

Advancing Snow Depth Monitoring with Machine Learning and L-band InSAR Data: A Case Study Using SnowEx 2017 Data

1 Ibrahim Olalekan Alabi^{1*}, Hans-Peter Marshall², Jodi Mead³, Ernesto Trujillo²

2 ¹Department of Computer Science, Boise State University, Boise, ID, USA

3 ²Department of Geoscience, Boise State University, Boise, ID, USA

4 ³Department of Mathematics, Boise State University, Boise, ID, USA

5 * **Correspondence:**

6 Ibrahim Olalekan Alabi

7 ibrahimolalekana@u.boisestate.edu

8 Abstract

9 Current snow depth mapping from space faces challenges in spatial coverage, revisit frequency, and
10 cost. Airborne lidar, although precise, incurs high costs and has limited geographical coverage, thereby
11 necessitating the exploration of alternative, cost-effective methodologies for snow depth estimation.
12 The forthcoming NASA-ISRO Synthetic Aperture Radar (NISAR) mission, with its 12-day global
13 revisit cycle and 1.25 GHz L-band frequency, introduces a promising avenue for cost-effective, large-
14 scale snow depth and snow water equivalent (SWE) estimation using L-band Interferometric SAR
15 (InSAR) capabilities. This study demonstrates InSAR's potential for snow depth estimation via
16 machine learning. Using 3 m resolution L-band InSAR products over Grand Mesa, Colorado, we
17 compared the performance of three machine learning approaches (XGBoost, ExtraTrees, and Neural
18 Networks) across open, vegetated, and the combined (open + vegetated) datasets using Root Mean
19 Square Error (RMSE), Mean Bias Error (MBE), and R^2 metrics. XGBoost emerged as the superior
20 model, with RMSE values of 9.85 cm, 10.46 cm, and 9.88 cm for open, vegetated, and combined
21 regions, respectively. Validation against in-situ snow depth measurements resulted in an RMSE of
22 approximately 16 cm, similar to in-situ validation of the airborne lidar. Our findings indicate that L-
23 band InSAR, with its ability to penetrate clouds and cover extensive areas, coupled with machine
24 learning, holds promise for enhancing snow depth estimation. This approach, especially with the
25 upcoming NISAR launch, may enable high-resolution (~10 m) snow depth mapping over extensive
26 areas, provided suitable training data are available, offering a cost-effective approach for snow
27 monitoring.

28 Keywords: Snow Depth, InSAR, Machine Learning, NISAR, Remote Sensing

29 1 Introduction

30 Accurately measuring snow depth is critical for applications in hydrologic science, water resource
31 management, and climate modeling (Lievens et al., 2022). Seasonal snowpacks act as natural
32 reservoirs, storing winter precipitation and gradually releasing meltwater in spring and summer as the
33 temperature rises (Simpkins, 2018; Livneh and Badger, 2020; Vano, 2020; Lievens et al., 2022).
34 Tracking snow accumulation and melt is instrumental in understanding how snowpack storage affects
35 water resources. For instance, in California, the Sierra Nevada seasonal snowpack contributes
36 approximately 70% of additional water storage to supplement the existing artificial reservoir system

37 (Dettinger and Anderson, 2015; Hedrick et al., 2018; Henn et al., 2020). However, monitoring snow
38 depth, especially on a large scale, is fraught with challenges. The diverse and rugged terrain of
39 mountainous regions makes it particularly difficult (Deems et al., 2006).

40 Traditionally, snow depth monitoring has relied on in-situ measurements and airborne observations
41 (Deems et al., 2013; Marshall et al., 2015; Dong, 2018; Webb et al., 2021). However, these methods
42 are often limited by geographical reach, temporal frequency, and financial resources. In-situ
43 measurements are known for their accuracy and reliability, as they are often conducted manually using
44 instruments such as snow probes, snow coring, or through automated snow telemetry stations
45 (SNOTEL) (Serreze et al., 1999). However, access to mountainous regions can be difficult during
46 winter when snow monitoring is essential. Additionally, the resources required for frequent and
47 widespread in-situ measurements, including financial and human capital, can be prohibitively high
48 (Deems et al., 2013). Moreover, in-situ measurements have limited support (i.e.; representativeness)
49 and cannot fully resolve the spatial heterogeneity of snowpack properties (e.g.; snow depth, snow water
50 equivalent (SWE), and snow density) across a landscape (Trujillo and Lehning, 2015; Bühler et al.,
51 2016), as snow has a spatial autocorrelation length on the order of 100m (e.g.; Trujillo et al. (2009)).

52 Spaceborne remote sensing technologies such as Light Detection and Ranging (lidar) provide high-
53 precision snow depth estimates, but the high cost limits comprehensive monitoring (Deems et al.,
54 2013). Airborne lidar and structure-from-motion (i.e.; photogrammetry using aerial photography)
55 deliver detailed snow depth maps but are geographically limited (Harder et al., 2020; Meyer et al.,
56 2022). Satellite passive microwave sensors estimate basin average SWE yet suffer from coarse
57 resolution (~25 km) and saturate at 150 mm SWE, preventing application in mountainous areas (Chen
58 and Wang, 2018; Taheri and Mohammadian, 2022). Optical remote sensing is weather-dependent and
59 unable to penetrate dense forest cover (Sinha et al., 2015; Aquino et al., 2021). Additionally, while
60 initiatives like the Airborne Snow Observatory (ASO) are making strides towards the broader use of
61 lidar for snow mapping, snow depth data is primarily collected in the western United States (US)
62 (Ferraz et al., 2018), leaving a significant geographical gap in global snow monitoring. Consequently,
63 the development of new spaceborne capabilities to map snow depth continues to be an active area of
64 research.

65 Amid the limitations of traditional snow monitoring methods and the challenges of lidar acquisition on
66 a global scale, the pursuit of alternative remote sensing technologies is imperative. The forthcoming
67 NASA-ISRO (NASA-Indian Space Research Organization) Synthetic Aperture Radar (NISAR)
68 mission, scheduled for launch in October 2024 or January 2025, appears promising in this regard.
69 NISAR, equipped with an L-band radar, is set to observe nearly all of Earth's terrestrial and ice surfaces
70 at an approximate resolution of 10 m, with a revisit frequency of twice every 12 days (Kellogg et al.,
71 2020; LAL et al., 2022). Operating within the 1-2 GHz frequency range (L-band) with wavelengths
72 between 15 to 30 cm, it can penetrate cloud cover and 10+ meters of snow, facilitating all-weather,
73 day-night snow monitoring in a wide range of conditions.

74 Guneriussen et al. (2001) first suggested the use of L-band InSAR for estimating SWE and depth, and
75 several others have demonstrated the approach using PALSAR (Deeb et al., 2011) and tower-based
76 observations (Leinss et al., 2015). During 2017-2021, the NASA SnowEx Mission performed repeat
77 InSAR surveys with UAVSAR, with initial physics-based retrievals showing promise (Marshall et al.,
78 2021). A recent study by Hoppinen et al. (2023) employing repeat pass InSAR for monitoring SWE
79 over Idaho found strong correlations between retrieved SWE changes from SAR images and both in-
80 situ and modeled results. Tarricone et al. (2023) used repeat-pass L-band InSAR to estimate SWE
81 changes in an environment with both snow accumulation and ablation, showcasing the capability of L-

band InSAR for tracking SWE changes. Studies have also shown promise in retrieving snow depth information from InSAR coherence, phase, and incidence angle data. Li et al. (2017) used repeat-pass InSAR measurements for estimating snow depth and SWE in the Northern Piedmont Region of the Tianshan Mountains and found the snow depth estimation to be consistent with field measurements. With the promising capabilities of InSAR, the NISAR mission may be able to provide high-resolution (~10-50 m) snow depth estimates over large areas if appropriate snow depth observations are available for training machine learning models. While physics-based retrievals have shown promise, machine learning techniques can be advantageous as they often require fewer assumptions about snow conditions, such as whether the snow is dry or wet, and can be effective in scenarios where traditional inversion methods may be limited by data availability.

Our goal in this work is to use a pair of repeat-pass L-band InSAR products from the 2017 NASA SnowEx campaign on Grand Mesa to estimate total snow depth (e.g.; bulk snowpack depth to the ground) using Machine Learning algorithms. While previous work has focused on physics-based retrievals, in 2017 we only have one lidar flight measuring total snow depth, and therefore physics-based inversion is not possible. While InSAR is inherently more related to changes in snow depth rather than total snow depth, snow accumulation patterns tend to exhibit consistency. As such, snow depth before the onset of melt displays similar patterns to snow depth changes (Mason, 2020). Snow distribution patterns have been shown to exhibit intra- and inter-seasonal consistency despite differences in weather patterns and seasonal snowfall amounts (Deems et al., 2008; Sturm and Wagner, 2010; Schirmer and Lehning, 2011; Pflug and Lundquist, 2020). While the actual snow depths may change, the locations of deeper and shallower snow areas generally tend to be consistent.

Since the 1990s, machine learning algorithms have gained prominence in environmental remote sensing (Reichstein et al., 2019; Zhang et al., 2019; Yuan et al., 2020) and, over time, have proliferated across various application areas such as snow depth retrieval (Tedesco et al., 2004; Wang et al., 2019; King et al., 2022a; Ofekeze et al., 2022, 2023; Alabi et al., 2023), snow density (Alabi et al., 2022; Feng et al., 2022), and SWE (Bair et al., 2018; Broxton et al., 2019) predictions. Hu et al. (2021) conducted a study using machine learning algorithms to fuse gridded snow depth datasets with inputs including geolocation, topography, and in-situ observations. The Random Forest algorithm proved to be the most proficient of the three learning machines tested. Liang et al. (2015) applied the Support Vector Machine (SVM) to estimate snow depth in northern Xinjiang using data from visible and infrared surface reflectance, brightness temperature, and auxiliary information. The SVM method outperformed the Artificial Neural Networks (ANN) utilized in Finland (Tedesco et al., 2004). King et al. (2022a) utilized a random forest model trained on vertical radar reflectivity profiles from the VertiX X-band radar instrument in Egbert, Ontario, and atmospheric temperature estimates from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA-5) for snow accumulation predictions resulting in a mean square error (MSE) of $\sim 1.8 \times 10^{-3} \text{ mm}^2$ when compared to collocated in-situ measurements.

These works and others have highlighted the potential of machine learning in producing improved snow depth predictions. However, a common limitation across many of these studies is the constrained spatial coverage of validation datasets, which can be attributed to the scarcity of global snow depth data and the prohibitive costs associated with acquiring lidar data globally. This study aims to develop a snow depth prediction system using L-band InSAR (which can be deployed over large areas) products and ML algorithms. Specifically, we will compare the performance of three machine learning algorithms: eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), Extremely Randomized Trees (ExtraTrees) (Geurts et al., 2006), and Artificial Neural Networks (ANN). We will also investigate the impact of vegetation on snow depth prediction accuracy.

128 The objective of this work is divided into three broad aspects:

- 129 1. to test the effectiveness of L-band InSAR products in estimating total snow depth using ML
130 algorithms,
- 131 2. to analyze the effect of vegetation on the performance of the ML models, and
- 132 3. to understand the relative importance of each input feature in snow depth estimation.

133 The study area is stratified into open areas and vegetated areas, using a 0.5 m threshold on lidar
134 vegetation height observations, to evaluate the impact of vegetation on model performance. Machine
135 learning algorithms can be used to extract snow depth information from L-band InSAR data in a robust
136 and efficient manner.

