https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

IceWatch: a multimodal deep-learning framework for monitoring glacial lake outburst floods, evaluated at Shisper Glacier, Karakoram

Zuha Fatima[1], **Muhammad Anser Sohaib**[1], **Muhammad Talha**[1], **Ayesha Kanwal**[1], **Sidra Sultana**[1], **and Nazia Perwaiz**[1]

[1]School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad, Pakistan

Correspondence: Nazia Perwaiz (nazia.perwaiz@seecs.edu.pk)

Abstract. Lakes dammed by surging glaciers can fill and drain within one season, often under cloud. We present IceWatch, which combines a convolutional classifier of Sentinel-2 scenes (RiskFlow), forecasts of glacier velocity (TerraFlow) and land-surface temperature (TempFlow), and a probabilistic fusion stage (FusionFlow). We test it at Shisper Glacier, Karakoram, where an ice-dammed lake drained four times between 2019 and 2022, with all data taken over the lake and its dam, and scenes labelled from a published chronology. Trained on earlier years and tested on 2022 to 2025, RiskFlow detected both post-outburst scenes of 2022 with 2 false alarms in 102 other scenes. With each year held out in turn, it detected 4 of 7 post-outburst scenes with no false alarm in 151 others, whereas a water-index rule detected all 7 but raised 26 false alarms. A model trained on 103 other ice-dammed lakes did not transfer. TerraFlow was 30 % more accurate than persistence, and TempFlow beat persistence and climatology two days ahead. Three of the four outbursts followed cooler-than-normal weeks, and neither an agreement rule nor the fusion improved on RiskFlow. IceWatch can confirm drainage from the next clear scene; anticipating outbursts will require observations of lake filling.

1 Introduction

Glacial lake outburst floods (GLOFs) are sudden releases of water from lakes impounded by glacier ice or moraines, typically triggered by glacier surging and retreat, dam degradation or accelerated melt (Osti and Egashira, 2009). A single event can release millions of cubic metres of water within hours, with peak discharges reaching the order of 10^4 m^3 s^{-1} (Richardson and Reynolds, 2000), and recent global assessments estimate that millions of people live within the potential reach of such floods (Taylor et al., 2023). The hazard is concentrated in high-mountain regions such as the Himalaya, the Karakoram and the Andes, where settlements, roads and hydropower infrastructure occupy narrow valley floors downstream of expanding glacial lakes (Bajracharya and Mool, 2009). Timely detection and, ideally, anticipation of outburst conditions are therefore central concerns of disaster risk reduction and climate adaptation in these environments (Emmer et al., 2022).

Operational monitoring still relies largely on hydrological breach models, threshold-based water and land indices such as the normalised difference water index (NDWI) and its modified counterpart (MNDWI), and expert interpretation of satellite

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

imagery (Worni et al., 2013; Anand et al., 2025; Javed et al., 2025), a lineage extending back to early remote-sensing hazard appraisals in the Himalaya (Quincey et al., 2007). Classical machine-learning classifiers, notably support vector machines and random forests, have improved glacial-lake mapping, including in the Karakoram (Abbas et al., 2025), but struggle with small lakes, cloud contamination and heterogeneous terrain. Deep segmentation networks such as U-Net and DeepLab raise spatial accuracy considerably, with intersection-over-union scores above 0.85 (Cao et al., 2024; Xie et al., 2022), and transformer-based models now generalise glacier mapping across large regions (Maslov et al., 2025). These systems, however, are mapping-centric: they characterise the present state of lakes and glaciers rather than the forward evolution of outburst risk (Shirsath and Bhojane, 2024), and vision-only pipelines integrate poorly with the physical glacier dynamics that ultimately govern dam failure, a limitation compounded by the opacity of deep models in safety-critical settings (Samek et al., 2017).

Recent reviews point in the same direction: multi-sensor data and deep architectures are transforming glacial-lake mapping (Nurakynov et al., 2025; Li et al., 2024; Sahu, 2025), yet forecasting remains the open problem. Deep time-series models offer the complementary capability: convolutional networks have accelerated fluvial-flood prediction relative to hydraulic models (Kabir et al., 2020), and learned approaches to ice-flow prediction suggest that forward modelling of glacier dynamics is feasible (Min and Mukkavilli, 2019), but these strands have not been coupled with visual evidence and tested against a series of documented outbursts.

In response, we propose IceWatch, a dual-stream framework for GLOF monitoring. A vision stream, RiskFlow, applies a convolutional neural network to Sentinel-2 multispectral scenes to detect the state of the lake after an outburst. A tabular stream models the physical state of the glacier with two modules: TerraFlow, a Transformer encoder that forecasts surface velocity from NASA ITS_LIVE records, and TempFlow, a stacked bidirectional long short-term memory (LSTM) network that forecasts Moderate Resolution Imaging Spectroradiometer (MODIS) land-surface temperature (LST). A fusion stage, FusionFlow, combines the two streams. The framework builds on GLOFNet (Fatima et al., 2025), our previously published multimodal dataset for Shisper Glacier, but for this study all three data streams were extracted again from the source archives over a common area that contains the lake, the ice dam and the lower glacier tongue, so that image, velocity and temperature describe the same place. Scenes are labelled from the outburst chronology published by Lovell and Muhammad (2025).

The objectives of this article are threefold: (i) to test whether a convolutional classifier recognises the state of the lake after an outburst in single Sentinel-2 scenes, compared with a simple water-index rule and with a model trained on other glacial lakes; (ii) to evaluate the velocity and temperature forecasts against held-out data and against persistence and climatology baselines; and (iii) to assess whether velocity and temperature add information about outbursts to the image stream, under the agreement rule of the original design and under a calibrated probabilistic fusion, across the four Shisper outbursts of 2019 to 2022.

The remainder of this article is organised as follows. Section 2 reviews related work. Section 3 describes the study area, the data and the IceWatch methodology. Section 4 presents the results, including the baselines and the 2022 outburst season. Section 5 discusses the outburst mechanism, implications for monitoring and limitations, and Sect. 6 concludes.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

2 Related work

2.1 Static GLOF mapping and classical machine-learning detection

Early GLOF assessment relied on manual mapping, empirical breach modelling and static lake inventories (Das et al., 2024; Anand et al., 2025), with foundational regional frameworks emerging from inventory work in the Himalaya (Bajracharya and Mool, 2009) and early satellite-based hazard recognition (Quincey et al., 2007). The methodological challenges of mapping glacier parameters from space are long documented (Racoviteanu et al., 2009), and the global scale of exposure has since been quantified (Taylor et al., 2023). Machine-learning classifiers such as support vector machines and random forests were subsequently applied to optical archives (Abbas et al., 2025) but generalised poorly across terrain and time, while object-based image analysis proved sensitive to noise and required extensive manual tuning (Cao et al., 2024).

2.2 Deep learning for glacier and glacial-lake segmentation

Convolutional architectures pushed glacier and lake delineation to intersection-over-union scores above 0.85 (Cao et al., 2024; Xie et al., 2022), with GlacierNet2 integrating hydrological post-processing to handle debris-covered ice (Xie et al., 2022). Attention-based and multi-sensor fusion networks incorporating Sentinel-1, Sentinel-2 and digital elevation models improved generalisation across regions (Diaconu et al., 2024; Holzmann et al., 2021); spectral-anomaly-oriented methods improved small-lake detection (Sharma et al., 2024); and globally scalable transformer segmentation now matches expert delineation accuracy (Maslov et al., 2025). Comprehensive reviews summarise this rapid progress (Sahu, 2025; Nurakynov et al., 2025). All of these systems, however, remain static: they segment the present, but do not forecast.

2.3 Time-series deep learning for glacier dynamics

Long short-term memory networks (Hochreiter and Schmidhuber, 1997) consistently outperform classical statistical models for cryospheric time series such as snow cover and temperature. Deep learning likewise improves glacier-velocity estimation from optical imagery (Zandler et al., 2025), fully convolutional networks track internal ice layers and thickness (Rahnemoonfar et al., 2020), and unsupervised approaches simulate ice-flow evolution (Min and Mukkavilli, 2019). These studies demonstrate the capacity of learned temporal models, but they stop short of integrating visual evidence or of testing their output against documented outbursts.

2.4 Hybrid and multimodal frameworks for GLOF risk

A small number of studies combine spatial and physical indicators. Bazai et al. (2024) correlated glacier surge velocity with lake volume to anticipate debris flows in the Karakoram; Abbas et al. (2025) coupled machine learning with process-based indicators; Nurakynov et al. (2025) merged InSAR-derived deformation, elevation-change metrics and lake-expansion measures in a multisensor hazard assessment; and Tom et al. (2025) applied machine learning to global lake inventories, albeit retrospectively. Reviews of the remote-sensing landscape retain a mapping emphasis (Li et al., 2024). What is still missing is

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

a direct test of whether physical channels such as glacier velocity and surface temperature carry information about outburst timing beyond what imagery already shows. IceWatch is designed to make that test explicit at a well-documented site.

3 Study area, data and methods

3.1 Study area

Shisper Glacier (also written Shishper or Shispare) is a surge-type glacier in the Hunza valley of the Karakoram, Pakistan (Fig. 1), whose recent surge activity impounded an ice-dammed lake that has produced repeated outburst floods affecting the Hassanabad valley and the Karakoram Highway. Its terminus lies about 4 km above the highway at Hassanabad. The glacier began to surge in 2017 and reached surface speeds of about 18 m d^{-1} in early June 2018 and again in late April and early May 2019 (Bhambri et al., 2020). The advancing terminus blocked the meltwater stream from the neighbouring Muchuhar Glacier, and an ice-dammed lake began to form in November 2018 (Bhambri et al., 2020). From Sentinel-2 imagery of 2017 to 2024, Lovell and Muhammad (2025) documented four lake seasons, from 2018/19 to 2021/22. In each of them the lake filled from late in the year, reached its largest extent in May and drained between May and July, usually within one to two days, and the remaining water had gone by June or July. The outbursts occurred on 21 to 23 June 2019, on 29 May 2020, between 10 and 20 May 2021 and on 7 May 2022, when the flood destroyed the bridge at Hassanabad. No lake formed in 2023 or up to September 2024 (Lovell and Muhammad, 2025). The glacier's documented transition from quiescence through surge to lake formation and outburst, together with dense multi-sensor coverage, makes it a demanding and well-observed testbed: four well-dated outbursts from one lake, a long velocity record that spans a full surge and a dense optical archive.

3.2 Data

Table 1 summarises the data sets, and Fig. 2 shows where they lie.

Sentinel-2. We took the acquisition dates of the GLOFNet Sentinel-2 archive between 17 December 2018 and 5 March 2025 (Fatima et al., 2025) and exported the corresponding Level-2A surface-reflectance scenes again from the Microsoft Planetary Computer for 74.50 to 74.62° E and 36.30 to 36.40° N, at 10 m in UTM zone 43N. This gave 158 scenes; two dates (25 January 2022 and 24 January 2025) could not be exported, and neither falls after an outburst. Six bands were kept (B4, B3, B2, B8, B11 and B12), and the reflectance offset introduced with processing baseline 04.00 was removed at export. RiskFlow sees a window of 5 by 5 km centred on the lake (36.3606° N, 74.5628° E, mapped as open water on 31 May 2019), which also contains the ice dam and the start of the outflow channel. The window was fixed from the imagery before any model was trained and is resampled to 128×128 pixels. To check whether clearer scenes existed than those in the archive, we screened all 454 Level-2A acquisitions over the window between December 2018 and March 2025 with the scene classification layer of the Sen2Cor processor (Main-Knorn et al., 2017). Of the 104 clear acquisitions (at most 10 % cloud or shadow), 97 were already in the archive and none of the other seven falls within 30 days after an outburst. All six acquisitions in the 30 days after the

