUSE MÄÄRAMINE SENTINEL-1
RADAR ANDMETEST EESTI RANNIKUMERES**

Magistritöö

Kaari Laanemäe

Juhendaja: Rivo Uiboupin, PhD

Tallinn

2017

TABLE OF CONTENTS

Abstract	6
Resümee	7
1 Introduction	8
2 Objectives.....	10
3 Literature review	11
3.1 Ice conditions in Estonian coastal sea.....	11
3.2 Radar scattering from sea ice.....	12
3.3 Machine learning algorithms	14
4 Data	16
4.1 Sentinel-1 SAR	16
4.2 Moderate Resolution Imaging Spectroradiometer data	17
4.3 Finnish Meteorological Institute Sea Ice Forecast.....	17
4.4 <i>In situ</i> observations	17
5 Method	18
5.1 Operational processing chain.....	18
5.1.1 Pre-processing	19
5.1.2 Segmentation	20
5.1.3 Segment properties	21
5.1.4 Segment classification	24
5.1.5 Projection.....	25

5.2	Comparison of properties.....	25
5.3	Ice map accuracy and comparison	25
6	Results.....	27
6.1	Segmentation accuracy	27
6.2	Properties selection	29
6.3	Class separation: ice, open ice, water	34
6.4	Machine learning algorithm comparison	36
7	Discussion	38
7.1	Accuracy of ice classification and false classification causes	38
7.1.1	Classification algorithm performance comparison.....	38
7.1.2	Errors caused by source data	41
7.1.3	Effect of meteorological conditions	42
7.2	Possible improvements	43
8	Conclusion	45
	Acknowledgments	47
	List of references	48
	Appendices	52
	Appendix 1. SAR data.....	52
	Appendix 2. <i>In situ</i> observations	54
	Appendix 3. Classifier algorithm accuracy	56
	Kokkuvõte	57

ABSTRACT

Current study is focused on Synthetic Aperture Radar (SAR) based sea ice and water classification using different machine learning methods for coastal area in Western Estonia. *K*-means based superpixel algorithm is chosen to separate SAR data into homogenous segments representing only ice or water. Statistical parameters from SAR backscattering field, Grey Level Covariance Matrix parameters (texture) and parameters describing segment geometry, are calculated for each segment and properties which best describe ice or water class are selected. Classification based on AdaBoost, Random Forest and Support Vector Machine algorithms is performed.

Accuracy of the classification algorithms is assessed. In cold conditions when water between islands and on the coast is frozen Random Forest provided best accuracy with 93.6% on ice and 82.2% on water. AdaBoost performed correctly on 90.7% ice segments and 80.1% water segments, Support Vector Machine accuracy on ice was 80.2% and on ice 70.6%. Comparison between Finnish Meteorological Institute SAR based ice charts and Random Forest based classification gives greater similarities in cold conditions (average difference 54%) and increases in warmer period when coastal area is clear of ice (average difference 214%). The algorithm performs poorly on warm conditions with little ice, as the algorithm is prone to overclassification of ice on the coastal area.

Key words: sea ice, synthetic aperture radar, backscattering, machine learning

RESÜMEE

Antud töö keskendub tehisavaradari (SAR) põhisele merejää ja vee klassifitseerimisele kasutades erinevaid masinõppe meetodeid Lääne-Eesti rannikumeres. SAR andmed jaotatakse homogeenseteks segmentideks, mis sisaldavad ainult vett või jääd, kasutades *k-means* põhist superpikslite algoritmi. Igale segmendile arvutatakse statistilised parameetrid, mis kirjeldavad SAR tagasipeegelduvust, tekstuuri (*Grey Level Covariance Matrix*) ning segmendi geomeetriat. Nende seast valitakse välja parameetrid, mis kirjeldavad ning eristavad vee ja jää klasse kõige paremini. Klassifikatsioon teostatakse kasutades *AdaBoost*, *Random Forest* ja *Support Vector Machine* algoritme.

Hinnatakse klassifikatsiooni algoritmide täpsust. Külmaes ilmingimustes, kui vesi on saarte vahel ja rannikul jäätunud, annab *Random Forest* parima tulemuse, 93.6% jää ja 82.2% veel. *AdaBoost*'il põhineva algoritmi täpsus on 90.7% jää ja 80.1% veel. *Support Vector Machine* andis täpsuseks jää 80.2% ja veel 70.6%. Võrdluses Soome meteoroloogia instituudi jääkaartidega annab *Random Forest*'il põhinev algoritm sarnasemaid tulemusi külmaes tingimustes (keskmine erinevus 54%), kuid erinevus suureneb soojades tingimustes, mil rannik on jääst vaba (keskmine erinevus 214%). Algoritm annab soojades tingimustes kehvi tulemusi, kuna algoritm üleklassifitseerib jää esinemist rannikul.

Märksõnad: merejää, tehisavaradar, tagasipeegeldus, masinõpe

INTRODUCTION

Sea ice extent is essential information for ice map formation. In current work sea ice extent is defined as area covered by ice or drifting ice. Accurate ice maps in turn offer valuable information for winter marine navigation by enabling to efficiently plan ice breakers work, shipping routes, thereby saving fuel. In addition, accurate knowledge of ice conditions increases marine safety by allowing to evaluate what type of vessels could safely navigate in given conditions and if ice breaking is needed.

Extensive *in situ* ice measurements provide most accurate information about ice conditions but are expensive to conduct and need considerable man power. Out of remote sensing methods optical high resolution satellite imagery supply one of most accurate and easy to interpret information, but only during daytime and in cloud free situations. Unfortunately during winter time in Estonia, there are few cloud free days. According to “Meteorological yearbook of Estonia 2016” (Loodla *et al.*, 2017) only 16% of daytime was sunny in January 2016, 15% in February 2016 and 41% in March 2016, measured in Tallinn-Harku meteorological station. Synthetic Aperture Radar (SAR) is an active remote sensing system that sends microwave to the surface of the Earth and reads the backscatter from the surface. This means it is independent from daylight and cloud cover.

However, SAR data is difficult to interpret and different ice conditions may produce extremely similar backscattering. As SAR data understanding requires detection of small changes in backscattering patterns, growing number of sea ice applications use machine learning methods in order to separate ice and water.

Most of ice classification algorithms are developed for Arctic Ocean with the aim of distinguishing between first-year ice, multi-year ice and water. Leigh and Wang (2014) deploy both unsupervised (iterative region growing using semantics algorithm)

and supervised (Support Vector Machine) algorithms in order to separate ice and water in Canadian Arctic region with 96.42% of overall accuracy. Zakhvatkina *et al.* (2012) classify Arctic ice using neural networks and Bayesian algorithm with accuracies for different ice types ranging 83% to 85% for neural networks and 68% to 96% for Bayesian algorithm.

There is lesser amount of research done about using machine learning algorithms to classify ice in small regional areas with ice appearing only seasonally. The most extensive research about first-year ice classification is done by Finnish Meteorological Institute. Karvonen (2004) uses hypothesis that SAR data distribution within segment is very close to Gaussian distribution. Estimating overall data distribution, six different distributions can be detected which respond to three different ice types. Karvonen (2012a) presents automatic ice concentration algorithm based on autocorrelation distribution which can be used as statistical texture measure. Image is segmented using simple *k*-means clustering and autocorrelation is calculated within one segment.

As ice cover area in Estonian coastal area is very small and topographical characteristics have to be taken in account, new algorithm is proposed and tested. Estonia also lacks extensive ice survey service which means ice extent algorithm based solely on SAR would provide cost effective solution for collecting information about ice conditions.

2 OBJECTIVES

The main objective of this study is to develop operational ice extent and type algorithm for the Baltic Sea and Estonian coastline with only SAR HH and HV polarization data as an input.

Specific tasks include (1) finding suitable segmentation algorithm and ensuring that the segments follow ice-water border; (2) collecting training data and selecting the best features that enable to separate different ice and water classes; (3) to train ice/water classification algorithm using three different machine learning methods (AdaBoost, Support Vector Machine and Random Forest); (4) and comparing the results of the ice/water classification algorithms and assessing their accuracy. In addition, suitability of simple machine learning classification based only on SAR data is investigated.

3 LITERATURE REVIEW

3.1 Ice conditions in Estonian coastal sea

According to report on Estonian ice conditions (Uiboupin and Pärn, 2016) ice appears every year at least in Väinameri and in Gulf of Pärnu (Figure 1). In cold winters all of Estonian marine waters is covered with ice.

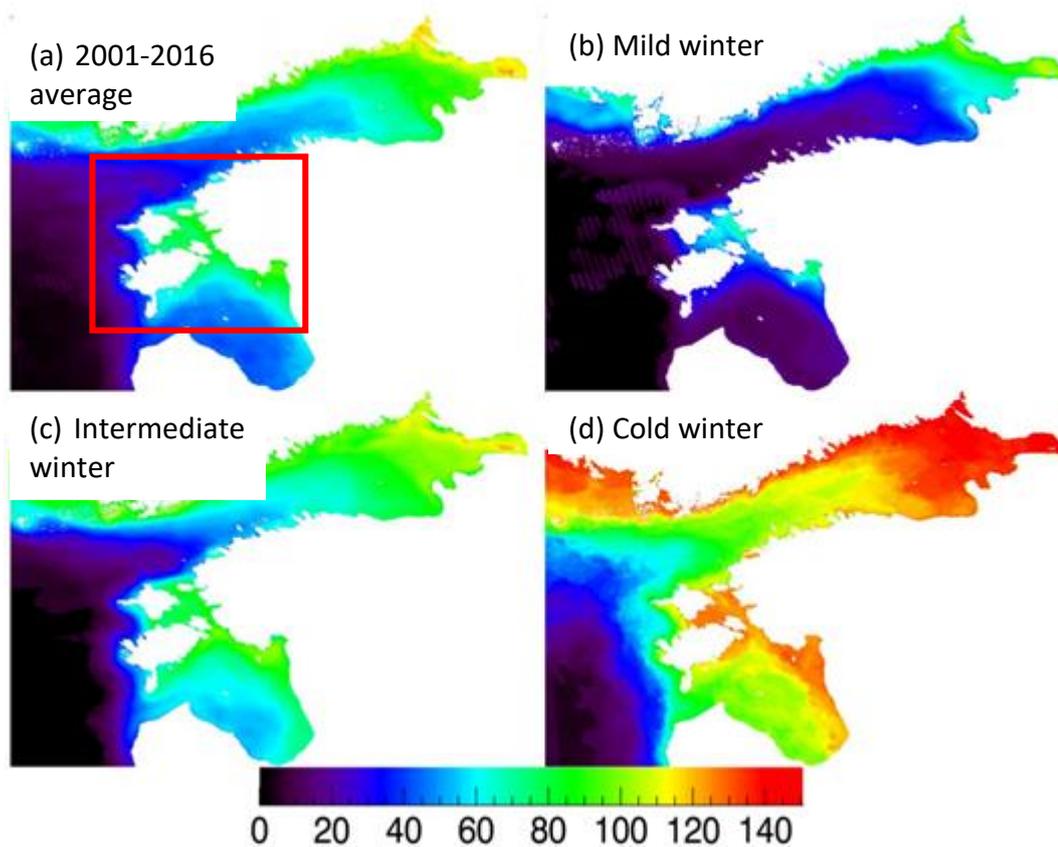


Figure 1. Length of ice period in days (a) during period 2001-2016; (b) during warm winter; (c) during intermediate winter; (d) during cold winter. (Uiboupin and Pärn, 2016). Red square indicates area of interest.

On average winter the ice cover appears for 70 to 110 days in Väinameri and Gulf of Pärnu. The probability of ice occurrence in period of 15th December to 1th of May is 40% - 70%. On cold winters ice appears for more than 110 days, with probability of ice occurrence over 70% during ice season.

3.2 Radar scattering from sea ice

Overview article “Sea Ice Monitoring by Synthetic Aperture Radar” by Dierking (2013) states that what is visible from SAR images is determined by various radar properties. The foremost one being radar frequency, most common frequencies are L-band (15-30 cm), C-band (3.8-7.5 cm) and X-band (2.4-3.8 cm), latter two allow to see objects of interest and surface roughness in fine detail. In addition, brightness of sea ice in radar images is influenced by incidence angle and polarization. Overview of SAR geometry is provided in Figure 2.

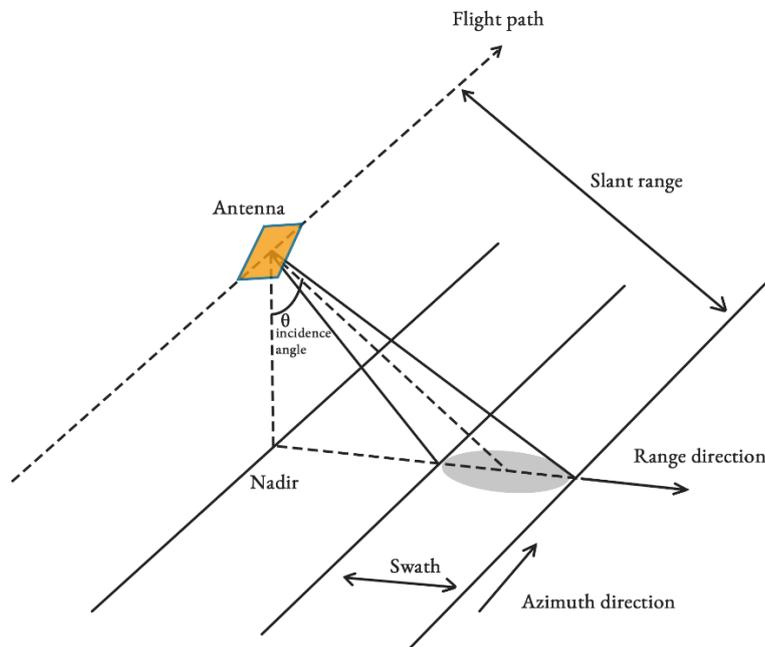


Figure 2. Schematic overview of SAR geometry.