137 We believe that our research has the potential to complement existing snow monitoring practices by
138 facilitating cost-effective, high-resolution, and extensive snow depth estimation. Our findings could
139 lead to the development of a global snow depth prediction system that provides valuable information
140 for water resource management, flood forecasting, and avalanche hazard assessment, provided that
141 accurate and representative training data is available.

142 **2 Data and Methods**

143 **2.1 Study Area**

144 This study was conducted on Grand Mesa, a flat-topped mesa in Western Colorado, USA (39.1°N,
145 107.9°W). Grand Mesa is one of the largest flat-topped mesas in the world (Chesnutt et al., 2017;
146 Gatebe et al., 2018), with elevations from 2,000 to 3,400 m above sea level (Kulakowski et al., 2004).
147 The climate is continental (Kulakowski et al., 2004), with the coldest, warmest, and windiest months
148 being January, July, and June, respectively (Time and Date website, 2024:
149 <https://www.timeanddate.com/weather/@5423575/climate>).

150 **2.2 Data**

151 The data for our study was sourced from NASA's 2017 SnowEx campaign (Kim et al., 2017). The 2017
152 campaign aimed to evaluate the sensitivity of snow remote sensing techniques through a gradient of
153 forest densities (NSIDC: <https://nsidc.org/data/snowex>). This campaign was primarily conducted in
154 two locations in Colorado, USA: Grand Mesa and Senator Beck Basin. Our focus is on Grand Mesa
155 (Figure 1). SnowEx datasets are publicly available through the National Snow and Ice Data Center
156 (NSIDC: <https://nsidc.org/data/snowex>).

157 The 2017 SnowEx campaign spanned from September 2016 to July 2017, with the Intense Observation
158 Period (IOP) taking place between February 6 and February 25, 2017 (Brucker et al., 2017a). During
159 this period, a multitude of measurements were collected, including data from cloud-absorption radar,
160 ground-penetrating radar, synthetic aperture radar, lidar, airborne video, Global Navigation Satellite
161 Systems (GNSS) measurements, and snow pit measurements (Brucker et al., 2017a; Kim et al., 2017).
162 In this study, we focus on the L-band InSAR products (phase change, coherence, amplitude, incidence
163 angle) and high-resolution airborne lidar data collected during the campaign (snow depth, bare ground
164 elevation, vegetation height). Our study site within Grand Mesa spans approximately 70 km², and all
165 data used in this study has a resolution of 3 m, which was the resolution of the airborne lidar. The
166 InSAR products were resampled to have the same resolution and align with the lidar raster. The details
167 of each data type's collection and use are outlined in the following sections.

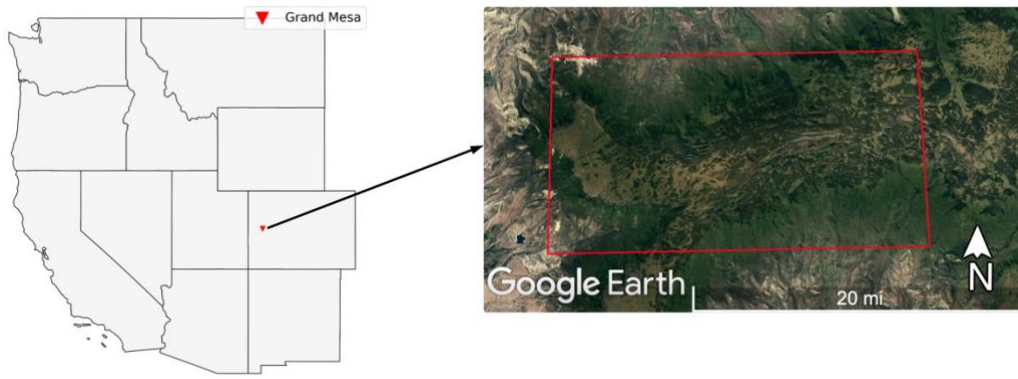


Figure 1. Study Site: Grans Mesa, Colorado. The large red rectangle outlines a 9.25 x 32 km target area that was imaged by all airborne sensors. The natural gradient of Snow Water Equivalent (SWE) increases from west to east, and the forest cover also naturally varies across the region (Kim et al., 2017).

2.2.1 Lidar Snow Depth Measurements

Lidar, an acronym for Light Detection and Ranging (Brucker et al., 2017b; King et al., 2022b), is a remote sensing method that uses the travel time of a laser pulse to measure distances (Killinger, 2014; Revuelto et al., 2014). The lidar data used in this study was collected by the NASA Airborne Snow Observatory (ASO) as part of the 2017 SnowEx campaign on two dates: September 20, 2016, and February 8, 2017. The September 20 flight was a "snow-free" flight, which means that the lidar data was collected before any snow had fallen on the ground. The February 8 flight was a "snow-on" flight, which means that the lidar data was collected when a snowpack had developed on the ground. The initial gridded ASO lidar product has a spatial resolution of 3 m and a vertical accuracy characterized by a root mean square difference of 8 cm when compared to near-coincident median values from 52 snow-probe transects (Currier et al., 2019). Snow depths were then calculated by differencing the two DEMs (snow-on minus snow-off). In addition to snow depth, bare ground elevation and vegetation height to the top of the canopy at matching 3-m resolution were produced from the lidar data (Figure 2).

For this work, we divided the data into three categories based on vegetation:

- Open areas: areas with no vegetation or vegetation heights < 0.5 meters.
- Vegetated areas: areas with vegetation height \geq 0.5 meters
- Combined dataset: a combination of both open and vegetated areas.

The resulting training dataset contains 3.174×10^6 , 2.166×10^6 , and 5.341×10^6 3-m resolution snow depths for open areas, vegetated areas, and a combination of both, respectively. Table 1 shows the snow depth statistics for each category in the training data. The mean snow depth is approximately 1.06 m in open areas, which is slightly lower than that of the vegetated areas at 1.10 m. The range of snow depth across all areas is from 0 m to a maximum of 3 m.

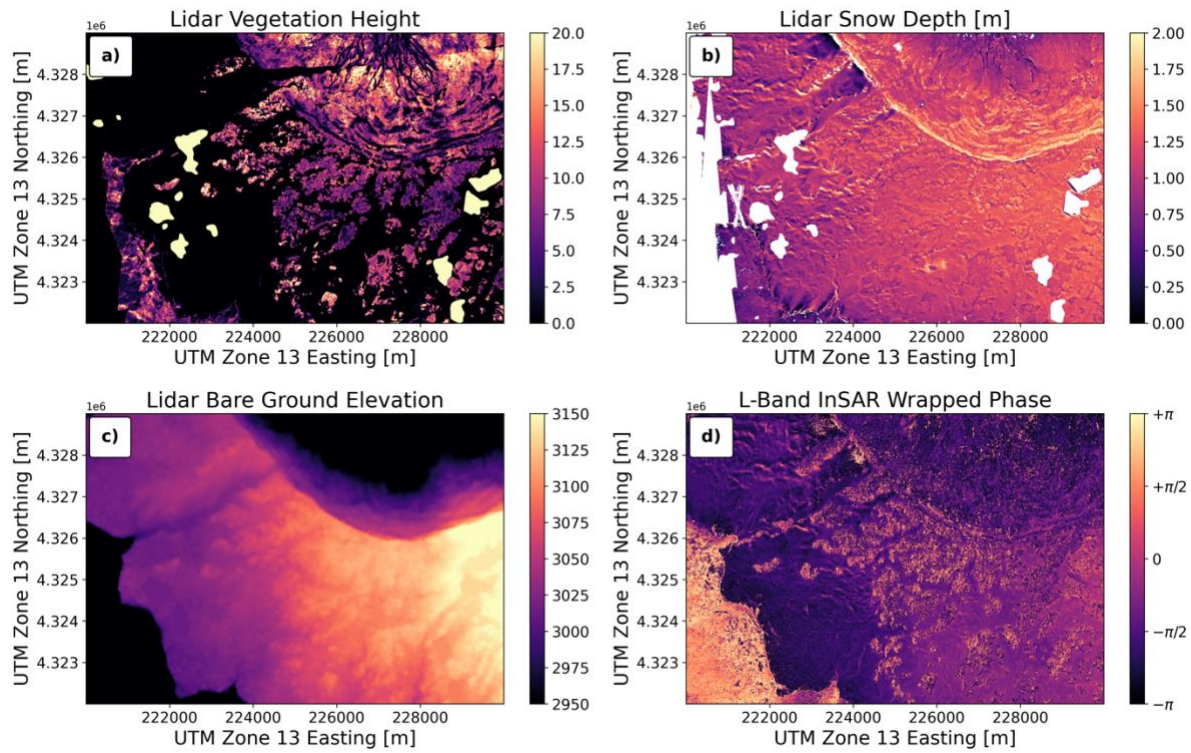


Figure 2. Lidar-derived products (a-c) from the study site. **a:** Vegetation height with a color range from 0 to 20 m, **b:** Snow depth with a color range from 0 to 2 m. **c:** Bare ground elevations with a color range spanning approximately 2950 to 3150 m. **d:** L-Band InSAR Wrapped Phase with a color scale ranging from $-\pi$ to $+\pi$.

Table 1. Summary Statistics of L-band InSAR products, lidar snow depths, bare ground elevation, and vegetation height (all at 3 m resolution) in the training set. The table provides the count, mean, standard deviation (Std), minimum (Min), 25th percentile (25%), median (50%), 75th percentile (75%), and maximum (Max) values for each product within open, vegetated, and the combined dataset (open + vegetated).

Subset	Product	Mean	Std	Min	25%	50%	75%	Max
Open Areas (n=3174348)	Amplitude	0.12	0.07	0.02	0.09	0.11	0.13	3.55
	Unwrapped Phase	-7.92	0.82	-15.22	-8.44	-8.01	-7.33	-0.06
	Coherence	0.64	0.13	0.003	0.56	0.66	0.73	0.98
	Incidence Angle	0.92	0.13	0.28	0.84	0.94	1	1.95
	Bare Earth DEM	3044.3	57.61	2492.03	3019.76	3035.7	3076.51	3157.45
	Wrapped Phase	-1.38	0.96	-3.14	-2.06	-1.54	-0.87	3.14
	Elevation (m)	3048.59	57.61	2501.13	3023.22	3039.78	3081.68	3168
	Snow Depth (m)	1.06	0.30	0	0.87	1.07	1.26	3.0
Vegetated Areas (n=2166210)	Amplitude	0.39	0.18	0.01	0.28	0.35	0.46	6.11
	Unwrapped Phase	-8.73	0.81	-15.26	-9.17	-8.67	-8.23	-0.47
	Coherence	0.48	0.15	0.006	0.36	0.47	0.59	0.97
	Incidence Angle	0.79	0.22	0.14	0.65	0.82	0.94	1.95
	Bare Earth DEM	2957.81	154.61	2490.96	2869.95	3004.13	3076.61	3157.38
	Vegetation Height (m)	8.43	5.26	0.5	4.09	8.12	12.06	34.62
	Wrapped Phase	-1.03	1.67	-3.14	-2.26	-1.69	-0.09	3.14
	Elevation (m)	2969.7	153.62	2500.16	2883.21	3016.08	3088.03	3168
	Snow Depth (m)	1.11	0.28	0	0.95	1.11	1.26	3.0
Combined Dataset (n=5340595)	Amplitude	0.23	0.18	0.01	0.10	0.15	0.33	6.11
	Unwrapped Phase	-8.25	0.91	-15.26	-8.76	-8.28	-7.68	-0.06
	Coherence	0.57	0.16	0.003	0.46	0.59	0.70	0.98
	Incidence Angle	0.87	0.18	0.14	0.77	0.90	0.99	1.95
	Bare Earth DEM	3009.17	116.11	2490.91	3007.21	3031.91	3076.55	3157.44
	Vegetation Height (m)	3.47	5.30	0	0	0.18	6.35	34.62
	Wrapped Phase	-1.24	1.3	-3.14	-2.14	-1.59	-0.75	3.14
	Elevation	3016.55	114.27	2500.20	3012.51	3036.44	3084.15	3168
	Snow Depth (m)	1.08	0.29	0	0.90	1.09	1.26	3.0

2.2.2 L-band InSAR Products

Interferometric Synthetic Aperture Radar (InSAR) is a radar technique that calculates changes in the distance between the radar antenna and the ground surface by comparing the phase difference of two or more radar images (Lu et al., 2007). Repeat-pass InSAR uses two or more InSAR images of the

same scene taken at different times to generate an interferogram, which measures the phase shift between the image acquisitions. InSAR can be used to measure a variety of surface changes, including deformation, subsidence, and ice sheet movement. More recently, repeat-pass InSAR has been used to measure SWE changes (Tarricone et al., 2023).