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

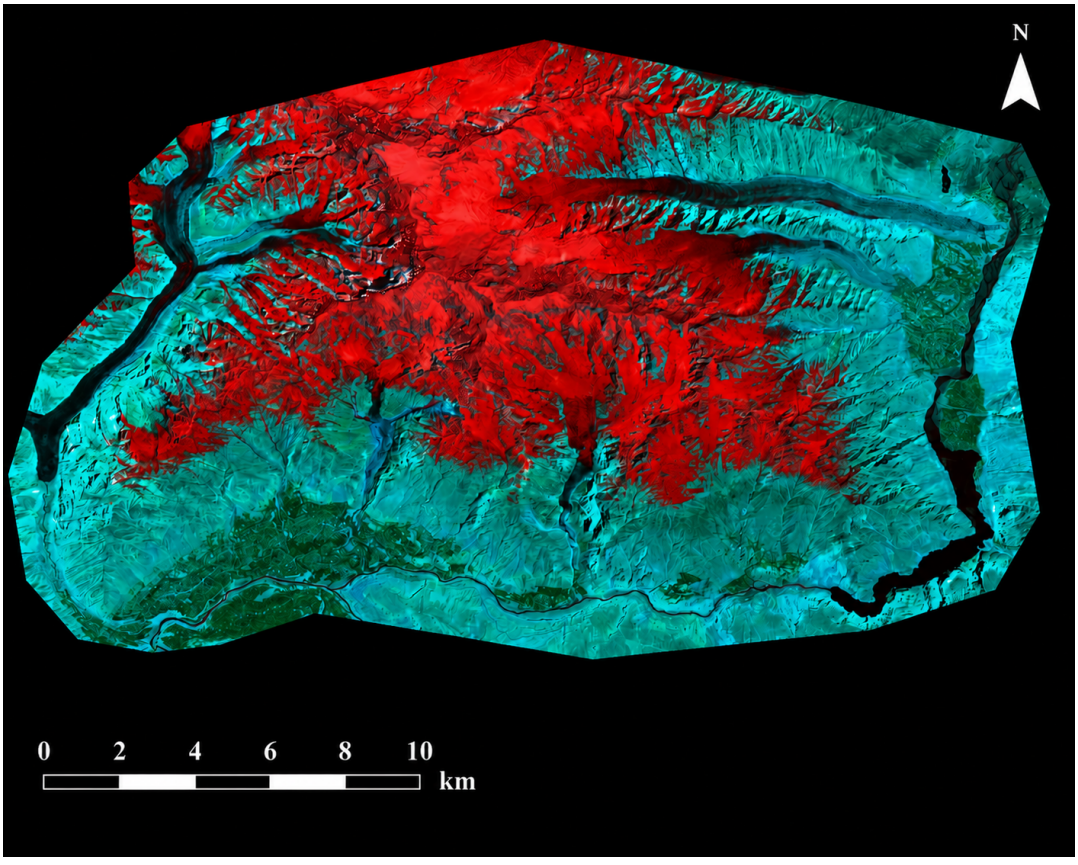


Figure 1. Study area: false-colour Sentinel-2 composite of Shisper Glacier, Hunza valley, Karakoram (acquired 5 July 2019). The footprints of the data used in this study are shown in Fig. 2.

2020 outburst were 66 to 100 % cloud covered over the lake. The same screening of the clear acquisitions from March 2025 to October 2026 found no lake at the dam site, so the record contains no later outburst of this kind.

Labels. A scene is labelled GLOF if it was acquired within 30 days after an outburst, using 22 June 2019, 29 May 2020, 15 May 2021 (the middle of the 10 to 20 May drainage) and 7 May 2022 as outburst dates (Lovell and Muhammad, 2025). Seven of the 158 scenes are GLOF scenes: 5, 15 and 20 July 2019, 4 and 9 June 2021, and 10 and 25 May 2022; the 2020 outburst has none because of cloud. A water test was applied to every scene: pixels within 1.5 km of the lake count as water when their NDWI (McFeeters, 1996) exceeds 0.15, their B8 reflectance is below 0.15 and their B11 reflectance is below 0.10. The test confirms that the seven GLOF scenes show a drained basin: open water covered 0.12 to 0.23 km^2 in the last clear scene before each outburst and at most 0.015 km^2 in all seven GLOF scenes.

Velocity. Image-pair velocities were taken from the ITS_LIVE version 2 data cube (Gardner et al., 2025) for every pixel of the cube's glacier mask within 74.54 to 74.63° E and 36.34 to 36.41° N (540 pixels) and for a 1.2 km margin around it (2493 pixels), because the mask predates the 2018 to 2019 advance and the margin contains the ice dam. The 49.4 million records

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1. Data used in this study.

Stream	Source	Area	Period and size	Used by
Sentinel-2 L2A	Microsoft Planetary Computer	5 by 5 km window at the lake	158 dates, Dec 2018 to Mar 2025	RiskFlow
ITS_LIVE version 2	data cube (Gardner et al., 2025)	3033 pixels on the tongue and ice dam	Dec 1988 to May 2025, 10.97 million pixel-days	TerraFlow
MODIS MOD11A1 v6.1	Terra daytime LST (Wan et al., 2021)	8×8 km grid at the lake (64 pixels)	Feb 2000 to Dec 2022	TempFlow
MODIS MOD11A1 v6.1	original TempFlow points	14 km north-west of the lake (9 pixels)	Feb 2000 to Dec 2022	TempFlow (comparison)
Sentinel-2 L2A, other lakes	GLOF database v4.2 (Lützow et al., 2023)	103 ice-dammed lakes	2017 to 2025, 3973 scenes	RiskFlow (transfer test)

from December 1988 to May 2025 were summarised to a daily mean and maximum speed for each pixel and observation day (10.97 million pixel-days; Fig. 3). Each velocity is dated by the middle of its image pair.

Temperature. Daytime LST was taken from the MODIS Terra MOD11A1 collection 6.1 product (Wan et al., 2021) at 1 km resolution on an 8×8 grid of points centred on the lake (64 distinct pixels on tiles h23v05 and h24v05), from 24 February 2000 to 1 January 2023; 41.6 % of pixel-days are cloud free. LST after 2022 was not used. Terra made its last inclination manoeuvre in 2020, and its overpass has since drifted earlier, passing 10:15 local time in autumn 2022 (NASA, 2026); an earlier overpass sees a surface that has had less time to warm, which would cool daytime LST for reasons unrelated to the glacier. For comparison, we also report TempFlow on the points of its original training data: 64 points at 36.456 to 36.489° N and 74.490 to 74.496° E, 12 to 16 km north-west of the lake, which fall in 9 distinct MODIS pixels.

Other lakes. For a transfer test, we selected outbursts of ice-dammed lakes from version 4.2 of the global GLOF database (Lützow et al., 2023): events between 2017 and 2025 dated to within 10 days, more than 30 km from Shisper, with at most six per lake. Sentinel-2 scenes from 60 days before to 60 days after each outburst were prepared in the same way as for Shisper (5 km window, same bands, at most 20 % cloud and at least 95 % valid data, and GLOF scenes defined as for Shisper), and scenes in the uncertain period of any recorded outburst at the lake were left out. Keeping only outbursts whose lake was visible before drainage (at least 0.02 km^2 of water in the preceding 60 days) gave 3973 scenes at 103 lakes, of which 925 scenes from 280 outbursts are GLOF scenes. Alaska supplies 61 % of the prepared scenes.

3.3 Framework overview

IceWatch is a dual-stream architecture (Fig. 4). The vision stream (RiskFlow, Sect. 3.4) assigns each Sentinel-2 scene a probability of GLOF conditions. The tabular stream models the physical state of the glacier with two modules: TerraFlow, which forecasts surface velocity (Sect. 3.6), and TempFlow, which forecasts surface temperature (Sect. 3.7). Their forecasts are

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

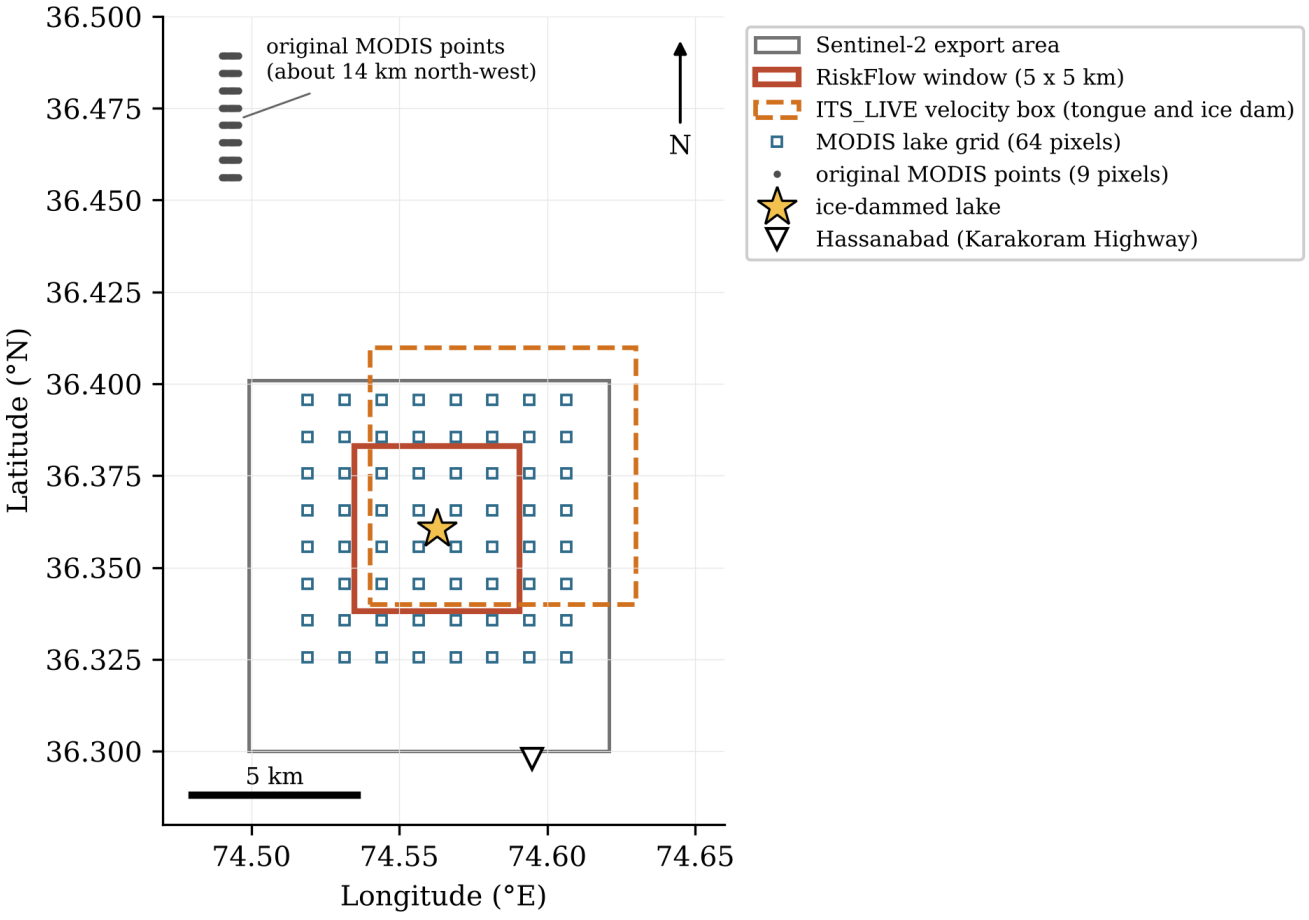


Figure 2. Footprints of the data at Shisper: the Sentinel-2 export area, the 5 by 5 km RiskFlow window centred on the ice-dammed lake, the box within which ITS_LIVE glacier and margin pixels were taken, the 8×8 MODIS lake grid and, for comparison, the original MODIS points about 14 km north-west of the lake. The triangle marks Hassanabad on the Karakoram Highway.

converted to anomalies relative to normal conditions, which a tabular model turns into a daily probability, and FusionFlow (Sect. 3.8) combines that probability with RiskFlow. In the original design of IceWatch, a high-confidence alert was issued only when the vision and tabular streams were simultaneously elevated (the agreement rule). We keep the agreement rule as a comparison and add a calibrated combination, so that the contribution of the physical streams is measured rather than assumed.