Incidence angle is the angle between the radar system’s line of sight and a local vertical vector with respect to the geoid. Polarization is marked as a combination of transmitted and received signals where “H” means “horizontal” and “V” means “vertical”. If transmitted and received signals are in same polarization, it is co-polarized method, if transmitted and received signals are in different polarization, it is cross-polarized method. HH-polarization is preferred for operational sea ice mapping as noise caused by water in radar signal is more suppressed at HH polarization than at VV. Level and deformed ice can be better identified from cross-polarization (HV or VH) as deformation zones have strong depolarization effects (Dierking, 2013).

The main scattering from sea ice is by surface and volume scattering (*Ibid.*). Radar waves are reflected from smooth surface relative radar wavelength as from a mirror, therefore appear very dark in SAR images (Figure 3a). Scattering from rougher surfaces is dependent on size of disruption compared to radar wavelength. The probability of rubble and ice ridges reflecting toward the radar is much higher than for level ice. Thus, deformations appear as areas of higher backscatter. Direction of scattering from surface with different roughness are depicted on Figure 3.

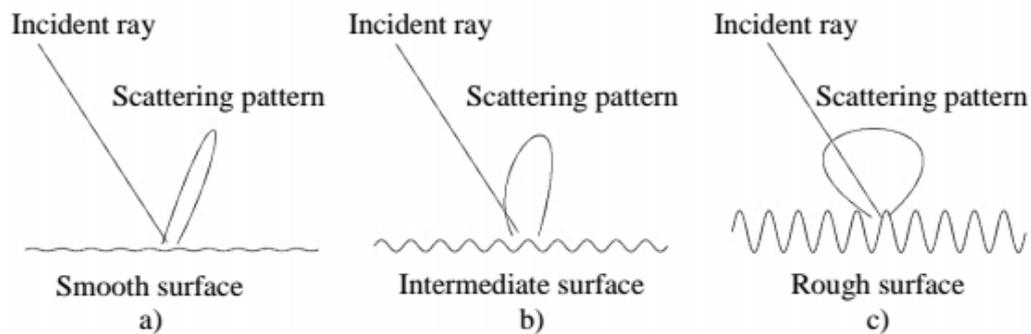


Figure 3. Direction of microwave scattering from a) a smooth surface, b) a roughened surface, c) a rough surface. (Robinson, 2004)

Volume scattering occurs when radar wave penetrates into the ice or snow and then is redirected to the surface (Dierking, 2013). Amount of volume scattering occurring is determined partly by dielectric constant and penetration depth of radar wave, which decreases with increasing radar frequency, ice salinity and temperature. In first year

ice in cold conditions X-band may have penetration depth from 3 to 15 cm, whereas in L-band it may range from 0.15 to 1 m. Ice characteristics, which influence volume scattering are volume fraction, size and shape of inclusion particles, such as air bubbles and brine. Brine concentration in sea ice is large during formation, but decreases with age and thickness growth. Volume scattering is reduced or non-existent when ice cover is wet due to melting.

Radar scattering from water is overviewed by Nekrasov (2014) according to whom scattering from water surface is dependent on incidence angle, which is highest near nadir and decreases with increasing incidence angle. Radar backscattering is predominantly due to presence of capillary-gravity wavelets, which are superimposed on large gravity waves on the sea surface. Wind speed and direction also affect backscattering as strong wind produces larger backscattering values at medium incidence angle and smaller backscattering values at small incidence angle compared to mild wind.

3.3 Machine learning algorithms

According to Michie *et al.* (1994) “Machine Learning is generally taken to encompass automatic computing procedures based on logical or binary operations, that learn a task from a series of examples”.

Machine learning can be divided into three groups, (1) supervised learning, (2) reinforcement learning and (3) unsupervised learning, based on how much information is given to the classifier about correct outcome (Russel and Norvig, 1995). If both inputs and outputs are defined by the user, the process is called supervised learning, these include popular classifying algorithms like decision trees and support vector machines. Reinforced learning occurs if the classifier receives some evaluation of its action but is not told the correct action. If there is no hint at all about the correct output, it is called unsupervised learning, such as segmentation algorithms, which group similar data points to clusters but do not tell the user what they are. In current work, both supervised and unsupervised machine learning methods are used.

A common unsupervised method is clustering, which discovers the compositions and structure of a given dataset (Castrounis, 2016). Clustering is a process of bundling data into clusters to see if or what groups emerge. One cluster can be characterised by set of features. Another important method is supervised classification (*Ibid.*) which assign a class to observation or estimating the probabilities that an observation belongs to each classes.

Machine learning algorithm implementation has been made accessible for non-computer sciences experts through Python data analysis modules as *Scikit-learn* (Pedregosa *et al.*, 2011).

4 DATA

4.1 Sentinel-1 SAR

C-band Sentinel-1 SAR data has been used in this study. Sentinel-1 satellite series is part of Europe's Copernicus program and combines two identical satellites orbiting as a constellation 180° apart (Torres *et al.*, 2012). One satellite images the Earth in 12 days, thus the two-satellite constellation repeats its orbit in 6 days, providing identical acquisition in this period. Combined with Estonia's high latitude, Sentinel-1 satellite series provides new images of Estonia at least twice as often.

Sentinel-1 data is distributed in Standard Archive Format for Europe (SAFE) and Level-1 Ground Range Detected (GRD) products are used in this study, which are in Extra Wide Swath (EW) SAR acquisition mode. According to Sentinel-1 Product Specification (Bourbigot *et al.*, 2016) the EW mode provides very large swath coverage with medium spatial resolution. The pixel resolution for EW mode GRD product in medium resolution (GRDM) is 40x40 m (SUHET, 2013, 65-69).

Acquisitions over Estonia are done in dual polarizations, providing one image with co-polarization (HH or VV) and one image with cross-polarization (respectively HV or VH). According to "Sentinel High Level Operations Plan (HLOP)" (CSC Mission Management Team, 2017) HH+HV polarizations are most suitable for sea-ice observations, thus GRDM products are acquired in HH+HV polarisations in the Baltic Sea during ice season.

Training data used for training supervised machine learning classifier is from 11 January 2017 to 6 February 2017 (list of files is located in Appendix 1). During this

period, 35 Sentinel-1 images over Estonian coastal region are acquired. Validation data was acquired between 13 January 2016 and 25 February 2016, in this period 17 images are randomly selected for validation.

4.2 Moderate Resolution Imaging Spectroradiometer data

Data from Moderate Resolution Imaging Spectroradiometer (MODIS) is used for visual estimation of sea ice extent. MODIS images have resolution of 250 m and are available as false colour image (Red: 859 nm; Green: 645 nm; Blue: 645 nm) on Department of Marine Systems website. In Training period of 11 January to 6 February 2016 there are nine relatively cloud free images available.

4.3 Finnish Meteorological Institute Sea Ice Forecast

Finnish Meteorological Institute (FMI) provides SAR based sea ice thickness charts using Sentinel-1 or Radarsat-1 data (Karvonen, 2012b) and combining this information with ice thickness history and thermodynamic ice model (Karvonen *et al.*, 2008). The maps have resolution of 500 m and are publically available as part of project Copernicus Marine Environmental monitoring Service (CMEMS). As the product is Sentinel-1 based, it has same temporal coverage as data in paragraph 4.1.

4.4 *In situ* observations

In situ observations were conducted in 2015 during three days and in 2016 during five days. In 2015 the observations were done on foot and in 2016 using TTÜ Department of Marine Systems drone. *In situ* observations enabled to estimate the texture of ice and how different ice types looked like from SAR images. Summary of *in situ* observations is presented in Appendix 2.

5 METHOD

5.1 Operational processing chain

In order to automatically classify SAR images, various machine learning techniques are used. Figure 4 depicts flowchart of processes to create ice maps. The most important steps are segmentation and classification, which are investigated thoroughly in current study.

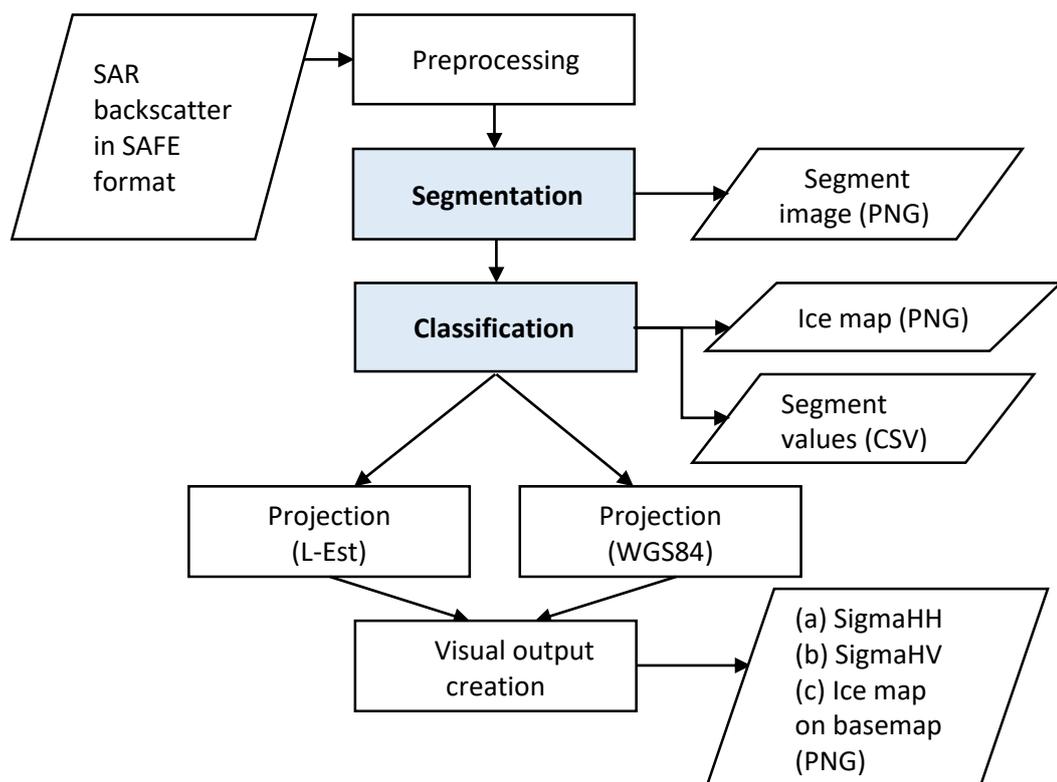


Figure 4. Flowchart of sea ice extent classification algorithm. Segmentation involves separating ice and water into homogenous divisions. Classification involves separating and defining segments as ice or water.

5.1.1 Pre-processing

Pre-processing is done using program ESA Sentinel Application Platform (SNAP). Sentinel-1 GRD products are calibrated and Land mask is applied using prewritten algorithms. Flowchart of pre-processing is visible in Figure 5.

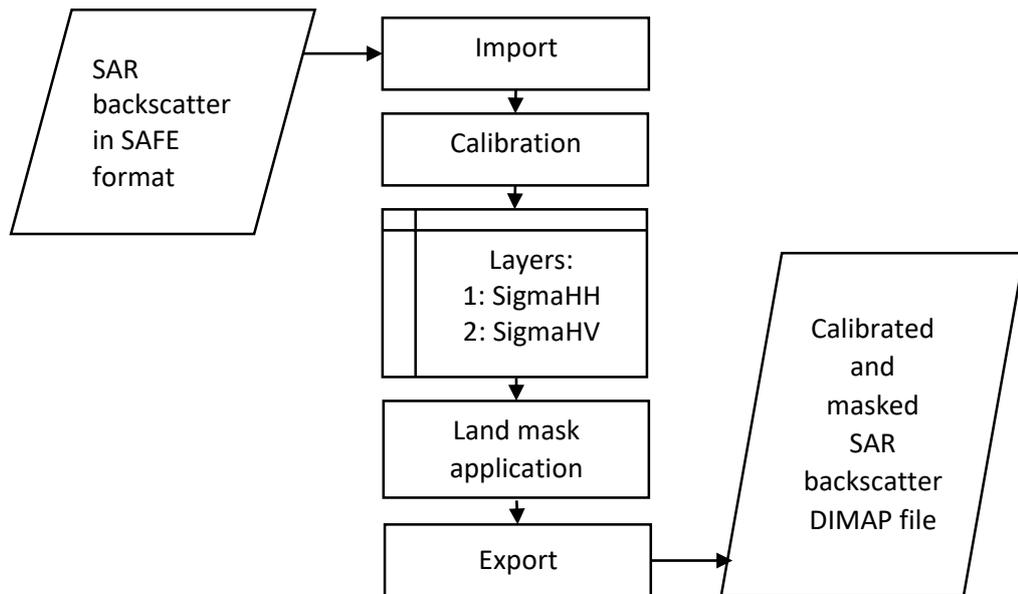


Figure 5. Flowchart of pre-processing in SNAP.

SAFE file distributed by ESA is imported to ESA SNAP. The first processing step is calibration which relates pixel values to the radar backscatter of the scene, thus converting digital pixel values to radiometrically calibrated backscatter (SNAP Sentinel-1 Toolbox Help). Although uncalibrated images can provide information for visual evaluation, in order to statistically analyse the data, calibration must be performed. Information needed for calibration is provided along with the SAR data. Calibrated backscatter values are defined as sigma nought (σ^0), also referred as variable *SigmaHH* and *SigmaHV* depending on the polarisation (HH or HV).