The L-band InSAR frequency range is 1 to 2 GHz, which corresponds to a wavelength range of 30 to 15 cm (NASA: <https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar>). L-band InSAR signals can penetrate clouds, snow, and vegetation. As the wave moves through the snow, it slows down, causing a phase shift. This phase shift can be used to estimate changes in snow depth and SWE. L-band SAR signals are also less sensitive to atmospheric moisture than other SAR bands, which makes the InSAR data more reliable in all weather conditions. As part of the 2017 SnowEx campaign, the NASA Jet Propulsion Laboratory (JPL) flew the L-Band UAVSAR sensor over Grand Mesa in February and March. A total of five flights were conducted on February 6, 22, 25, and March 8 and 31. For this work, we used the pair of flights closest to the lidar acquisition. Since the lidar was acquired on February 8, we used the HH polarization for February 6 and 22 as the InSAR features for our ML models. UAVSAR products come at a native resolution of 2 m. However, we resampled the UAVSAR features to 3 m to match the lidar resolution.

The following L-band InSAR products, all of which are resampled to 3-m resolution, were used in this study:

- **Coherence (CO):** Coherence, also known as interferometric correlation, is a measure of the accuracy of determining the interferometric phase. It ranges from 0 to 1, with higher coherence indicating greater confidence in the phase changes.
- **Incidence Angle (IA):** The incidence angle is the angle between the radar beam and the snow surface. The incidence angle affects the backscatter signal from the snow surface and the path of the signal through snow, and it must be considered when interpreting InSAR data.
- **Amplitude (AM):** Amplitude is one of the two fundamental characteristics of radar signals. It is the magnitude/energy of the backscattered signal, and it records information on the surface roughness. For example, surfaces that are rough at the scale of the radar wavelength or have a higher dielectric constant (such as wet surfaces) will generally backscatter more of the radar signal and thus have a higher amplitude. Conversely, smoother surfaces or those with a lower dielectric constant will backscatter less of the signal, resulting in a lower amplitude. Forests typically have larger amplitude responses compared to open areas.
- **Wrapped Phase (WP):** Phase is the second fundamental property of radar signals. The phase change image records information on changes in sensor-to-target time of flight between the two acquisitions. Each pixel in a SAR image has a phase value that is related to the relative time of flight between the satellite sensor and the ground target (snow ground surface), modulated by the wavelength of the radar signal. When two or more SAR images are acquired at different times in repeat-pass interferometry, the phase difference between corresponding pixels in the two images can be calculated. This phase difference, referred to as the interferometric phase or wrapped phase, contains valuable information about changes in the target surface, such as elevation change or displacement. The term "wrapped" refers to the fact that the phase values are cyclic, i.e., a phase value of 2π is equivalent to a phase value of 0. Thus, the phase values "wrap" around at intervals of 2π .
- **Unwrapped Phase (UW):** The unwrapped phase is the phase difference between the two SAR images after it has been processed to remove the cycle "wrapping". It provides a continuous measure of the time of flight between the radar antenna and the ground surface, which is crucial for measuring changes in the snow depth.

InSAR measures changes in time of flight to the ground surface. This technology uses radar signals to precisely detect alterations in the Earth's surface, providing valuable insights into various geophysical phenomena. Therefore, to use InSAR for estimating total snow depth, we assume that the total snow depth patterns before melts starts within a season are relatively consistent (Sturm and Wagner, 2010) with the patterns of snow depth change detected by the InSAR instrument being representative of the distribution of total snow depth. This assumption is generally valid, but there are instances and locations where this may not be the case depending on weather patterns, landscape and vegetation changes, and other processes affecting snow distribution. L-band InSAR data will be available from the NISAR mission at a 12-day temporal resolution, which has the potential to complement existing snow depth measurement practices. In addition, with a continuously operating InSAR, 12-day snow depth/SWE changes can be summed to estimate total depth (Oveisgharan et al., 2024). The InSAR products used in this work are shown in Figure 3, and their summary statistics are included in Table 1.

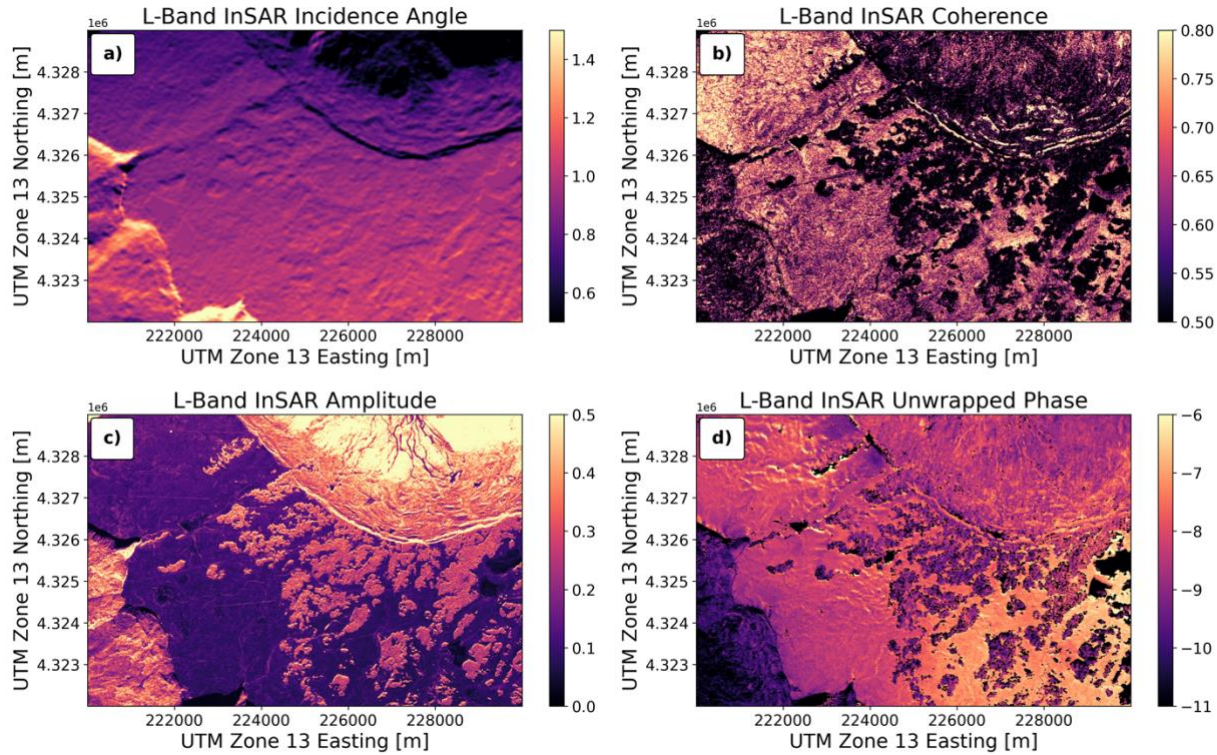


Figure 3. L-band InSAR products: (a) Incidence angles with values from 0.5 to 1.5, (b) Coherence metrics with values from 0.50 to 0.80, (c) Amplitudes with signal intensities between 0.0 and 0.5, and (d) Unwrapped Phase with phase shift values from ~ -11 to -6 .

2.3 Methods

Our methodology for estimating snow depth from InSAR products using machine learning models comprised several key steps, which are outlined below.

2.3.1 Data Preprocessing

Data preprocessing was the foundational step in our methodology, preparing the dataset for compatibility with our Machine Learning algorithms. The raw remote sensing data spanned numerous raster layers at a 3-m resolution. Each raster layer represents approximately a 7 km x 10 km area on the ground, with each pixel representing an average value over a 3 m x 3 m area. We first restructured

282 the raster into pixel-wise tabular data frames to enable vector-based preprocessing. After concatenating
283 all variables, we inspected for noise and outliers that could potentially affect the models' learning.

284 Noise in the data was identified through pixel values coded as -9999, which are standard no-data flags,
285 and these values were dropped from the analyses. Outliers were defined based on the physical
286 plausibility of measurements, with snow depth values greater than 3 m classified as outliers. This
287 threshold was chosen after inspecting the in-situ snow depth distribution for the day (February 8),
288 which showed a maximum snow depth of 3 m. Extreme values (e.g., snow depths of 13 m) were
289 unrealistic and likely due to errors such as the lidar laser hitting the top of vegetation during the snow-
290 on flight.

291 No missing value imputation was done as we have an abundance of data points (~ 6.7 million
292 observations after data cleaning). Imputation is a process where missing data points are filled in using
293 various statistical methods, which can sometimes introduce uncertainties. By avoiding imputation, we
294 maintained dataset integrity. These data cleaning steps resulted in a dataset of 6,675,745 usable
295 observations.

296 After data cleaning, we split our data into three disjoint subsets:

- 297 • Training Set: 80% of the data was used for training the models.
- 298 • Tuning Set: 10% of the data was used for hyperparameter tuning to optimize model
299 performance and to prevent overfitting during model training. Hyperparameter tuning involves
300 adjusting the parameters of the learning algorithm itself to find the best configuration. This
301 process can help improve predictive accuracy and robustness. Overfitting occurs when a model
302 learns the training data too well, capturing noise and details that do not generalize to new data.
303 By using a tuning set, we can ensure the model generalizes well and performs effectively on
304 unseen data. Note that using a tuning set for hyperparameter optimization is an alternative to
305 cross-validation for large datasets.
- 306 • Testing Set: The remaining 10% of the data was used to test the models' performance.

307 Following the partitioning, all input features were converted into a common dimensionless scale with
308 a mean of zero and a standard deviation of one using the z-score standardization approach following
309 Eq. (4). This will prevent potential domineering effects from variables with larger raw value ranges
310 during model training. We have chosen the z-score standardization approach because it ensures outliers
311 are handled more properly (Ozdemir, 2022). The normalized version of every observation $X_{i,j}$ in
312 feature X_j is obtained by:

$$X_{i,j}^z = \frac{X_{i,j} - \mu_j}{\sigma_j} \quad (1)$$

313 Where $X_{i,j}^z$ is the standardized value, μ_j and σ_j are the mean and standard deviation of variable X_j from
314 the training set.