3.4 RiskFlow: the vision stream

RiskFlow is a compact convolutional classifier operating on $128 \times 128 \times 6$ Sentinel-2 patches. Convolutional networks exploit local connectivity: each layer applies a bank of learnable filters that respond to localised patterns such as edges and textures and, at deeper layers, composite glacier and lake structures. For an input image I and a kernel K, the convolution operation is

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

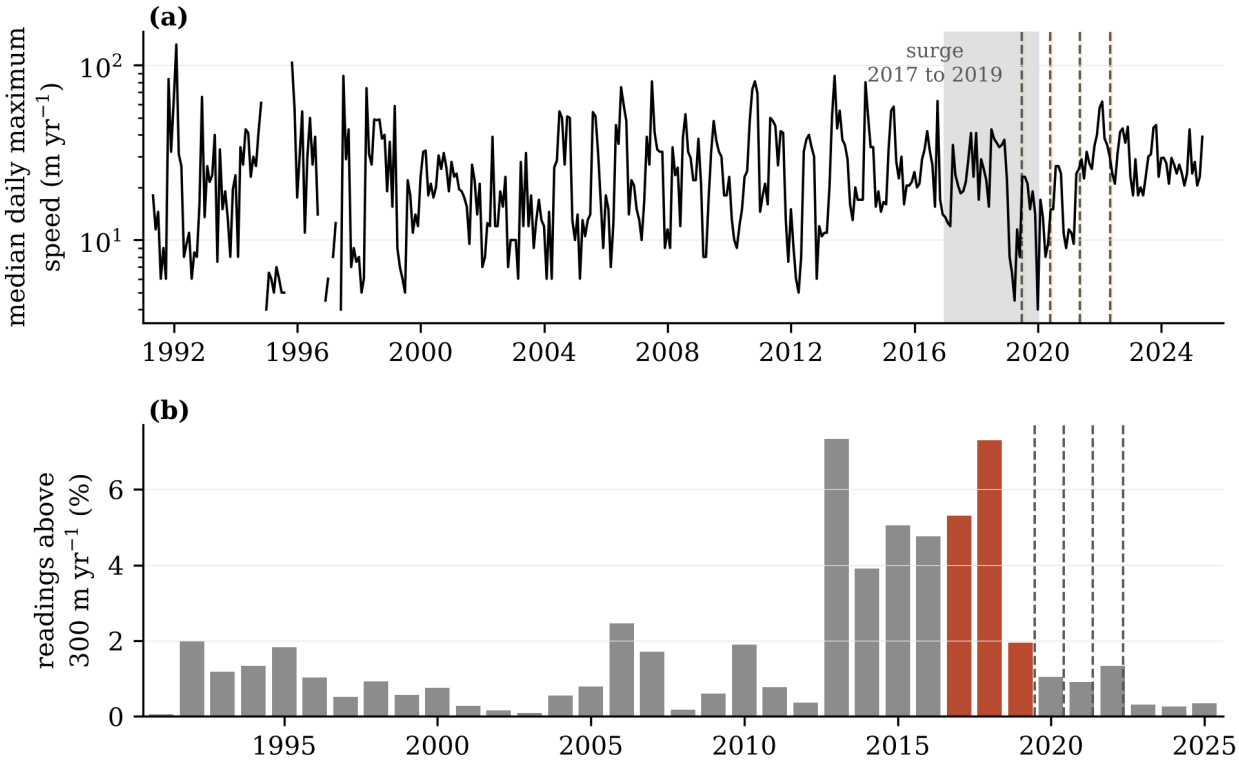


Figure 3. Glacier speed in the new ITS_LIVE extraction. (a) Median daily maximum speed over all pixels, by month (logarithmic axis); grey shading marks the surge of 2017 to 2019 described by Bhambri et al. (2020). (b) Share of pixel-days with a daily maximum above 300 m yr^{-1} in each year; the surge years are shown in red. Dashed lines mark the four outbursts. Fast readings are sparse, and the surge is not clearly separated from earlier years in this record.

defined as

$$(I * K)(x,y) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} I(x-m, y-n)\, K(m,n), \tag{1}$$

where (x,y) are pixel coordinates and (m,n) index the kernel window. Representational capacity is introduced by the rectified linear unit (ReLU),

$$f(x) = \max(0,x), \tag{2}$$

and 2×2 max-pooling after each convolutional block reduces spatial dimensionality while improving translational invariance.

The network comprises two convolutional blocks (32 and 64 filters of size 3×3, each followed by max-pooling), a flattening stage, a fully connected layer of 64 units with ReLU activation, and a single output neuron with sigmoid activation,

$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-z}}, \tag{3}$$

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

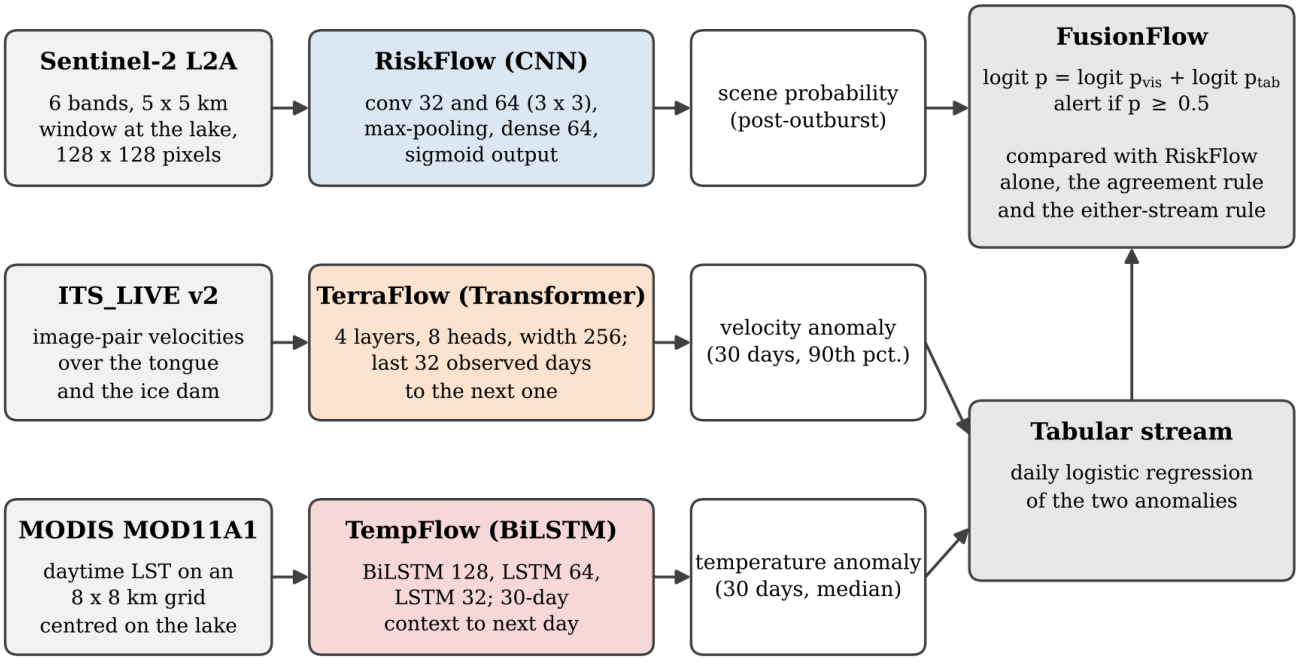


Figure 4. The IceWatch framework. RiskFlow gives each Sentinel-2 scene a probability of showing the lake after an outburst. TerraFlow and TempFlow forecast velocity and temperature, and their forecasts are summarised as anomalies over the preceding 30 days, which a daily logistic regression turns into the tabular probability. FusionFlow adds the two streams on the log-odds scale and is compared with RiskFlow alone, the agreement rule and an either-stream rule.

where z is the final linear activation, so that $\hat{y} \in (0,1)$ is interpreted as the probability of GLOF conditions, that is, that the scene was acquired within 30 days after an outburst (Sect. 3.2). The model totals 3.71 million trainable parameters (Table 2). This deliberately small architecture reflects the size of the available labelled corpus.

Given the binary nature of the task, training minimised the binary cross-entropy loss

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right], \tag{4}$$

where $y_i \in \{0,1\}$ is the ground-truth label of sample i, $\hat{y}_i$ the corresponding predicted probability and N the number of samples, using the Adam optimiser (Kingma and Ba, 2015). Adam maintains exponentially decaying estimates of the first and second moments of the gradient g_t and adapts the step size for every parameter:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\, g_t, \tag{5}$$

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 2. RiskFlow architecture and parameter counts.

Layer	Type	Output shape	Trainable parameters
conv2d	Conv2D (32 filters, 3×3, ReLU)	$126 \times 126 \times 32$	1 760
max_pooling2d	MaxPooling2D (2×2)	$63 \times 63 \times 32$	0
conv2d_1	Conv2D (64 filters, 3×3, ReLU)	$61 \times 61 \times 64$	18 496
max_pooling2d_1	MaxPooling2D (2×2)	$30 \times 30 \times 64$	0
flatten	Flatten	57 600	0
dense	Dense (ReLU)	64	3 686 464
dense_1	Dense (sigmoid)	1	65
Total			3 706 785

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\, g_t^2, \tag{6}$$

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}, \tag{7}$$

where g_t is the gradient at step t, β_1 and β_2 are the exponential decay rates of the first- and second-moment estimates m_t and v_t, $\hat{m}_t$ and $\hat{v}_t$ are their bias-corrected counterparts, η is the learning rate and ε a small constant ensuring numerical stability. Training used a batch size of 16 and a learning rate of 10^{-3}, following the time-based protocol of Sect. 3.9. Band values are scaled between the 2nd and 98th percentiles of the training scenes. The number of epochs is chosen by early stopping (patience 10, at most 60 epochs) on a validation year that is not used for training, and the model is then retrained on all training years for that number of epochs. Class imbalance was handled at the data level through augmentation rather than through loss weighting: augmentation is applied to the training scenes only, after the split, and each GLOF scene contributes 40 and each other scene 4 augmented copies (random flips, rotations by multiples of 90°, crops of 85 to 100 % of the window resized back to full size, and brightness changes of up to 10 %). Because training on a GPU is not bit-reproducible, every configuration is trained with five random seeds and the five probabilities are averaged. The classification threshold on the sigmoid output was 0.5, with 0.2 as the only alternative examined.

3.5 Reference detectors for RiskFlow

As a reference that needs no training, a water-index rule flags a scene when open water within 1.5 km of the lake covered at least 0.05 km^2 in a clear scene of the preceding 60 days and covers at most 0.02 km^2 in the current scene. The thresholds and the 60-day look-back were fixed before the rule was applied.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

To test whether outbursts elsewhere can teach the classifier what a drained lake looks like, RiskFlow was also trained with the other lakes of Sect. 3.2 in three configurations: other lakes only, applied to every Shisper scene; other lakes plus the Shisper training years, tested on Shisper from 2022 onward; and other lakes plus all but one Shisper year, for each held-out year in turn. Training used class-balanced batches with augmentation applied on the fly (flips, rotations, crops and brightness changes), 100 batches of 32 scenes per epoch, early stopping with patience 6 and at most 30 epochs, and the average of three seeds; 15 % of the other lakes were held out to choose the number of epochs together with the Shisper validation year, and the model was then retrained on all of them. Before running, we required at least 5 of the 7 Shisper GLOF scenes to be caught when years are held out, a false-alarm rate below 10 % and both 2022 scenes to be caught in the main test; otherwise the Shisper-only model would be kept.

3.6 TerraFlow: glacier-velocity forecasting

TerraFlow is a Transformer-encoder regressor (Vaswani et al., 2017) over the ITS_LIVE velocity records. Each 32-day input window consists of the previous 32 observed days of a pixel, with daily feature vectors comprising longitude, latitude, year, cyclical month and day encodings, the daily mean and maximum surface speed, and the time before the target day; the model regresses the daily maximum speed, the surge-relevant quantity, on the next observed day of the same pixel. Longitude, latitude and year are standardised on the training period, speeds are given in units of 100 m yr^{-1} and the time before the target in months. No information from the target day enters the input, so TerraFlow forecasts the kinematic state rather than reconstructing it. ITS_LIVE observes a pixel on irregular days, so the next observation is usually one to a few days ahead, and results are reported by forecast horizon. The encoder has four layers with eight attention heads each, an embedding dimension of 256, learnable positional encodings and dropout of 0.2, followed by a linear head, totalling approximately 5.5 million trainable parameters.

The input window is first projected into the model space,

$$Z_0 = XW_E + P, \tag{8}$$

where X is the input sequence, W_E the embedding matrix and P the learnable positional-encoding matrix that preserves temporal order. Each encoder layer applies multi-head self-attention (MHA), which relates every day in the window to every other and thereby captures both short- and long-range dependencies in the velocity record. For query, key and value matrices Q, K and V, scaled dot-product attention is computed as

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{9}$$

where d_k is the dimensionality of the key vectors. Multi-head attention runs h such operations in parallel over learned projections and concatenates the results,

$$\mathrm{MHA}(Q,K,V) = \mathrm{Concat}(\mathrm{head}_1,\ldots,\mathrm{head}_h)\,W^O, \tag{10}$$

where W^O projects the concatenated heads back to the model dimension, and each head is defined as

$$\mathrm{head}_i = \mathrm{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right). \tag{11}$$

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

in which W_i^Q, W_i^K and W_i^V are the learned query, key and value projection matrices of head i. The output of each attention sub-layer passes through a residual connection, layer normalisation and a position-wise feed-forward network (FFN),

$$\mathrm{FFN}(x) = \max(0, xW_1 + b_1)\,W_2 + b_2. \tag{12}$$

Stacking four such encoder layers enables TerraFlow to capture the interacting temporal and spatial structure of glacier velocity, and a final linear head maps the encoded representation of the window's last day to the scalar velocity prediction $\hat{y}$.

