MACHINE LEARNING BASED CLASSIFICATION OF LAKE ICE AND OPEN WATER FROM SAR ALTIMETRY WAVEFORM PARAMETERS

Jaya Sree Mugunthan[1,2], Claude R. Duguay[1,2] and Elena Zakharova[3,4]

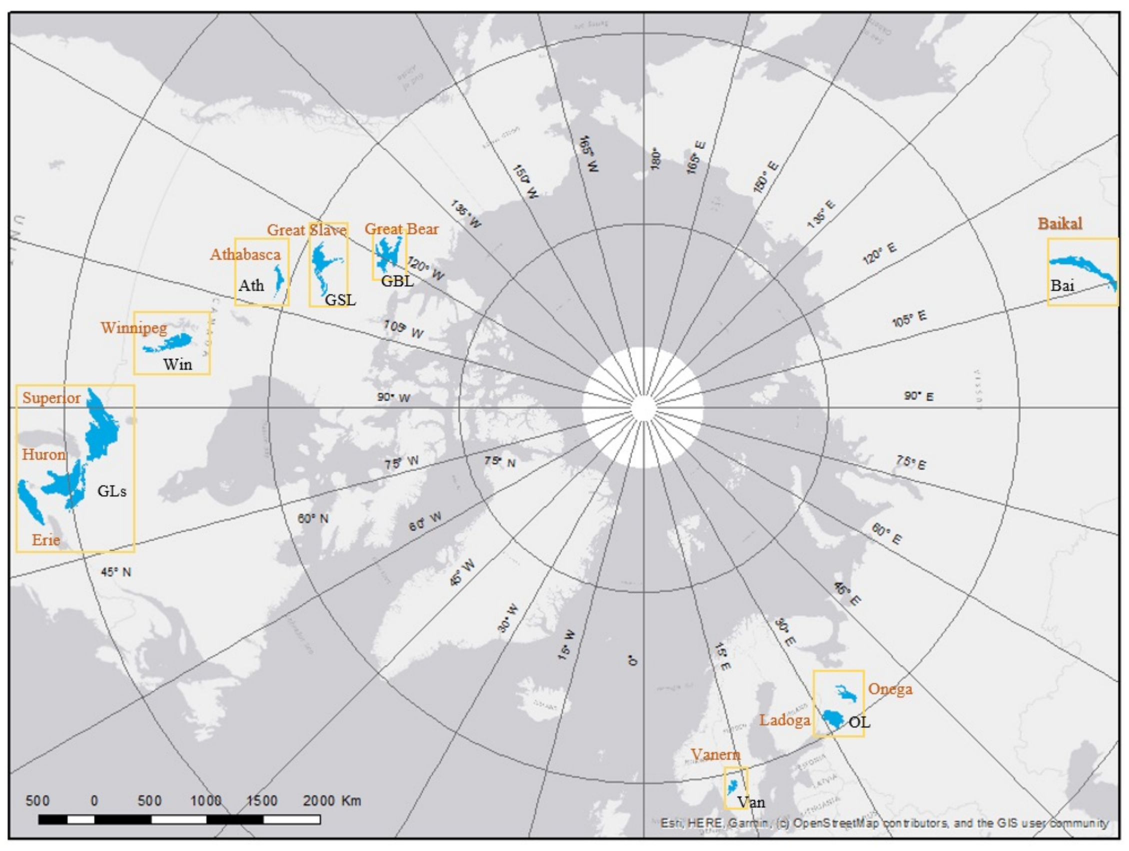
[1] University of Waterloo, Waterloo Ontario, Canada
[2] H2O Geomatics, Waterloo Ontario, Canada
[3] EOLA, Toulouse, France
[4] Water Problem Institute of RAS, Moscow, Russia

University of Waterloo | Department of Geography and Environmental Management
H2O Geomatics
EOLA

INTRODUCTION

- Lakes cover 15-40% of the arctic and sub-arctic regions and play a key role in regulating weather and climate (Duguay et al., 2003).
- Satellite radar altimetry has been used in many hydrological and cryosphere applications for several years. However, recent studies show that the retrieval of water level and ice thickness may be limited or not possible in the presence of ice of varying properties. Hence, information on surface types associated with each altimetry observation is important.
- This study addresses the research gap in finding an optimal approach to classify lake ice types and open water using altimetry data.