Creating Land Mask turns any pixel on land into no data value. ESA SNAP will automatically download Shuttle Radar Topographic Mission (SRTM) 5 minutes digital elevation model (DEM) which allows to quickly determine if a pixel is on land or in water. Applying Land Mask allows to reduce file size by removing land pixel values

and segregate land pixels in segmentation process and causing the segments to follow the coastline.

The images produced in ESA SNAP are exported in BEAM-DIMAP Data format which can be read into other applications. In addition, part of pre-processing is calculating longitude and latitude fields.

5.1.2 Segmentation

Segmentation in current work is done by using superpixels method which is developed for image classification and figure detection by Ren and Malik (2003). The aim of superpixels is to produce larger areas of pixels that contain pixels of similar values based on texture, brightness, sudden boundaries in the image and proximity of pixels.

Superpixel method has many benefits compared to traditional sliding window approach for classification. Sliding windows robustly take the mean values of side-by-side pixels and classifies the pixels together without taking into consideration whether the sliding pixels are similar or not. This creates inaccurate results in boundary areas. In order to produce satisfactory results, the sliding window has to contain a small number of pixels, which increases computing time. In comparison, segments contain significantly larger number of pixels that can be classified together due to their homogeneous properties.

In this work, simple linear iterative clustering (SLIC) superpixel algorithm is used (Achanta *et al.*, 2012). SLIC algorithm uses adaptation of k -means, but instead of searching the entire image for similar values, only region proportional to the segment size is searched. The number of segments desired on the image is controlled by the size of each pixel. The first step is to place cluster centres on regular grid spaced S pixels apart. Each pixel is associated to the nearest cluster centre. Then for each pixel new cluster centre is assigned based on brightness and distance. For large segments, spatial distances outweigh similarity in brightness, for small segments, the converse is true.

For this work, the optimal segment size is chosen to be 40 000 pixels. For regular grid, this translates into a square size of 200 by 200 pixels. In the real world, using resolution of Sentinel-1 GRDM product, 8 km by 8 km is covered with one regular segment. The actual size of segment can vary from 2000 to 110 000 pixels depending on the boundaries and different ice types in the image being processed. The minimal width of any segment is eight pixels.

5.1.3 Segment properties

For each segment 30 properties are calculated from sigma naught field of HH and HV polarization and the segment field itself. From HH and HV polarizations HH and HV field value parameters and texture parameters based on Grey-Level Co-occurrence Matrix (GLCM) are calculated. Geometrical parameters are calculated from segment field.

Value parameters are following:

- *Mean_HH* – Mean value of HH polarization sigma naught
- *Stddev_HH* – Standard deviation value of HH polarization sigma naught
- *Variance_HH* – Variance value of HH polarization sigma naught
- *Mean_HV* – Mean value of HV polarization sigma naught
- *Stddev_HV* – Standard deviation value of HV polarization sigma naught
- *Variance_HV* – Variance value of HV polarization sigma naught
- *HH-HV_mean* – HH and HV polarization sigma naught difference
- *HH_div_HV_mean* – HH divided by HV polarization sigma naught values

GLCM is a simple texture analysis devised by Haralick *et al.* (1973) which is based on assumption that texture information in an image is contained in the overall spatial relationship which grey levels of neighbouring pixels have to one another. 22 different texture features can be derived from GLCM, but usually only Contrast, Homogeneity, Dissimilarity, Energy and Entropy are considered as parameters of importance. GLCM parameters are calculated separately for both co-polarized and cross-polarized images,

with 50 levels (range between 0.01 and 0.51 with step 0.01) for co-polarized data and 18 levels (range between 0.001 and 0.01 with step 0.0005) for cross-polarized data.

GLCM value (texture) parameters are following:

- *GLCM_mean* – the mean of rows and columns mean (Soh and Tsatsoulis, 1999)

$$\mu = \frac{\mu_x + \mu_y}{2} = \frac{\sum_{i,j} i \cdot p(i,j) + \sum_{i,j} j \cdot p(i,j)}{2} \quad (1)$$

- *GLCM_variance* – measures how far each GLCM element is from mean (Haralick *et al.*, 1973)

$$Variance = \sum_{i,j} (i - \mu)^2 p(i,j) \quad (2)$$

- *GLCM_correlation* – measures the linear dependency of grey levels on those of neighbouring pixels. (*Ibid.*)

$$Correlation = \frac{\sum_{i,j} (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3)$$

where μ_x and μ_y represent means of rows and columns and σ_x and σ_y represent standard deviations of rows and columns.

- *GLCM_entropy* – represents spatial disorder (Gebejes and Huertas, 2013)

$$Entropy = \sum_{i,j} p(i,j) \log(p(i,j)) \quad (4)$$

- *GLCM_homogeneity* – measures uniformity of the non-zero entries in the GLCM (*Ibid.*).

$$Homogeneity = \sum_{i,j} \frac{1}{1 + (i - j)^2} p(i,j) \quad (5)$$

- *GLCM_energy* – a measure of local homogeneity, which enables to estimate how uniform the texture is (*Ibid.*).

$$Energy = \sum_{i,j} p(i,j)^2 \quad (6)$$

- *GLCM_contrast* – local grey level variation in the GLCM (*Ibid.*).

$$Contrast = \sum_{i,j} |i - j|^2 p(i,j) \quad (7)$$

- *GLCM_dissimilarity* – defines variation of grey level pairs in an image (*Ibid.*).

$$Dissimilarity = \sum_{i,j} |i - j| p(i,j) \quad (8)$$

Geometrical parameters are following:

- *Pixels* – Number of pixels in segment
- *X_length* – Segment length along x axis
- *Y_length* – Segment length along y axis
- *Ratio* – Shorter segment side length (*x_length* or *y_length*) ratio to longer segment side length (correspondingly *y_length* or *x_length*)
- *Difference* – Shorter segment side length and longer segment side length difference
- *Dif_to_pix* – Number of pixel in segment divided by area of square where the segment would fit (*x_length*y_length*)

Image of backscattering at HH polarization band with segment borders and segment identification is saved to PNG file. In addition, values of each segment are saved in csv file along with label whether the segment covered area is water, newly formed or melting open ice or full ice. The labelling is done by studying the SAR image texture and comparing the SAR image with optical satellite image. This information is later used for training the classifier.

Images in training dataset are segmented and parameters from each segment are calculated. Then segments are labelled by human as ice, open ice and water by comparing SAR data with optical MODIS imagery. In most cases of sea ice classification there are no simultaneous ground truth data available, thus visual classification based on SAR data or other remote sensing data is used (Karvonen *et al.*, 2005). Where ice is visible on optical images, the segments are labelled as ice. Where no ice is visible, the segments are labelled as water. If ice is not visible from optical

images, but changes in roughness can be detected from SAR, the segments are labelled as open ice.

5.1.4 Segment classification

Segments are classified using supervised machine learning algorithms AdaBoost, Random Forest and Support vector machine.

AdaBoost (AB) (Freund and Schapire, 1997) is a method that combines series of weak classifiers that are only slightly better than guessing and incorporates single prediction rule. The booster is presented with set of properties associated with certain label. In our case, the properties are described in paragraph 5.1.3 and the label is visually assigned based on SAR image texture or optical satellite data. The booster devises a distribution using naïve Bayes method over the set of examples and produces a weak hypothesis based on each property. The booster starts with unweighted training sample and accuracy of each weak hypothesis is calculated. If a training data point is misclassified, the weight of that training data point is boosted. This procedure is repeated and the final classifier is defined as the linear combination of the classifiers from each stage. In this work AdaBoost-SAMME algorithm that can successfully solve multi-class predictions is used (Zhu *et al.*, 2009).

Random Forest algorithm (RF) (Ho, 1998), (Breiman, 2001) constructs multiple decision trees in randomly selected subspaces with random features. Individual decision is made by using all the decision trees to classify the object and the most common outcome of all the decision trees is selected as the label. Decision trees used in current work are formed using Classification And Regression Tree (CART) algorithm.

Support vector machine (SVM) (Cortes and Vapnik, 1995) maps input vector to a high-dimension feature space where linear decision surface is constructed. Decision surface is calculated by finding the hyperplane that gives the largest minimum distance to the training examples – finding the maximum margin. The trade-off between minimizing the training error and the complexity of the decision function is controlled

by a kernel and kernel parameters (Friedrichs and Igel, 2005). The current work uses Radial Basis Function for kernel, which is controlled mainly by parameters γ and C . γ defines how far the influence of a single training element reaches, with low meaning 'far' and high values 'close'. C parameter adjusts the relationship between misclassification of training samples with simplicity of decision surface, with low values making decision surface smooth.

5.1.5 Projection

When segments have been formed and each pixel has been assigned value in ice map creation, then the image is converted into L-EST97 (EPSG:3301) and WGS84 (EPSG:4326) projection to create visual output.

5.2 Comparison of properties

The amount of information provided by certain property is estimated by performing analysis of variance (ANOVA), also known as F-test. An ANOVA determines if and independent variable, in this case property, had a significant impact on dependent variable, in current work – class (Scott MacKenzie, 2013). Higher F-scores indicate greater separability between classes.

Correlation matrix is constructed using all segment parameters to eliminate parameters which are duplicating each other. From parameter pairs with correlation higher than 0.9 or lower than -0.9, the least representative based on visual inspection is removed.

5.3 Ice map accuracy and comparison

Ice map accuracy with three classification algorithms is measured by comparing ice map results retrieved by implementing the algorithm on independent validation dataset, with MODIS imagery and visual estimation of source SAR image to

determine if each segment is correctly classified. Segments are divided into four groups: correctly classified ice segments, falsely classified ice segments, correctly classified water segments and falsely classified water segments. This process is analogous to training data collection described in 5.1.3 Classification accuracy is calculated for each class and overall accuracy is calculated using all evaluated segments.

In addition, comparison with FMI ice chart is performed. The area covered by ice in square kilometres is measured by counting the number of pixels defined as ice in both FMI ice chart and ice area chart produced in this study.

Ice map accuracy is only measured for Väinamere region, around Hiiumaa and Saaremaa, and for northern part of gulf of Riga where ice appears every year. As winter of 2016 and 2017 were mild, this are also the only regions where ice appeared in those winters. The area of interest is bordered with latitude 57.75 degrees from south, latitude 59.5 degrees from north, longitude 21.0 degrees from west and longitude 25.0 degrees from east.

6 RESULTS

6.1 Segmentation accuracy

Segmentation accuracy is determined by visual inspection comparing segment border with MODIS images and drone observations. Superpixel based segmentation generally offers good separation between ice and water and between different ice types.

On Figure 6 it is visible that superpixel based segmentation separates two different ice classes (Figure 6b and Figure 6c) into different segments (Figure 6a). Visual inspection also shows that water and ice are divided into different segments. At the same time the segments are large enough for reasonable computing time.

Training data was compiled using supporting image (Figure 7) where radar backscattering is overlaid with segment borders and segment identification is written in the centre of the segment, as well as other source data. In total 14 725 segments were separated of which 626 were ice, 298 were open ice and 13 801 were water. In order to prevent the classifier to be biased toward water, as it was overrepresented, 601 segments were randomly selected out of all water segments.

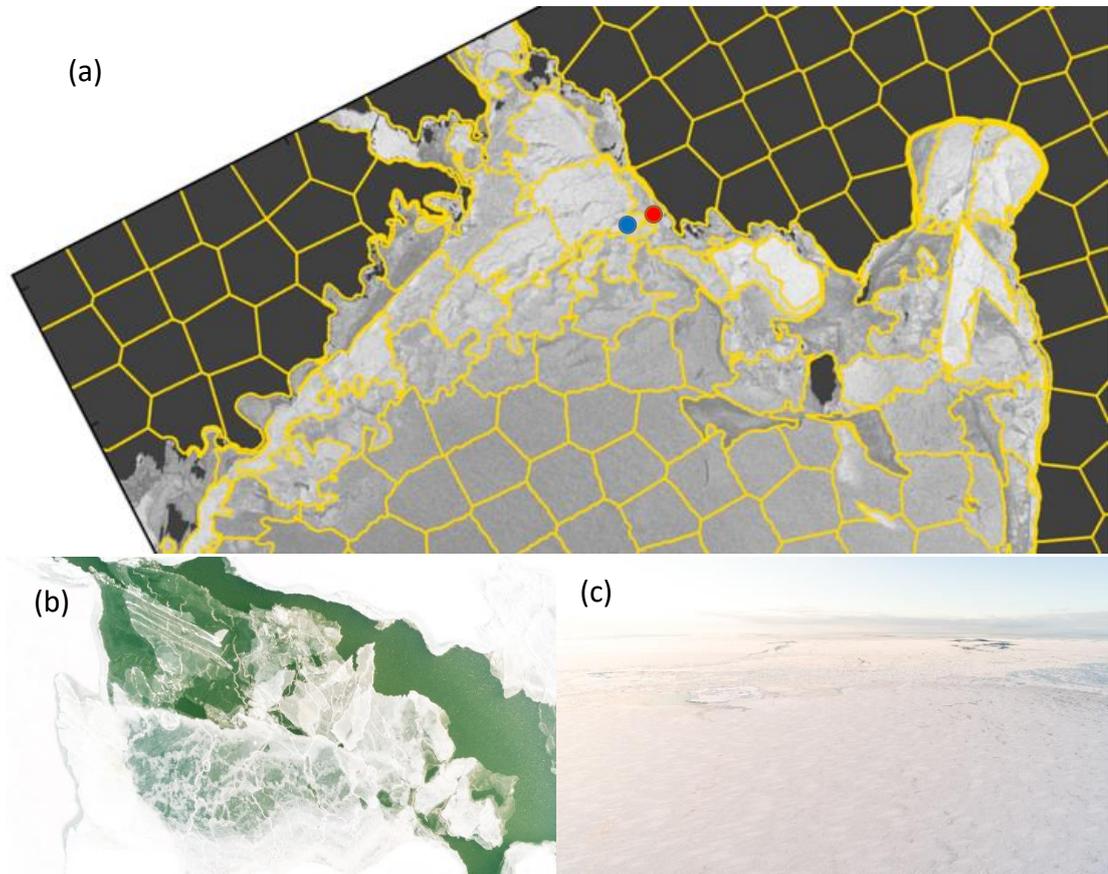


Figure 6. Ice conditions on 15 January 2016 (a) Sentinel-1 SAR backscattering at HH polarization with segment borders (yellow) where blue dot corresponds to image (b) and red dot to image (c); (b) drone photo of closed ice; (c) drone photo of fast ice.

