GPS Solutions (2022) 26:117
https://doi.org/10.1007/s10291-022-01307-2

ORIGINAL ARTICLE

Machine learning-based snow depth retrieval using GNSS signal-to-noise ratio data

Cemali Altuntas[1] · Muzaffer Can Iban[2] · Erman Şentürk[3] · Utkan Mustafa Durdag[4] · Nursu Tunalioglu[1]

Received: 11 November 2021 / Accepted: 21 July 2022

Abstract

GNSS-IR enables the extraction of environmental parameters such as snow depth by analyzing signal-to-noise ratio, indicating the strength of the GNSS signal. We propose a machine learning (ML) classification approach for snow depth retrieval using the GNSS-IR technique. ML classifier algorithms were studied to classify the strong and weak ground reflections using input parameters (azimuth angle, satellite elevation angle, day of year, amplitude of reflected signal, epoch number, etc.) as independent variables. GPS data collected by UNAVCO AB39 and daily snow depth data from SNOTEL Fort Yukon for a 6-year period (2015–2020) were considered. The first 4-year data were trained by some well-known ML classifiers to weight the input data and then used to classify the strong and weak signals. Tree-based classifiers, Random Forest, AdaBoost, and Gradient Boosting overperformed the other classifiers since they have more than 70% accuracy, so we performed our analysis with these three methods. The last 2-year data were used to validate both trained models and snow depth retrievals. The results show that ML classifier algorithms perform better results than traditional GNSS-IR snow depth retrieval; they improve the correlations by up to 19%. Moreover, the root-mean-square errors decrease from 15.4 to 4.5 cm. This study has a novel approach to the use of ML techniques in GNSS-IR signal classification, and the proposed methods provide a critical improvement in accuracy compared to the traditional method.

Keywords Classification · GNSS-IR · Machine learning · Snow depth retrieval · SNR

Introduction

GNSS satellites continuously transmit electromagnetic signals, and these carrier signals can reach GNSS receivers in line-of-sight (LOS) or by reflection from one or more reflective surfaces. Direct and reflected signals interfere at the antenna phase center (APC) of the receiver, which has a characteristic effect on both the signal-to-noise ratio (SNR) data and the geometry-free linear combination of carrier phase observations (L4) (Ozeki and Heki 2012). By analyzing the effects of the reflected signals on the SNR or L4 data, some information related to the radiometric or geometric properties of the reflection surface can be obtained. This technique is referred to as GNSS Interferometric Reflectometry (GNSS-IR). Compared to terrestrial, airborne, and spaceborne methods, the GNSS-IR has several advantages, such as a more optimum temporal and spatial resolution, more affordable instrument costs, and a global network of continuously operating reference stations (Larson et al. 2009).

Snow depth retrieval based on SNR data from geodetic grade GNSS receiver in a standard orientation was first revealed by Larson et al. (2009) and strong correlations between retrievals and in situ measurements were reported. After that, several researchers implemented SNR-based GNSS-IR studies. Gutmann et al. (2012) validate GNSS-IR snow depth retrievals depending on SNR L2 frequency with manual and laser ranging snow depth measures and airborne LIDAR surveys using 8-month observation data. During the snow season, GNSS-IR retrievals yielded an RMSE of 13 cm as noted 10 cm of bias compared to laser data.

✉ Cemali Altuntas
cemali@yildiz.edu.tr

[1] Department of Geomatic Engineering, Yildiz Technical University, Istanbul, Turkey

[2] Department of Geomatics Engineering, Mersin University, Mersin, Turkey

[3] Department of Geomatics Engineering, Kocaeli University, Kocaeli, Turkey

[4] Department of Geomatic Engineering, Artvin Coruh University, Artvin, Turkey

Published online: 08 August 2022

Larson and Nievinski (2013) expand their studies to determine available GPS sites from Plate Boundary Observatory (PBO) network for snow depth retrievals. Moreover, Hefty and Gerhatova (2014) analyzed both SNR data of L1 and L2 signals and L4 and reported that compared with the manual measurements, snow depth retrievals provide consistency better than 5 cm, although in some period biases reach to 10 cm. To assess the performance and feasibility of SNR L1 data on snow depth retrieval studies, Larson and Small (2016) analyzed L1 SNR data provided from 23 sites for 5-year period and compared retrieval results with SNR L2C results and in situ measurements. SNR L1 results compared to SNR L2C results provide a mean bias of 1 cm with a correlation of 0.95. Apart from the studies using single signal frequency or dual-frequency combination, Yu et al. (2015) proposed an approach using SNR combination of GPS triple frequency (i.e., L1, L2 and L5) signals. They defined the relation between reflector height change and spectral peak frequency and found improvement in results compared to current studies. Then, Zhou et al. (2019) proposed a method to improve the results using SNR combination of GPS triple frequency signals by modeling to reduce the random errors. Results from two GPS sites show strong agreement with in situ measurements with correlations of 99% and 97%.

In recent years, machine learning (ML)-based retrievals have been conducted by several researchers. More accurate results can be obtained by ML, as they are data-based and able to build robust modeling of the relation between input and output data (Chu et al. 2020). Studies in the literature especially focus on ML-based retrievals with GNSS-reflectometry (GNSS-R) method. Liu et al. (2019) applied an ocean surface wind speed estimation with GNSS reflections and trained delay-Doppler Map observables using a multi-hidden layer neural network. According to the results, ML-based retrieval brings a better solution than the conventional wind speed retrieval method. Wang et al. (2020) conducted a study for snow depth retrieval based on deep learning methods by combining GNSS-R estimates, in situ data, and satellite observations with 25 GNSS-R stations in Alaska from 2008 to 2017. Jia et al. (2020) implied an ML approach for solving the uncertainty of physical characteristics of sites, complexity, and nonlinearity of the inversion process in the retrieval of the soil moisture content with GNSS-R data acquired from airborne and in situ observations. This study proves that ML-based soil moisture retrieval from GNSS-R data gives effective and encouraging solutions. Zhan et al. (2022) proposed a snow depth retrieval approach by combining existing satellite data using back propagation neural network (BPNN).

Excluding the research by Wang et al. (2020) and Zhan et al. (2022), ML-based studies using reflected GNSS signals have focused specifically on analyzing or fusing spaceborne and airborne data. However, Wang et al. (2020) used snow depth products provided by the PBO H2O ground-based GNSS-R network (Larson and Nievinski 2013) as true values and combined them with in situ data to increase the station density of the sample data used for deep learning-based snow depth retrieval. Here, they used the deep learning approach not for the analysis of GNSS signals, but for the combination of GNSS-R products with other data. Similarly, Zhan et al. (2022) used BPNN for the fusion of snow depths obtained with data from different satellites.