TerraFlow is trained with the pinball (quantile) loss (Koenker and Bassett, 1978). For a chosen quantile $\tau \in (0, 1)$,

$$\mathcal{L}_\tau = \frac{1}{N}\sum_{i=1}^{N} \rho_\tau(y_i - \hat{y}_i)\,, \tag{13}$$

with the check function

$$\rho_\tau(u) = u\left(\tau - \mathbf{1}_{\{u<0\}}\right), \tag{14}$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. We set $\tau = 0.5$, for which Eqs. (13) and (14) reduce to one half of the absolute error, the L1 objective, and the regression targets the conditional median. This choice is deliberate: glacier speeds at Shisper follow a strongly right-skewed, approximately log-normal distribution with rare extreme surges (Fatima et al., 2025), and median regression is substantially more robust to such heavy tails than mean squared error, which is dominated by the rare extremes it cannot fit. Validation and test skill were tracked with the mean absolute error (MAE),

$$\mathcal{L}_{\mathrm{MAE}} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|\,. \tag{15}$$

The model was trained in PyTorch with mixed precision, using the AdamW optimiser (Loshchilov and Hutter, 2019), that is Adam (Eqs. 5 to 7) with decoupled weight decay, at a learning rate of 1×10^{-4}, a batch size of 2048, a plateau-based learning-rate schedule, one million training samples drawn per epoch and early stopping after five epochs without improvement on the validation year. The split is by target year: targets up to 2020 for training (7.62 million samples), 2021 for validation (a fixed subset of 200 000 of its 0.73 million samples) and 2022 onward for testing (2.53 million samples). The baselines are persistence, the last observed daily maximum of the pixel, and the typical speed of the pixel, the median of its training days; they play the roles of the persistence and climatology reference forecasts (Murphy, 1992).

3.7 TempFlow: surface-temperature forecasting

TempFlow forecasts daily MODIS land-surface temperature from a 30-day input window of eight daily features: the pixel coordinates, the calendar year, the month and the day of the year as sine and cosine pairs, and the land-surface temperature of that day; the target is the next day's LST of the same pixel. Missing days in the LST record are filled by linear interpolation, at most seven in a row, and only windows without remaining gaps are used; the target is always a real cloud-free retrieval, never an interpolated value. The LST of the window is the decisive input: without it, the network could only learn the seasonal cycle from the date and location features. The architecture is a stacked recurrent network: a bidirectional LSTM of 128 units

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

(returning sequences) with dropout 0.4; an LSTM of 64 units (returning sequences) with dropout 0.4; an LSTM of 32 units with dropout 0.3; and a single linear output unit. At each time step t, given the input vector x_t, the previous hidden state h_{t-1} and the previous cell state c_{t-1}, an LSTM cell updates its gates and states as

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)\,, \tag{16}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)\,, \tag{17}$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)\,, \tag{18}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{19}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)\,, \tag{20}$$

$$h_t = o_t \odot \tanh(c_t), \tag{21}$$

where σ is the logistic function of Eq. (3) and $\odot$ denotes element-wise multiplication. The matrices W_f, W_i, W_c and W_o project the input vector x_t; the matrices U_f, U_i, U_c and U_o project the previous hidden state h_{t-1}; and b_f, b_i, b_c and b_o are the corresponding bias vectors, all learned during training. The forget, input and output gates f_t, i_t and o_t regulate what the cell state c_t discards, admits and exposes, $\tilde{c}_t$ is the candidate cell state proposed at step t, and h_t is the hidden state passed to the next step. This gated memory mitigates the vanishing-gradient problem of standard recurrent networks and preserves information across the multi-week dependencies that characterise the thermal record (Hochreiter and Schmidhuber, 1997), which is why LSTMs were preferred over feed-forward and classical time-series alternatives. The bidirectional first layer lets the encoder relate each day to both earlier and later context within the window, improving the isolation of short-lived anomalies embedded in the strong seasonal cycle, while the tapering stack consolidates daily variability into the final estimate.

Training minimised the mean-squared-error loss,

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\,, \tag{22}$$

with the AdamW optimiser (learning rate 1×10^{-3}, weight decay 4×10^{-3}), a batch size of 512, and early stopping on validation loss (patience 8, with 15 % of the training sequences held out) with plateau-based learning-rate reduction (halving

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

after four epochs without improvement). Inputs and targets were standardised with scalers fitted on the training period only, and predictions inverse-transformed for evaluation. Temperatures are reported in kelvin. The data are split chronologically at 7 June 2018, so every scene date used by FusionFlow lies in the test period. TempFlow is compared on identical targets with the two standard reference forecasts (Murphy, 1992), persistence and climatology. Persistence is the last LST value of the context, and two climatologies are fitted on the training period: the mean seasonal cycle of each pixel (annual and semi-annual harmonics), which is the stronger and main baseline, and a single day-of-year mean over the whole grid. To measure skill beyond the next day, a separate model with the same architecture and settings is trained for each lead time of 2, 3 and 4 days, with persistence taken as the last value of the context.

3.8 FusionFlow and the decision rule

FusionFlow works on the outputs of the three modules. For every day, a velocity anomaly and a temperature anomaly summarise the 30 days up to that day, the same window that defines a GLOF label. The velocity anomaly is the 90th percentile, over all pixel-days in the window, of $\ln[(\hat{v}+10)/(\bar{v}+10)]$, where $\hat{v}$ is the TerraFlow forecast and $\bar{v}$ the median speed of the pixel in the same calendar month over its training years, both in $\mathrm{m\,yr^{-1}}$. A high percentile is used because the surge and the ice dam affect a minority of pixels, and it is insensitive to single false matches. The temperature anomaly is the median, over all pixel-days in the window, of the TempFlow forecast minus the seasonal cycle of the pixel from the training period.

A logistic regression of the two anomalies (L2 penalty with $C = 1$, class weights that balance outburst and other days, inputs standardised, missing anomalies set to the training mean) is trained on every day of the training years, using the label rule of Sect. 3.2; it can therefore also learn from the 2020 outburst, which has no clear post-outburst scene. Each day is given a weight of $1/31$, because consecutive days share almost the whole window; otherwise four outbursts would act as 124 independent cases and the model would be overconfident. The result is a daily tabular probability p_{tab}.

FusionFlow adds the evidence of the two streams on the log-odds scale,

$$\operatorname{logit} p_{\mathrm{fus}} = \operatorname{logit} p_{\mathrm{vis}} + \operatorname{logit} p_{\mathrm{tab}}, \qquad \operatorname{logit} p = \ln \frac{p}{1-p}, \tag{23}$$

where p_{vis} is the RiskFlow probability of the scene and both probabilities are kept within 0.01 and 0.99. Because RiskFlow and the tabular model are both trained with approximately balanced classes, Eq. (23) is the combination that would be optimal if the two streams were conditionally independent given the label, and it fits nothing on top of them, which matters with only seven GLOF scenes. An alert is raised when $p_{\mathrm{fus}} \geq 0.5$. We compare FusionFlow with RiskFlow alone, the tabular stream alone, the agreement rule (an alert only when $p_{\mathrm{vis}} \geq 0.5$ and $p_{\mathrm{tab}} \geq 0.5$) and an either-stream rule (an alert when either probability is at least 0.5). All settings, including the window, the percentile, the penalty, the day weight, the clipping and the threshold, were fixed before the first run on the Shisper data.

3.9 Evaluation protocol

Each module is evaluated on data withheld from training and model selection, always from years the model did not see. In the main split, models are trained on the scenes up to 2021 (54 scenes, 5 of them GLOF scenes, with 2021 as the validation

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

year) and tested once on the 104 scenes from 2022 onward (2 GLOF scenes). In the leave-one-year-out test, each of seven folds (2018 to 2019, which contains the single December 2018 scene, and each year from 2020 to 2025) is held out in turn; the model is trained on all other years, with the number of epochs chosen on the latest outburst year outside the held-out fold, so that every scene, including all seven GLOF scenes, is predicted once by a model that never saw its year. The tabular stream is trained on the days of the same training years. Detection is summarised by the number of GLOF scenes caught and the number of false alarms among the other scenes, and, given the small number of underlying positive scenes, 95 % Wilson confidence intervals (Wilson, 1927) accompany these counts. For an observed proportion $\hat{p}$ over n samples, the Wilson interval is

$$p \in \frac{1}{1+z^2/n}\left(\hat{p}+\frac{z^2}{2n} \pm z\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}+\frac{z^2}{4n^2}}\right), \tag{24}$$

with $z = 1.96$ at the 95 % level; unlike the normal approximation, it remains well behaved for the small positive-class counts encountered here. We also report the area under the receiver operating characteristic curve (AUC). TerraFlow and TempFlow are evaluated by MAE, and TempFlow also by the root-mean-square error (RMSE) and the coefficient of determination (R^2). Before RiskFlow was run on the re-exported scenes, we required at least 4 of the 7 GLOF scenes to be caught when years are held out, with false alarms below about 1 in 10 other scenes.

4 Results

4.1 Component-level performance

4.1.1 Vision stream (RiskFlow)

With the five-seed average and a threshold of 0.5, RiskFlow caught both GLOF scenes of the main test (10 and 25 May 2022) and raised 2 false alarms among the other 102 test scenes (AUC 0.99; Table 3). One false alarm fell on 14 June 2022, 38 days after the outburst, when the basin was still drained but the scene lay just outside the 30-day label window; the other fell on 31 December 2022. Individual seeds caught 1 or 2 of the 2 scenes and raised 2 to 8 false alarms (AUC 0.94 to 0.99), so the average is more stable than any single run.

When each year was held out in turn, RiskFlow caught 4 of the 7 GLOF scenes and raised no false alarm among the 151 other scenes (AUC 0.88; Fig. 5), which passes the bar set before the run. The four scenes of 2021 and 2022 were caught by four or five of the five seeds (mean probabilities 0.80 to 0.93), while the three scenes of July 2019 were missed (mean probabilities 0.13, 0.00 and 0.00). Lowering the threshold to 0.2 recovered none of them and added five false alarms, so the misses are not a threshold effect. They are not a cloud effect either: the scenes of 5 and 20 July 2019 had about 3 % cloud in the lake window. When 2019 is held out, the model learns only from the outbursts of 2021 and 2022, and the drained basin of July 2019 evidently looked different enough from those to be missed. The three years without a lake, 2023 to 2025, produced no false alarm, so RiskFlow did not mistake the empty basin for a drained one.

The water-index rule caught all 7 GLOF scenes, including the three from 2019, but raised 26 false alarms among the 151 other scenes (17 %). Twenty-one of them were winter scenes between December and March, when snow and shadow pass the

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 3. Detection of post-outburst (GLOF) scenes at Shisper. RiskFlow probabilities are the mean of five seeds with a threshold of 0.5 unless stated otherwise; 95 % Wilson intervals are given in parentheses. The water-index rule needs no training, so it is evaluated on the same scenes without held-out years. The multi-site model was trained with 280 outbursts at 103 other lakes (Sect. 3.5).

Test	Detector	GLOF scenes caught	False alarms	AUC
Main split (test 2022 onward)	RiskFlow	2 of 2 (0.34 to 1.00)	2 of 102 (0.01 to 0.07)	0.99
	RiskFlow, single seeds	1 to 2 of 2	2 to 8 of 102	0.94 to 0.99
	Water-index rule	2 of 2 (0.34 to 1.00)	22 of 102 (0.15 to 0.31)	
	Multi-site RiskFlow	2 of 2 (0.34 to 1.00)	78 of 102 (0.67 to 0.84)	0.57
Leave one year out (all years)	RiskFlow	4 of 7 (0.25 to 0.84)	0 of 151 (0.00 to 0.02)	0.88
	RiskFlow, threshold 0.2	4 of 7 (0.25 to 0.84)	5 of 151 (0.01 to 0.08)	0.88
	RiskFlow, single seeds	3 to 5 of 7	0 to 6 of 151	0.81 to 0.91
	Water-index rule	7 of 7 (0.65 to 1.00)	26 of 151 (0.12 to 0.24)	
	Multi-site RiskFlow	1 of 7 (0.03 to 0.51)	5 of 151 (0.01 to 0.08)	0.73
All Shisper scenes	Other lakes only	7 of 7 (0.65 to 1.00)	149 of 151 (0.95 to 1.00)	0.37

water test, and five showed the drained lake 33 to 56 days after an outburst, just outside the label window. From 2022 onward the rule raised 22 false alarms in 102 scenes, against 2 for RiskFlow. RiskFlow is thus far more specific than the rule, and the two make different errors.

The transfer test failed the bar set before running. Trained on the other lakes only, RiskFlow gave nearly every Shisper scene a high probability: it caught all 7 GLOF scenes but also flagged 149 of the 151 other scenes (AUC 0.37). Adding the Shisper training years did not remove the problem in the main test (2 of 2 caught, 78 false alarms in 102 scenes, AUC 0.57), and with years held out the combined model caught 1 of 7 GLOF scenes with 5 false alarms (AUC 0.73). We therefore keep the Shisper-only model.