315 For each data stratum (open, vegetated, and their combination), we trained three models:

- 316 • Model 1: The first model uses only bare ground elevation as input, and this model serves as
317 the baseline model in this work. Mathematically, we write:
318

$$SD = f(BGE) \quad (2)$$

- Model 2: This model uses InSAR products and bare ground elevation as input features. The model is represented mathematically as follows:

$$SD = f(CO, IA, BGE, WP, AM, UP) \quad (3)$$

- Model 3: This model combines the InSAR products, bare ground elevation, and vegetation height as features. This configuration is used to evaluate the effect of vegetation height on the model performance.

$$SD = f(CO, IA, BGE, WP, AM, UP, VH) \quad (4)$$

where f is the various learning algorithms compared, VH is the vegetation height, BGE is the bare ground elevation, and SD is the snow depth. Note that the open areas only have models 1 and 2 because, in the open areas, vegetation height is 0 or buried under the snow cover (<0.5 m).

2.3.2 Model Selection and Training

Once the data were appropriately prepared, we proceeded to the next phase: model selection and training. We trained and compared three machine learning algorithms: Extremely Randomized Trees (Extra Trees), eXtreme Gradient Boosting (XGBoost), and artificial neural networks (ANN). These models were chosen because they have been shown to perform well on a variety of remote sensing data (Maxwell et al., 2018; Meloche et al., 2022). For hyperparameter tuning, we used the Optuna framework (Akiba et al., 2019). Optuna is an open-source hyperparameter optimization framework in Python that is designed to optimize the hyperparameters for machine learning models. Unlike traditional grid search, Optuna utilizes a Tree-structured Parzen Estimator (TPE) algorithm (Zhao and Li, 2018), a Bayesian optimization algorithm, which tends to find optimal hyperparameters faster and with fewer function evaluations compared to grid search. This efficient search approach proved to be especially advantageous given the large dataset involved in our study. The details of the hyperparameters tuned for each model can be found in Table A1 in the appendix.

2.3.2.1 Extremely Randomized Trees

Extra Trees is an ensemble algorithm that averages predictions across a pre-defined number of randomized decision trees to improve accuracy and control overfitting. The "extra" in Extra Trees stands for extremely randomized, indicating that at each split in the learning process, the features and cut points are chosen in a random manner, hence reducing the variance of the model. Due to this extreme randomization, Extra Trees are faster than Random Forest and hence suitable for large datasets. Mathematically, the prediction $\hat{f}_{EXT}(\mathbf{x})$ from an extra tree regressor can be expressed as:

$$\hat{f}_{EXT}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T \hat{f}^t(\mathbf{x}) \quad (1)$$

Where T is the total number of trees in the ensemble and $\hat{f}^t(\mathbf{x})$ is the prediction of the t -th tree for the input vector \mathbf{x} .

In our analysis, the optimal hyperparameters for the Extra Trees model were identified using Optuna. The optimal number of trees in the forest was found to be 150. The maximum depth of the trees was

set to None, indicating that the nodes are expanded until they contain fewer than the minimum samples required to split, allowing the trees to grow to their full depth. Finally, the mean squared error (MSE) was used as the measure of split quality at each node. These hyperparameter settings were found to provide the best performance on the tuning set. The details of these hyperparameters can be found in Table A1 in the appendix.

2.3.2.2 eXtreme Gradient Boosting

XGBoost is another ensemble algorithm, but unlike Extra Trees, it builds a sequence of decision trees instead of a forest of decision trees. XGBoost operates by sequentially constructing weak learners (decision trees), with each tree aiming to correct the errors made by the previous one. This process of sequential error correction is known as Boosting. At each iteration, a weak learner is trained to approximate the gradient of the loss function (the residual errors). Boosting is then achieved by iteratively updating the residual errors when a new learner is added to the ensemble. This methodology of leveraging the gradient of the loss function to guide the boosting process is known as Gradient Boosting. XGBoost uses a variant of the Gradient Boosting algorithm called Newton boosting, which attaches weights to the residuals through the Hessian (second-order derivative of the loss function). With this, observations with larger errors have more weight. XGBoost takes Newton boosting to the extreme by regularizing the loss functions and introducing efficient tree learning with parallelizable implementation. This “extreme” attribute of XGBoost makes it suitable for large datasets. Moreover, XGBoost can also benefit from GPU (CUDA support only) acceleration, making it suitable for large datasets.

In our analysis, the optimal hyperparameters for the XGBoost model were identified using Optuna, as detailed in Table A1 in the appendix. The objective was set to minimize the MSE between the predicted and true snow depth values. We used a learning rate of 0.05 to control the step size at each iteration while moving toward a minimum of the loss function. The depth of the trees was set to grow unrestricted to allow the model to learn complex relationships in the data. A total of 1000 trees were grown using a histogram-based training method to accelerate the training process.

2.3.2.3 Artificial Neural Network

Artificial Neural Network (ANN) is an ensemble of linked artificial neurons organized into input, hidden, and output layers. Each neuron receives inputs, performs mathematical operations on these inputs, and passes the output to the next layer. Every connection between nodes has an associated weight; the optimal weights are learned during training. The input layer holds the features, with one node per input feature, and the output layer holds the network’s prediction. The number of neurons in the hidden layers is determined through hyperparameter tuning. In this study, we employed a Feed-forward Neural Network (FNN), a type of ANN where connections between nodes do not form a cycle. Specifically, we used the FNN because it is the gold standard for tabular datasets.

We designed a five-layer FNN using the PyTorch framework for our analysis. The architecture comprises one input layer, three densely connected hidden layers with 2048, 1500, and 1000 nodes, respectively, and one output layer. Rectified Linear Unit (ReLU) activation functions were used in the hidden layers to introduce non-linearity, while a linear activation function was employed in the output layer for snow depth estimation. The model was trained using the Adam (Kingma and Ba, 2015) optimization algorithm with a learning rate of 0.0001 to minimize the Mean Squared Error Loss (MSELoss) between the predicted and lidar-derived snow depths. The training was conducted over 15 epochs with a batch size of 128. The hyperparameters for the FNN model were optimized using Optuna, with the details of the tuned hyperparameters presented in Table A1 in the appendix.

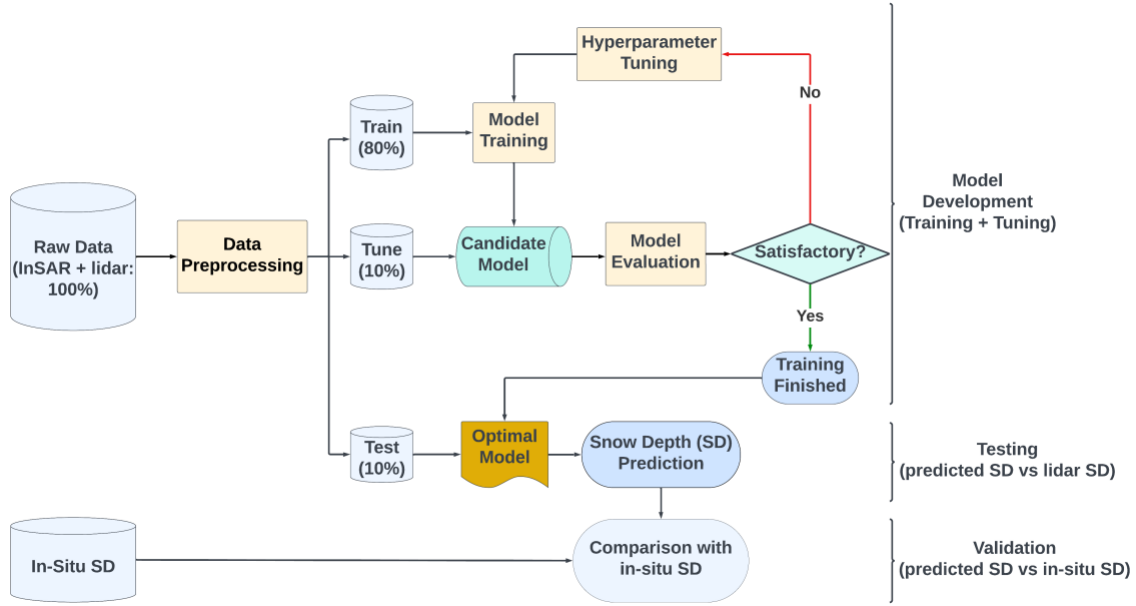


Figure 4. A flowchart illustrating the step-by-step methodology of our model training.

Model performance was evaluated using root mean squared error (RMSE), mean bias error (MBE), and coefficient of determination (R^2). RMSE measures the average magnitude of the errors between the predicted and observed values, providing insight into the model's overall accuracy. MBE assesses the average bias in the predictions, indicating whether the model systematically overestimates or underestimates the true values. R^2 measures how well the predicted values match the true values using a 1:1 line, with higher values indicating a better fit. RMSE has a lower bound of zero, with smaller values indicating better performance. MBE is unbounded and can be positive or negative, with values closer to zero in absolute value being preferable. Note that R^2 , in this case, is not the square of the correlation coefficient; hence, it can be negative for a bad model.

2.4 Feature Importance

To quantify the contribution of individual features to snow depth prediction, we employed two complementary methods: the gain metric from the XGBoost model and SHapley Additive exPlanations (SHAP) (Zhang et al., 2012). The *gain*, derived from the XGBoost framework, measures the average contribution of each feature to reducing the mean squared error loss across all trees within the model. This method offers an initial insight into the relative importance of features based on their utility in constructing the predictive model. However, gain-based importance can be misleading for high cardinality (many unique values) features and may not fully capture the nuanced interactions between features. To address this, we also utilized SHAP values, which provide a more comprehensive and stable measure of feature importance by considering the contribution of each feature to every possible combination of features in the dataset (Man and Chan, 2021). SHAP originates from concepts in cooperative game theory, and it computes the importance of a feature by distributing the predictive contribution among features in a model, akin to dividing payoffs among collaborating players. Although SHAP analysis is computationally expensive, particularly for large datasets, its ability to offer clear and consistent interpretations of the features' impact on the model's output makes it a valuable tool for in-depth feature importance analysis.

427 **3 Results**

428 **3.1 Snow Depth Estimation**

429 Comparing the performance of Models 1, 2, and 3 allowed us to evaluate the incremental value of
430 remote sensing predictors for improving snow depth estimation beyond using bare ground elevation
431 alone. Across most configurations and areas, XGBoost consistently outperformed other models (Table
432 2). Therefore, our discussion will focus on the XGBoost results.

433 The model comparison revealed distinct performance across different model configurations. The
434 baseline model (Model 1) provided a foundational understanding but demonstrated limited predictive
435 power with R^2 values ranging from 0.21 to 0.33 across the open, vegetated, and mixed areas. However,
436 incorporating InSAR-derived features (Model 2) markedly increased the precision of snow depth
437 estimates.

438

Table 2: Comparative Performance of FNN, Extra Trees, and XGBoost across Vegetated, Open, and Mixed Areas on the Held-out Test Set. The most preferred model is highlighted in bold fonts.