The ground-based GNSS-R has some challenges related to the characteristics of the station such as the effect of the site environment on the signals. Considering that a site-specific assessment is required, as the multipath is very dependent on the characteristics of the receiver and the near environment, choosing the appropriate variables such as elevation and azimuth angle limits, and minimum number of epochs is very crucial but complicated in the traditional GNSS-IR analysis procedure. Due to the nonlinearity of the relationship between these variables and results, strong and weak reflective signals (hereafter referred to as 'good' and 'bad', respectively) are difficult to classify by taking all these variables into account. Considering its ability to easily identify trends and patterns in multidimensional and multivariate data, it may be possible for ML algorithms to classify SNR signal sections, i.e., GNSS-IR signals, as good and bad. With this motivation, we developed an ML-based signal classification approach using eight variables, which are related with the properties of GNSS-IR signals, are used as inputs. First, we examined the training performance of eight ML algorithms using 4-year (2015–2018) L1 SNR data from AB39 station in the UNAVCO network. Three ML algorithms, Random Forest (RF), AdaBoost (AB), and Gradient Boosting (GB) outperformed the others. Hence, we validated these algorithms with 2019 and 2020 data. Moreover, we compared GNSS-IR snow depths estimated by signals classified as 'good' by these ML algorithms for the same years, with both traditional GNSS-IR results and SNOTEL snow depth measurements. The results show that ML-based classification approach significantly improves GNSS-IR snow depth estimations.

First, the SNR-based GNSS-IR snow depth retrieving procedure and the inputs used in the ML-based classification approach are given. The study area and data are described in the following section. Then, traditional and proposed approaches are represented, and their results are discussed in terms of classification and snow depth estimation performance. Finally, the study is concluded.

Snow depth retrieval with GNSS-IR

The GNSS SNR data can be expressed with the following equation (Bilich et al. 2008; Larson and Nievinski 2013).

$$\mathrm{SNR}^2 \equiv A_c^2 = A_d^2 + A_m^2 + 2A_dA_m \cos\Delta\varphi \quad (1)$$

where A_c is the amplitude of the composite signal, A_d and A_m are the amplitudes of direct and reflected signals, respectively, and $\Delta\varphi$ is the relative multipath phase. The multipath is dominant at low elevation angles, and as GNSS satellites move on the sky, SNR is expressed as a function of satellite elevation angle. Since $A_d \gg A_m$, the contribution of the direct signal on SNR data can be eliminated by a low order polynomial. After eliminating the direct signal contribution over SNR, the remaining, namely detrended SNR (dSNR), can be formulated as

$$\mathrm{dSNR} = A_m \cos\left(\frac{4\pi h}{\lambda}\sin\varepsilon + \phi\right) \quad (2)$$

where A_m and ϕ are the amplitude and phase offset of the reflected signal, respectively; λ is the wavelength; h is the reflector height; and ε is the elevation angle. The h is related to the dominant frequency of dSNR data, which can be expressed as $f = 2h/\lambda$. By using the sine of the elevation angle as the independent variable, the dominant frequency can be estimated by Lomb–Scargle periodogram (LSP) analysis. Once f is obtained, the reflector height can be computed from $h = f\lambda/2$. Then, snow depth can be retrieved using reflector heights.

In the SNR-based GNSS-IR analysis, several variables about the GNSS signals are considered. Since multipath is dominant at low elevation angles, fluctuations on SNR data can be generally tracked in low angle ranges such as 5–25, 5–30, 0–25, 0–30 degrees. The other angle mask that should be implemented to data is the azimuth angle, which shows the direction of the receiving signal. If the information about the site environment is available, azimuth ranges that are free of obstructions (e.g., buildings, trees) in signal transmission with the satellite and facing a reflection area of interest (sea, ice, soil, vegetation, etc.) are preferred. It is usually reasonable to manually select azimuth mask when there is direct access to the station or when there are auxiliary materials, such as photographs, satellite imagery, and digital elevation model, of the station. However, the selection of the appropriate azimuth mask is also problematic, as these materials are unavailable for some stations, or they are not sufficient to use. Furthermore, significant structural changes may occur in the site environment. The other important variable is amplitude. The amplitude of the multipath signal is related to the antenna gain pattern, roughness, and dielectric constant of the reflecting surface (Larson et al. 2009). Additionally, day-of-year (DoY) is used to determine the snowless and snowy days (Larson and Nievinski 2013). For daily estimations, the first epoch of SNR indicates the related day to be involved in the analysis (Larson and Nievinski 2013). In addition, reporting all satellites as ascending and descending arc segments (Bilich et al. 2008) enables to achieve availability of the satellites. In each arc segment, enough epochs should be reported to compute the LSP with desired spectral resolution (Larson and Nievinski 2013). Another important issue is the comparison value, called peak-to-background noise (PBN) ratio, which is a criterion used in many studies. It can be defined as the ratio of the peak amplitude in the periodogram to the background noise. The choice of PBN ratio is rather arbitrary, i.e., 2 in Löfgren et al. (2014), 2.5 in Larson et al. (2013), 3 in Wang et al. (2019), and 4 in Li et al. (2021). Although the signals meet the PBN condition, they may cause incorrect estimations compared to in situ measurements. Therefore, it cannot be said that a signal classification based on the PBN is successful in all conditions.

Study area and data description

The study area covers the AB39 GPS station ($\varphi = 66.55935°$N; $\lambda = 145.21263°$W; $h = 147.693$ m) located in the state of Alaska, USA, operated by UNAVCO and its

Fig. 1 GPS and SNOTEL sites: AB39—south view (top), Fort Yukon (bottom)

surroundings. GPS data collected by AB39 between 2015 and 2020 were evaluated. To verify the GNSS-IR snow depth estimates, in situ snow depth measurements from the Fort Yukon climate station ($\varphi = 66.57050°$N; $\lambda = 145.24553°$W; $h = 131.064$ m) in the SNOTEL network were used. Figure 1 shows the GPS and SNOTEL sites (https://www.unavco.org/ and https://wcc.sc.egov.usda.gov/).