STUDY AREA



- Eleven lakes distributed across the Northern Hemisphere were selected to collect sample data for the machine learning (ML) classifiers.
- These lakes are situated in different geographical locations and experience a wide range of ice-cover conditions during winter.

Figure 1. Locations of the lakes selected in the study to assess the capability of ML classifiers in classifying ice and open water across years. Yellow rectangles with black-coloured labels indicate regions of spatial clusters

MATERIALS AND METHODS

DATA

- *Altimetry data*: Sentinel-3 SRAL Level 2 data.
- *Lake mask data*: Distance-to-land dataset derived using ESA CCI Land Cover Map.
- *Labelling data*: MODIS Aqua/Terra, Sentinel-1 SAR, Sentinel-2 Multispectral and ERA5 reanalysis data (2-m near-surface air temperature).
  - The sample collected across three ice seasons (2018-2019, 2019-2020 and 2020-2021) consisted of **104,558 waveforms** (open water: 29,131; young ice: 22,258; growing ice: 25,920; and melting ice: 26,249).

PARAMETERIZATION OF WAVEFORMS

Depending on the surface types of the lake, four ML classes were considered in this study: 1) Open water (OW), 2) Young ice (YI), 3) Growing ice (GI) and 4) Melting ice (MI).

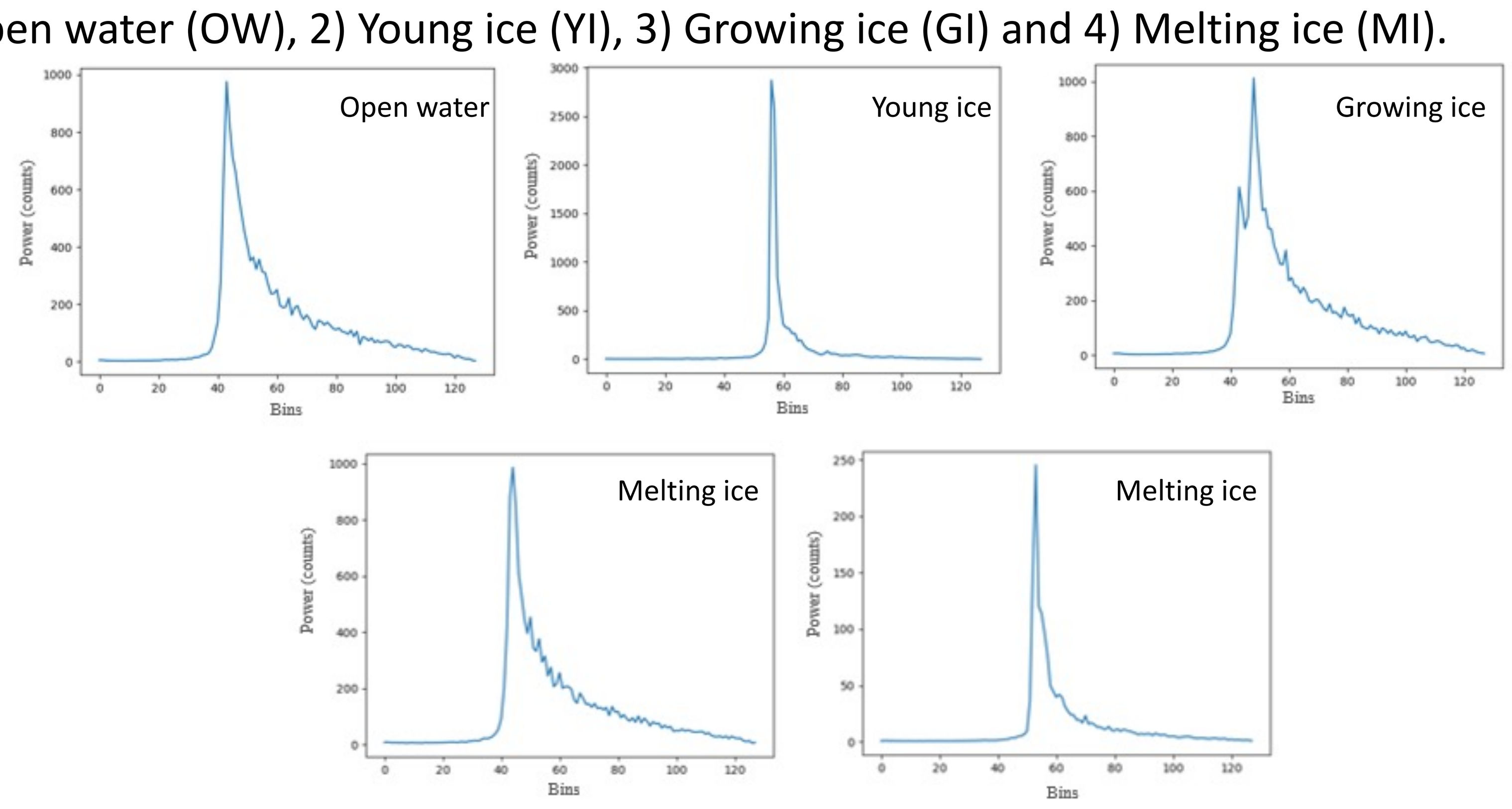


Figure 2. SRAL waveforms observed on Great Slave Lake (Canada)