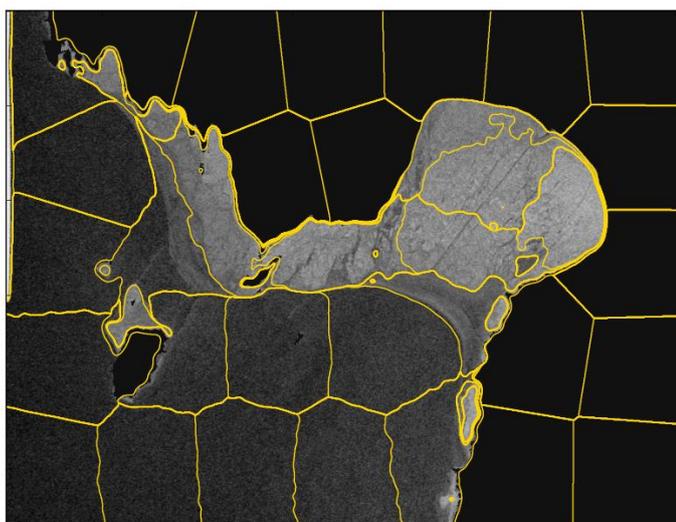


Figure 7. Segment borders (yellow) laid over radar backscattering image at HH polarization.

6.2 Properties selection

Aim of property selection is to reduce the number of properties that describe the different classes/segments. The benefit of this is reduced calculation time and better performing classifier attributes that describe noise are removed. To estimate wheatear a property gives any valuable information for classification – has significantly different values for different classes, an analysis of variance (F-test) is performed. F-values for each property is depicted in Figure 8. *p*-values for all properties were less than 0.05, thus being statistically significant. Properties *stddev_HH*, *variance_HH*, *variance_HV*, *pixels*, *glcm_energy_HH* and *glcm_entropy_HH* have significantly lower F-values than other properties, thus these properties do not describe the difference between different classes. Visual interpretation of the data confirms the hypothesis, as depicted on Figure 9. The features that best describe the difference between the classes are *dif_to_pix*, *glcm_correlation_HV*, *glcm_correlation_HH*, *HH_div_HV_mean* and *glcm_variance_HV*. In Figure 10 are examples of scatterplots where water and ice classes are separable.

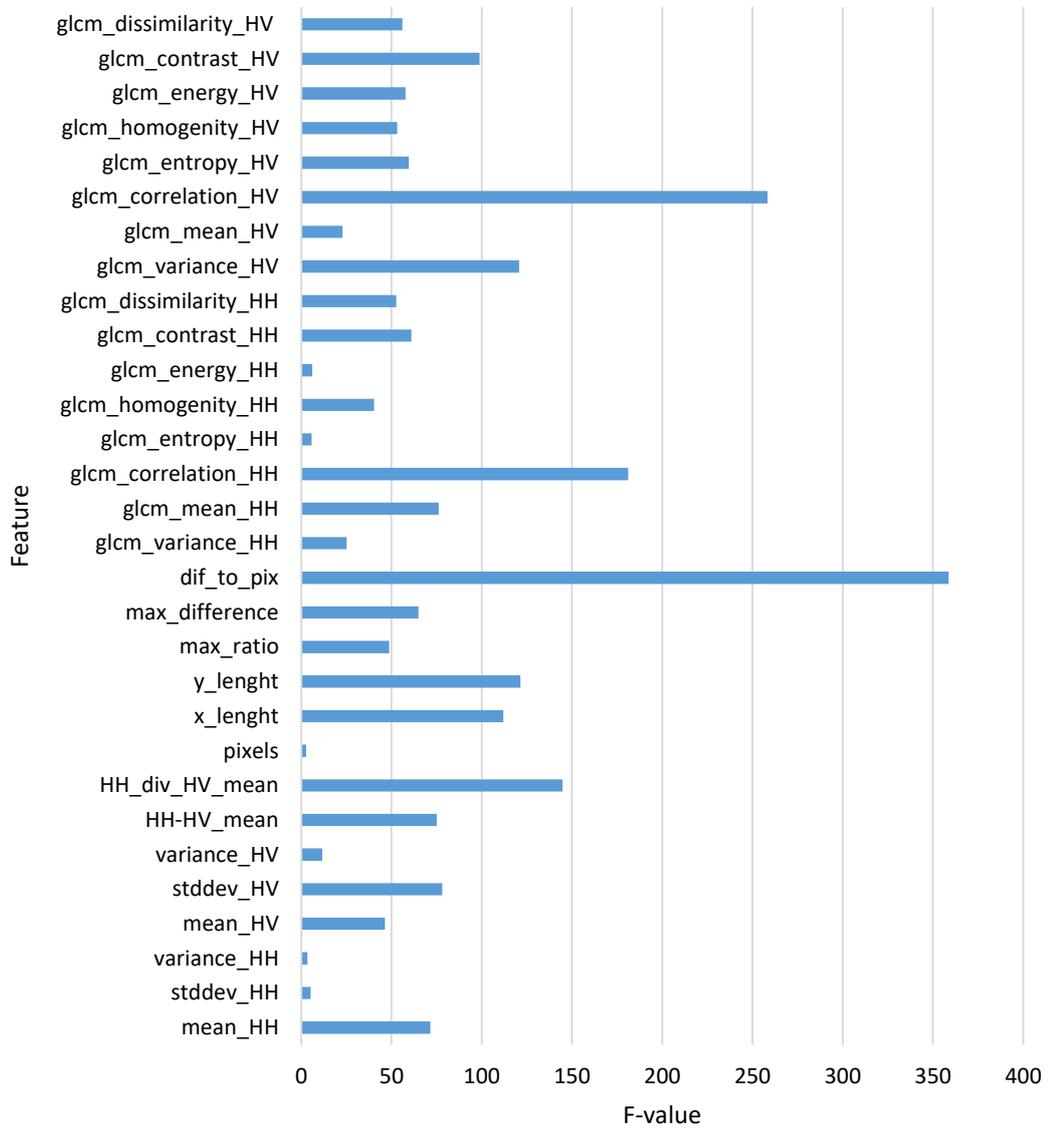


Figure 8. Analysis of variance for each feature.

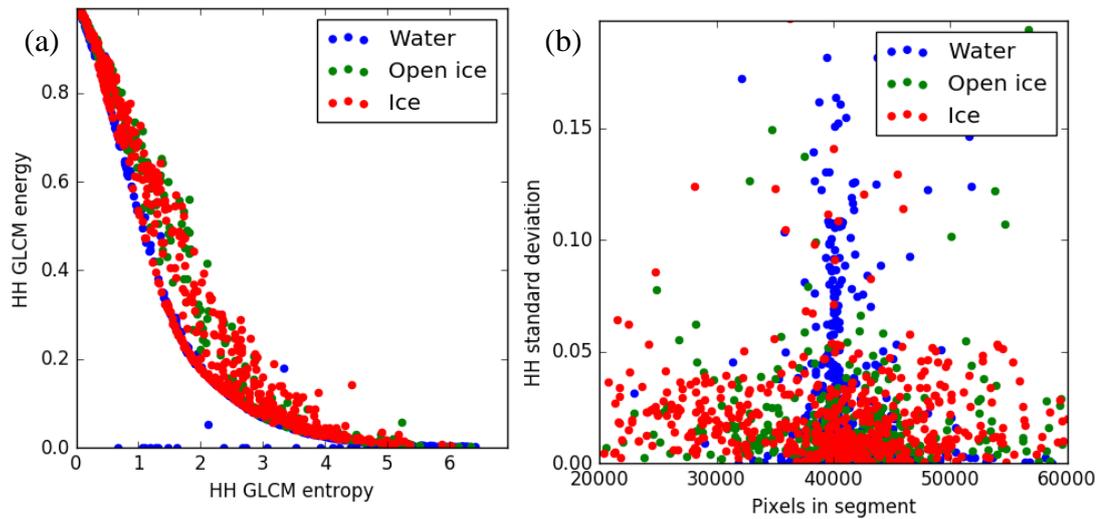


Figure 9. Properties with similar values for different classes (a) HH polarization GLCM energy value of segment and HH polarization GLCM entropy value of segment; (b) HH polarization standard deviation of segment and pixels in segment

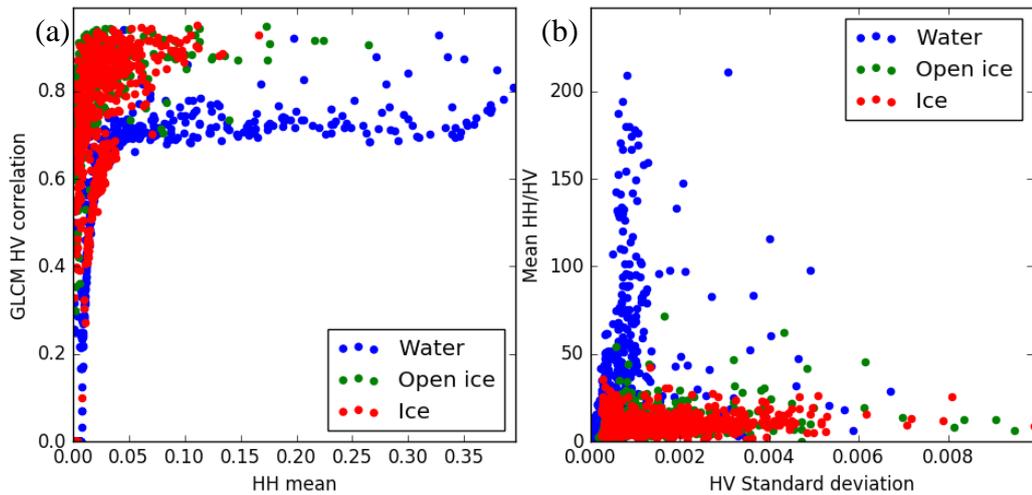


Figure 10. Properties with different values for different classes (a) HH polarization GLCM correlation of segment and HH mean value of segment; (b) Mean value of HH divided by HV and HV standard deviation of segment

In order to further reduce the numbers of properties, the ones that duplicate each other are also excluded. This is done by calculating correlation between each property (Figure 11). *glcm_dissimilarity_HH* has over 0.9 absolute correlation value with

glcm_mean_HH (Figure 12a), *glcm_homogeneity_HH* and *glcm_contrast_HH* while *glcm_dissimilarity_HV* has absolute correlation over 0.9 with *glcm_mean_HV*, *glcm_entropy_HV*, *glcm_energy_HV* and *glcm_contrast_HV*. Therefore, it is concluded that GLCM dissimilarity in both polarizations may be removed. In HV polarization GLCM homogeneity, entropy and energy are all highly correlating with each other as also visible on Figure 12b, thus only one of the parameters is necessary. *glcm_homogeneity_HV* is chosen as a representative texture parameter, as it provided visually the best data separation.

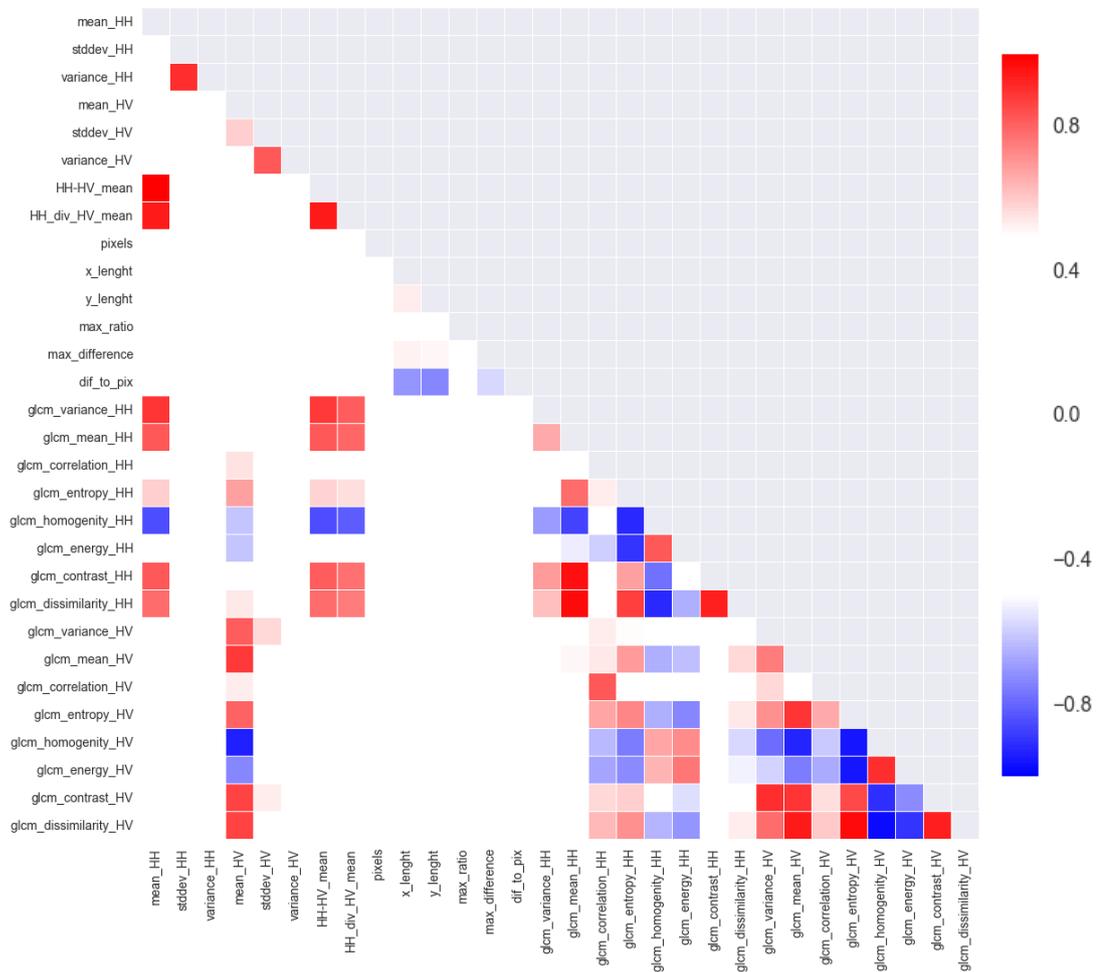


Figure 11. Correlation matrix of all properties.