4.1.2 Glacier-velocity forecasting (TerraFlow)

Over the 2.53 million test forecasts from 2022 onward, TerraFlow had a mean absolute error of 28.6 m yr^{-1}, against 40.7 m yr^{-1} for persistence and 30.9 m yr^{-1} for the typical speed of the pixel (Table 4). One day ahead the error was 28.0 m yr^{-1}, and it grew slowly with the horizon, to 29.6 m yr^{-1} for two to five days and 34.8 m yr^{-1} for 6 to 16 days. The error was similar on the glacier outline (29.9 m yr^{-1}) and in the margin (28.3 m yr^{-1}), and during the outburst season of April to June 2022 it was 32.6 m yr^{-1}, against 51.0 m yr^{-1} for persistence. TerraFlow did not forecast fast readings: for observations above 300 m yr^{-1} its error (448 m yr^{-1}) exceeded that of persistence (425 m yr^{-1}). At the test pixel with the most fast readings, these were isolated one-day spikes of up to about 2000 m yr^{-1} spread through 2022 rather than clustered near the 7 May outburst (Fig. 6), which most likely reflects false matches in the image pairs rather than glacier motion; the forecast follows the slow background flow. FusionFlow also uses TerraFlow forecasts for 2018 to 2020, which lie in its training period; their error (41.8 m yr^{-1} over

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

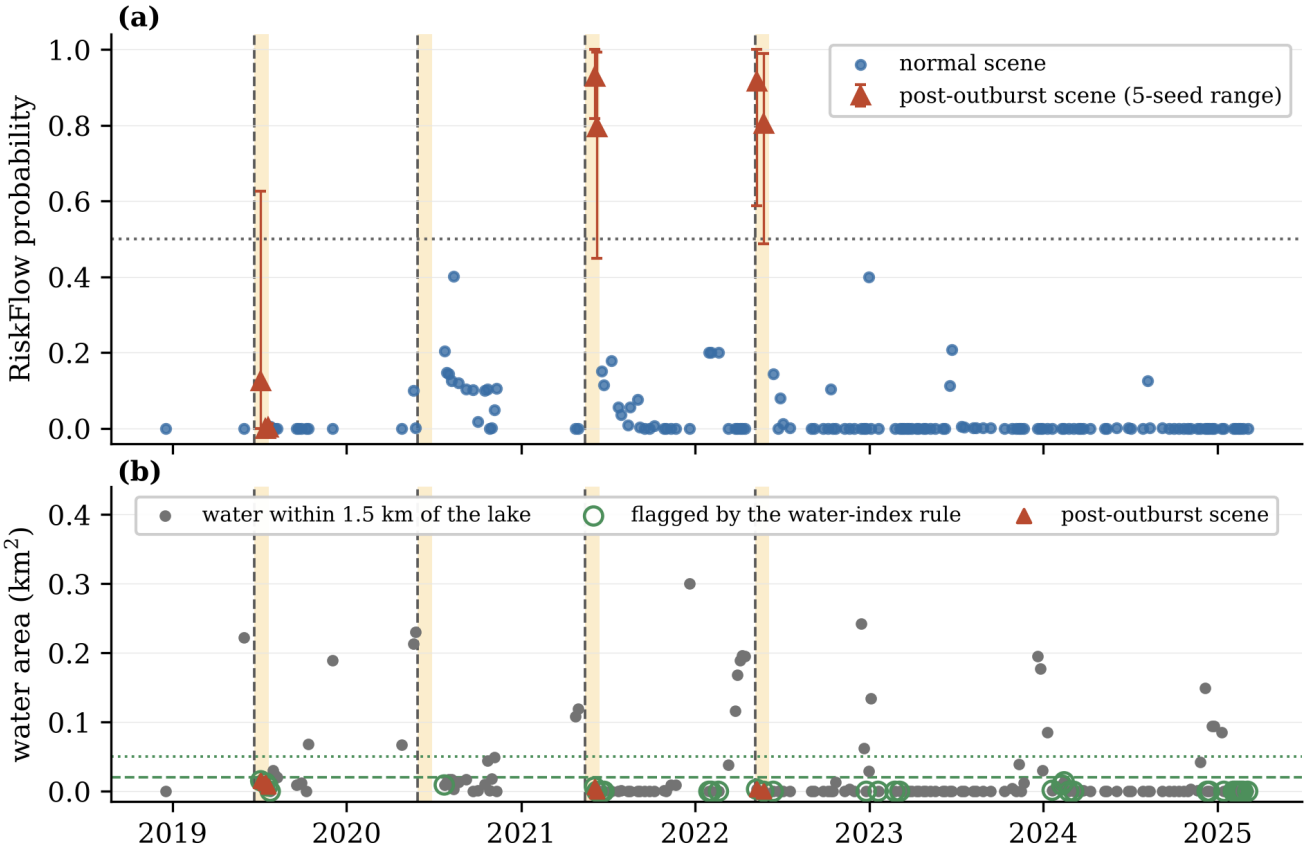


Figure 5. RiskFlow with each year held out in turn. (a) Probability for every scene (mean of five seeds); bars on the GLOF scenes show the range over the seeds, and the dotted line is the threshold of 0.5. (b) Open water within 1.5 km of the lake in each scene, with the scenes flagged by the water-index rule circled; the dotted and dashed lines are the 0.05 and 0.02 km^2 thresholds of the rule. Dashed vertical lines mark the outbursts and shading the 30 days after each.

all training years) is larger than in the test years, so these forecasts are not unrealistically accurate. Readings above 300 m yr^{-1} were most common between 2013 and 2018 and rare after 2022 (Fig. 3b), but the documented surge of 2017 to 2019 (Bhambri et al., 2020) is not clearly separated from earlier years in this record, consistent with fast flow being sampled sparsely and noisily.

4.1.3 Surface-temperature forecasting (TempFlow)

At the lake grid, TempFlow forecast next-day LST on 29 581 cloud-free test targets with a mean absolute error of 4.81 K, a root-mean-square error of 6.11 K and $R^2 = 0.75$ (Fig. 7; Table 5). Persistence was slightly better one day ahead (4.75 K), and both were better than the pixel climatology (5.10 K) and the day-of-year mean over the grid (6.62 K). For the continuous

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 4. Mean absolute error (m yr^{-1}) of the next-observation velocity forecasts on the test years (2022 onward).

Subset	Forecasts	TerraFlow	Persistence	Typical speed
All	2 531 429	**28.6**	40.7	30.9
Next day	1 616 766	**28.0**	41.4	30.1
2 to 5 days ahead	908 611	**29.6**	39.5	32.2
6 to 16 days ahead	5 901	**34.8**	47.2	37.3
More than 16 days ahead	151	47.4	**45.0**	47.2
Glacier outline	443 329	**29.9**	40.0	34.0
Margin	2 088 100	**28.3**	40.9	30.2
Observed above 300 m yr^{-1}	15 738	448.5	**425.1**	460.3
Observed above 1000 m yr^{-1}	793	1543.8	**1532.4**	1553.9
April to June 2022	200 205	**32.6**	51.0	34.4

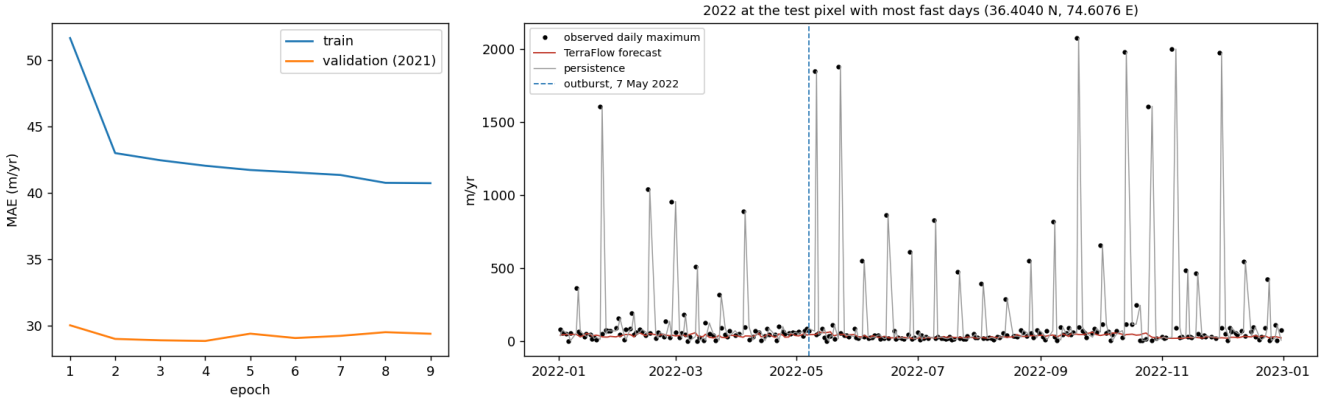


Figure 6. TerraFlow. Left: training and validation error by epoch; the model of epoch 4 was kept. Right: observed daily maximum speed, TerraFlow forecast and persistence through 2022 at the test pixel with the most readings above 300 m yr^{-1} (36.4040° N, 74.6076° E). The fast readings are isolated one-day spikes spread through the year.

evaluation of 2022, TempFlow was scored on every day of the year with at least one test target, which gave 152 days and 5384 targets. The error over these targets was 5.05 K, and the grid-mean forecast followed the observed record through the spring warming and the 7 May outburst (Fig. 8). The remaining days of 2022 were cloudy over the whole grid or lacked a complete 30-day context.

Two days ahead, TempFlow was the most accurate forecast at the lake grid (4.83 K, against 4.95 K for persistence and 5.10 K for climatology); at three and four days the pixel climatology was best, and TempFlow was worse than both baselines (Fig. 9a). At the original points, the same code gave an error of 2.50 K (root-mean-square error 3.24 K, $R^2 = 0.905$) on 5161 targets, against 2.25 K for persistence and 3.08 and 3.45 K for the two climatologies. There, persistence was better one and two days

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 5. Mean absolute error (K) of TempFlow and the baselines by lead time, at the lake grid and at the original points. The climatology is the seasonal cycle of each pixel; the day-of-year mean over the grid gave 6.62 to 6.64 K at the lake grid and 3.45 to 3.57 K at the original points. The best value in each row and site is in bold.

	Lake grid (64 pixels)				Original points (9 pixels)			
Lead (days)	Targets	TempFlow	Persistence	Climatology	Targets	TempFlow	Persistence	Climatology
1	29 581	4.81	**4.75**	5.10	5161	2.50	**2.25**	3.08
2	29 363	**4.83**	4.95	5.10	5147	2.75	**2.63**	3.11
3	29 196	5.80	5.72	**5.12**	5155	**2.92**	3.11	3.10
4	29 156	6.33	5.74	**5.13**	5155	**3.09**	3.21	3.20

ahead, and TempFlow was the most accurate forecast three and four days ahead, by 0.18 and 0.11 K (Fig. 9b). The useful lead time of TempFlow therefore depends on the site: three to four days at the original points, and two days at the lake grid, where day-to-day variability is about twice as large (persistence error 4.75 K against 2.25 K).

4.2 Fusion of the streams

Velocity and temperature anomalies did not distinguish the outburst periods. Over all days, with each year held out in turn, the tabular stream reached an AUC of 0.39, below chance, and it flagged 44 % of the days that lay neither within 30 days after nor within 30 days before an outburst. On their own, the velocity anomaly gave an AUC of 0.39 and the temperature anomaly 0.53. The fitted coefficients changed between folds: the temperature weight ranged from -0.26 to $+0.70$, and the velocity weight was negative in every fold. The tabular stream flagged all 31 days after the 2020 outburst, 6 of the 31 days after the 2021 outburst and none after the 2019 or 2022 outbursts.

At the scene level (Table 6), the tabular stream alone caught 1 of 7 GLOF scenes with 43 false alarms in 151 scenes when years were held out. The agreement rule kept RiskFlow free of false alarms but caught only 1 of 7, the either-stream rule caught the same 4 of 7 as RiskFlow but with 43 false alarms, and FusionFlow caught 2 of 7 with no false alarm (AUC 0.80). With years held out, the two streams disagreed on 46 of the 158 scenes, 43 of them normal scenes flagged by the tabular stream alone (Fig. 10). In the main test, RiskFlow alone caught both 2022 scenes, FusionFlow one and the agreement rule none.

4.3 Case study: the 2022 Shisper Glacier outburst season

The 2022 season shows why the physical streams did not help (Table 7, Fig. 11). The outburst came during a warm spell, 2 to 3 K above normal, and with high velocity anomalies, whereas in the training years the outbursts had come with cooler conditions and low velocity anomalies. The tabular stream therefore gave the two GLOF scenes of 2022 probabilities of only 0.09 and 0.16 and pulled the RiskFlow probability of 0.54 for 10 May below the threshold. RiskFlow's false alarm of 14 June, 38 days after the outburst when the basin was still drained, stayed just below the threshold after fusion (0.46).

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.