First, LSP analysis of S1 data from 5 to 25 degrees elevation angles was performed. No azimuth mask or PBN condition was applied. GNSS-IR Analysis Software (GIRAS), which is open-source MATLAB-based software, was utilized to estimate the frequency, amplitude, and phase of GNSS-IR signals (Altuntas and Tunalioglu 2022). The frequencies were converted to reflector height

Table 1 Training and validation data sets derived from AB39

	Years	GC range (cm)	Number of GC samples	Number of BC samples	Total number of samples
Input data sets to train the classifier (3 sets)	2015–2018	0–10	40,740	115,163	155,903
		0–15	55,503	100,400	155,903
		0–20	67,112	88,791	155,903
Validation data sets to assess and compare the predictive performance of the trained classifiers (6 sets)	2019	0–10	11,037	27,663	38,700
	2020	0–10	11,151	27,526	38,677
	2019	0–15	15,074	23,626	38,700
	2020	0–15	15,358	23,319	38,677
	2019	0–20	18,348	20,352	38,700
	2020	0–20	18,764	19,913	38,677

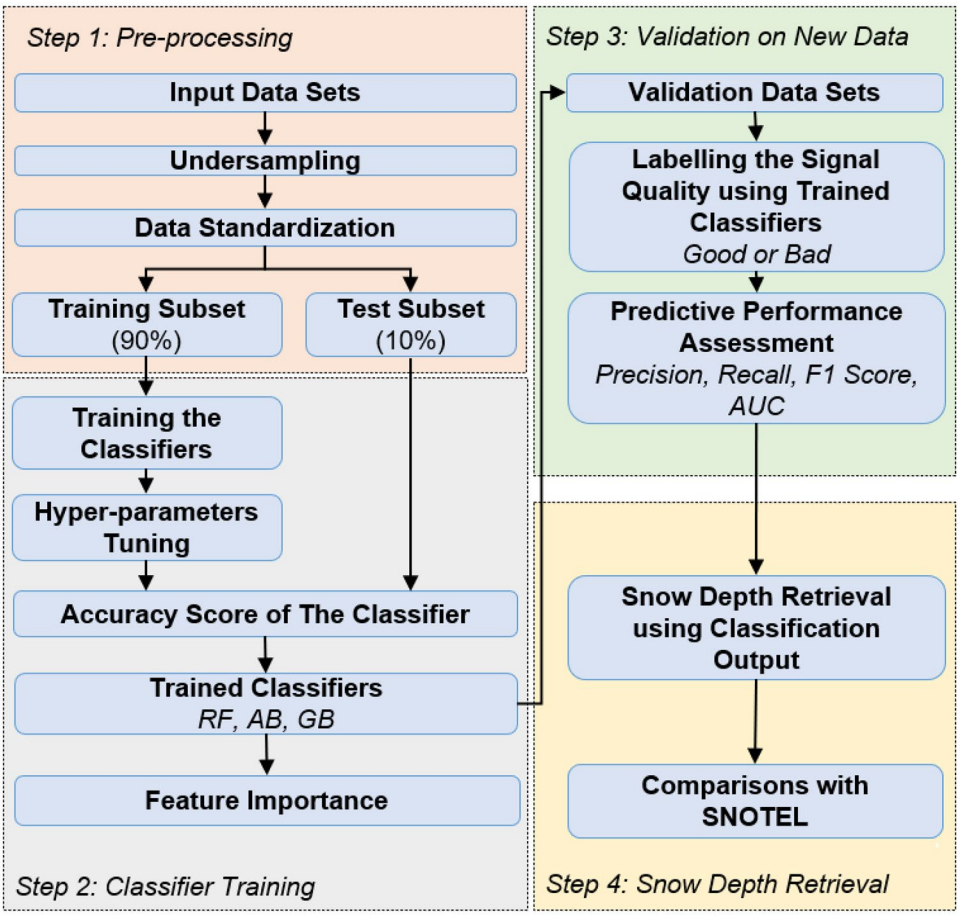


Fig. 2 Flowchart of the proposed ML-based work. Data pre-processing involves data preparation, undersampling for class balance, data standardization, and train-test splitting. Classifier training helps obtain trainer models with tuned hyperparameters. The validation step entails performing new classification tasks on new data sets. The snow depth retrieval step involves the procedure of obtaining GNSS-IR snow depths using classification outputs and comparing them with SNOTEL measurements

and then snow depth. Absolute differences between each GNSS-IR snow depth estimate and the SNOTEL measurement were calculated. These values were used to classify GNSS-IR signals as 'good' or 'bad'. Considering that the snow depth estimations obtained with GPS SNR data during the snow season may have biases of 10 cm (Gutmann et al. 2012), and additionally, there is an overall bias of 8.5 cm between the snow depths obtained from the AB39 and Fort Yukon station data, three different ranges for the 'good' class (GC) were selected: (1) 0–10 cm; (2) 0–15 cm; and (3) 0–20 cm. Signals whose estimates had absolute differences outside of these ranges were classified as 'bad' (BC). Finally, three alternative input data sets using three ranges were prepared for classification purposes. We used a collection of signals from AB39 covering the period of 2015–2018 to train our ML classifiers. Since we selected three ranges for classifying the GC and BC, three different model training were performed. The trained models were separately validated with signals covering the years 2019 and 2020 and compared to each other. Ultimately, there are six different validation data sets with three different ranges covering two different years. Table 1 gives the details about the input data sets to train the classifiers and validation data sets to assess and compare the predictive performance of the trained classifiers.

The signal quality (BC or GC) is the target variable for the classification task. Since we have only 'good' or 'bad' classes for the signal quality, the task becomes a bi-class classification. The independent variables are the satellite ID, first epoch, number of epochs, minimum elevation angle, elevation angle range, mean azimuth angle, day of year, and peak amplitude of the periodogram. Once the classifier models are trained using the input data sets, the trained classifiers are used to predict the output signal quality labels on validation data. In the end, these predicted signal quality classes are compared with the results obtained from the traditional GNSS-IR estimation. The flowchart of our ML-based work is demonstrated in Fig. 2.

Methods and materials

In this section, the analyzing strategies of both approaches, traditional GNSS-IR and ML, were introduced. First, criterions selected depending on azimuth and satellite elevation angles and PBN for the Traditional Approach (TA) were briefly explained. Then, ML approach was discussed after assessments of the classification, test, and validation stages.