To characterize the altimetry waveforms, seven waveform parameters including Sigma0, Max, Pulse Peakiness (PP), Leading Edge Width (LEW), OCOG Width (OCOG_W), Late Tail to Peak Power (LTPP) and Early Tail to Peak Power (ETPP) were extracted from each waveform. Sigma0 (OCOG retracker based) was obtained directly from the SRAL product.

1) Max: maximum power of the waveform.
2) $PP = 128 * \frac{P_{max}}{\sum_{i=1}^{128} P_i}$ (Ricker et al., 2014)
3) LEW: the distance between the first bin position containing equal to or greater than 10% of the power maximum and the bin position of the maximum waveform power.
4) $OCOG\_W = \frac{\left(\sum_{i=1}^{128} P_i^2\right)^2}{\sum_{i=1}^{128} P_i^4}$ (Wingham et al., 1986)
5) $LTPP = \frac{\frac{1}{21} * \sum_{i=\max+50}^{\max+70} P_i}{P_{max}}$ (Rinne & Similä, 2016)
6) $ETPP = \frac{\frac{1}{6} * \sum_{i=\max+1}^{\max+6} P_i}{P_{max}}$ (Rinne & Similä, 2016)

Max, ETPP and LTPP were removed from the subsequent analysis. The first two were eliminated due to their least contribution to the classification performance. LTPP was removed as the algorithm fails to estimate LTPP values during the ice season (especially young ice).

With the remaining parameters, different parameter configurations were created to test the ML classifiers including Support Vector Machine (SVM), K- Nearest Neighbors (KNN), Random Forest (RF) and Gradient Booster Trees (GBT).