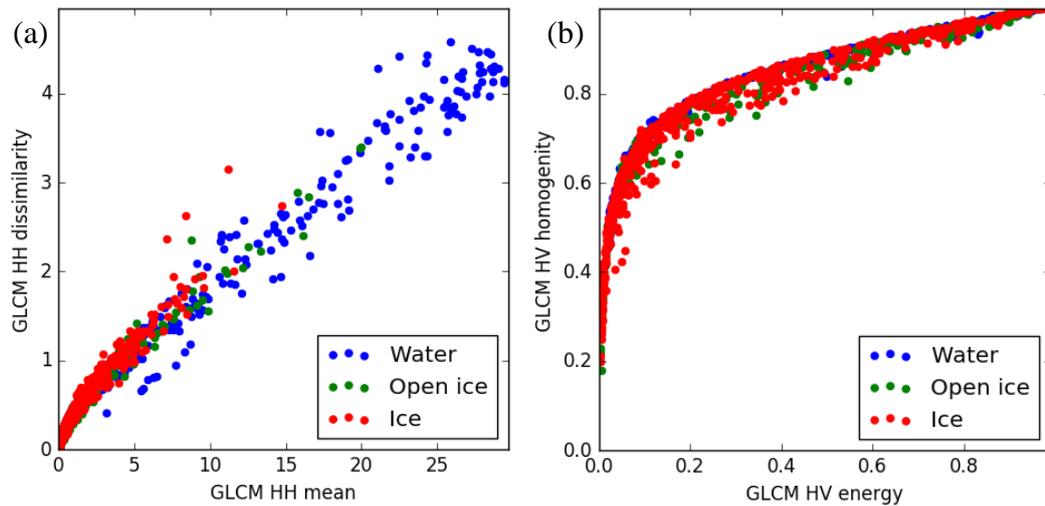


Figure 12. High correlation of properties (a) HH polarization GLCM dissimilarity and HH polarization GLCM mean; (b) HV polarization GLCM homogeneity and HV polarization GLCM energy

The F-test and correlation analysis results in reduction of properties used for classification. The number of properties was reduced from 20 to 10. The remaining properties are:

- *mean_HH*
- *mean_HV*
- *stddev_HV*
- *HH-HV_mean*
- *HH_div_HV_mean*
- *x_lenght*
- *y_lenght*
- *max_ratio*
- *max_difference*
- *dif_to_pix*
- *glcm_variance_HH*
- *glcm_mean_HH*
- *glcm_correlation_HH*
- *glcm_homogeneity_HH*
- *glcm_contrast_HH*
- *glcm_variance_HV*
- *glcm_mean_HV*
- *glcm_correlation_HV*
- *glcm_homogeneity_HV*
- *glcm_contrast_HV*

6.3 Class separation: ice, open ice, water

The data has been divided into three classes: ice, open ice and water. Analysis of variance with different variables was conducted (Figure 13) to assess if these groups are statistically separable before machine learning algorithms are deployed. In Figure 13 it is visible that all F-values between ice and open ice are extremely low. Therefore, it may be stated that ice and open ice are statistically similar with each other and therefore inseparable from each other during classification. In most of the properties, the best separability manifests between ice and water, with the exception of *dif_to_pix*, *x_lenght* and *y_lenght*, where open ice and ice offer better division. It is concluded that open ice class should be neglected and viewed as part of ice class.

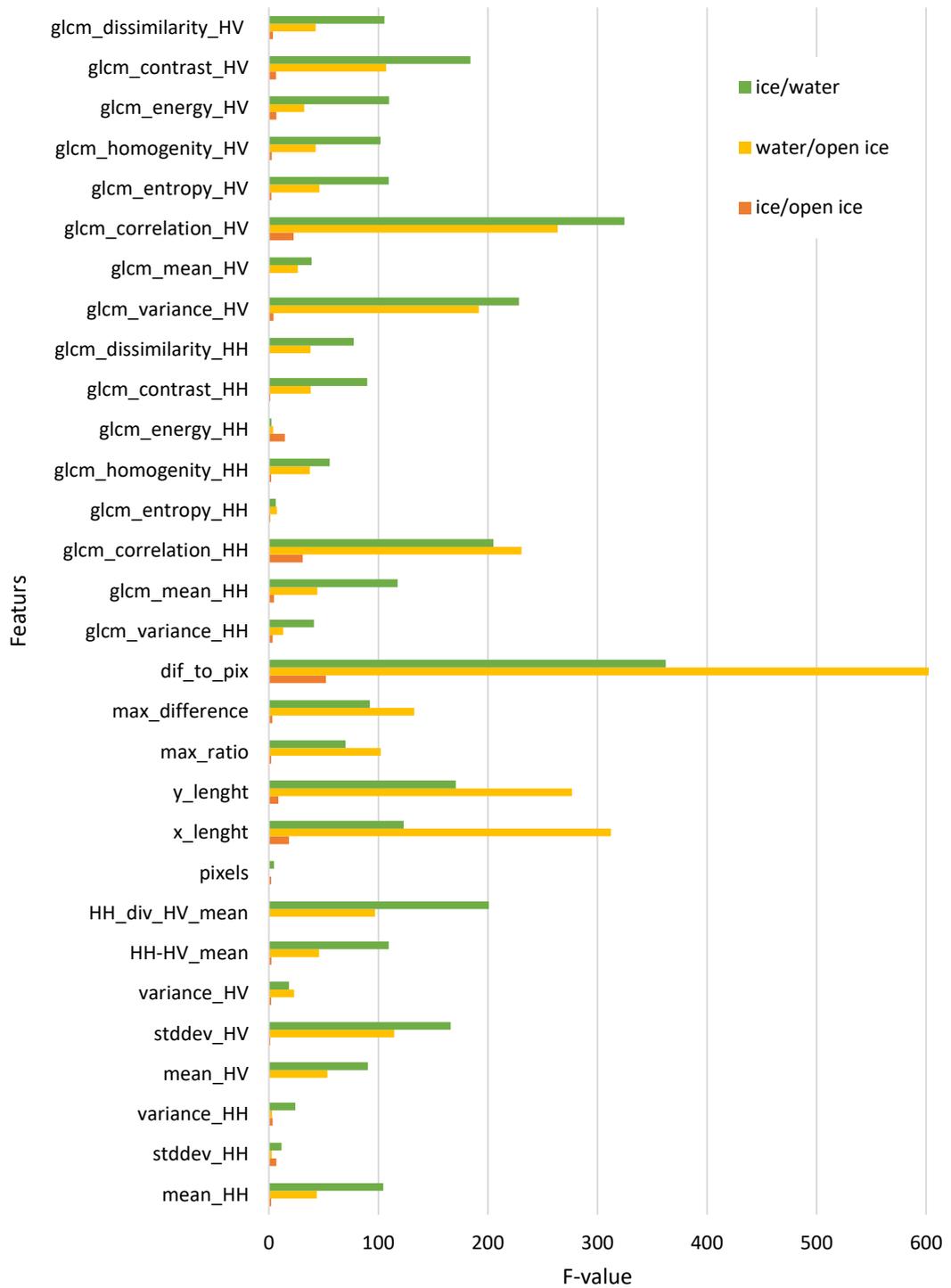


Figure 13. Analysis of variance for different features between all class pairs (ice/water; water/maybe; ice/maybe)

6.4 Machine learning algorithm comparison

Three machine learning algorithms (AdaBoost, Random Forest, Support Vector Machine) are used for sea ice extent calculation. For accuracy measurement 1505 segments from 8 randomly selected training images acquired between 13 January 2016 and 1 February 2016 were evaluated. Ice and water class accuracy by image is available on Table 1.

Table 1. Accuracy of sea ice extent algorithm using different machine learning algorithms for each test image.

Date	Time	AdaBoost		Random Forest		Support Vector Machine	
		Ice class accuracy	Water class accuracy	Ice class accuracy	Water class accuracy	Ice class accuracy	Water class accuracy
13.01.2016	04:57	93%	91%	100%	88%	83%	79%
13.01.2016	16:11	92%	75%	89%	75%	80%	75%
20.01.2016	16:03	96%	88%	98%	88%	77%	78%
25.01.2016	04:57	91%	72%	94%	75%	83%	61%
25.01.2016	04:58	100%	33%	100%	49%	100%	34%
25.01.2016	16:12	63%	85%	81%	90%	63%	61%
27.01.2016	04:42	91%	90%	89%	87%	84%	76%
1.02.2016	16:03	94%	85%	100%	89%	78%	83%
Total		90.7%	80.1%	93.6%	82.2%	80.2%	70.6%
Overall		82.7%		85.0%		73.0%	

The most accurate result was acquired using Random Forest (RF) for classification with overall accuracy of 85%. Because of that, RF based algorithm was chosen for comparison with FMI ice charts in period between 13 January 2016 and 25 February 2016. In Table 2, it is visible that on images with more ice, compared to other images, the RF based algorithm produces more similar results to FMI ice charts. On images where there is less ice, the RF algorithm greatly overestimates the amount of ice in the scene. In case of large ice extent the difference between FMI and RF based algorithm was between 20-85% (on average 54%) and in case of small ice cover the difference was 113-424% (on average 214%). In Figure 14 it is visible that FMI and RF based

algorithm follow similar trend. However when there is more ice, the error is less significant as on images with little ice where the overestimation happens on the coast.

Table 2. Comparison with FMI ice chart.

Date	Time	FMI (km ²)	RF (km ²)	Difference
13.01.2016	04:57	4837	6998	45%
13.01.2016	16:11	3585	6649	85%
20.01.2016	16:03	4266	5890	38%
25.01.2016	04:57	6141	8724	42%
25.01.2016	04:58	1408	4810	242%
25.01.2016	16:12	2180	2883	32%
27.01.2016	04:42	4330	5180	20%
1.02.2016	16:03	1651	3514	113%
6.02.2016	04:57	1947	4569	135%
6.02.2016	16:11	871	1753	101%
8.02.2016	04:42	2094	3233	54%
8.02.2016	15:55	667	1644	147%
13.02.2016	16:03	650	2293	253%
18.02.2016	04:57	593	2119	257%
20.02.2016	04:42	669	2523	277%
20.02.2016	15:55	348	1826	424%
25.02.2016	16:03	453	2188	383%

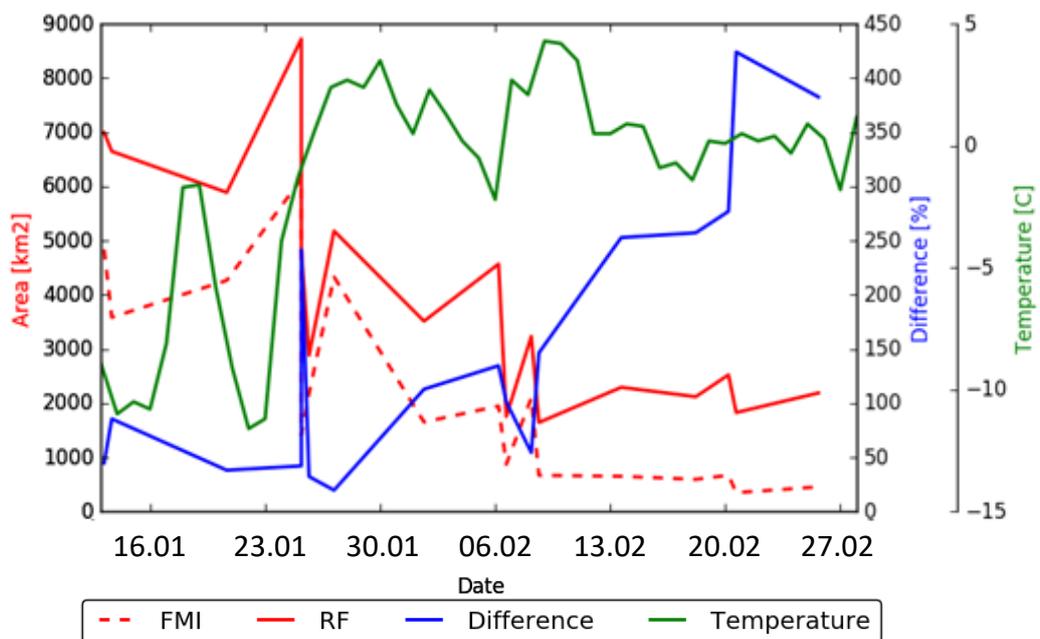


Figure 14. Comparison between FMI ice chart and RF based ice map from 13th Jan to 28th Jan 2016. Both FMI ja RF area decreases in time. The difference between FMI and RF increases. Meteorological data from Loodla et al. (2016)

7 DISCUSSION

7.1 Accuracy of ice classification and false classification causes

According to Table 1, the best overall accuracy is produced by Random Forest based algorithm. The second best result is produced by AdaBoost. Both are ensemble methods, which means that several base estimators are combined to make a prediction. This means that they are easy to use and require little parameters tuning to make the algorithm work for any problem. Zhu (2008) stated that inexperienced users usually produce better results with ensemble methods, rather than kernel methods (as is SVM), even if the problem is better suited for kernel method. This is due to the fact, that kernel methods like SVM require thorough testing and are very sensitive to parameters set up. Inadequate parameterisation could be the reason why SVM performed poorly.