Traditional GNSS-IR approach

Azimuth and elevation angle masks are of great importance in the analysis of GNSS SNR data. Generally, signals from low satellite elevation angles, where the multipath is intense, are evaluated. The azimuth mask is applied to consider only reflections from surfaces of interest, excluding angles, where reflections from the ground may be irregular. Another important criterion in the TA is the PBN ratio of the LSP. This ratio is used to distinguish data with strong reflection. Since some azimuth ranges cannot be used due to natural or manmade obstacles (Larson and Small 2016), in the implementation of the TA, the azimuth ranges were selected as 0–80 and 95–360 degrees by interpreting the photos and satellite images of the AB39 GPS site. Moreover, the satellite elevation angle range was selected as 5–25 degrees by following Larson et al. (2009). SNR data sections with a satellite elevation angle range of less than 5 degrees

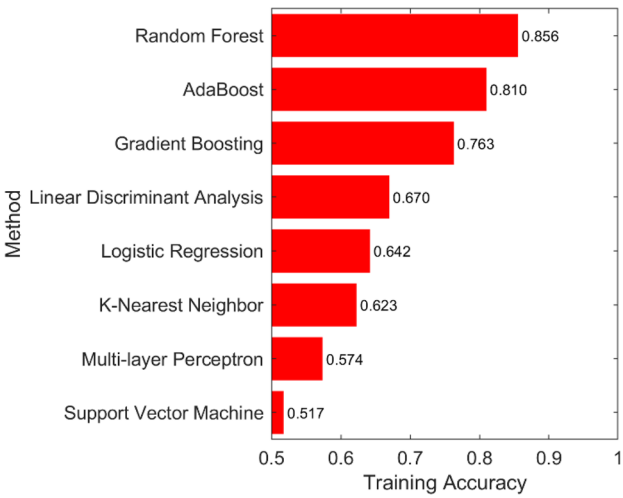


Fig. 3 Initial mean training accuracies of tested ML classification algorithms for 4-year period (70% of bias considered for performance assessment)

Table 2 Best combination of hyperparameter values derived from GridSearchCV

Algorithm	Hyperparameter	Value
RF	n_estimators	300
	max_features	3
	max_depth	None
	criterion	Gini
GB	learning_rate	0.1
	n_estimators	300
	max_features	0.3
	criterion	Mean squared error
AB	random_state	7
	learning_rate	0.001
	n_estimators	100
	algorithm	SAMME

were excluded from the evaluation. The minimum PBN ratio was selected as 3.

Machine learning approach

Step 1—pre-processing: ML algorithms tend to work well on data sets with equally distributed classes. In most cases, there is an imbalance problem in class distribution, and the raw data needs to be resampled (Japkowicz and Stephen 2002). We implemented a random undersampling process on our three alternative input data sets. As seen in Table 1, while the raw data sets consist of more BC samples, the undersampled data sets have the same number of samples for both classes after randomly omitting bad signal samples. Furthermore, to assess the training accuracy of any ML model, typically, the input data set is split into two sub-sets as training and test sets. We preferred to use 90% of the samples for stronger model training and 10% of the samples for testing the accuracy of the model.

Step 2—classifier training: We implemented a quick classification scheme using multiple algorithms at once to evaluate the initial fitting behavior of each selected algorithm. We tested support vector machine, adaptive boosting (AdaBoost, AB), random forest (RF), gradient boosting (GB) classifiers, K-nearest neighbors, logistic regression, and linear discriminant analysis. In three

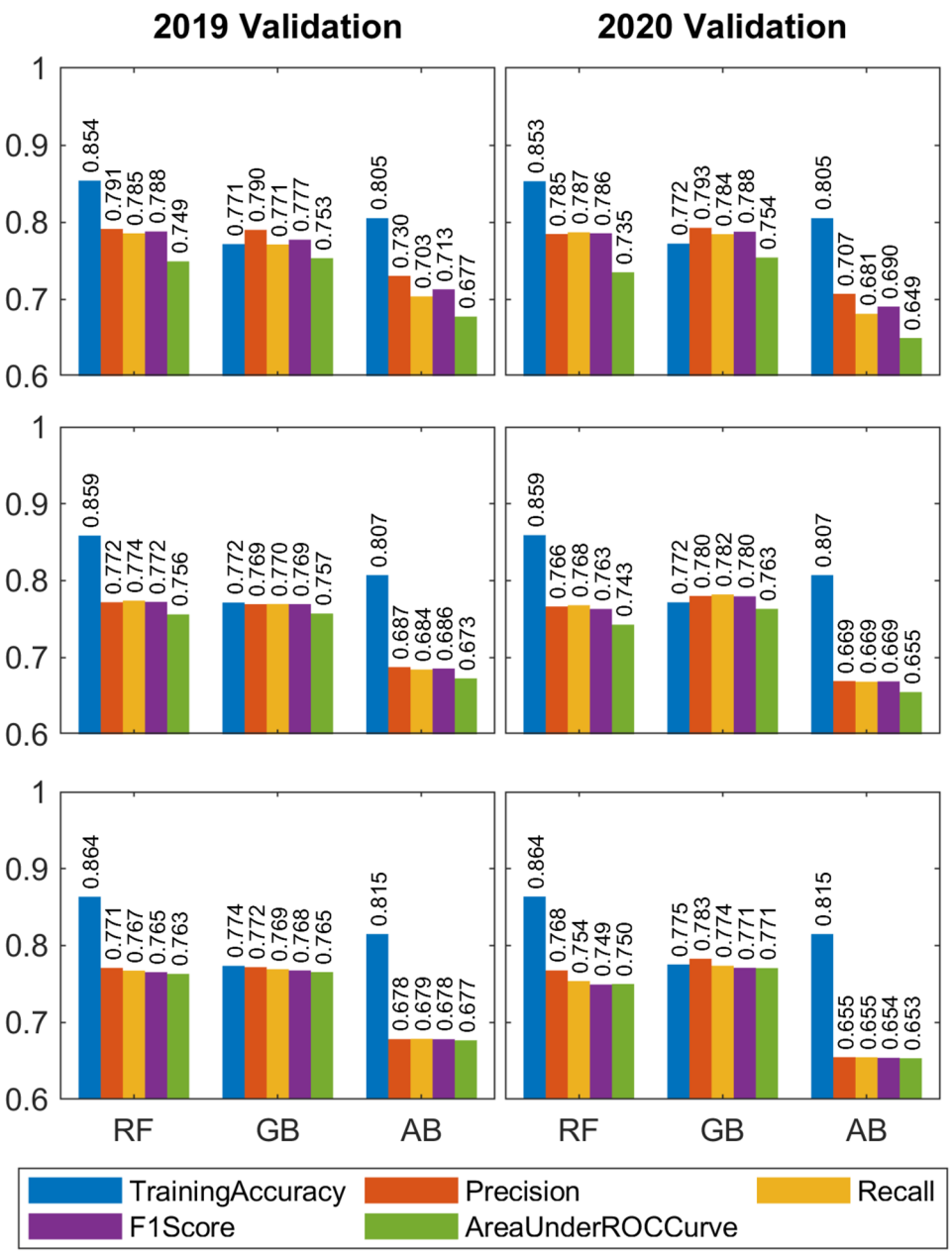


Fig. 4 Statistical performance comparisons of validation results of selected ML algorithms for different GC ranges: 0–10 cm (top), 0–15 cm (middle), 0–20 cm (bottom)

alternative data sets, RF, AB, and GB classifiers showed more than 70% mean accuracy (Fig. 3). Therefore, we decided to proceed with the classification task with these three classifier algorithms. These three selected algorithms are examples of tree-based ensemble methods that produce a series of classifiers and then classify all data samples by taking a weighted score of their iterative class predictions (Dietterich 2000). As aforementioned, we collected data created from a mix of observations covering the period between 2015 and 2018 to train our classifiers. Once the classifier is trained and its accuracy is assessed, this classifier can be used for predicting the new unknown values. Hence, we made predictions on observations in 2019 and 2020 separately to discover which observations belong to the GC or BC.