Figure 3. Methodology

RESULTS

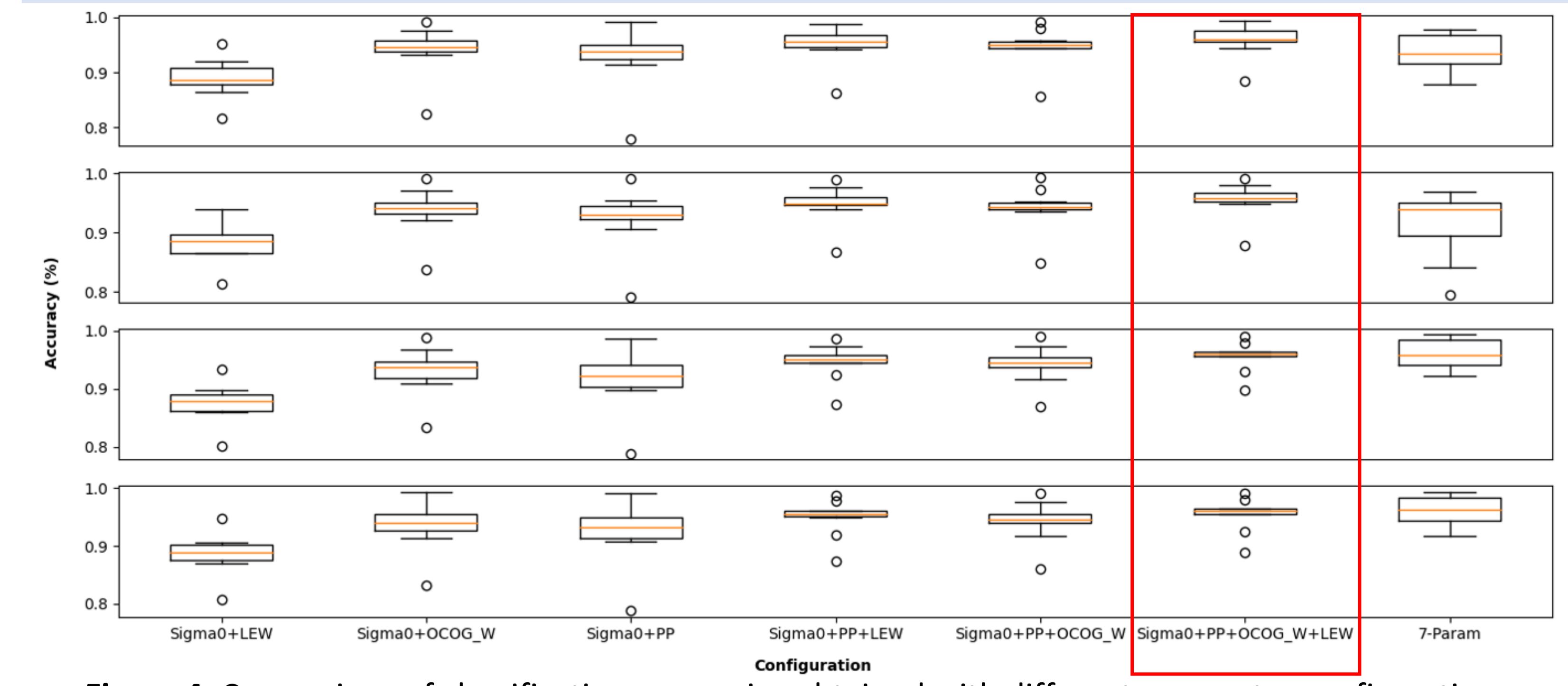


Figure 4. Comparison of classification accuracies obtained with different parameter configurations on 10-fold cross-validation method across all classifiers