Model	Data	Configuration	RMSE (m)	MBE (m)	R^2
FNN	Open Areas	Model 1	0.2493	0.0003	0.321
		Model 2	0.1646	-0.0035	0.704
	Combined	Model 1	0.2613	0.001	0.2047
		Model 2	0.1738	0.0095	0.6481
		Model 3	0.1617	-0.004	0.6952
	Vegetated Areas	Model 1	0.2425	-0.0025	0.2368
		Model 2	0.1802	-0.0094	0.5787
		Model 3	0.1611	0.0071	0.6631
Extra Trees	Open Areas	Model 1	0.2365	0.0001	0.3891
		Model 2	0.104	0.0001	0.8818
	Combined	Model 1	0.2483	0.0003	0.2815
		Model 2	0.1085	0.0002	0.8629
		Model 3	0.1052	0.0003	0.871
	Vegetated Areas	Model 1	0.2201	-0.0004	0.3712
		Model 2	0.1039	0.00004	0.8441
		Model 3	0.1118	0.0008	0.8379
XGBoost	Open Areas	Model 1	0.2483	0.0007	0.3265
		Model 2	0.0985	-0.0001	0.894
	Combined	Model 1	0.2606	0.0004	0.2087
		Model 2	0.104	-0.000009	0.8739
		Model 3	0.0988	0.0001	0.8863
	Vegetated Areas	Model 1	0.2421	0.0001	0.2393
		Model 2	0.1098	-0.0002	0.8436
		Model 3	0.1046	0.0005	0.8579

Model 2 increased the R^2 value from 0.33 to 0.89 in the open areas, from 0.24 to 0.84 in the vegetated areas, and from 0.21 to 0.87 for the combined data, with the addition of the InSAR data. This improvement underscores the potential of InSAR data in capturing key variables influencing snow depth. The highest R^2 was observed in open areas. The performance in vegetated areas also improved, which indicates that InSAR products contribute valuable information even in vegetated landscapes. For the combined (open + vegetated) dataset, Model 2 attained an R^2 of 0.87, which is also a substantial increase from the baseline performance (0.21).

Introducing vegetation height as an additional predictor alongside InSAR products in Model 3 yielded minimal performance gains in model accuracy across vegetated and combined areas (since open areas have no vegetation or they are buried under the snow over). In vegetated areas, Model 3 increased the R^2 from 0.84 to 0.86, while for the combined dataset, the R^2 improved from 0.87 to 0.89. Despite these incremental improvements, the overall best performance was observed in open areas (Model 2:

RMSE=9.85 cm, $R^2=0.894$), followed closely by the combined dataset (Model 3: RMSE=9.88 cm, $R^2=0.886$). Vegetated areas exhibited the least performance (Model 3: RMSE=10.46 cm, $R^2=0.858$), although still maintaining a high level of accuracy.

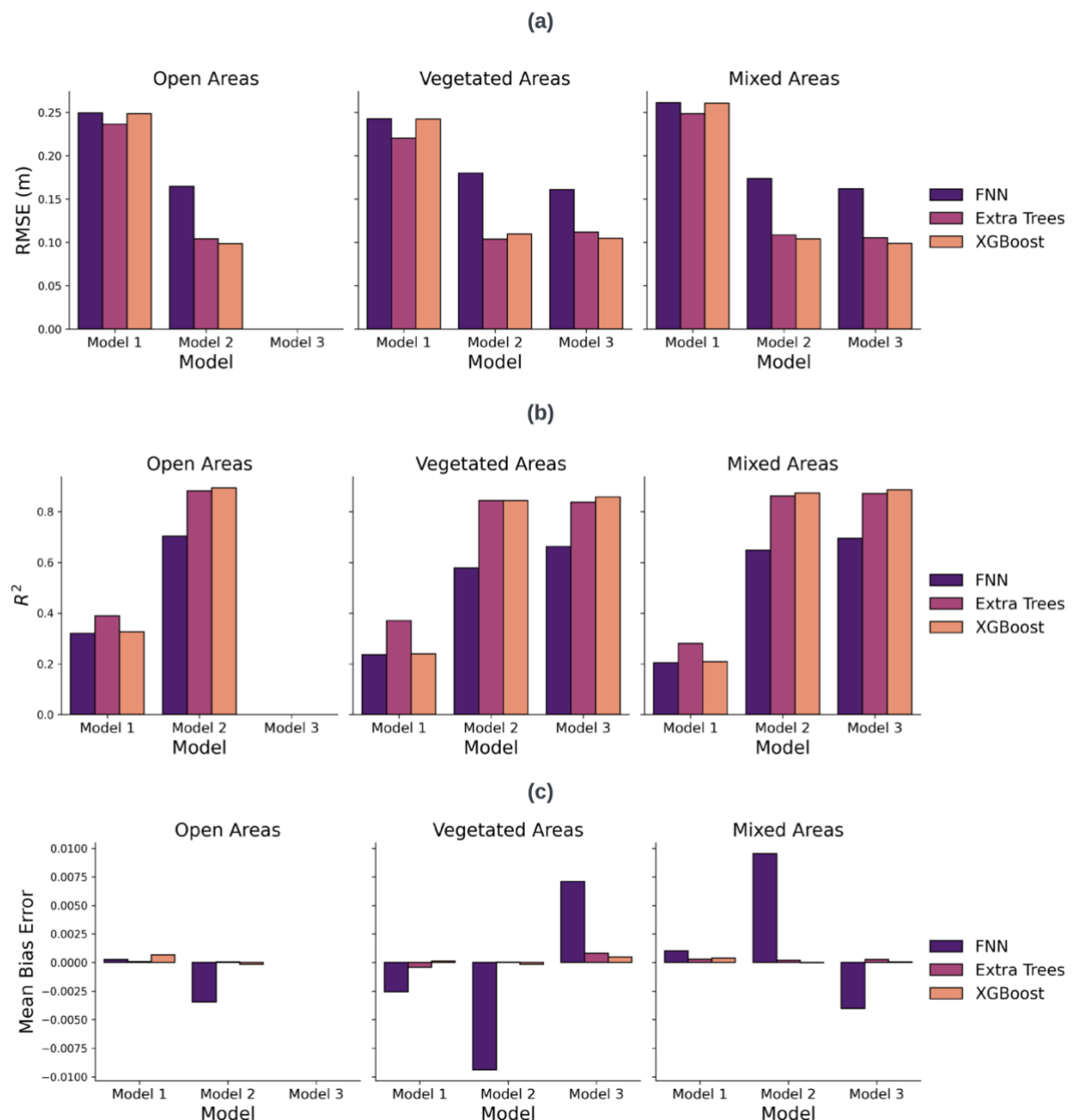


Figure 5: Comparative performance metrics of snow depth estimation Models on the testing set. Panel (a) displays the RMSE across all models and data strata, illustrating the accuracy of snow depth predictions. Panel (b) presents the R^2 values, and Panel (c) presents the MBE.

Based on the performance comparison, Model 3 for the combined dataset appears to be the optimal choice. This model demonstrated comparable results (Table 2) to those obtained in open areas (Model 2), effectively eliminating the need for separate models for open and vegetated regions. Adopting a single model approach offers several advantages, including saving time and resources while allowing

more focus on refining and optimizing this unified model. Figures 6, 7, 8, and 9b and Table 3 are based on Model 3 for the combined dataset, as Model 3 provides a comprehensive and efficient solution for our application.

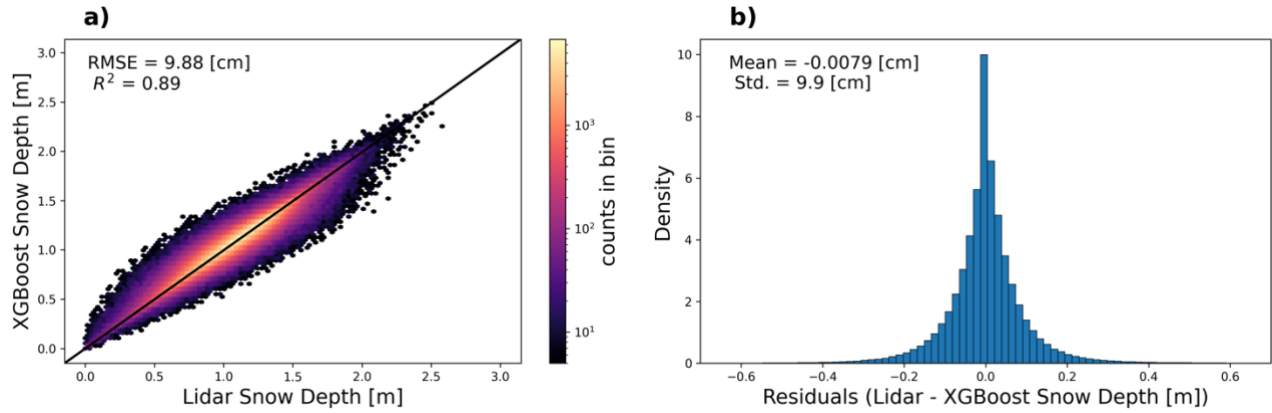


Figure 6. XGBoost Residual Analysis (Model 3 on the combined dataset). (a) Hexbin density plot comparing lidar measured snow depths with predictions from the XGBoost model. (b) Histogram of the residuals (lidar minus XGBoost predicted snow depths)

The XGBoost modeled snow depths showed strong correspondence with lidar snow depths at the 3-m scale (RMSE=9.88 cm, $R^2=0.89$; Figure 6a), illustrating the model's ability to reproduce the observed snow depth variability. Similarly, the residuals (i.e., lidar minus XGBoost predicted snow depth; Figure 6b) are concentrated around zero with a mean of -7.9×10^{-3} cm and a standard deviation of 9.9 cm, indicating minimal errors and biases, and an error similar to that reported for airborne lidar.

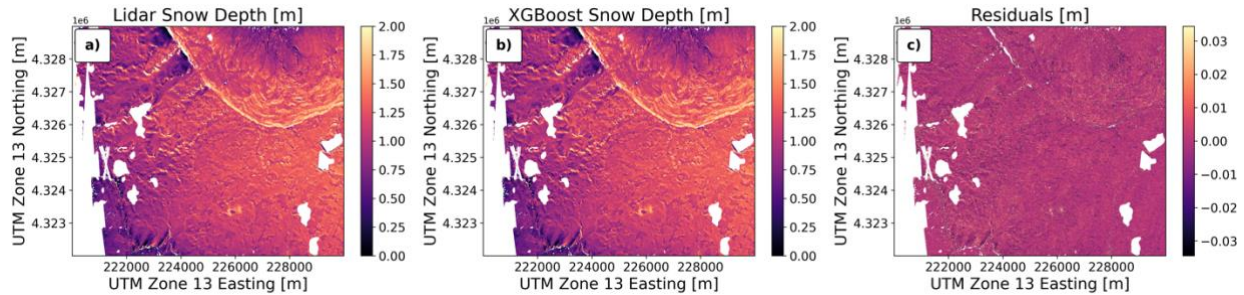


Figure 7. Spatial Residual Analysis. (a) Lidar-derived snow depth. (b) The XGBoost predicted snow depth. (c) The prediction error (lidar minus predicted snow depths).

Figure 7b (predicted depths) visually appears to capture the spatial variability observed in the lidar data (Figure 7a), suggesting good model performance. Figure 7c quantifies the prediction error, visually encoded to emphasize areas of underestimation or overestimation by the model. Figure 7c indicates minimal bias based on the color map.

3.2 Feature Importance

Using SHAP values (Figure 8a), the most influential variables were unwrapped phase, bare ground elevation, and amplitude. Gain-based importance (Figure 8b) also showed unwrapped phase, bare

ground elevation, and amplitude as the top 3 predictors, aligning with SHAP. This indicates their high explanatory power for estimating snow depth from InSAR.