RF, developed by Breiman (2001), is a variation of decision trees, and it generalizes the model by training different combinations of the data set on different features. Therefore, RF is found to be very robust to overfitting. The AB classifier algorithm was developed by Freund and Shapire (1997). AdaBoost, short for "Adaptive Boosting", boosts the classification process by combining a group of decision trees known as weak classifiers to form a stronger one. After each round, the samples are reweighted to detect misclassified samples in the previous round. This sequence stops at a predefined threshold, and a weighted group of weak classifiers forms a stronger classifier. GB is another supervised ML algorithm developed by Friedman (2001). Both AB and GB classifiers are based on the concept of producing multiple classifiers to average their performance to find out the best one. AB tries to up-weight misclassified samples from the previous rounds, while GB tries to detect the samples with larger residuals in each round.

Hyperparameters are the non-constant variables which should be tuned before building the model for the best prediction results (Schratz et al. 2019). We tuned the hyperparameters of three selected classifiers by implementing a GridSearchCV method. It utilizes different combinations of selected hyperparameters, computes the model with each combination, and illustrates the best one (Pedregosa et al. 2011). Table 2 shows the values of tuned hyperparameters.

The feature importance analysis explains the contribution of each independent variable to the inner performance of the classification (Grömping 2009). To do so, we implemented Permutation Feature Importance (PFI), which was introduced by Breiman (2001) for tree-based models. The model accuracy in PFI analysis here is measured only with R^2 for test data.

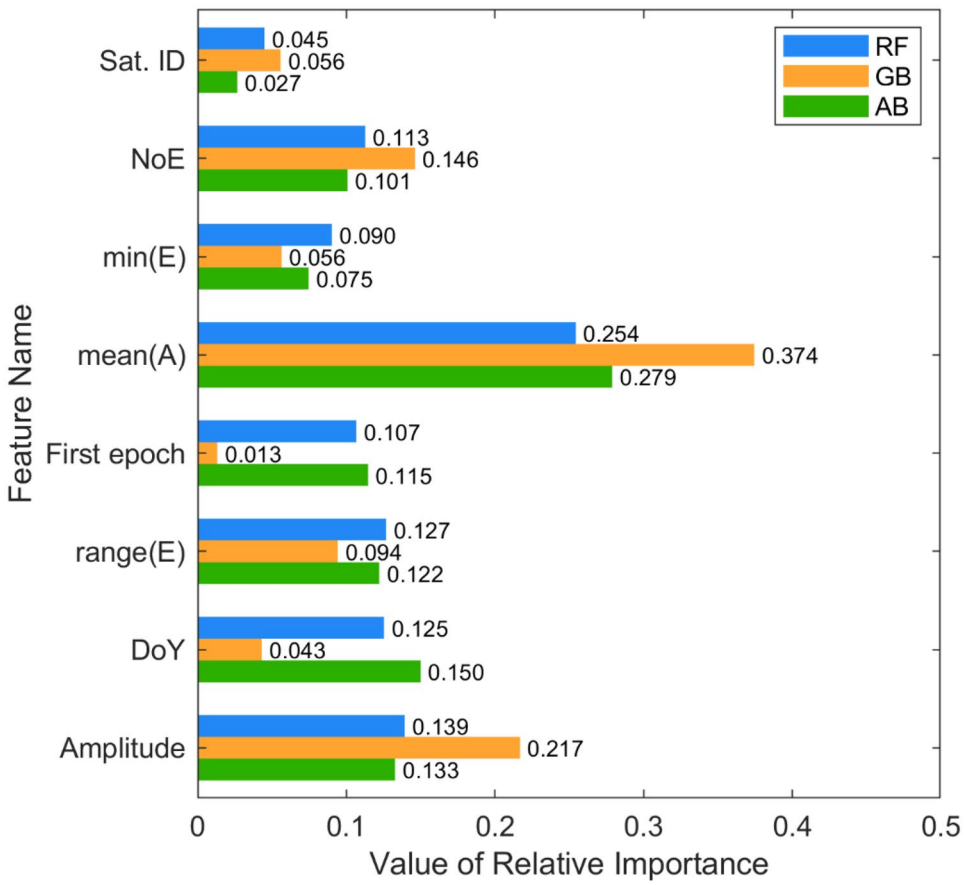


Fig. 5 Feature importance for selected ML algorithms. Sat. ID, NoE, min (E), mean (A), range (E), and DoY indicate the satellite ID, number of epochs, minimum satellite elevation angle, mean satellite azimuth angle, satellite elevation angle range, and day of year of the GNSS-IR signal, respectively

To assess the performance of each ML algorithm on our training data, an accuracy score is utilized, as formulized below:

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1\left(\hat{y}_i = y_i\right) \tag{3}$$

where y_i is the true class value and $\hat{y}_i$ is the predicted class value of any sample.

Step 3—validation on new data: After the classifier is trained, new predictions for observations in 2019 and 2020 individually are to be done to determine the classes of the signal quality. To assess the prediction performance of proposed classification algorithms on validation data, we used four other performance metrics: F1-score, area under ROC curve, precision, and recall. To calculate these metrics, a confusion matrix, which presents the core for assessing the performance, is required. The confusion matrix consists of four diagonal elements: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). The precision, recall, and F1-score are computed as follows:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \tag{4}$$