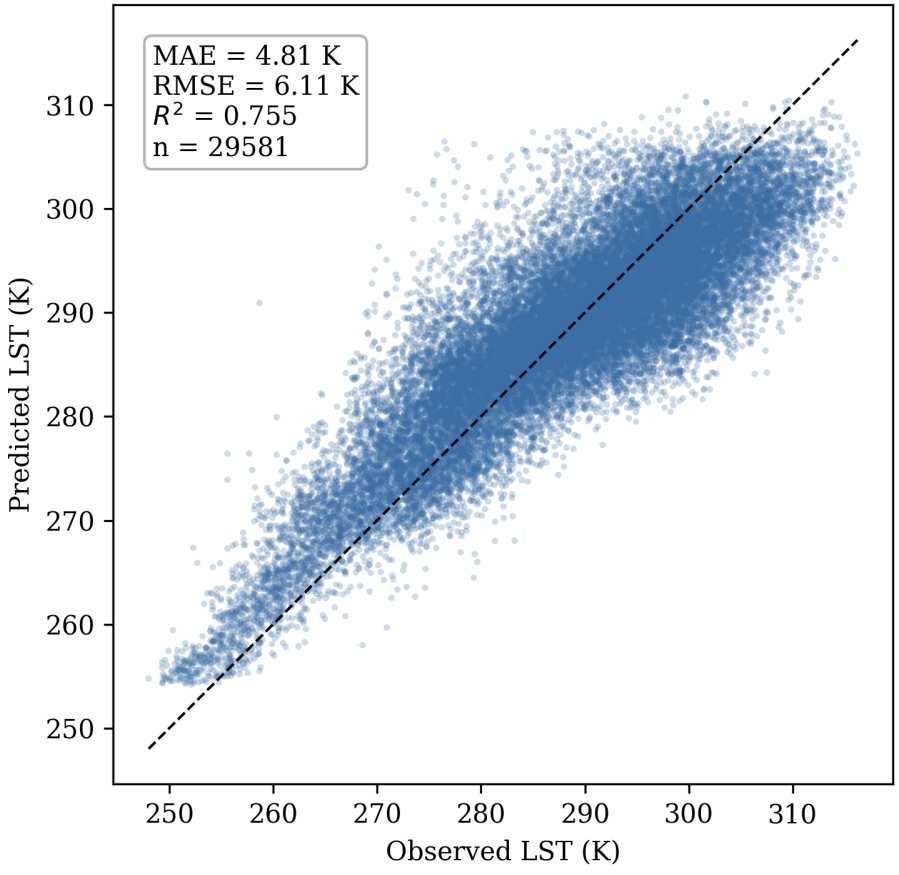


Figure 7. Forecast against observed next-day LST for the 29 581 cloud-free test targets at the lake grid (dashed line: perfect forecast).

5 Discussion

5.1 The outburst mechanism at Shisper

The Shisper record fits the classical picture of drainage from an ice-dammed lake. The surge of 2017 to 2019 advanced the terminus across the Muchuhar meltwater stream and created the dam (Bhambri et al., 2020). In each of the four seasons the lake then filled through winter and spring, reached its maximum in May and typically drained within one to two days (Lovell and Muhammad, 2025). Rapid, repeated drainage of a lake that refills behind the same dam is the behaviour expected when the water becomes deep enough to lift the ice and open a path beneath it, at a depth of roughly nine tenths of the ice thickness (Glen, 1954), after which the channel is enlarged by the heat of the flowing water (Nye, 1976; Tweed and Russell, 1999). In this picture the timing of an outburst is set mainly by how full the lake is relative to the dam, and only indirectly by the weather of the preceding weeks.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

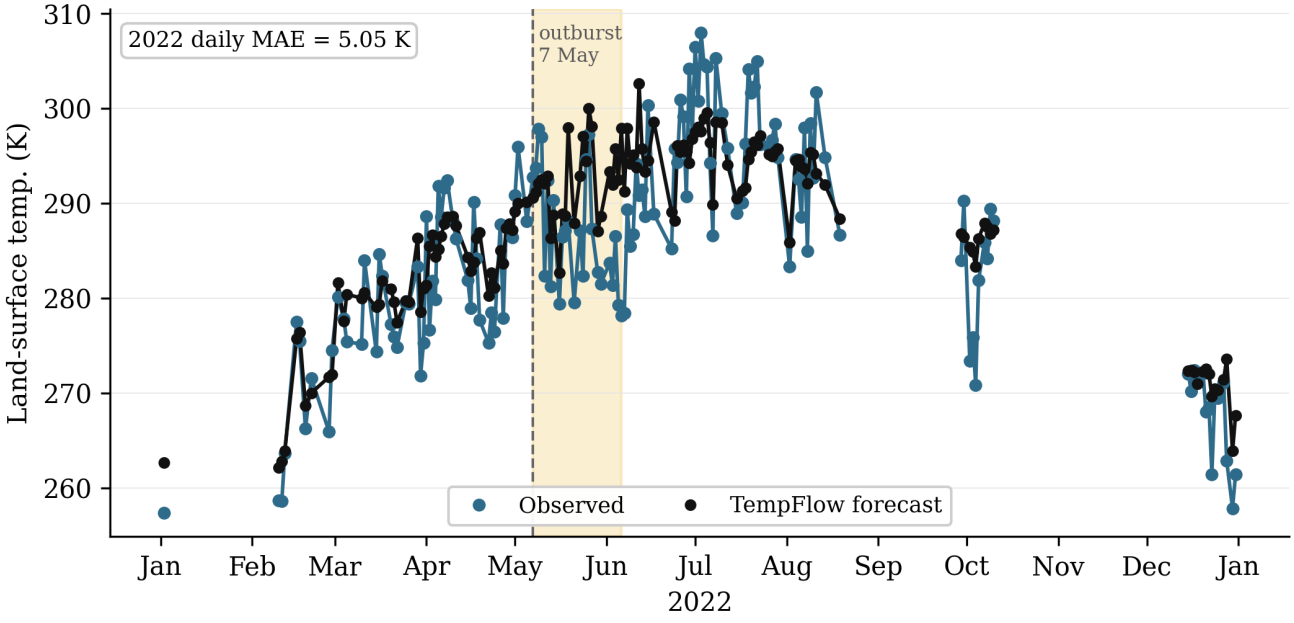


Figure 8. Day-by-day evaluation of TempFlow through 2022 at the lake grid: next-day forecast (black) and observed LST (blue), averaged over the test targets of each day. A point is shown for every day with at least one target, and lines are broken across gaps longer than 12 days. The dashed line marks the outburst of 7 May 2022 and the shading the following 30 days.

Table 6. Detection of GLOF scenes by RiskFlow, the tabular stream and three ways of combining them; 95 % Wilson intervals in parentheses. The AUC is not defined for the two binary rules.

Test	Decision	GLOF scenes caught	False alarms	AUC
Leave one year out	RiskFlow alone	4 of 7 (0.25 to 0.84)	0 of 151 (0.00 to 0.02)	0.88
	Tabular stream alone	1 of 7 (0.03 to 0.51)	43 of 151 (0.22 to 0.36)	0.33
	Agreement rule (both at least 0.5)	1 of 7 (0.03 to 0.51)	0 of 151 (0.00 to 0.02)	
	Either-stream rule	4 of 7 (0.25 to 0.84)	43 of 151 (0.22 to 0.36)	
	FusionFlow	2 of 7 (0.08 to 0.64)	0 of 151 (0.00 to 0.02)	0.80
Main split (test 2022 onward)	RiskFlow alone	2 of 2 (0.34 to 1.00)	2 of 102 (0.01 to 0.07)	0.99
	Tabular stream alone	0 of 2 (0.00 to 0.66)	15 of 102 (0.09 to 0.23)	0.13
	Agreement rule (both at least 0.5)	0 of 2 (0.00 to 0.66)	0 of 102 (0.00 to 0.04)	
	Either-stream rule	2 of 2 (0.34 to 1.00)	17 of 102 (0.11 to 0.25)	
	FusionFlow	1 of 2 (0.09 to 0.91)	1 of 102 (0.00 to 0.05)	0.97

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

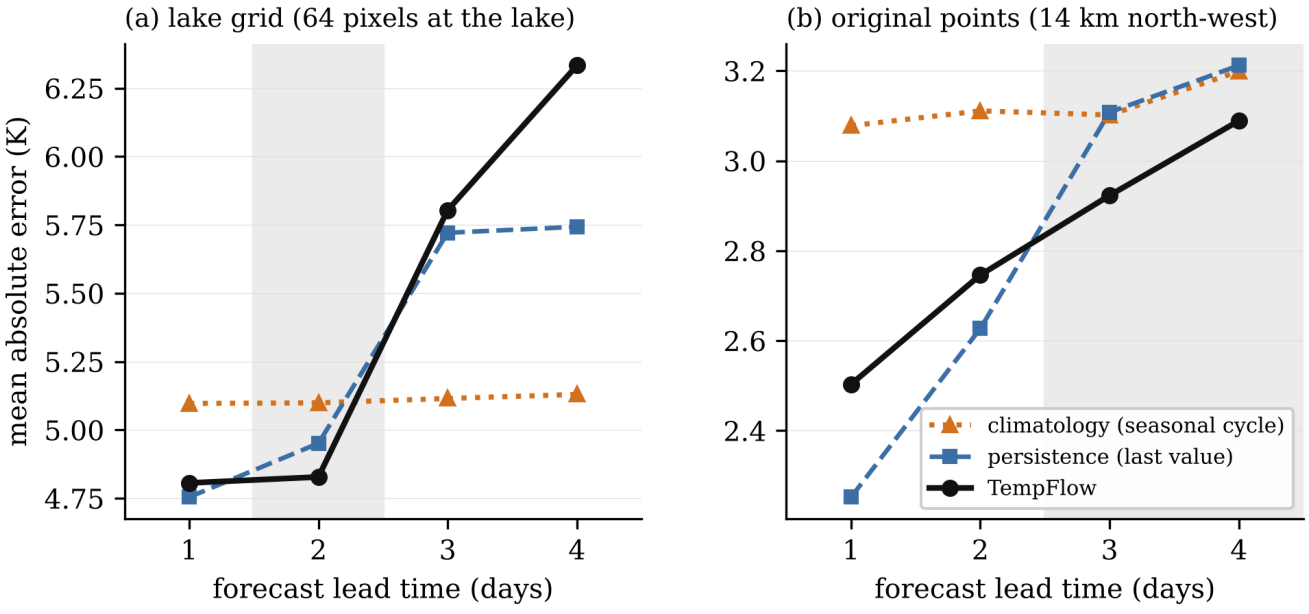


Figure 9. Mean absolute error of TempFlow, persistence and the pixel climatology by lead time, (a) at the lake grid and (b) at the original points about 14 km north-west of the lake. Shading marks the lead times at which TempFlow is the most accurate forecast.

Our data agree with that picture. Over the 30 days before each outburst, the temperature anomaly at the lake grid was -1.4 K in 2019, -0.6 K in 2020, -1.1 K in 2021 and $+2.0$ K in 2022. Three of the four outbursts therefore followed cooler-than-normal weeks, and only the 2022 event came during a warm spring, when faster melt may have brought the lake to its threshold sooner. Velocity gives no clearer signal. The velocity anomaly over the same 30 days was 0.03, 0.00, 0.39 and 0.84, and in 2022 it was falling, not rising, in the two months before the outburst (Table 7); the fast readings in the record are sparse and partly spurious (Sect. 4.1.2). A precursor pairing of accelerated flow with surface temperatures near the melting point, which motivated the agreement rule, is therefore not supported once all four outbursts are considered. The negative weights that the tabular model gave to velocity, and the temperature weight that changed sign between folds, are the statistical expression of this: the record shows no stable relation between these anomalies and the outburst dates.

The modelling choices carry physical meaning as well. Median regression for velocity concedes the exact magnitude of the rarest surges in exchange for stable, unbiased tracking of the bulk dynamics from which anomalies are measured, which is how TerraFlow behaved on the fast readings (Sect. 4.1.2). The 30-day thermal context mirrors the timescale over which the surface energy balance integrates into melt.

5.2 IceWatch in the context of existing approaches

Most deep-learning contributions to glacial-lake research remain mapping-centric (Cao et al., 2024; Xie et al., 2022; Maslov et al., 2025): they characterise the present state rather than the forward risk. Hybrid frameworks have begun to couple physical

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 7. The 2022 season in the main test (models trained on the years up to 2021): probabilities of RiskFlow, the tabular stream and FusionFlow for every scene between March and July, with the two anomalies over the 30 days up to the scene. An alert requires a probability of at least 0.5.