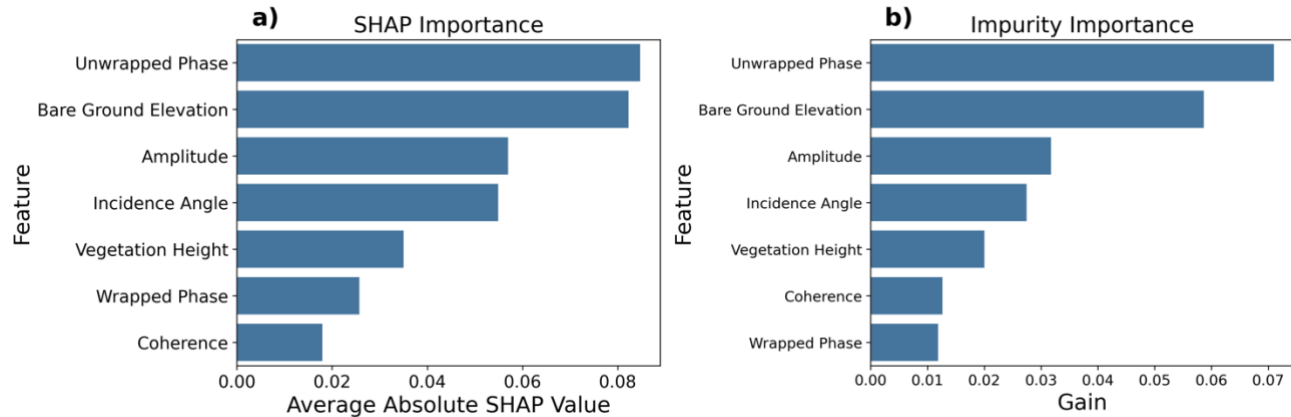


Figure 8. Feature Importance for Model 3 on the combined dataset: (a) SHAP importance (b) XGBoost impurity-based importance.

Amplitude backscatter ranked 3rd, providing information on surface roughness and structure. When stratified into open and vegetated areas, the feature importance ranks were the same as Figure 8 in open areas. However, for vegetated areas, bare ground elevation was the leading feature, highlighting the complex relationship between the topography and snow accumulation beneath the canopy. Vegetation height also became more important in forests, ranking 3rd compared to 5th overall. This confirms vegetation's role in attenuating radar signals in dense cover. Feature importance plots for forested and open areas can be found in Figure A1 in the appendix.

3.3 In-Situ Validation

To provide an independent assessment of model accuracy, predicted snow depths were validated against in-situ observations collected as part of the 2017 NASA SnowEx campaign (Brucker et al., 2018). Manual measurements were taken using either a standard, handheld 3-m snow probe or a shorter GPS-equipped 1.2-m MagnaProbe. The GPS technology in the MagnaProbe provides a position accuracy of ± 2.5 m (Sturm and Holmgren, 2018a). During the intense observation period (February 6-25), a total of 27,081 snow depth measurements were taken at intervals of approximately 3 m.

Validation was done using a 3-m buffer approach, where the average snow depth (both from lidar and our predictions) within a 3-m radius of each in-situ observation was calculated. We used lidar snow depth data from February 8, 2017, for developing our models, so our validation focused on in-situ measurements taken on the same date to ensure temporal consistency. Initially, we had 1777 in-situ snow depth measurements from February 8. However, the lidar data didn't cover the entire area where in-situ measurements were taken. Hence, we considered only points where the 3-meter buffer contained lidar/modeled depths.

This spatial filtering reduced the number of usable in-situ measurements to 234, as only these points were within the lidar swath coverage area. These 234 measurements were used for comparison against the average of the lidar dataset and the model predictions (training or testing) within their respective 3-m buffers.

When compared to in-situ measurements (Table 3), the lidar-derived snow depths achieved an RMSE of 15.85 cm and R^2 of 0.332, while snow depths predicted by XGBoost achieved a slightly worse performance (RMSE = 15.90 cm and R^2 = 0.328).

Table 3: In-situ validation results

Snow Depth Source	RMSE (m)	R^2
Lidar	0.1585	0.332
XGBoost	0.1590	0.328

Overall, the in-situ validation provides confidence in the modeling framework, with XGBoost predictions showing accuracy approaching the lidar training data.

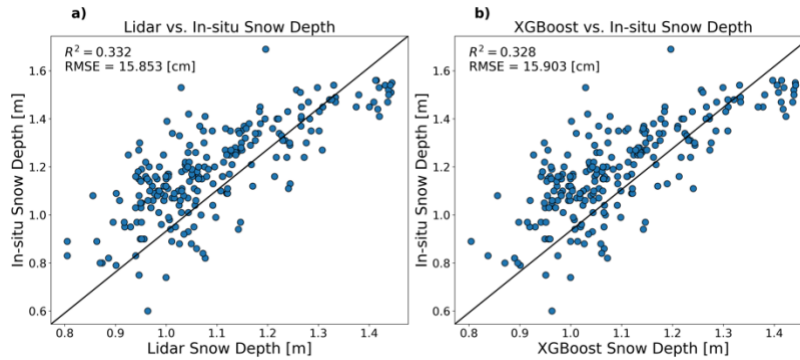


Figure 9. Snow depth validation using in-situ measurements: The closer the data aligns with the diagonal line, the more accurate the snow depth estimates are in comparison to in-situ measurements. (a): Lidar-derived depths versus in-situ depths. (b) XGBoost predicted depths versus in-situ depths.

4. Discussion

4.1 Model Performance

This study explored the potential of machine learning for snow depth prediction using data from NASA JPL's UAVSAR sensor, which employs L-band radar for InSAR measurements. We developed and compared three machine learning algorithms – Extra Trees, XGBoost, and FNN, each selected for their robustness in handling complex datasets and their fast training and prediction time, with XGBoost and FNN capable of benefiting from GPU acceleration. XGBoost consistently outperforms other algorithms. Hence, the results are based on the XGBoost model. To investigate the potential differences in model performance across open and vegetated areas and assess the feasibility of using a single model for snow depth prediction across the entire study area, we partitioned the study area into three distinct categories: vegetated regions, open regions, and a combination of both.

Within each category, we developed three distinct models: a baseline model using only elevation data (model 1), a second model incorporating InSAR parameters and elevation (model 2), and a third model combining InSAR parameters, elevation, and vegetation height (model 3). The addition of vegetation height improved model performance for both vegetated and mixed areas (assuming open areas have no vegetation or are buried beneath the snow cover). In vegetated areas, including vegetation height as a

feature reduced the RMSE from 10.98 cm to 10.46 cm, while for the combined dataset, the RMSE reduced from 10.40 cm to 9.88 cm.

These findings suggest that while vegetation height contributes to the model's predictive power, its impact is less pronounced compared to the inclusion of InSAR-derived features. In open areas, adding InSAR-derived features reduced the RMSE from 24.83 cm to 9.85 cm compared to the baseline; in vegetated areas, the RMSE reduced from 24.21 cm to 10.98 cm; and for the combined dataset, the RMSE reduced from 26.06 cm to 10.40 cm. The marginal improvements in model accuracy from model 2 to model 3 indicate that the InSAR parameters capture a larger portion of the variability in snow depth, even in the presence of vegetation.

The model performed better in open areas than in vegetated areas. This is likely because forests have a more complex structure than open areas, which can scatter and attenuate the InSAR signal. Additionally, trees can intercept snowfall, making it more difficult to measure snow depth beneath the canopy accurately. Although our models performed better in open areas, the results from the vegetated areas also maintained high accuracy. For model 2, the difference in RMSE between open and vegetated areas is 1.13 cm. The high accuracy in vegetated areas can be attributed to the fact that L-band data is less sensitive to the forest structure. This finding is consistent with the work of Hosseini (Hosseini and Garestier, 2021), where he found the L-band to be less sensitive to forest structure than the P-band.

Considering the performance metrics across different areas, Model 3 for the combined dataset emerges as the best choice, offering a balance between accuracy and practicality. By incorporating both InSAR-derived attributes and vegetation height, Model 3 effectively captures the variability in snow depth across open and vegetated areas, achieving an impressive R^2 of 0.89 and an RMSE of 9.88 cm. This single model approach for mixed areas eliminates the need to develop separate models for open and vegetated regions, saving time and resources while maintaining high predictive power. The inclusion of vegetation height in Model 3 successfully differentiates between open and vegetated areas, as evidenced by the comparable performance metrics obtained when the combined dataset's predictions are divided into their respective subsets (vegetated areas: RMSE=10.46 m, R^2 =0.86; open areas: RMSE=9.5 cm, R^2 =0.90). These results further validate the robustness and versatility of Model 3 in accurately estimating snow depth across vegetated and non-vegetated areas. This enhanced performance implies the potential of a single and robust model to reliably estimate snow depth in both vegetated and open areas, which can enhance the efficiency of large-scale snow monitoring efforts.

4.2 Feature Importance

We conducted feature importance using SHAP and gain metrics from XGBoost. This is because the SHAP feature importance provides a more reliable and stable ranking, and the gain metrics can be misleading for features with high cardinality. We found an agreement between the importance of the feature from both SHAP's and XGBoost's gain (Figure 8). Unwrapped phase, bare earth elevation, and amplitude are the top three influential features. Unwrapped phase emerged as the most influential predictor, as it is a direct indicator of snow depth changes, hence its strong influence on the model's predictive power. The bare ground elevation's importance is similarly intuitive; it represents the underlying topography, which is fundamental in understanding snow accumulation patterns. Topography relates to snow depth through factors like elevation, slope, and aspect (Trujillo et al., 2007; Hojatimalekshah et al., 2021). Higher elevations experience more snowfall, north-facing slopes retain snowpack longer in the northern hemisphere, and leeward areas develop drifts. The bare ground elevation provides the terrain context to model these topological influences on snow accumulation. Amplitude, ranking third in importance, suggests that the backscatter intensity, which is affected by surface characteristics, including roughness and snow density, is also an important predictor. This

aligns with findings in literature where backscatter properties have been directly correlated with snow depth. King et al. (2015) found a strong correlation between Ku-band backscatter and snow depth in the tundra. The remaining features, like incidence angle and coherence, showed lower importance. Further analysis is warranted to fully explain the relative feature contributions. Additionally, from a remote sensing perspective, SAR imagery captures details about the Earth's surface by recording both the strength (amplitude) and timing (phase) of the backscattered radar signal. Given the rich information contained in the amplitude and phase components of SAR images, it is reasonable to expect that these features would rank high on the feature importance scale when used in various remote sensing applications, such as snow depth estimation.

When the data was stratified by open versus vegetated areas, some notable differences emerged in the feature importance (Figure A1). In open areas, unwrapped phase and bare earth elevation remained the top two predictors, followed by incidence angle and then amplitude. This aligns with the overall importance rankings and reinforces the primacy of phase and topography for snow depth estimation in open regions. However, in vegetated regions, bare earth elevation was the top performer, followed by incidence angle and vegetation height. Unwrapped phase dropped to fourth. The decreased ranking of the unwrapped phase in forests is likely because dense vegetation attenuates and scatters the radar signal, reducing the phase's sensitivity to snow depth variations below the canopy. Meanwhile, vegetation height became the third most important variable, confirming its role in correcting for signal attenuation effects in vegetated areas. These land cover-specific feature importance findings provide useful insights. The unwrapped phase appears most valuable in open areas where radar penetration is uncompromised, while bare earth elevation and vegetation structure take precedence in forests. These results point to potential pathways for improving InSAR snow depth retrieval through optimal parameterization tailored to different land cover regimes. Additionally, these results can guide future research and data collection efforts, making them more focused, and may lead to more sophisticated physics-based models and physics-informed machine learning.