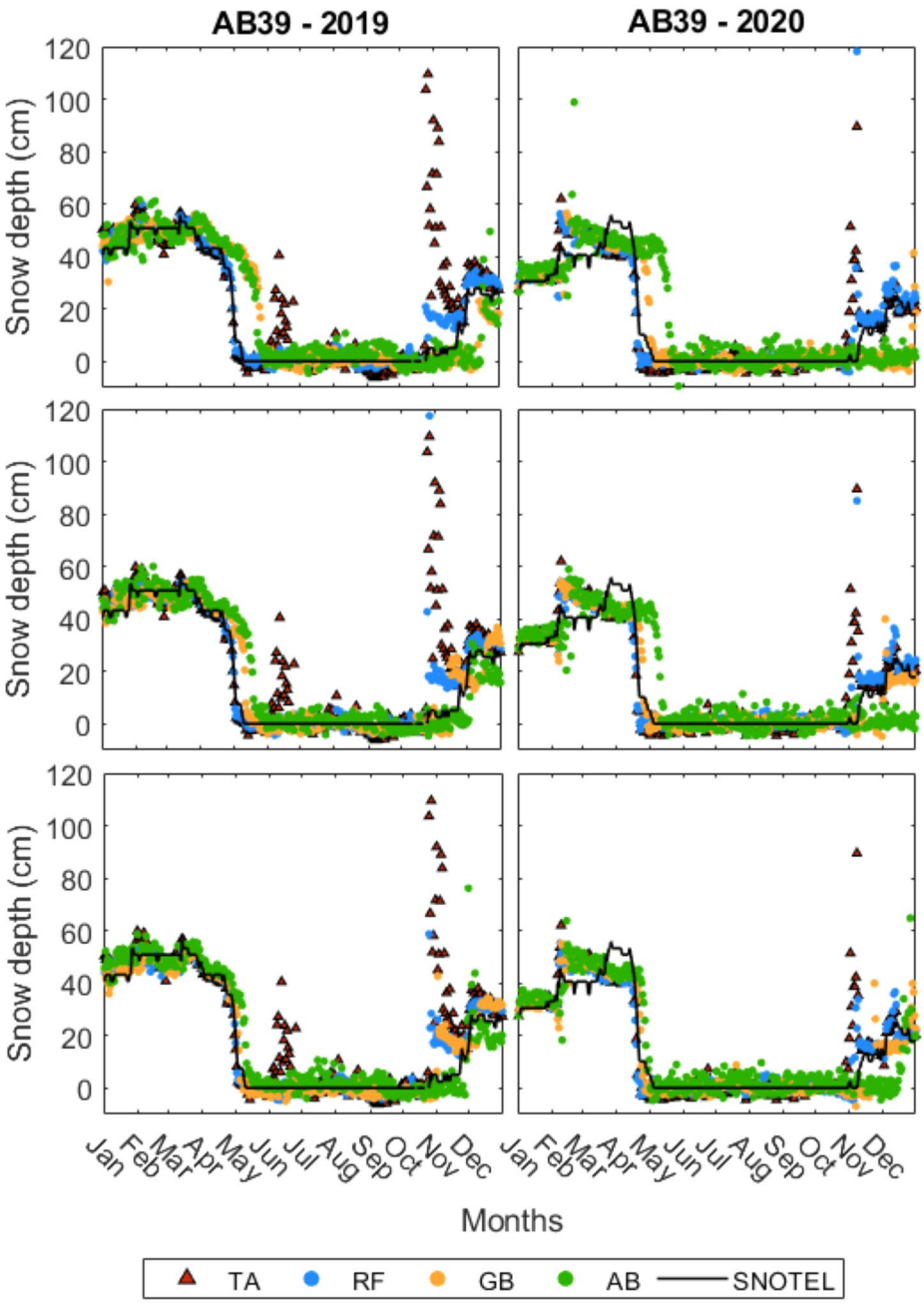


Fig. 6 Snow depth retrievals obtained using AB39 GPS L1 SNR data for the years 2019–2020 for different GC ranges: 0–10 cm (top), 0–15 cm (middle), 0–20 cm (bottom)

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \tag{5}$$

$$\text{F1-Score} = 2/\left(\frac{1}{\text{precision}} + \frac{1}{\text{recall}}\right) \tag{6}$$

The last metric we used is the area under ROC curves (AUC). AUC is a final measure to see the capability of our classifiers to distinguish between two classes.

Step 4—snow depth retrieval: Frequency values of GNSS-IR signals were converted to snow depth values. Daily median values of GNSS-IR snow depths estimated by signals labeled as 'good' by ML classifiers were taken as daily GNSS-IR snow depths. Finally, GNSS-IR snow depths obtained with different approaches were compared with in situ measurements in terms of correlation and RMSE.

Results and discussion

This section addresses the results in two subsections. First, validation results of ML classifiers, RF, GB, and AB, were expressed according to three signal quality classes. Then, feature importance values for model training were reported. In the next subsection, daily snow depth retrievals from traditional GNSS-IR and ML-based approaches were compared with in situ measurements. The performance of approaches was discussed with statistical analysis, i.e., correlations, RMSEs and slope values.

Classification results

Figure 4 shows the results of classification tasks using validation data sets from 2019 and 2020. Alternative validation data sets do not affect model training because both data sets with three distinct GC ranges have nearly the same training accuracy (around 0.85). Therefore, we did not notice any effect of the year and GC range selection on model training performance, possibly due to symmetry in the variables of both years. RF outperforms AB and GB for training accuracy. On the other hand, other indicators can be used to analyze the quality of the classification performance. With GB and RF classifiers, precision, recall, and F-1 scores reach around 0.75, whereas AB fluctuates between 0.65 and 0.70. Such precision and recall scores demonstrate the ability of ML classifiers to predict good signals correctly. Similarly, F-1 scores show that ML classifiers are not quite vulnerable to class distribution in validation data sets that are not evenly distributed. The AUC ratings reflect overall accuracy, and it appears that ML classifiers accurately predicted three-quarters of the signal quality classes. Figure 5 shows the feature importance values for model training. According to