7.1.1 Classification algorithm performance comparison

All used algorithms overestimated the amount of ice in the coastal area as segments by the coast follow the coastal line thus producing segments with irregular size, which in turn is more characteristic of an ice segment. In Figure 15, it is visible that AB and RF greatly overestimates the ice on the coast compared to FMI ice chart. SVM performs slightly better on the coast, but also considers some ice segments as water.

Therefore the algorithm produces better results if coastal area is frozen (Figure 16), especially Väinamere region. This is also illustrated when comparing air temperature data with FMI ice chart and RF based algorithm results (Figure 14). In February 2016 started a warm period, which melted the ice. This coincides with increase of difference with FMI ice chart.

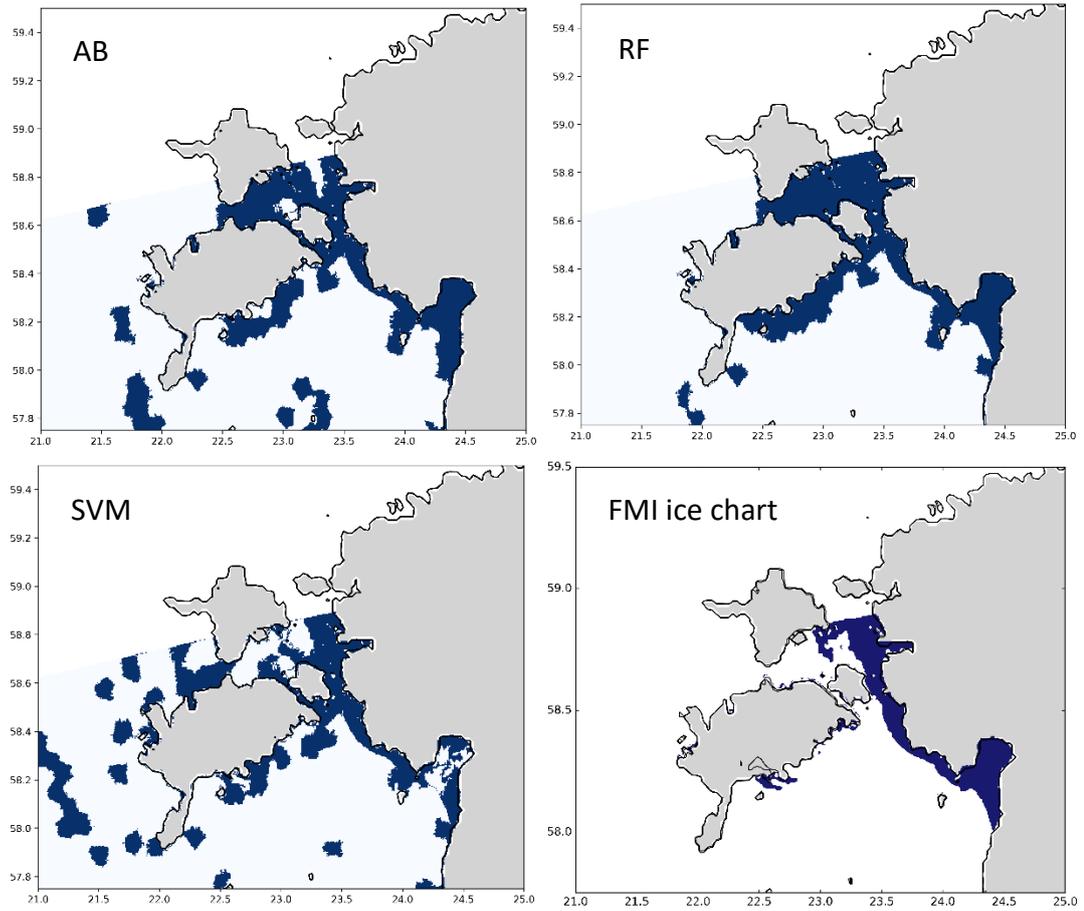


Figure 15. Icemaps produced by Adaboost (AB), Random Forest (RF), Support Vector Machine (SVM) and FMI ice chart.

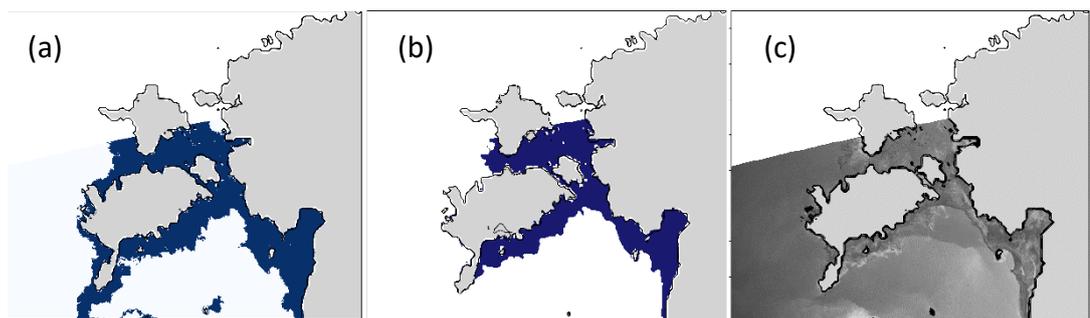


Figure 16. Example of good ice classification result. (a) Random Forest based ice map; (b) FMI ice chart; (c) Radar backscattering at HH polarization on 20th January 2016.

Random Forest based ice classification has shown good ice-water border detection where the influence of coastal area is reduced (Figure 17). This is possible because segmentations produces segment which only contain ice or water class. The false classification of ice in Figure 17a is caused by homogenous nature of ice in the northern coast of Saaremaa, which is more common to water surface.

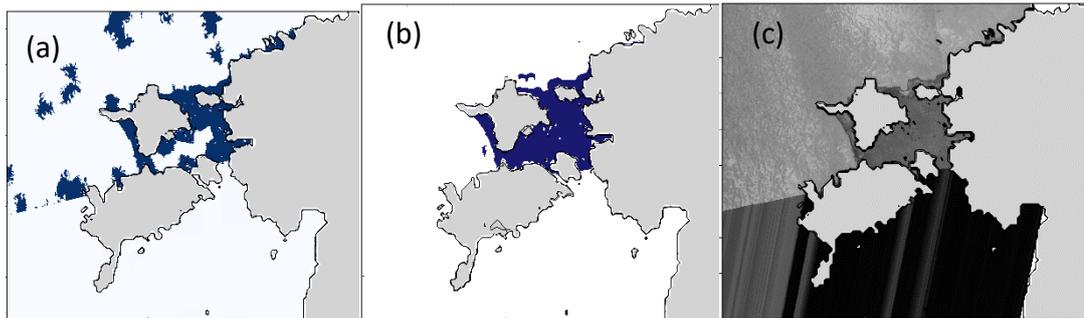


Figure 17. (a) Example of good ice-water border detection. Random Forest based ice map; (b) FMI ice chart; (c) Radar backscatter at HH polarization on 25th January 2016.

It has to be noted, that FMI ice charts also contain errors, although in smaller scale than algorithm produced in current study. In Figure 18 it is visible that FMI ice chart shows existence of ice in area where radar backscatter gives no reason to suspect the presence of ice. This error is present even though ice charts are produced using ice thickness history and visual inspection of the product is usually done.

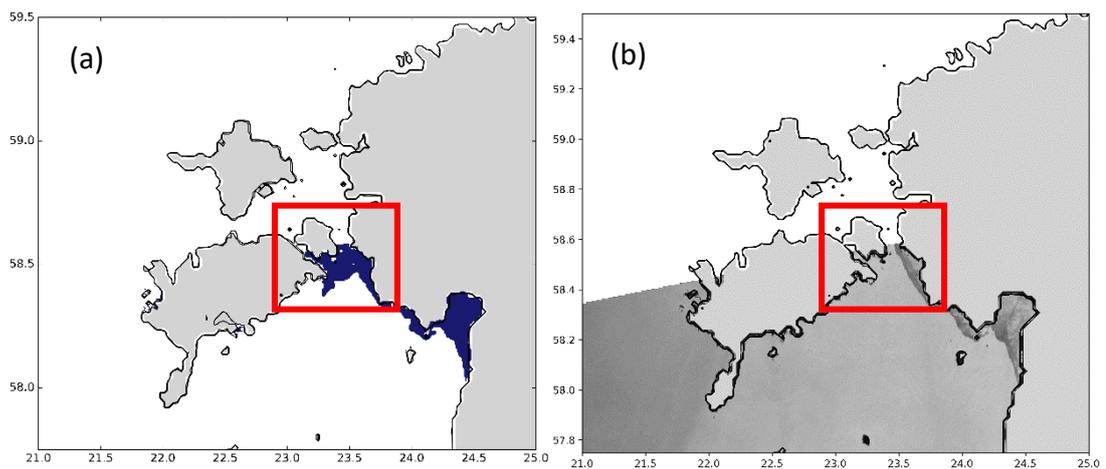


Figure 18. Error in FMI ice chart (red square); (a) FMI ice chart (b) Radar backscatter at HH polarization on 6th January 2016.

7.1.2 Errors caused by source data

Scan lines from HV polarization backscattering may cause false classification (Figure 19). As segments are calculated from HH backscattering field, segmentation is not affected by the scan lines. Scan lines are similar to change in ice type or discontinuity in ice, which smooth and homogeneous water surface do not have. This problem has also been noted by Leigh *et al.* (2014).

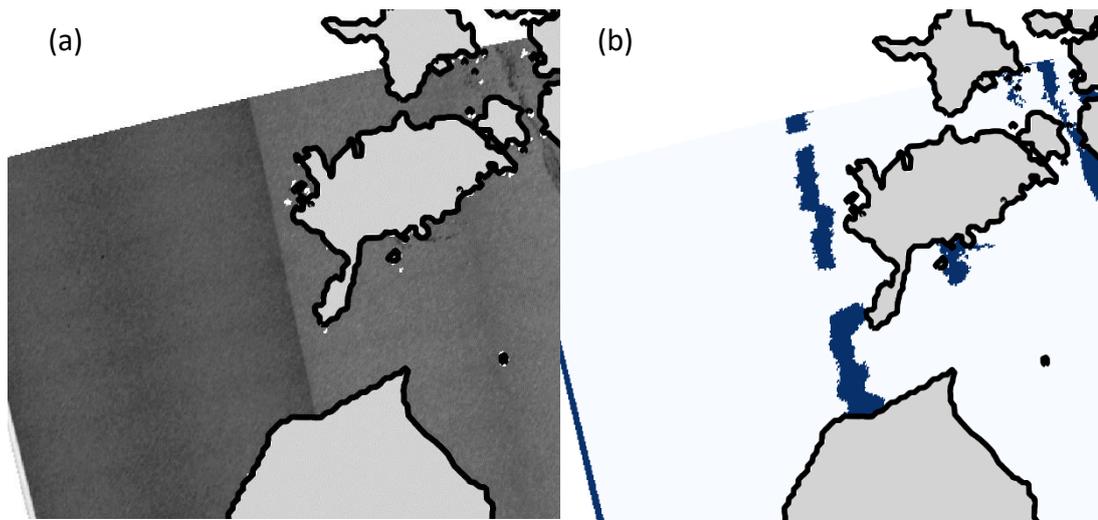


Figure 19. (a) Backscattering at HV polarization; (b) icemap (water - light blue; ice - dark blue) false classification caused by scan line in HV polarization backscattering using SVM.

However, HH polarization is more susceptible to influence of incidence angle on backscatter (Figure 20), where areas scanned with small incidence angles appear brighter on backscattering field. This in turn makes water and ice separation difficult, as brighter backscattering values are common for rough ice surface. Therefore HV polarization enables to distinguish between water and ice

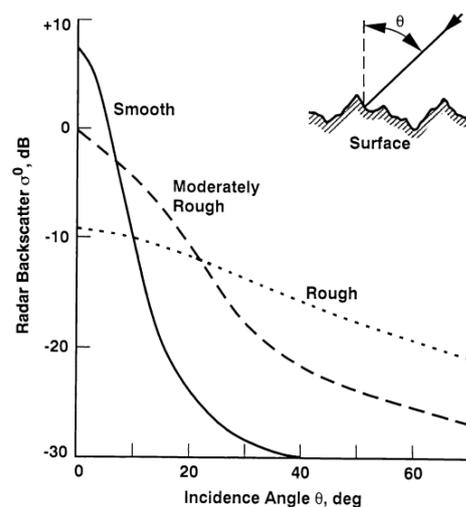


Figure 20. Radar backscatter as a function of incidence angle for representative surfaces. (Farr, 1993)

without incidence angle correction on HH polarization backscattering field.