Date (2022)	Days from outburst	GLOF scene	RiskFlow	Tabular	FusionFlow	Velocity anomaly	Temperature anomaly (K)
11 March	−57	no	0.00	0.04	0.00	1.17	0.23
26 March	−42	no	0.00	0.05	0.00	1.09	0.72
31 March	−37	no	0.00	0.05	0.00	1.07	0.77
5 April	−32	no	0.00	0.06	0.00	1.04	1.07
10 April	−27	no	0.00	0.05	0.00	1.02	1.81
15 April	−22	no	0.00	0.05	0.00	1.01	1.91
10 May	3	yes	**0.54**	0.09	0.11	0.81	2.16
25 May	18	yes	**0.98**	0.16	**0.89**	0.61	2.71
14 June	38	no	**0.71**	0.26	0.46	0.46	1.35
24 June	48	no	0.00	0.36	0.01	0.35	0.76
29 June	53	no	0.47	0.40	0.37	0.32	0.14
4 July	58	no	0.05	0.37	0.03	0.32	0.84
19 July	73	no	0.00	0.40	0.01	0.34	−0.52

indicators with learned models, such as surge velocity against lake volume in the Karakoram (Bazai et al., 2024) and multisensor deformation and lake-expansion metrics in hazard assessment (Nurakynov et al., 2025), yet these operate retrospectively or at inventory scale. RiskFlow shares the first limitation, since it recognises the appearance of the basin after drainage. It does so with few false alarms, far fewer than the water-index rule, which is easily misled by snow and shadow in winter, but it can only confirm an outburst once a clear scene is available. At Shisper the first scene in the archive came 3 days after the 2022 outburst, 13 days after the 2019 outburst and 20 days after the reference date of the 2021 drainage, and none came within 30 days of the 2020 outburst. The detection is also specific to the site. A model trained on 925 post-outburst scenes from 103 other lakes flagged almost every Shisper scene. A likely reason is that at most of those lakes the normal state is a full lake, so that an empty basin means recent drainage, whereas at Shisper the basin is empty in most scenes, including all of 2023 to 2025. Transfer between lakes would need a model that compares each scene with the recent state of the same lake rather than with the typical appearance of other lakes. The complementary errors of RiskFlow and the water-index rule, which caught the three July 2019 scenes that RiskFlow missed, suggest such a combination, but it would have to be designed and tested on events not used here.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

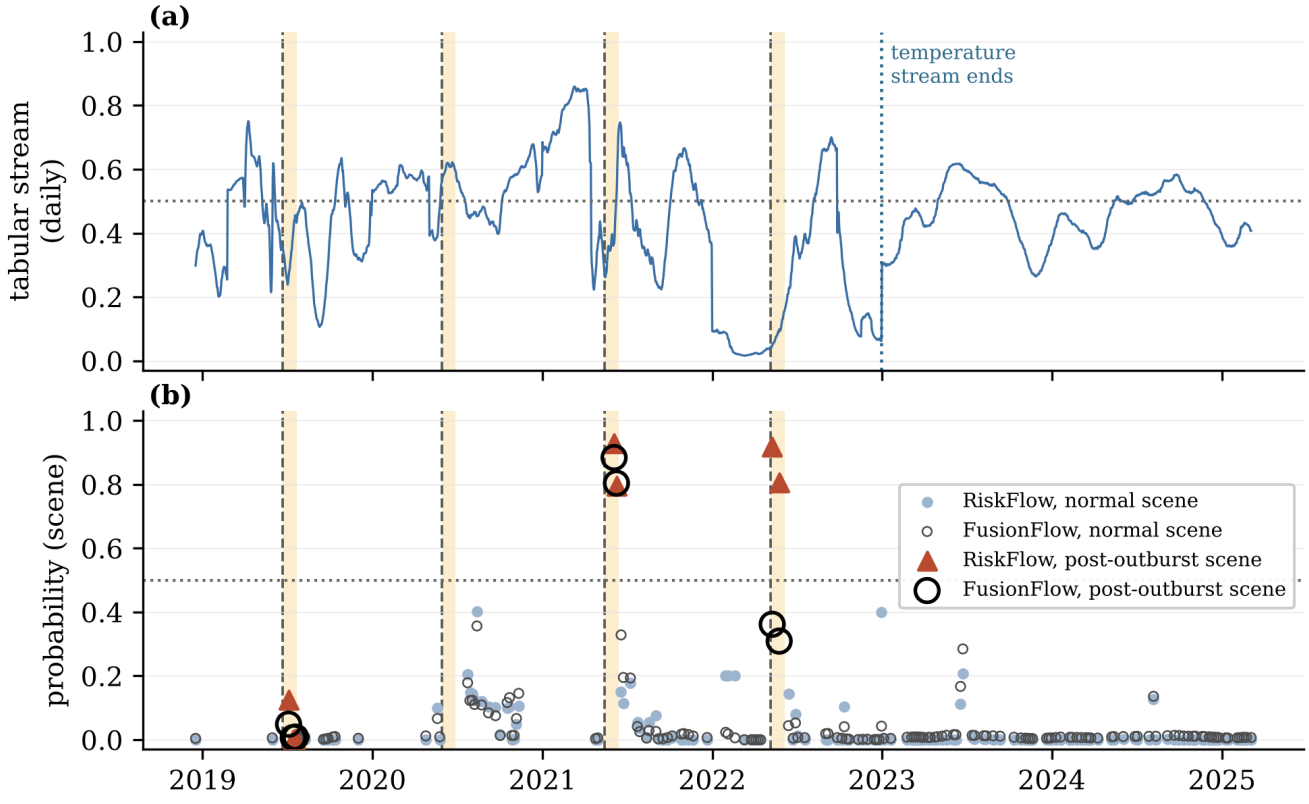


Figure 10. All years with each year held out in turn. (a) Daily tabular probability; the temperature stream ends with 2022. (b) RiskFlow and FusionFlow probabilities of every scene, with the GLOF scenes highlighted. Dashed lines mark the outbursts and shading the 30 days after each.

5.3 Implications for disaster risk reduction

For disaster risk reduction at Shisper, IceWatch is a monitoring tool rather than an early-warning system. It can confirm from the next clear scene that the lake has drained, with a low false-alarm rate, and its velocity and temperature forecasts describe the state of the glacier one to a few days ahead; neither stream anticipated an outburst. Traditional breach models and static inventories typically identify hazard once structural failure has begun, leaving downstream communities hours at best, and an anticipatory margin would have direct value in the Hunza valley given the repeated impacts of Shisper outbursts on the Hassanabad crossing of the Karakoram Highway. Because drainage depends on how full the lake is, the most promising route to such a margin is to observe lake filling directly, for example lake area and level from optical imagery and from radar, which sees through cloud. Lovell and Muhammad (2025) showed that the lake reached its maximum size every May before draining. A filling stream could be evaluated within the same framework and with the same leave-one-year-out test. We are careful,

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

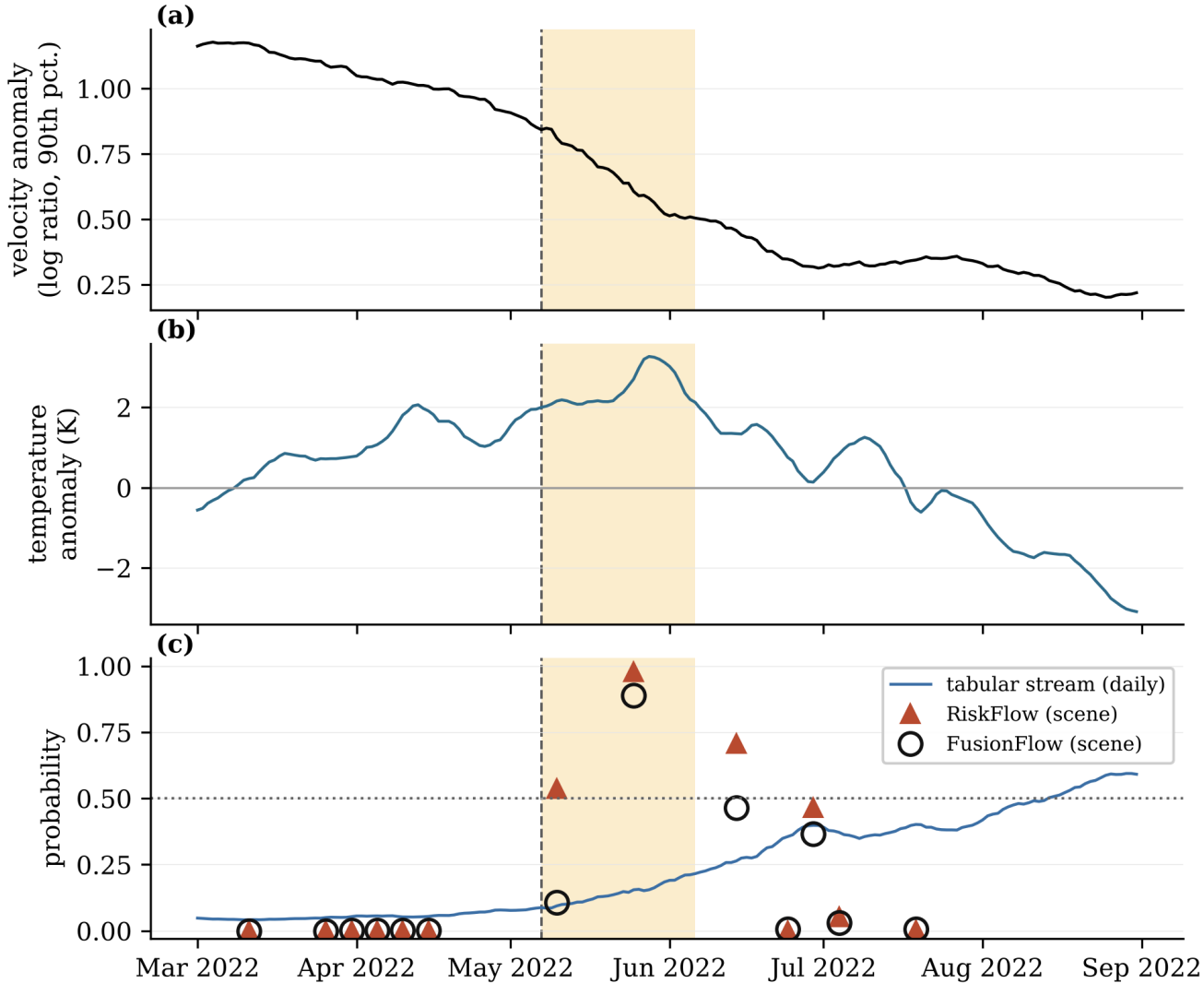


Figure 11. The 2022 season in the main test. (a) Velocity anomaly and (b) temperature anomaly over the 30 days up to each day. (c) Daily tabular probability, with the RiskFlow and FusionFlow probabilities of each Sentinel-2 scene; the dotted line is the threshold of 0.5. The dashed line marks the outburst of 7 May 2022 and the shading the following 30 days.

however, not to over-generalise: these implications derive from a single glacier and must be confirmed across additional events and basins.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

5.4 Limitations and future work

Several limitations qualify the results. The evaluation rests on four outbursts at one lake and on seven post-outburst scenes, so all detection rates carry wide intervals (Table 3); assembling a multi-basin evaluation corpus of surge-dammed lakes in the Karakoram and Himalaya is the most important next step. The 30-day label window is a convention: scenes 31 to 60 days after an outburst still show a drained basin but count as normal, which counts some correct detections by RiskFlow and by the water-index rule as false alarms. ITS_LIVE velocities are dated by the middle of the image pair, so a value dated before a scene can include imagery acquired after it, and fast readings are sparse and partly false matches. Median regression deliberately trades accuracy in the extreme surge tail for robust baseline tracking; augmenting TerraFlow with upper-quantile heads (for example $\tau = 0.9$ and 0.95) would provide explicit tail-risk estimates and prediction intervals at negligible architectural cost. The temperature stream ends in 2022 because of the orbital drift of Terra, and MODIS LST at 1 km is limited by cloud, with 42 % of pixel-days clear. The water test confuses snow and shadow with water in winter. Training on a GPU is not bit-reproducible, which we address by averaging five seeds and reporting their spread. Finally, the log-odds sum of FusionFlow assumes that the streams are conditionally independent; with seven GLOF scenes, a combination fitted on top of the streams could not be estimated reliably.

Operationally, persistent multi-week cloud cover during the monsoon can occlude both the optical stream and the 30-day thermal context; integrating Synthetic Aperture Radar would maintain uninterrupted sequences and is a natural extension, as is improved spatial downscaling of the thermal observations (Fatima et al., 2025). Finally, while the modelling pipeline is fully realised, real-world deployment introduces engineering constraints outside the present scope: edge inference on site, power management in sub-zero environments and reliable telemetry from remote valleys remain open challenges for operationalisation.

6 Conclusions

This article presented IceWatch, a dual-stream deep-learning framework that couples visual evidence from Sentinel-2 imagery with forecasts of glacier velocity and surface temperature for GLOF monitoring, and tested it at Shisper Glacier on four outbursts, with all three data streams sampled over the lake, the ice dam and the glacier tongue and with labels from a published outburst chronology. Three findings stand out. First, RiskFlow detects the drained lake with few false alarms when a clear scene exists: it caught both post-outburst scenes of 2022 in the main test, and 4 of 7 post-outburst scenes with no false alarm when each year was held out, whereas a water-index rule caught all seven at the cost of many false alarms. A model trained on other lakes did not transfer, probably because the normal state of the Shisper basin is empty. Second, TerraFlow and TempFlow are useful short-range forecasts: TerraFlow is 30 % more accurate than persistence for next-day speed, and TempFlow is the best forecast two days ahead at the lake and three to four days ahead at the original temperature points, where day-to-day variability is lower. Third, velocity and temperature anomalies did not mark the outbursts, three of which followed cooler-than-normal weeks, so neither the agreement rule nor a calibrated fusion improved on the image stream. At Shisper the timing of an outburst appears to be set by lake filling rather than by the weather or the speed of the glacier in the preceding weeks, and warning

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

systems for surge-dammed lakes should observe the lake itself. These findings are scoped by an evaluation record of four outbursts at one glacier, and we make no claim about other lakes or other hazards.