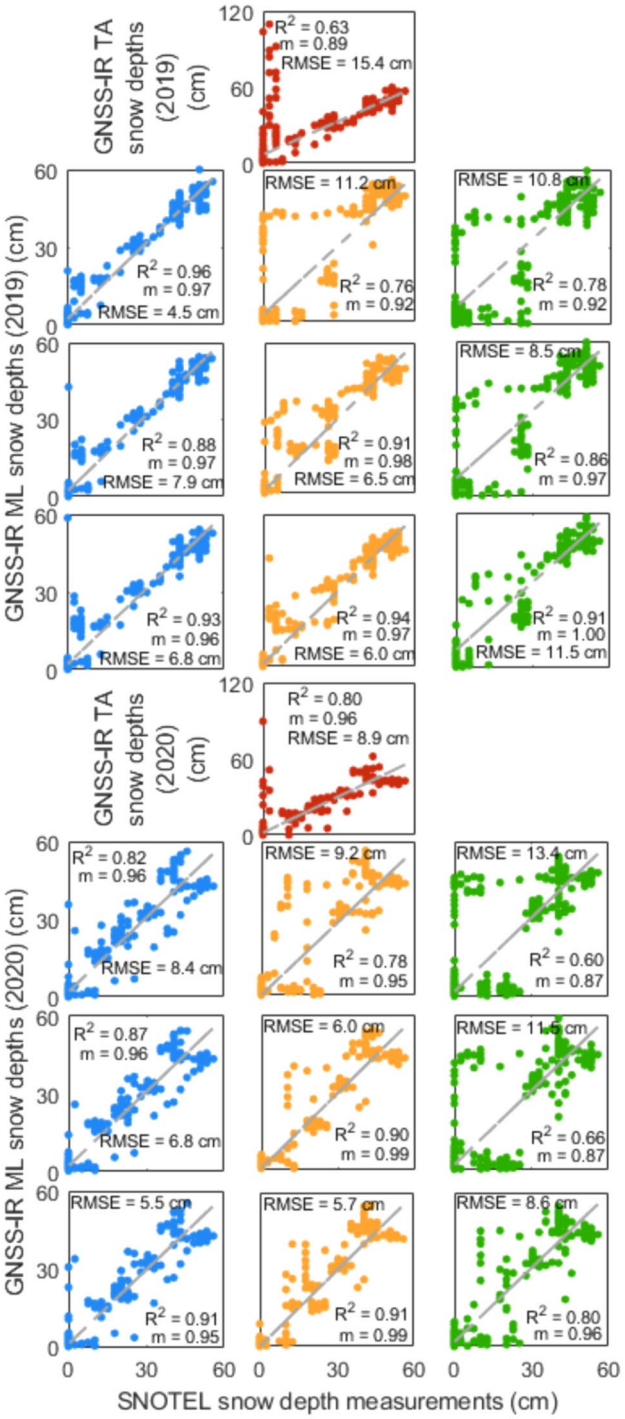


Fig. 7 Comparison of snow depths retrieved with different classification conditions using AB39 GPS data for the years 2019–2020. The figure consists of two subsections showing the 2019 results in the first four rows and the 2020 results in the last four rows. GNSS-IR ML snow depths are shown in both subsections for three GC ranges: 0–10 cm (rows 2 and 6), 0–15 cm (rows 3 and 7), and 0–20 cm (rows 4 and 8). Snow depths from GNSS-IR TA are shown in rows 1 and 5. The red, blue, orange, and green dots represent the results of TA and the RF, GB, and AB algorithms, respectively. The gray dashed lines show the linear regressions

the results, the mean azimuth degree is the most contributing feature class for the training of ML classifiers.

GNSS-IR results

The time series of snow depths obtained with both ML-based approach and TA, and measurements of the SNOTEL site (i.e., in situ values) are shown in Fig. 6. For the GC: 0–10 cm range, ML algorithms (with RF giving the best results) provide more compatible estimations with SNOTEL measurements compared to TA. For the same GC condition, the GB and AB algorithms could not detect the snow melt in May, and the estimates were decreased to around 0 cm with a delay of about one month (Fig. 6, top). These algorithms also failed to yield consistent results with the increase in snow depth in the last two months of 2020 (Fig. 6, top right). However, changing the GC condition from 0–10 cm to 0–20 cm enables the GB and AB algorithms to detect snow melt in May (Fig. 6, left column). Furthermore, this change enables the GB algorithm to detect the snow depth increase at the end of 2020, while the AB algorithm cannot do so (Fig. 6, right column). It can also be seen in Fig. 6 that although there are large errors in the estimations of the TA for June and November 2019 and November 2020, all three ML algorithms provide much better results for the same period. Thus, it can be said that ML algorithms are more effective in excluding individual estimations that worsen daily ones.

Comparisons of GNSS-IR snow depth estimates were obtained with TA and ML algorithms, and SNOTEL measurements, including coefficient of determination (R^2), slope (m), and RMSE values, are shown in Fig. 7. It is seen that a slope close to 1, high correlation, and low RMSE values are obtained with GC: 0–20 cm using the GB and AB algorithms. A general summary of the correlation and RMSE results is given in Table 3. Accordingly, the RF algorithm provided the highest correlation and lowest RMSE values as 98.1% and 4.5 cm, respectively, with 2019 data for GC: 0–10 cm. This represents an improvement of up to 19% in correlation and a decrease of about 11 cm in RMSE compared to the TA. Moreover, it can be said that each of the ML algorithms gives better results than TA in terms of correlation and RMSE when GC: 0–20 cm is selected.

Conclusion

We investigated the feasibility of ML-based snow depth retrieval using GNSS SNR data. The ML-based estimates provided from three classifiers (RF, GB, and AB) are compared with the traditional GNSS-IR approach for 2-year data of one site by validating the results with in situ snow depth observations. The estimation performance of the methods is determined in terms of the correlation coefficient, slope, and RMSE values. Three GC ranges, 0–10 cm, 0–15 cm, and 0–20 cm are used to assess the predictive performance. ML algorithms trained with GC: 0–20 cm produced better results compared to the results when the other two GC ranges were selected. Each ML classifier generally provided more accurate results than the TA, but we can say that the RF and GB methods were somewhat more successful than the AB method.

In the present study, we demonstrated the capabilities of ML algorithms for snow depth retrieval with GNSS-IR. The artificial intelligence approach, which is the general concept of ML, will be one of the cutting-edge topics in the near future for estimating soil moisture, vegetation water content, sea level, sea/lake ice, and snow depth/accumulation, with the GNSS-IR. Here we used the ML approach for signal classification for the first time and

Table 3 Correlations and RMSEs of snow depths estimated by the methods followed with SNOTEL measurements

	Year	TA	GC range (cm)	ML algorithms		
				RF	GB	AB
ρ (%)	2019	79.2	0–10	98.1	87.1	88.1
	2020	89.4		90.4	88.4	77.2
RMSE (cm)	2019	15.4		4.5	11.2	10.8
	2020	8.9		8.4	9.2	13.4
ρ (%)	2019	79.2	0–15	93.7	95.5	92.7
	2020	89.4		93.3	94.8	81.5
RMSE (cm)	2019	15.4		7.9	6.5	8.5
	2020	8.9		6.8	6.0	11.5
ρ (%)	2019	79.2	0–20	96.3	96.8	95.4
	2020	89.4		95.4	95.3	89.7
RMSE (cm)	2019	15.4		5.9	5.4	6.8
	2020	8.9		5.5	5.7	8.6

showed that tree-based ML methods are quite successful compared to the TA. Our next study will tend to deep learning methods that we think may provide comparable estimation results thanks to hidden layers in weight optimization.