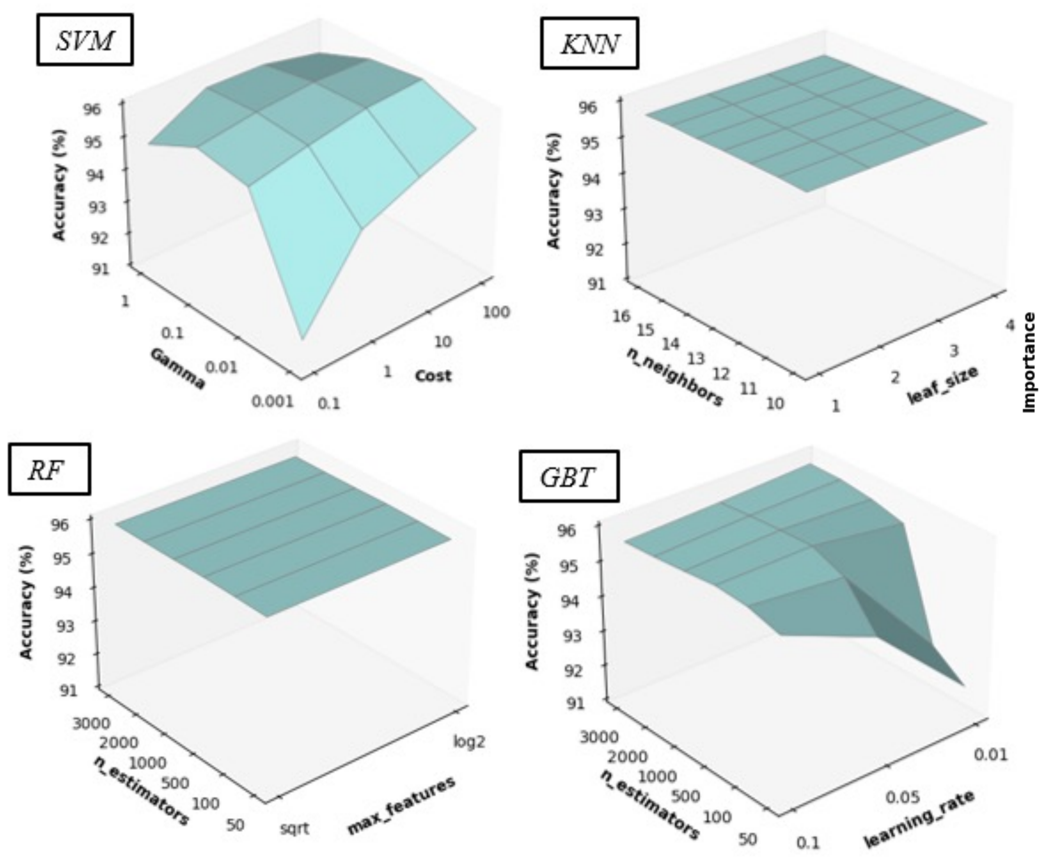


Figure 5. Comparison of classification accuracies with different hyperparameter values

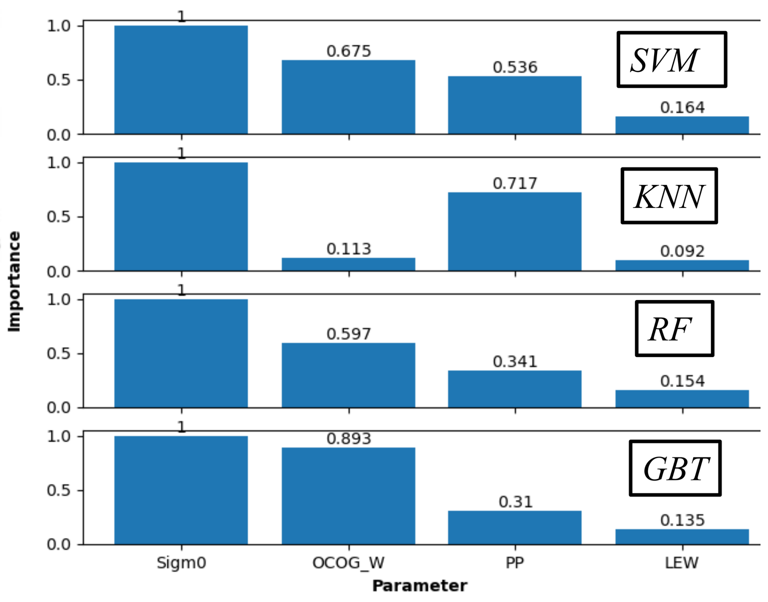


Figure 6. Permutation-based feature importance

Clusters	SVM (%)	KNN (%)	RF (%)	GBT (%)
Ath	94.88	94.5	95.08	95.09
Bai	97.86	97.74	97.74	97.9
GBL	97.85	97.95	97.51	97.31
GLs	95.59	92.51	91.08	92.59
GSL	94.23	94.06	94.87	94.61
OL	99.02	98.59	98.47	98.60
Van	86.66	84.9	83.58	83.87
Win	92.48	89.98	92.51	91.87
Mean accuracy	**94.82**	**93.78**	**93.86**	**93.98**

Table 1. Spatial cross-validation accuracy of the lake clusters across all classifiers (Yellow rectangles in Figure 1 represent the lake clusters considered in the study)