4.3 In-situ Validation

Validation against in-situ snow depth measurements provides an independent assessment of model accuracy. The lidar training data achieved 15.9 cm RMSE versus in-situ points. For the optimized XGBoost model, we achieved 16 cm RMSE compared to in-situ data. We observe from Figures 9a and 9b that the lidar data and XGBoost's predictions generally underestimated the in-situ snow depth measurements. Currier et al. (2019) also reported this negative bias between the lidar and in-situ probe measurements. The close alignment between lidar and ML model errors suggests that our model effectively learned the patterns present in the lidar data. However, this also means that the ML model inherited the biases present in the lidar measurements (Figure 9). We observed a general underestimation of snow depth by lidar, which was consequently reflected in our ML predictions. This underscores a fundamental principle in machine learning: the model's performance is inherently limited by the quality of its training data.

Moreover, the estimated ± 5 cm uncertainty in in-situ snow depth probe measurements, arising from factors such as probe penetration into the soil and low vegetation impacts (Sturm and Holmgren, 2018b; Currier et al., 2019), sets a practical limit on the achievable accuracy of our model validations. Additionally, potential geolocation errors from the probe introduce spatial discrepancies when comparing point-based in-situ measurements with grid-based lidar or model predictions. These factors introduce a margin of error that must be considered when interpreting the validation results.

Our findings show a larger discrepancy between lidar and in-situ snow depths compared to the work of Currier et al. (2019) on the same dataset. While we found an RMSE of 15.9 cm, they reported a root-

mean-square difference of 8 cm between the lidar-derived snow depths and in-situ measurements. This difference in error magnitudes can be attributed to methodological distinctions. The difference might be due to Currier et al. (2019) comparing median values from 52 snow-probe transects collected over four days (February 8, 9, 16, and 17) to near-coincident lidar data from February, while we compared 234 individual point measurements with lidar from February 8 alone. Their use of transect medians likely smoothed out local variations and reduced the influence of outliers, potentially reducing the overall error.

In summary, while the ML model successfully replicated the patterns in the lidar data, it also replicated its biases. This serves as a cautionary note for the application of ML in snow monitoring: while ML can be a powerful tool for pattern recognition and prediction, it is not a panacea for underlying data quality issues. Enhancing the accuracy of lidar measurements would likely improve the model's predictive capabilities. Addressing geolocation errors and understanding the inherent measurement uncertainties can further refine the validation process and the interpretation of model performance.

5. Limitation of the Study

While this study provides valuable insights into the use of InSAR products for estimating total snow depth, there is a limitation that must be acknowledged. A key assumption of this approach is that the spatial patterns of total snow depth remain consistent between the time of snow accumulation and the subsequent InSAR acquisition. In other words, the snow depth change detected by the InSAR instrument is assumed to be representative of the total snow depth distribution. However, if external factors, such as strong winds from an unusual direction, alter the snow accumulation pattern during this time interval, the accuracy of the total snow depth estimates may be compromised. Consequently, the effectiveness of this approach relies heavily on the stability of snow depth patterns between InSAR acquisitions.

6. Conclusion and Future Work

This study serves as a proof of concept for the potential of machine learning to estimate snow depth using L-band InSAR data. The XGBoost model demonstrated promising performance, and the feature importance analysis provided valuable insights into the relationships between L-band InSAR features and snow depth. The upcoming NISAR mission, with its global L-band InSAR coverage, presents a unique opportunity to further advance this approach. With the availability of NISAR data, we can expand the training dataset, incorporate additional polarizations, and explore alternative machine learning approaches, potentially leading to even more accurate snow depth estimation. The success of this approach hinges on the availability of representative training data. With sufficient, representative, and diverse training data, machine learning models can effectively capture the complex relationships between snow depth and its influencing factors, enabling accurate snow depth prediction. With the advent of NISAR and continued research efforts, we can harness the power of machine learning to potentially improve water resource management.

Moving forward, we recommend that future work focus on:

- Exploring model transferability: Can we develop models that perform well on new datasets without the need for retraining?
- Analyzing model performance at 10-50 meter resolution: Can we achieve similar performance using lower-resolution data, which is more widely available?

676 • Incorporating advanced computer vision approaches: Can we improve performance using 2D
677 estimation methods such as Convolutional Neural Networks (O'Shea and Nash, 2015) and
678 Vision Transformers (Dosovitskiy et al., 2020) instead of pixel-wise approaches?

679 Addressing these questions will allow us to further advance the field of machine learning-based snow
680 depth prediction and contribute to more accurate and reliable snow monitoring and forecasting systems.

7. References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631. doi: 10.1145/3292500.3330701
- Alabi, I. O., Marshall, H.-P., Mead, J., and Ofekeze, E. (2022). How Transferable are Our Models? A Case Study of Idaho SNOTEL Sites., in *AGU Fall Meeting Abstracts*, C52C--0365.
- Alabi, I. O., Marshall, H.-P., Mead, J., Trujillo, E., and Ofekeze, E. (2023). Harnessing L-Band InSAR and Lidar Data Through Machine Learning for Accurate Snow Depth Estimation in Grand Mesa, Colorado. *AGU23*.
- Aquino, C., Mitchard, E., McNicol, I., Carstairs, H., Burt, A., Vilca, B. L. P., et al. (2021). Using Experimental Sites in Tropical Forests to Test the Ability of Optical Remote Sensing to Detect Forest Degradation at 0.3–30 M Resolutions., in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, (IEEE), 677–680.
- Bair, E. H., Abreu Calfa, A., Rittger, K., and Dozier, J. (2018). Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *Cryosphere* 12, 1579–1594.
- Broxton, P. D., Van Leeuwen, W. J. D., and Biederman, J. A. (2019). Improving snow water equivalent maps with machine learning of snow survey and lidar measurements. *Water Resour Res* 55, 3739–3757.
- Brucker, L., Gsfc, N., Gestar, U., Hiemstra, C., Marshall, H., Uni, B. S., et al. (2017a). A FIRST OVERVIEW OF SNOWEX GROUND-BASED REMOTE SENSING ACTIVITIES DURING THE WINTER 2016–2017 Colorado ; Richard Kelly , Uni . Waterloo ; Jason Kraft , NASA GSFC ; Alexandre Langlois , Uni . Sherbrooke ; Daniel McGrath , Colorado State Uni . ; Chelsea Merr. 1391–1394.
- Brucker, L., Hiemstra, C., Marshall, H., and Elder, K. (2018). SnowEx17 Community Snow Depth Probe Measurements, Version 1. doi: 10.5067/WKC6VFMT7JTF
- Brucker, L., Hiemstra, C., Marshall, H.-P., Elder, K., De Roo, R., Mousavi, M., et al. (2017b). A first overview of SnowEx ground-based remote sensing activities during the winter 2016--2017., in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1391–1394.
- Bühler, Y., Adams, M. S., Bösch, R., and Stoffel, A. (2016). Mapping snow depth in alpine terrain with unmanned aerial systems (UASs): potential and limitations. *Cryosphere* 10, 1075–1088.
- Chen, L., and Wang, L. (2018). Recent advance in earth observation big data for hydrology. *Big Earth Data* 2, 86–107.

716 Chen, T., and Guestrin, C. (2016). Xgboost: A scalable tree boosting system., in *Proceedings of the*
717 *22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

718 Chesnutt, J. M., Wegmann, K. W., Cole, R. D., and Byrne, P. K. (2017). Landscape Evolution
719 Comparison between Sacra Mensa, Mars and the Grand Mesa, Colorado, USA., in *AGU Fall*
720 *Meeting Abstracts*, EP53B-1690.

721 Currier, W. R., Pflug, J., Mazzotti, G., Jonas, T., Deems, J. S., Bormann, K. J., et al. (2019).
722 Comparing aerial lidar observations with terrestrial lidar and snow-probe transects from
723 NASA’s 2017 SnowEx campaign. *Water Resour Res* 55, 6285–6294.

724 Deeb, E. J., Forster, R. R., and Kane, D. L. (2011). Monitoring snowpack evolution using
725 interferometric synthetic aperture radar on the North Slope of Alaska, USA. *Int J Remote Sens*
726 32, 3985–4003.

727 Deems, J. S., Fassnacht, S. R., and Elder, K. J. (2006). Fractal distribution of snow depth from
728 LiDAR data. *J Hydrometeorol* 7, 285–297.

729 Deems, J. S., Fassnacht, S. R., and Elder, K. J. (2008). Interannual consistency in fractal snow depth
730 patterns at two Colorado mountain sites. *J Hydrometeorol* 9, 977–988.

731 Deems, J. S., Painter, T. H., and Finnegan, D. C. (2013). Lidar measurement of snow depth: a review.
732 *Journal of Glaciology* 59, 467–479.

733 Dettinger, M. D., and Anderson, M. L. (2015). Storage in California’s reservoirs and snowpack in
734 this time of drought. *San Francisco Estuary and Watershed Science* 13.

735 Dong, C. (2018). Remote sensing, hydrological modeling and in situ observations in snow cover
736 research: A review. *J Hydrol (Amst)* 561, 573–583.

737 Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020).
738 An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint*
739 *arXiv:2010.11929*.

740 Feng, T., Zhu, S., Huang, F., Hao, J., Mind’je, R., Zhang, J., et al. (2022). Spatial variability of snow
741 density and its estimation in different periods of snow season in the middle Tianshan Mountains,
742 China. *Hydrol Process* 36, e14644.

743 Ferraz, A., Saatchi, S., Bormann, K. J., and Painter, T. H. (2018). Fusion of NASA Airborne Snow
744 Observatory (ASO) lidar time series over mountain forest landscapes. *Remote Sens (Basel)* 10,
745 164.

746 Gatebe, C., Li, W., Chen, N., Fan, Y., Poudyal, R., Brucker, L., et al. (2018). Snow-covered area
747 using machine learning techniques., in *IGARSS 2018-2018 IEEE International Geoscience and*
748 *Remote Sensing Symposium*, (IEEE), 6291–6293.

749 Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach Learn* 63, 3–42.