Acknowledgements We would like to thank UNAVCO for providing GNSS data and station information. We would also like to the National Water and Climate Center for snow depth data.

Data availability The GNSS data were obtained from the UNAVCO Data Center (https://data.unavco.org/archive/gnss/rinex/obs/) and the snow depth measurements from the National Water and Climate Center (https://wcc.sc.egov.usda.gov/nwcc/site?sitenum=961).

References

Altuntas C, Tunalioglu N (2022) GIRAS: an open-source MATLAB-based software for GNSS-IR analysis. GPS Solut 26(1):1–8

Bilich A, Larson KM, Axelrad P (2008) Modeling GPS phase multipath with SNR: case study from the Salar de Uyuni, Boliva. J Geophys Res Solid Earth 113(B4):B04401

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Chu X, He J, Song H, Qi Y, Sun Y, Bai W, Li W, Wu Q (2020) Multimodal deep learning for heterogeneous GNSS-R data fusion and ocean wind speed retrieval. IEEE J Sel Top Appl Earth Observ Remote Sens 13:5971–5981

Dietterich TG (2000) Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1–15). Springer, Berlin, Heidelberg.

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189–1232.

Grömping U (2009) Variable importance assessment in regression: linear regression versus random forest. Am Stat 63(4):308–319

Gutmann ED, Larson KM, Williams MW, Nievinski FG, Zavorotny V (2012) Snow measurement by GPS interferometric reflectometry: an evaluation at Niwot Ridge, Colorado. Hydrol Process 26(19):2951–2961

Hefty J, Gerhatova LU (2014) Using GPS multipath for snow depth sensing-first experience with data from permanent stations in Slovakia. Acta Geodyn Geomater 11(1):53–63

Japkowicz N, Stephen S (2002) The class imbalance problem: a systematic study. Intell Data Anal 6(5):429–449

Jia Y, Jin S, Savi P, Yan Q, Li W (2020) Modeling and theoretical analysis of GNSS-R soil moisture retrieval based on the random forest and support vector machine learning approach. Remote Sens 12(22):3679

Larson KM, Nievinski FG (2013) GPS snow sensing: results from the EarthScope Plate Boundary Observatory. GPS Solut 17(1):41–52

Larson KM, Small EE (2016) Estimation of snow depth using L1 GPS signal-to-noise ratio data. IEEE J Sel Top Appl Earth Observ Remote Sens 9(10):4802–4808

Larson KM, Löfgren JS, Haas R (2013) Coastal sea level measurements using a single geodetic GPS receiver. Adv Space Res 51(8):1301–1310

Larson KM, Gutmann ED, Zavorotny VU, Braun JJ, Williams MW, Nievinski FG (2009) Can we measure snow depth with GPS receivers?. Geophysical Research Letters, 36(17).

Li Z, Chen P, Zheng N, Liu H (2021) Accuracy analysis of GNSS-IR snow depth inversion algorithms. Adv Space Res 67(4):1317–1332

Liu Y, Collett I, Morton YJ (2019) Application of neural network to GNSS-R wind speed retrieval. IEEE Trans Geosci Remote Sens 57(12):9756–9766

Löfgren JS, Haas R, Scherneck HG (2014) Sea level time series and ocean tide analysis from multipath signals at five GPS sites in different parts of the world. J Geodyn 80:66–80

Ozeki M, Heki K (2012) GPS snow depth meter with geometry-free linear combinations of carrier phases. J Geod 86(3):209–219

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825–2830.

Schratz P, Muenchow J, Iturritxa E, Richter J, Brenning A (2019) Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol Model 406:109–120

Wang X, He X, Zhang Q (2019) Evaluation and combination of quad-constellation multi-GNSS multipath reflectometry applied to sea level retrieval. Remote Sens Environ 231:111229

Wang J, Yuan Q, Shen H, Liu T, Li T, Yue L, Shi X, Zhang L (2020) Estimating snow depth by combining satellite data and ground-based observations over Alaska: a deep learning approach. J Hydrol 585:124828

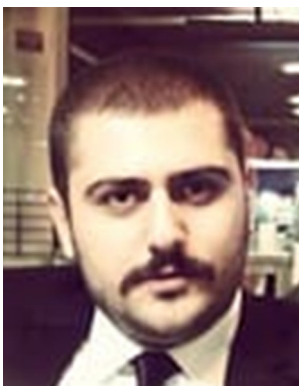
Yu K, Ban W, Zhang X, Yu X (2015) Snow depth estimation based on multipath phase combination of GPS triple-frequency signals. IEEE Trans Geosci Remote Sens 53(9):5100–5109

Zhan J, Zhang R, Tu J, Lv J, Bao X, Xie L, Li S, Zhan R (2022) GNSS-IR snow depth retrieval based on the fusion of multi-satellite SNR data by the BP neural network. Remote Sens 14(6):1395

Zhou W, Liu L, Huang L, Yao Y, Chen J, Li S (2019) A new GPS SNR-based combination approach for land surface snow depth monitoring. Sci Rep 9(1):1–20

Cemali Altuntas received his B.Sc. and M.Sc. degrees in geomatics from Yildiz Technical University, Istanbul, Türkiye. He works as a research assistant in the Department of Geomatic Engineering at the same university. His research is mainly focused on GNSS Interferometric Reflectometry (GNSS-IR) and its implementation.

Muzaffer Can Iban received his B.Sc. degree from Istanbul Technical University, M.Sc. degree from Polytechnic University of Milan, Italy, and Ph.D. degree from Istanbul Okan University in land use and land management. Currently, he works as an assistant professor at the Department of Geomatics Engineering, Mersin University. His research interests involve land use, housing, natural hazards and data science in the geospatial domain.



Erman Şentürk received his B.Sc., M.Sc., and Ph.D. degrees from the Institute of Science under Kocaeli University, Kocaeli, Türkiye, in 2010, 2014, and 2018. He is an Associate Professor at the Department of Geomatics Engineering, Kocaeli University. He is currently working on the effects of natural hazards on the earth's atmosphere and the state-of-the-art approaches of Artificial Intelligence techniques in atmospheric science.

Utkan Mustafa Durdag received his Ph.D. from Yildiz Technical University Istanbul, in the field of reliability of deformation analysis models for geodetic networks. Since 2019, he is employed at Artvin Coruh University, where his current research interests involve the statistical analysis of geodetic measurements.

Nursu Tunalioglu received her Ph.D. degree in geomatic engineering from Yildiz Technical University, Istanbul, Türkiye, in 2011. She is currently working as a professor at the same university. Her research interests include engineering surveys, highway planning, GNSS and high-accurate GNSS positioning, and GNSS-IR.