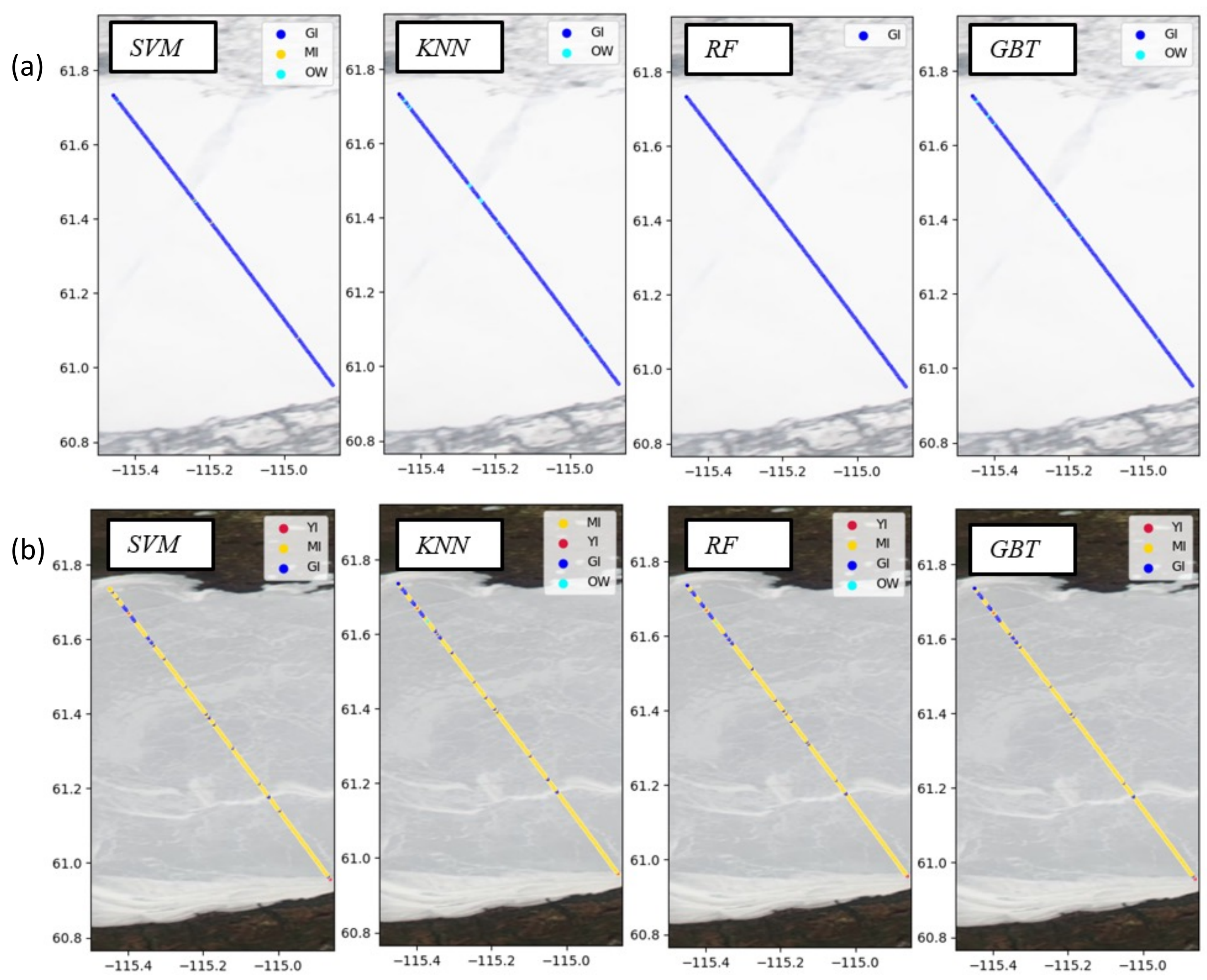


Figure 7. Prediction results of the four ML classifiers for (a) 23 March 2017 and (b) 16 May 2017

CONCLUSIONS

- All four classifiers provide comparable results in classifying lake ice and open water. However, RF and KNN are found to be a better fit for global lake ice mapping as they are less sensitive to internal hyperparameters and have faster processing speeds compared to SVM and GBT.
- Sigma0, OCOG_W and PP are the most important waveform parameters.
- Sigma0+OCOG_W+PP+LEW is the optimal parameter configuration that provides the best classification performance in all classifiers.

References

- Duguay, C. R., Flato, G. M., Jeffries, M. O., Ménard, P., Morris, K., & Rouse, W. R. (2003). Ice-cover variability on shallow lakes at high latitudes: model simulations and observations. *Hydrological Processes*, *17*(17), 3465-3483.
- Ricker, R., Hendricks, S., Helm, V., Skourup, H., & Davidson, M. (2014). Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation. *The Cryosphere*, *8*(4), 1607-1622.
- Wingham, D. J., Rapley, C. G., & Griffiths, H. (1986, September). New techniques in satellite altimeter tracking systems. In *Proceedings of IGARSS* (Vol. 86, pp. 1339-1344).
- Rinne, E., & Similä, M. (2016). Utilisation of CryoSat-2 SAR altimeter in operational ice charting. *The Cryosphere*, *10*(1), 121-131.