- 750 Guneriussen, T., Hogda, K. A., Johnsen, H., and Lauknes, I. (2001). InSAR for estimation of changes
751 in snow water equivalent of dry snow. *IEEE Transactions on Geoscience and Remote Sensing*
752 39, 2101–2108.
- 753 Harder, P., Pomeroy, J. W., and Helgason, W. D. (2020). Improving sub-canopy snow depth
754 mapping with unmanned aerial vehicles: lidar versus structure-from-motion techniques.
755 *Cryosphere* 14, 1919–1935.
- 756 Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., et al. (2018).
757 Direct insertion of NASA Airborne Snow Observatory-derived snow depth time series into the
758 iSnobal energy balance snow model. *Water Resour Res* 54, 8045–8063.
- 759 Henn, B., Musselman, K. N., Lestak, L., Ralph, F. M., and Molotch, N. P. (2020). Extreme runoff
760 generation from atmospheric river driven snowmelt during the 2017 Oroville Dam spillways
761 incident. *Geophys Res Lett* 47, e2020GL088189.
- 762 Hojatimalekshah, A., Uhlmann, Z., Glenn, N. F., Hiemstra, C. A., Tennant, C. J., Graham, J. D., et al.
763 (2021). Tree canopy and snow depth relationships at fine scales with terrestrial laser scanning.
764 *Cryosphere* 15, 2187–2209.
- 765 Hoppinen, Z., Oveisgharan, S., Marshall, H., Mower, R., Elder, K., and Vuyovich, C. (2023). Snow
766 Water Equivalent Retrieval Over Idaho , Part B : Using L-band UAVSAR Repeat-Pass
767 Interferometry. 1–24.
- 768 Hosseini, S., and Garestier, F. (2021). Pol-InSAR sensitivity to hemi-boreal forest structure at L-and
769 P-bands. *International Journal of Applied Earth Observation and Geoinformation* 94, 102213.
- 770 Hu, Y., Che, T., Dai, L., and Xiao, L. (2021). Snow depth fusion based on machine learning methods
771 for the Northern Hemisphere. *Remote Sens (Basel)* 13, 1250.
- 772 Kellogg, K., Hoffman, P., Standley, S., Shaffer, S., Rosen, P., Edelstein, W., et al. (2020). NASA-
773 ISRO synthetic aperture radar (NISAR) mission., in *2020 IEEE Aerospace Conference*, (IEEE),
774 1–21.
- 775 Killinger, D. K. (2014). “Lidar (light detection and ranging),” in *Laser spectroscopy for sensing*,
776 (Elsevier), 292–312.
- 777 Kim, E., Gatebe, C., Hall, D., Newlin, J., Misakonis, A., Elder, K., et al. (2017). OVERVIEW OF
778 SNOWEX YEAR 1 ACTIVITIES NASA Goddard Space Flight Center , 2 USRA , 3 Aerospace
779 Corp ., 4 ATA Aerospace , 5 US Forest Service ., *Igarss*, 1388–1390.
- 780 King, F., Duffy, G., and Fletcher, C. G. (2022a). A centimeter-wavelength snowfall retrieval
781 algorithm using machine learning. *J Appl Meteorol Climatol* 61, 1029–1039.
- 782 King, F., Kelly, R., and Fletcher, C. G. (2022b). Evaluation of lidar-derived snow depth estimates
783 from the iPhone 12 pro. *IEEE Geoscience and Remote Sensing Letters* 19, 1–5.
- 784 King, J., Kelly, R., Kasurak, A., Duguay, C., Gunn, G., Rutter, N., et al. (2015). Spatio-temporal
785 influence of tundra snow properties on Ku-band (17.2 GHz) backscatter. *Journal of Glaciology*
786 61, 267–279.

Kingma, D. P., and Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Kulakowski, D., Veblen, T. T., and Drinkwater, S. (2004). The persistence of quaking aspen (*Populus tremuloides*) in the Grand Mesa area, Colorado. *Ecological Applications* 14, 1603–1614.

LAL, P., Singh, G., Das, N. N., and Entekhabi, D. (2022). A data-driven snapshot algorithm for high-resolution soil moisture retrievals for the upcoming NISAR mission., in *AGU Fall Meeting Abstracts*, H42G--1379.

Leinss, S., Wiesmann, A., Lemmetyinen, J., and Hajnsek, I. (2015). Snow water equivalent of dry snow measured by differential interferometry. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8, 3773–3790.

Li, H., Wang, Z., He, G., and Man, W. (2017). Estimating snow depth and snow water equivalence using repeat-pass interferometric SAR in the northern piedmont region of the Tianshan Mountains. *J Sens* 2017. doi: 10.1155/2017/8739598

Liang, J., Liu, X., Huang, K., Li, X., Shi, X., Chen, Y., et al. (2015). Improved snow depth retrieval by integrating microwave brightness temperature and visible/infrared reflectance. *Remote Sens Environ* 156, 500–509. doi: 10.1016/j.rse.2014.10.016

Lievens, H., Brangers, I., Marshall, H.-P., Jonas, T., Olefs, M., and De Lannoy, G. (2022). Sentinel-1 snow depth retrieval at sub-kilometer resolution over the European Alps. *Cryosphere* 16, 159–177.

Livneh, B., and Badger, A. M. (2020). Drought less predictable under declining future snowpack. *Nat Clim Chang* 10, 452–458.

Lu, Z., Kwoun, O., and Rykhus, R. (2007). Interferometric synthetic aperture radar (InSAR): its past, present and future. *Photogramm Eng Remote Sensing* 73, 217.

Man, X., and Chan, E. (2021). The best way to select features? comparing mda, lime, and shap. *The Journal of Financial Data Science Winter* 3, 127–139.

Marshall, H., Hawley, R. L., and Tedesco, M. (2015). Field measurements for remote sensing of the cryosphere. *Remote Sensing of the Cryosphere*, 345–381.

Marshall, H.-P., Deeb, E., Forster, R., Vuyovich, C., Elder, K., Hiemstra, C., et al. (2021). L-band InSAR depth retrieval during the NASA SnowEx 2020 campaign: Grand mesa, Colorado., in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, (IEEE), 625–627.

Mason, M. A. (2020). Snow Depth Distribution Patterns and Consistency from Airborne Lidar Time Series.

Maxwell, A. E., Warner, T. A., and Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *Int J Remote Sens* 39, 2784–2817. doi: 10.1080/01431161.2018.1433343

824 Meloche, J., Langlois, A., Rutter, N., McLennan, D., Royer, A., Billecocq, P., et al. (2022). High-
825 resolution snow depth prediction using Random Forest algorithm with topographic parameters:
826 A case study in the Greiner watershed, Nunavut. *Hydrol Process* 36. doi: 10.1002/hyp.14546

827 Meyer, J., Deems, J. S., Bormann, K. J., Shean, D. E., and Skiles, S. M. (2022). Mapping snow depth
828 and volume at the alpine watershed scale from aerial imagery using Structure from Motion.
829 *Front Earth Sci (Lausanne)* 10, 989792.

830 Ofekeze, E., Marshall, H.-P., Mead, J., and Alabi, I. O. (2022). A Machine Learning Framework for
831 Active and Passive Microwave Observation from Snow Water Equivalent Synthetic Aperture
832 Radar And Radiometer (SWESARR) for Snow Depth Estimation., in *AGU Fall Meeting*
833 *Abstracts*, C52C--0364.

834 Ofekeze, E., Marshall, H.-P., Mead, J., Trujillo, E., and Alabi, I. O. (2023). Advancing Snow Water
835 Equivalent Estimation with SWESARR. *AGU23*.

836 O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint*
837 *arXiv:1511.08458*.

838 Oveisgharan, S., Zinke, R., Hoppinen, Z., and Marshall, H. P. (2024). Snow water equivalent
839 retrieval over Idaho—Part 1: Using Sentinel-1 repeat-pass interferometry. *Cryosphere* 18, 559–
840 574.

841 Ozdemir, S. (2022). *Feature Engineering Bookcamp*. Simon and Schuster.

842 Pflug, J. M., and Lundquist, J. D. (2020). Inferring distributed snow depth by leveraging snow
843 pattern repeatability: Investigation using 47 lidar observations in the Tuolumne watershed,
844 Sierra Nevada, California. *Water Resour Res* 56, e2020WR027243.

845 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N. (2019). Deep
846 learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204.

847 Revuelto, J., López-Moreno, J. I., Azorin-Molina, C., Zabalza, J., Arguedas, G., and Vicente-
848 Serrano, S. M. (2014). Mapping the annual evolution of snow depth in a small catchment in the
849 Pyrenees using the long-range terrestrial laser scanning. *J Maps* 10, 379–393.

850 Schirmer, M., and Lehning, M. (2011). Persistence in intra-annual snow depth distribution: 2. Fractal
851 analysis of snow depth development. *Water Resour Res* 47.

852 Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., and Pulwarty, R. S. (1999).
853 Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL)
854 data. *Water Resour Res* 35, 2145–2160.

855 Simpkins, G. (2018). Snow-related water woes. *Nat Clim Chang* 8, 945.

856 Sinha, S., Jeganathan, C., Sharma, L. K., and Nathawat, M. S. (2015). A review of radar remote
857 sensing for biomass estimation. *International Journal of Environmental Science and Technology*
858 12, 1779–1792.

859 Sturm, M., and Holmgren, J. (2018a). An Automatic Snow Depth Probe for Field Validation
860 Campaigns. *Water Resour Res* 54, 9695–9701. doi: 10.1029/2018WR023559

861 Sturm, M., and Holmgren, J. (2018b). An automatic snow depth probe for field validation campaigns.
862 *Water Resour Res* 54, 9695–9701.

863 Sturm, M., and Wagner, A. M. (2010). Using repeated patterns in snow distribution modeling: An
864 Arctic example. *Water Resour Res* 46.

865 Taheri, M., and Mohammadian, A. (2022). An overview of snow water equivalent: Methods,
866 challenges, and future outlook. *Sustainability* 14, 11395.

867 Tarricone, J., Webb, R. W., Marshall, H. P., Nolin, A. W., and Meyer, F. J. (2023). Estimating snow
868 accumulation and ablation with L-band interferometric synthetic aperture radar (InSAR).
869 *Cryosphere* 17, 1997–2019. doi: 10.5194/tc-17-1997-2023

870 Tedesco, M., Pulliainen, J., Takala, M., Hallikainen, M., and Pampaloni, P. (2004). Artificial neural
871 network-based techniques for the retrieval of SWE and snow depth from SSM/I data. *Remote*
872 *Sens Environ* 90, 76–85.

873 Trujillo, E., and Lehning, M. (2015). Theoretical analysis of errors when estimating snow
874 distribution through point measurements. *Cryosphere* 9, 1249–1264.

875 Trujillo, E., Ramirez, J. A., and Elder, K. J. (2007). Topographic, meteorologic, and canopy
876 controls on the scaling characteristics of the spatial distribution of snow depth fields. *Water*
877 *Resour Res* 43.

878 Trujillo, E., Ramirez, J. A., and Elder, K. J. (2009). Scaling properties and spatial organization of
879 snow depth fields in sub-alpine forest and alpine tundra. *Hydrological Processes: An*
880 *International Journal* 23, 1575–1590.

881 Vano, J. A. (2020). Implications of losing snowpack. *Nat Clim Chang* 10, 388–390.

882 Wang, J., Yuan, Q., Li, T., Shen, H., and Zhang, L. (2019). Estimating snow-depth by fusing satellite
883 and station observations: A deep learning approach., in *IGARSS 2019-2019 IEEE International*
884 *Geoscience and Remote Sensing Symposium*, (IEEE), 4109–4112.

885 Webb, R. W., Marziliano, A., McGrath, D., Bonnell, R., Meehan, T. G., Vuyovich, C., et al. (2021).
886 In situ determination of dry and wet snow permittivity: improving equations for low frequency
887 radar applications. *Remote Sens (Basel)* 13, 4617.

888 Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al. (2020). Deep learning in environmental
889 remote sensing: Achievements and challenges. *Remote Sens Environ* 241, 111716.

890 Zhang, B., Chen, Z., Peng, D., Benediktsson, J. A., Liu, B., Zou, L., et al. (2019). Remotely sensed
891 big data: Evolution in model development for information extraction [point of view].
892 *Proceedings of the IEEE* 107, 2294–2301.

893 Zhang, K., Zhang, Y., and Wang, M. (2012). A Unified Approach to Interpreting Model Predictions
894 Scott. *Nips* 16, 426–430.

895 Zhao, M., and Li, J. (2018). Tuning the hyper-parameters of CMA-ES with tree-structured Parzen
896 estimators. *Proceedings - 2018 10th International Conference on Advanced Computational*
897 *Intelligence, ICACI 2018*, 613–618. doi: 10.1109/ICACI.2018.8377530
898