Code and data availability. The GLOFNet dataset, including preprocessed Sentinel-2 imagery, ITS_LIVE velocity sequences and MODIS LST records, is publicly available as described in Fatima et al. (2025); the data used in this study were extracted again from the source archives (Sect. 3.2). Implementation code for the IceWatch components, including the data extraction, the transfer test and FusionFlow, is available at https://github.com/ZuhaFatima96/IceWatch-A-Multimodal-Deep-Learning-Framework-for-Early-Warning-of-GLOFs. Sentinel-2 Level-2A and MODIS MOD11A1 data were accessed through the Microsoft Planetary Computer (https://planetarycomputer.microsoft.com), velocities from the ITS_LIVE version 2 data cubes (Gardner et al., 2025) and outburst records from the global GLOF database (Lützow et al., 2023).

Author contributions. ZF and MAS designed the study and developed the models; MT and AK curated the data and performed the experiments; SS contributed to validation and visualisation; NP supervised the work; ZF and MAS prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research was supported by the resources and guidance provided by the authors' supervisors and by the National University of Sciences and Technology (NUST), Islamabad, Pakistan. Sentinel-2 data are provided by the European Union's Copernicus programme. An AI assistant (Claude, Anthropic) was used for language editing and structural suggestions during manuscript revision; all scientific content, analyses and conclusions are the authors' own.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Abbas, F., Cai, Z., Abbas, F., Abbas, A., Muhammad, S., Shoaib, M., Alrefaei, A., and Albeshr, M.: Glacial lake outburst flood susceptibility mapping in Central Karakoram National Park: a comparative study of machine learning, deep learning, and hybrid models, ResearchGate [preprint], https://doi.org/10.13140/RG.2.2.26767.91044, 2025.

Anand, V., Gandhi, F., and Songara, J. C.: Assessing and managing glacial lake outburst flood risks in the Indian Himalayas: a comprehensive review, Natural Hazards, 121, 17 253–17 281, https://doi.org/10.1007/s11069-025-07497-5, 2025.

Bajracharya, S. R. and Mool, P.: Glaciers, glacial lakes and glacial lake outburst floods in the Mount Everest region, Nepal, Annals of Glaciology, 50, 81–86, https://doi.org/10.3189/172756410790595895, 2009.

Bazai, N., Carling, P., Cui, P., et al.: Predicting the risk of glacial lake outburst floods in Karakoram, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-565, 2024.

Bhambri, R., Watson, C. S., Hewitt, K., Haritashya, U. K., Kargel, J. S., Shahi, A. P., Chand, P., Kumar, A., Verma, A., and Govil, H.: The hazardous 2017–2019 surge and river damming by Shispare Glacier, Karakoram, Scientific Reports, 10, 4685, https://doi.org/10.1038/s41598-020-61277-8, 2020.

Cao, Y., Pan, R., Pan, M., Lei, R., Du, P., and Bai, X.: Refined glacial lake extraction in a High-Asia region by deep neural network and superpixel-based conditional random field methods, The Cryosphere, 18, 153–168, https://doi.org/10.5194/tc-18-153-2024, 2024.

Das, S., Das, S., Mandal, S. T., Sharma, M. C., and Ramsankaran, R.: Inventory and GLOF susceptibility of glacial lakes in Chenab basin, Western Himalaya, Geomatics, Natural Hazards and Risk, 15, https://doi.org/10.1080/19475705.2024.2356216, 2024.

Diaconu, C. A., Heidler, K., Bamber, J., and Zekollari, H.: Multi-sensor deep learning for glacier mapping, arXiv [preprint], https://doi.org/10.48550/arXiv.2409.12034, 2024.

Emmer, A., Allen, S., Carey, M., et al.: Progress and challenges in glacial lake outburst flood research (2017–2021): a research community perspective, Natural Hazards and Earth System Sciences, 22, 3041–3061, https://doi.org/10.5194/nhess-22-3041-2022, 2022.

Fatima, Z., Sohaib, M. A., Talha, M., Sultana, S., Kanwal, A., and Perwaiz, N.: GLOFNet – a multimodal dataset for GLOF monitoring and prediction, in: Proceedings of the International Conference on Digital Futures and Transformative Technologies (ICoDT2), pp. 1–8, https://doi.org/10.1109/ICoDT269104.2025.11360730, 2025.

Gardner, A. S., Fahnestock, M., Greene, C. A., Kennedy, J. H., Liukis, M., López, L. A., and Scambos, T.: MEaSUREs ITS_LIVE Regional Glacier and Ice Sheet Surface Velocities, Version 2, NASA National Snow and Ice Data Center Distributed Active Archive Center [data set], https://doi.org/10.5067/JQ6337239C96, 2025.

Glen, J. W.: The stability of ice-dammed lakes and other water-filled holes in glaciers, Journal of Glaciology, 2, 316–318, https://doi.org/10.3189/S0022143000025132, 1954.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Computation, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Holzmann, M., Davari, A., Seehaus, T., et al.: Glacier calving front segmentation using Attention U-Net, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3483–3486, https://doi.org/10.1109/IGARSS47720.2021.9555067, 2021.

Javed, M., Böhner, J., and Hasson, S.: Mapping glacial lakes in the Western Himalayas using an enhanced breakpoint method and CubeSat imagery, PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 93, 401–419, https://doi.org/10.1007/s41064-025-00345-y, 2025.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.



Kabir, S., Patidar, S., Xia, X., Liang, Q., Neal, J., and Pender, G.: A deep convolutional neural network model for rapid prediction of fluvial flood inundation, Journal of Hydrology, 590, 125 481, https://doi.org/10.1016/j.jhydrol.2020.125481, 2020.

Kingma, D. P. and Ba, J.: Adam: a method for stochastic optimization, in: Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015.

Koenker, R. and Bassett, G.: Regression quantiles, Econometrica, 46, 33–50, https://doi.org/10.2307/1913643, 1978.

Li, X., Liu, Y., Jiang, L., Huang, R., Zhou, Z., and Pang, X.: A review on remote sensing monitoring of glacial lake change and glacial lake outburst floods in mountain glacier regions, Journal of Geo-information Science, 26, 1019–1039, https://doi.org/10.12082/dqxxkx.2024.230458, 2024.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, in: International Conference on Learning Representations (ICLR), New Orleans, USA, 2019.

Lovell, H. and Muhammad, S.: Multiple phases of ice-dammed lake formation and drainage associated with a surge of Shisper Glacier, western Karakoram, Journal of Glaciology, 71, e10, https://doi.org/10.1017/jog.2024.80, 2025.

Lützow, N., Veh, G., and Korup, O.: A global database of historic glacier lake outburst floods, Earth System Science Data, 15, 2983–3000, https://doi.org/10.5194/essd-15-2983-2023, 2023.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., and Gascon, F.: Sen2Cor for Sentinel-2, in: Image and Signal Processing for Remote Sensing XXIII, vol. 10427 of *Proc. SPIE*, p. 1042704, https://doi.org/10.1117/12.2278218, 2017.

Maslov, K. A., Persello, C., Schellenberger, T., et al.: Globally scalable glacier mapping by deep learning matches expert delineation accuracy, Nature Communications, 16, 43, https://doi.org/10.1038/s41467-024-54956-x, 2025.

McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17, 1425–1432, https://doi.org/10.1080/01431169608948714, 1996.

Min, Y. and Mukkavilli, S. K.: Predicting ice flow using machine learning, arXiv [preprint], https://doi.org/10.48550/arXiv.1910.08922, 2019.

Murphy, A. H.: Climatology, persistence, and their linear combination as standards of reference in skill scores, Weather and Forecasting, 7, 692–698, https://doi.org/10.1175/1520-0434(1992)007<0692:CPATLC>2.0.CO;2, 1992.

NASA: Terra orbital drift information, https://terra.nasa.gov/about/terra-orbital-drift-information, last access: 29 October 2026, 2026.

Nurakynov, S., Sydyk, N., Baygurin, Z., and Balakay, L.: Advancements in remote sensing for monitoring and risk assessment of glacial lake outburst floods, Geosciences, 15, 211, https://doi.org/10.3390/geosciences15060211, 2025.

Nye, J. F.: Water flow in glaciers: jökulhlaups, tunnels and veins, Journal of Glaciology, 17, 181–207, https://doi.org/10.3189/S002214300001354X, 1976.

Osti, R. and Egashira, S.: Hydrodynamic characteristics of the Tam Pokhari glacial lake outburst flood in the Mt. Everest region, Nepal, Hydrological Processes, 23, 2943–2955, https://doi.org/10.1002/hyp.7405, 2009.

Quincey, D. J., Richardson, S. D., Luckman, A., et al.: Early recognition of glacial lake hazards in the Himalaya using remote sensing datasets, Global and Planetary Change, 56, 137–152, https://doi.org/10.1016/j.gloplacha.2006.07.013, 2007.

Racoviteanu, A., Paul, F., Raup, B., Khalsa, S. J. S., and Armstrong, R.: Challenges and recommendations in mapping of glacier parameters from space: results of the 2008 Global Land Ice Measurements from Space (GLIMS) workshop, Annals of Glaciology, 50, 53–69, https://doi.org/10.3189/172756410790595804, 2009.

Rahnemoonfar, M., Varshney, D., Yari, M., and Paden, J.: Deep ice layer tracking and thickness estimation using fully convolutional networks, arXiv [preprint], https://doi.org/10.48550/arXiv.2009.00191, 2020.

https://doi.org/10.5194/egusphere-2026-4316
Preprint. Discussion started: 3 November 2026
© Author(s) 2026. CC BY 4.0 License.



Richardson, S. D. and Reynolds, J. M.: An overview of glacial hazards in the Himalayas, Quaternary International, 65–66, 31–47, https://doi.org/10.1016/S1040-6182(99)00035-X, 2000.

Sahu, R.: Conventional and deep learning approaches for glacial lake mapping using remote sensing data: a comprehensive review, International Journal of Remote Sensing, 46, https://doi.org/10.1080/01431161.2025.2496001, 2025.

Samek, W., Wiegand, T., and Müller, K.-R.: Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models, ITU Journal: ICT Discoveries, 1, 1–10, 2017.

Sharma, A., Prakash, C., and Thakur, D.: Glacier lake detection utilizing remote sensing integration with satellite imagery and advanced deep learning methods, Applied Geomatics, 16, https://doi.org/10.1007/s12518-024-00594-3, 2024.

Shirsath, N. and Bhojane, V.: Glacier flood detection using deep learning through satellite images, SSRN [preprint], https://doi.org/10.2139/ssrn.4832583, 2024.

Taylor, C., Robinson, T. R., Dunning, S., Carr, J. R., and Westoby, M.: Glacial lake outburst floods threaten millions globally, Nature Communications, 14, 487, https://doi.org/10.1038/s41467-023-36033-x, 2023.

Tom, M., Odermatt, D., David, C. H., et al.: Monitoring Earth's glacial lakes from space with machine learning, SSRN [preprint], https://doi.org/10.2139/ssrn.5200494, 2025.

Tweed, F. S. and Russell, A. J.: Controls on the formation and sudden drainage of glacier-impounded lakes: implications for jökulhlaup characteristics, Progress in Physical Geography, 23, 79–110, https://doi.org/10.1177/030913339902300104, 1999.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need, in: Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 2017.

Wan, Z., Hook, S., and Hulley, G.: MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061, NASA EOSDIS Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/MODIS/MOD11A1.061, 2021.

Wilson, E. B.: Probable inference, the law of succession, and statistical inference, Journal of the American Statistical Association, 22, 209–212, https://doi.org/10.1080/01621459.1927.10502953, 1927.

Worni, R., Huggel, C., and Stoffel, M.: Glacial lakes in the Indian Himalayas – from an area-wide glacial lake inventory to on-site and modeling based risk assessment of critical glacial lakes, Science of the Total Environment, 468–469, S71–S84, https://doi.org/10.1016/j.scitotenv.2012.11.043, 2013.

Xie, Z., Haritashya, U. K., Asari, V. K., Bishop, M. P., Kargel, J. S., and Aspiras, T. H.: GlacierNet2: a hybrid multi-model learning architecture for alpine glacier mapping, arXiv [preprint], https://doi.org/10.48550/arXiv.2204.05818, 2022.

Zandler, H., Abermann, J., Robson, B. A., Maschler, A., Scheiber, T., Carrivick, J. L., and Yde, J. C.: Deep learning outperforms existing algorithms in glacier surface velocity estimation with high-resolution data: the example of Austerdalsbreen, Norway, Frontiers in Remote Sensing, 6, 1586 933, https://doi.org/10.3389/frsen.2025.1586933, 2025.