An alternative to using HV polarization for separating ice and water is to perform incidence angle correction on backscatter at HH polarization. This however may prove to be problematic as backscatter relation to incidence angle is also influenced by the surface roughness (Figure 20), which is information that is not available. Menges *et al.* 2001 suggested a method that can be used without knowledge about surface based on assumption that each line in azimuth direction contains similar composition (histogram) and backscatter frequency distribution of each azimuth line can be used to correct the effect of variations in incidence angle. However, this means that target properties must not vary significantly from one azimuth band to another. Lang *et al.* 2016 proposes class based incidence angle correction for sea ice data, by using *k*-means clustering to separate different ice types and then normalizing the backscatter values within the class.

7.1.3 Effect of meteorological conditions

A problem frequently encounter in water-ice separation is roughening of water surface due to meteorological conditions (Leigh *et al.* 2014) (Scheuchl *et al.*, 2001). In current work it was noted that on 25 January 2016 in three images water surface displayed uncharacteristic patterns (Figure 21), which are likely caused by atmosphere effect. During images acquisitions in Kihnu meteorological station the temperature ranged from -1.6 to -1.4°C with wind speed less than 1 m/s, the sky was fully covered with clouds. It has been noted that on C-band radar clouds effect may show on radar images (Danklmayer, 2009). The patterns on water surface cause irregularly shaped segments and texture parameters similar to ice.

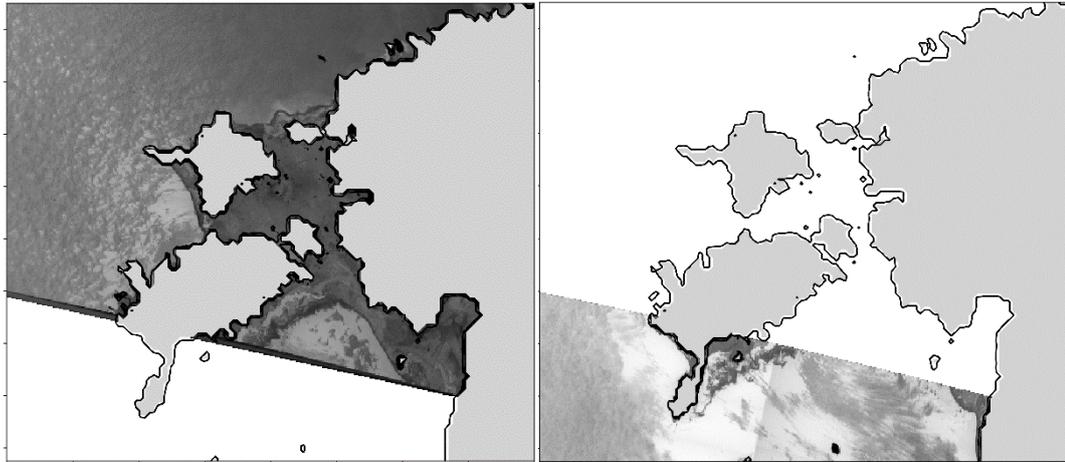


Figure 21. Patterns on water surface from radar backscattering at HH polarization in 25th January 2016.

7.2 Possible improvements

Combining SAR data with information about coastal borders could improve ice extent classification. Ice usually starts to form in the coastal areas and moves outward to the open sea (Granskog *et al.*, 2006). Therefore it could be assumed that areas classified as ice that are far from coast and where neighbouring areas are not covered with ice are falsely classified and could be corrected in an automatic process. This only solves false classification if only few segments are falsely classified. If a larger area is falsely classified, due to wind pattern or other ice look-a-likes, it would remain falsely classified. Similar tactics is employed by Karvone *et al.* (2005) who filter out small segments.

This could be partly solved by using ice maps produced by Estonian Weather Service which are manually produced every day during winter season using different inputs. If area classified as ice is far from ice areas on ice maps, they are probably falsely classified. Areas that are covered with ice on the ice maps and are not located at ice/water boarder, could be classified as ice with high confidence. Machine learning based SAR data classification would still remain as the main source of information at ice/water boarder. Any additional information to the SAR image will create alongside

benefits additional errors. Therefore, the trade-off of any additional information has to be evaluated.

The main problem in the coastal area is overclassification of ice. Due to coastal segments are in irregular shape, which is characteristic to ice segment. By reducing the influence of segment shape as classification parameter in the coastal area and increasing the influence of texture parameters, this could be reduced.

CONCLUSION

Superpixel algorithm based segmentation provides good separation of ice and water into different segments which provides basis for ice classification. Segmentation allows to calculate multiple characterizing parameters for area which provide additional information to radar backscattering at HH and HV polarizations.

For each segment 30 properties including 8 value parameters, 16 texture parameters and 6 geometrical parameters were calculated and analysed using analysis of variance and correlation. It was determined that backscatter standard deviation, GLCM energy, GLCM entropy calculated from HH polarization backscattering and variance calculated from HH and HV backscattering and number of pixels in segment offer little for class separation and may be removed. Good class separation was provided by GLCM correlation calculated at HV and HH polarization, variance from HV backscatter, HH backscatter divided by HV backscatter and number of pixels in segment relative to the area where the segment fits. Many GLCM parameters correlate strongly with each other as they are calculated from same field. Very strong correlation is between homogeneity, entropy and energy, thus two of the latter may be removed in order to avoid duplicating information. Number of properties used for operational processing and validation was reduced to 20.

Possibility of class separation between ice and open ice was investigated. Due to strong similarities between the datasets, it was concluded that ice and open ice are inseparable using current method. Therefore, only ice and water classification was attempted.

AdaBoost, Random Forest and Support Vector Machine algorithms were tested for ice classification. Out of tested algorithms, Random Forest allowed to achieve most

accurate results in cold conditions when most of the coastal area was frozen, providing accuracy of 93.6% on ice and 82.2% on water.

Comparison between FMI ice charts and Random Forest based classification gives greater similarities in cold conditions (difference 20-85%) and increases in warmer period when coastal area is clear of ice (difference 113-424%). This is induced by overclassification of ice near the coast due to segments irregular shape. If coast effect is reduced, the algorithm shows good ability to detect ice-water border.

Segmentation of SAR backscatter allows separating ice and water into homogenous divisions that provide good basis for ice-water classification. Current ice classification algorithm provides information about ice location in cold conditions when water by the coast is mostly frozen. In difficult conditions (warm weather and small amount of ice) the algorithm is unreliable and needs further investigation in order to improve classification on the coast.

ACKNOWLEDGMENTS

Sentinel-1 data was provided by ESA, MODIS imagery by NASA and FMI ice charts by Copernicus Marine Environmental monitoring Service (CMEMS).

The author would like to acknowledge Sander Rikka, Laura Siitam and Siim Pärt for taking part of ice data collection expeditions. Age Arikas for providing processed MODIS imagery. Dmitry Fisun for assisting with making the ice classification algorithm operational. Sander Rikka for providing GLCM texture parameter calculation code.

LIST OF REFERENCES

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (2012) SLIC Superpixels Compared to State-of-the-art Superpixel Methods. – *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 34 (11), 2274-2282.
- Bourbigot, M., Johnsen, H., Piantanida, R. (2016) Sentinel-1 Product Specification: S1-RS-MDA-52-7441.
- Breiman, L. (2001) Random Forest. – *Machine Learning*, 45(1), 5-32.
- Castrounis, A. (2016) Machine Learning: An In-Depth Guide - Overview, Goals, Learning Types, and Algorithms [WWW]
<http://www.innoarchitech.com/machine-learning-an-in-depth-non-technical-guide/> (02.05.2017)
- Cortes, C., Vapnik, V. (1995) Support-Vector Networks. – *Machine Learning*, 20, 273-297.
- CSC Mission Management Team (2017) Sentinel High Level Operations Plan (HLOP): COPE-S1OP-EOPG-PL-15-0020.
- Danklmayer, A. (2009) Propagation Effects in Satellite mounted Radar Remote Sensing – *Proceedings of Wave Propagation and Scattering in Communications, Microwave Remote Sensing and Navigation (WFMN): November 25-27, 2009, Chemnitz, Deutschland*.
- Dierking, W. (2013) Sea ice monitoring by synthetic aperture radar. – *Oceanography*, 26 (2), 100-111.
- Gebejes, A., Huertas R. (2013) Texture Characterization based on Grey-Level Co-occurrence Matrix – *Proceedings in Conference of Informatics and Management Sciences: March , 25-29. 2013, Slovakia*, 375-378.

- Granskog, M., Kaartokallio, H., Kuosa, H., Thomas, D. N., Vainio, J. (2006) Sea ice in the Baltic Sea – A review. – *Estuarine, Coastal and Shelf Science*, 70, 145-160.
 EPSG:3301. [WWW] <http://spatialreference.org/ref/epsg/3301/> (30.01.2017)
 EPSG:4326. [WWW] <http://spatialreference.org/ref/epsg/wgs-84/> (30.01.2017)
- Farr, T. G. (1993) Radar Interactions With Geological Surfaces. - *Guide to Magellan Image Interpretation*. JPL Publications: Pasadena, 45-56.
- Freund, Y., Schapire, R. (1997) A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. – *Journal of Computer and System Sciences*, 55, 119-139.
- Friedrichs, F., Igel, C. (2005) Evolutionary tuning of multiple SVM parameters. – *Neurocomputing*, 64, 107-117.
- Haralick, R., Shanmugam, K., Dinstein, I. (1973) Texture Features for Image Classification. – *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3 (6), 610-621.
- Ho, T. K. (1998) The Random Subspace Method for Constructing Decision Forests. – *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (8) 832-844.
- Karvonen, J. (2004) Baltic Sea Ice SAR Segmentation and Classification Using Modified Pulse-Coupled Neural Networks - *IEEE Transactions on Geoscience and Remote Sensing*, 42 (7), 1566-1574.
- Karvonen, J., Similä, M., Mäkynen, M. (2005) Open Water Detection From Baltic Sea Ice Radarsat-1 SAR Imagery. – *IEEE Geoscience and Remote Sensing Letters*, 2 (3) 275-279.
- Karvonen, J., Cheng, B., Similä, M., Hallikainen, M. (2008) Baltic sea ice thickness charts based on thermodynamic snow/ice model, C-band SAR classification and ice motion detection. – *Proceedings Geoscience and Remote Sensing Symposium, 2008 (IGARSS-08)*, 173-176.
- Karvonen, J. (2012a) Baltic Sea Ice Concentration Estimation Based on C-Band HH-Polarized SAR Data. – *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5 (6), 1874-1884.

- Karvonen, J. (2012b) Operational SAR-based sea ice drift monitoring over Baltic Sea. – *Ocean Science*, 8, 473-483.
- Lang, W., Zhang, P., Wu, J., Shen, Y., Yang, X. (2016) Incidence Angle Correction of SAR Sea Ice Data Based on Locally Linear Mapping. - *IEEE Transactions on Geoscience and Remote Sensing*, 54 (6), 3188-3199.
- Leigh, S., Wang, Z. (2014) Automated Ice-Water Classification Using Dual Polarization SAR Satellite Imagery. – *IEEE Transactions on Geoscience and Remote Sensing*, 52 (9), 5529-5539.
- Loodla, K., Tillmann, E., Kallis, A., Pärn, R., Vint, K., Juust, E., Krabbi, M. (2017) Eesti meteoroloogia aastaraamat 2016. Tallinn: Keskkonnaagentuur.
- Menges, C. H., Van Zyl, J. J., Hill, G. J., Ahmad, W. (2001) A procedure for correction of the effect of variation in incidence angle on AIRSAR data. – *Remote Sensing*, 22 (5), 829-841.
- Michie, D., Spiegelhalter, D.J., Taylor. C.C., (1994) Introduction. - *Machine learning, neural and statistical classification*, 1-5.
- Nekrasov. A (2014) Water Surface Backscattering and Wind Retrieval. – *Foundations for Innovative Application of Airborne Radars*. Springer, 5-11.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011) Scikit-learn: Machine Learning in Python. – *Journal of Machine Learning Research*, 12, 2825-2830.
- Ren, X., Malik, J. (2003) Learning a Classification Model for Segmentation. – *Proceedings Ninth IEEE International Conference on Computer Vision*, 1, 10-17.
- Robinson, I. (2004) Radars, sea surface roughness, and scatterometry - *Measuring the Oceans from Space*, Springer-Verlag: Berlin, 385-445.
- Russel, S., Norvig, P. (1995) Learning from observation – *Artificial Intelligence: A Modern Approach*, Prentice-Hall: New Jersey. 525-562.

- Scheuchl, B., Caves, R., Cumming, I., Staples, G. (2001) Automated sea ice classification using spaceborne polarimetric SAR data. – *Proceedings Geoscience and Remote Sensing Symposium (IGARSS-01)*. 3117-3119.
- Scott MacKenzie, I. (2013) Hypothesis Testing. – *Human-Computer Interaction: An Empirical Research Perspective*. Elsevier: Waltham, 191-227.
- SNAP Sentinel-1 Toolbox Help (2014) [computer program] version 4.0.0, Array Systems Computing Inc, DLR, OceanDataLab, Brockmann Consult.
- Soh, L., Tsatsoulis, C. (1999) Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. – *IEEE Transactions on Geoscience and Remote Sensing*, 37(2), 780-795.
- SUHET (2013) Sentinel-1 User Handbook: GMES-S1OP-EOPG-TN-13-0001.
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F. (2012) GMES Sentinel-1 mission. – *Remote Sensing of Environment*, 120, 9-24.
- Uiboupin, R., Pärn, O. (2016) Mereala planeeringu alusuuring: jääolude analüüs ja kaartide koostamine. TTÜ Meresüsteemide Instituut: Tallinn.
- Zakhvatkina, N., Alexandrov, V., Johannessen, O., Sandven, S., Frolov, I. (2012) Classification of Sea Ice Types in ENVISAT Synthetic Aperture Radar Images. - *IEEE Transactions on Geoscience and Remote Sensing*, 51 (5), 2587-2600.
- Zhu, J., Zou, H., Rosset, S., Hastie, T. (2009) Multi-class AdaBoost. – *Statistics and Its Interface*, 2, 349-360.
- Zhu, M. (2008) Kernels and Ensembles: Perspectives on Statistical Learning. – *The American Statistician*, 62 (2) 97-109.

APPENDICES

Appendix 1. SAR data

Table 1. Training data

No.	Date	Name
1	13.01.2017	S1B_EW_GRDM_1SDH_20170113T045702_20170113T045802_003825_00693E_9AEF
2	13.01.2017	S1B_EW_GRDM_1SDH_20170113T161111_20170113T161216_003832_006972_064C
3	15.01.2017	S1B_EW_GRDM_1SDH_20170115T044042_20170115T044142_003854_006A20_568D
4	15.01.2017	S1B_EW_GRDM_1SDH_20170115T155507_20170115T155611_003861_006A4C_848D
5	17.01.2017	S1B_EW_GRDM_1SDH_20170117T042409_20170117T042513_003883_006AFA_6380
6	17.01.2017	S1B_EW_GRDM_1SDH_20170117T153914_20170117T154018_003890_006B24_4826
7	18.01.2017	S1B_EW_GRDM_1SDH_20170118T050513_20170118T050613_003898_006B6A_56A1
8	18.01.2017	S1B_EW_GRDM_1SDH_20170118T050613_20170118T050714_003898_006B6A_BB24
9	18.01.2017	S1B_EW_GRDM_1SDH_20170118T161938_20170118T162042_003905_006B9C_0567
10	20.01.2017	S1B_EW_GRDM_1SDH_20170120T044850_20170120T044950_003927_006C4A_831A
11	20.01.2017	S1B_EW_GRDM_1SDH_20170120T044950_20170120T045051_003927_006C4A_A34D
12	20.01.2017	S1B_EW_GRDM_1SDH_20170120T160315_20170120T160419_003934_006C7D_BF4F
13	22.01.2017	S1B_EW_GRDM_1SDH_20170122T043221_20170122T043325_003956_006D35_3C8C
14	22.01.2017	S1B_EW_GRDM_1SDH_20170122T043325_20170122T043404_003956_006D35_E7F9
15	22.01.2017	S1B_EW_GRDM_1SDH_20170122T154655_20170122T154759_003963_006D63_0FCB
16	25.01.2017	S1B_EW_GRDM_1SDH_20170125T045702_20170125T045802_004000_006E7E_8510
17	25.01.2017	S1B_EW_GRDM_1SDH_20170125T045802_20170125T045902_004000_006E7E_30E7
18	25.01.2017	S1B_EW_GRDM_1SDH_20170125T161111_20170125T161215_004007_006EB1_FDCE

19	27.01.2017	S1B_EW_GRDM_1SDH_20170127T044042_20170127T044142_004029_006F62_CAF3
20	27.01.2017	S1B_EW_GRDM_1SDH_20170127T155506_20170127T155611_004036_006F92_6314
21	29.01.2017	S1B_EW_GRDM_1SDH_20170129T042409_20170129T042512_004058_00703D_955B
22	29.01.2017	S1B_EW_GRDM_1SDH_20170129T153914_20170129T154018_004065_007069_08AF
23	30.01.2017	S1B_EW_GRDM_1SDH_20170130T050513_20170130T050613_004073_0070A9_0BB9
24	30.01.2017	S1B_EW_GRDM_1SDH_20170130T050613_20170130T050713_004073_0070A9_F64C
25	30.01.2017	S1B_EW_GRDM_1SDH_20170130T161937_20170130T162042_004080_0070D7_831D
26	01.02.2017	S1B_EW_GRDM_1SDH_20170201T044850_20170201T044950_004102_007185_8BEB
27	01.02.2017	S1B_EW_GRDM_1SDH_20170201T044950_20170201T045050_004102_007185_C40B
28	01.02.2017	S1B_EW_GRDM_1SDH_20170201T160314_20170201T160419_004109_0071B5_E062
29	03.02.2017	S1B_EW_GRDM_1SDH_20170203T043220_20170203T043325_004131_007260_9FEA
30	03.02.2017	S1B_EW_GRDM_1SDH_20170203T043325_20170203T043403_004131_007260_999A
31	03.02.2017	S1B_EW_GRDM_1SDH_20170203T154655_20170203T154759_004138_00728E_ABB9
32	05.02.2017	S1B_EW_GRDM_1SDH_20170205T153105_20170205T153209_004167_00735C_06CA
33	06.02.2017	S1B_EW_GRDM_1SDH_20170206T045701_20170206T045801_004175_0073A4_22AE
34	06.02.2017	S1B_EW_GRDM_1SDH_20170206T045801_20170206T045902_004175_0073A4_4853
35	06.02.2017	S1B_EW_GRDM_1SDH_20170206T161111_20170206T161215_004182_0073DB_225B

Appendix 2. *In situ* observations

Table 1. Observations in 2015 and 2016.

Date	Location	Latitude	Longitude	Ice types
16.01.2015	Keemu	58.746111	23.672222	Rough ice, melted
16.01.2015	Saastna	58.738778	23.544972	Rough ice
16.01.2015	Puise Nina	58.763722	23.447639	Pressure ridge
16.01.2015	Suur-Holm	58.960722	23.521306	Rough and level ice
29.01.2015	Tahkuranna	58.249116	24.469733	Small pressure ridges, old fast ice, new level ice
29.01.2015	Uulu mole	58.299166	24.572916	Rough and level ice
29.01.2015	Pärnu port	58.366333	24.464233	Pressure ridge, level ice
29.01.2015	Liu	58.278183	24.271100	Pancake ice
17.02.2015	Uulu mole	58.298900	24.568150	Level ice
17.02.2015	Pärnu port	58.360475	24.468780	Pressure ridge
17.02.2015	Liu	58.271311	24.270020	Pancake ice
07.01.2016	Pedase	59.282355	23.897493	Closed ice
07.01.2016	Puise Nina	58.760672	23.470222	Very closed ice
07.01.2016	Kesselaid	58.632407	23.495728	Fast ice
07.01.2016	Matsi beach	58.371367	23.732642	Fast ice, water-ice border
07.01.2016	Munalaid	58.22742	24.122478	Fast ice
08.01.2016	Häädemeeste	58.071577	24.46929	Fast ice, water-ice border
08.01.2016	Liu	58.284848	24.272082	Fast ice
08.01.2016	Munalaid	58.219837	24.092083	Very closed ice, fast ice
08.01.2016	Matsi beach	58.370418	23.723205	Open ice
08.01.2016	Saastna peak	58.73554	23.503622	Fast ice
12.01.2016	Valaste fall	59.447875	27.334773	Drift ice, water-ice border
12.01.2016	Toila	59.427575	27.513597	Open ice, water-ice border
12.01.2016	Sillamäe	59.40077	27.778492	Open ice, water-ice border
15.01.2016	Pedase	59.280778	23.894747	Fast ice
15.01.2016	Ramsi cape	59.034562	23.403993	Fast ice
15.01.2016	Virtsu	58.554988	23.488383	Fast ice
15.01.2016	Muriste	58.508693	23.659612	Fast ice
15.01.2016	Matsi beach	58.370363	23.7253	Fast ice, very closed ice

15.01.2016	Manilaid	58.219415	24.144065	Fast ice
22.01.2016	Valaste fall	59.452825	27.337312	Fast ice
22.01.2016	Sillamäe	59.400248	27.7847	Fast ice
22.01.2016	Narva-Jõesuu	59.453817	28.014553	Fast ice

Appendix 3. Classifier algorithm accuracy

Table 1. Adaboost based algorithm accuracy.

Date	Time	Ice correct	Ice false	Water correct	Water false	Ice class accuracy	Water class accuracy	Overall accuracy
13.01.2016	04:57	55	4	171	17	93%	91%	91%
13.01.2016	16:11	59	5	86	29	92%	75%	81%
20.01.2016	16:03	55	2	142	19	96%	88%	90%
25.01.2016	04:57	71	7	107	41	91%	72%	79%
25.01.2016	04:58	9	0	29	60	100%	33%	39%
25.01.2016	16:12	17	10	123	21	63%	85%	82%
27.01.2016	04:42	50	5	87	10	91%	90%	90%
1.02.2016	16:03	17	1	168	29	94%	85%	86%

Table 2. Random Forest based algorithm accuracy.

Date	Time	Ice correct	Ice false	Water correct	Water false	Ice class accuracy	Water class accuracy	Overall accuracy
13.01.2016	04:57	59	0	165	23	100%	88%	91%
13.01.2016	16:11	57	7	85	29	89%	75%	80%
20.01.2016	16:03	63	1	135	19	98%	88%	91%
25.01.2016	04:57	73	5	111	37	94%	75%	81%
25.01.2016	04:58	9	0	44	45	100%	49%	54%
25.01.2016	16:12	22	5	130	14	81%	90%	89%
27.01.2016	04:42	49	6	84	13	89%	87%	88%
1.02.2016	16:03	18	0	176	21	100%	89%	90%

Table 3. Support Vector Machine based algorithm accuracy.

Date	Time	Ice correct	Ice false	Water correct	Water false	Ice class accuracy	Water class accuracy	Overall accuracy
13.01.2016	04:57	49	10	149	39	83%	79%	80%
13.01.2016	16:11	51	13	85	29	80%	75%	76%
20.01.2016	16:03	49	15	120	34	77%	78%	78%
25.01.2016	04:57	65	13	90	58	83%	61%	69%
25.01.2016	04:58	9	0	30	59	100%	34%	40%
25.01.2016	16:12	17	10	88	56	63%	61%	61%
27.01.2016	04:42	46	9	74	23	84%	76%	79%
1.02.2016	16:03	14	4	163	34	78%	83%	82%

KOKKUVÕTE

Superpikslitel põhinev segmenteerimise algoritm annab hea tulemuse vee ja jää eristamiseks eraldi osadeks, mis on aluseks jää klassifitseerimisele. Segmenteerimine võimaldab arvutada mitmeid ala iseloomustavaid parameetreid, mis annavad radari tagasipeegeldumisele HH ja HV polarisatsioonil lisa informatsiooni.

Igale segmendile arvutatakse 30 omadust, mille seas on 8 väärtusparameetrit, 16 tekstuuriparameetrit ning 6 geomeetrilist parameetrit. Parameetreid analüüsitakse kasutades variatsiooni analüüsi ja korrelatsiooni. Kuna HH polarisatsiooni tagasipeegelduvuse standardhälve, GLCM energia, GLCM entroopia, HH ja HV polarisatsiooni tagasipeegelduvuse varieeruvus ning pikslite arv segmendis annab vähe lisainformatsiooni klasside eristamiseks, siis need eemaldatakse. Head klasside eristamist võimaldasid GLCM korrelatsioon HH ja HV polarisatsioonis, HV tagasipeegelduvuse varieeruvus, HH tagasipeegelduvus jagatuna HV tagasipeegelduvusega ning pikslite arvu suhe alaga, kuhu segment mahuks. Mitmed GLCM parameetrid korreleeruvad üksteisega, kuna need arvutatakse ühest ja samast väljast. Eriti tugev korrelatsioon on homogeensuse, entroopia ja energia vahel, mille tõttu kaks viimast eemaldatakse, et vähendada dubleerivat informatsiooni. Omaduste hulk, mida operatiivses algoritmis ja valideerimises kasutati, vähendati kahekümneni.

Hinnati jää, lahtise jää ja vee eristamise võimalikkust. Andmehulkade sarnasusest tulenevalt on antud meetodiga jää ja lahtise jää klassid üksteisest eristamatud. Selle tõttu viidi läbi vaid jää ja vee klassifitseerimine.

Jää klassifitseerimiseks testiti *AdaBoost*, *Random Forest* ja *Support Vector Machine* algoritme. Valitud algoritmidest pakkus *Random Forest* kõige täpsemat tulemust külmades ilmatingimustes, mil enamik rannikualast oli jäätunud (93.6% jää ja 82.2% veel).

FMI jääkaardi ja *Random Forest*'il põhineva algoritmi tulemuse võrdlus annab sarnasemaid tulemusi külmades tingimustes (erinevus 20-85%), erinevus suureneb soojal perioodil, kui rannik on jäävaba (erinevus 113-424%). Selle põhjuseks on rannikul jää üleklassifitseerimine tulenevalt segmentide ebaregulaarsest kujust. Kui ranniku mõju vähendada, pakub algoritm head jää ja vee piiri eristamist.

Tehisavaradari tagasipeegelduvuse segmenteerimine võimaldab jää ja vee eristamist homogeenseteks aladeks, mis on hea põhi jää-vee eristamiseks. Antud jää klassifitseerimise algoritm annab informatsiooni jää asukoha kohta külmades tingimustes, kui rannik on jäätunud. Keerulistes oludes (soe ilm ja vähe jääd) on algoritm ebausaldusväärne ning vajab edasisi uuringuid, et parandada klassifitseerimist rannikul.