

FORECASTING AVALANCHES IN DATA-SPARSE CENTRAL ASIA USING SNOWPACK AND WEATHER MODELING WITH AVALANCHE OBSERVATIONS

Itai Sheleg^{*1}, Doug Chabot², John Snook³, Jaime Peters¹, Walter Steinkogler⁴, Ron Simenhois³

¹Lake County High School, Leadville, CO, USA

²Latok, LLC, Bozeman, MT, USA

³Colorado Avalanche Information Center, CO, USA

⁴Wyssen Avalanche Control AG, Switzerland

ABSTRACT: Avalanches in Central Asia, particularly in Afghanistan, Tajikistan, and Pakistan, are a persistent hazard, causing hundreds of fatalities in severe winters and routinely destroying homes, infrastructure, and livestock essential for survival. To address this risk, in 2015, Aga Khan Agency for Habitat established a remote avalanche forecasting program supported by external expertise. Following its implementation, it became clear that a primary constraint on forecasting capability is the limited availability of snowpack, weather, and avalanche observations, driven by both geographic remoteness and government restrictions on data sharing. We address this gap by developing a proof-of-concept framework that fuses heterogeneous, publicly available data sources to support avalanche forecasting in data-sparse regions. Specifically, we combine numerical weather prediction (NWP)-forced SNOWPACK simulations (4 km grid spacing) with opportunistically extracted avalanche observations from social media posts shared by local residents. This approach treats informal reports as weakly labeled observations, enabling the reconstruction of avalanche occurrence in the absence of systematic records. We compiled 69 valley-floor avalanche observations from social media and AKAH field logs across two winters, and matched each to the nearest SNOWPACK station sharing its predominant aspect, reconstructing the associated meteorological and snowpack conditions at 30 virtual stations. Five daily variables drive the model: total snow depth, 24-hour new snow, maximum air temperature (TA_{max}), the whole-profile minimum natural stability index ($Sn38$), and the burial depth of the weakest layer. To distinguish avalanche from non-avalanche days, we fit a hierarchical Bayesian logistic regression that gives each station its own baseline avalanche rate while sharing the influence of the five variables across the network. Trained on 2024–2025 and evaluated on the unseen 2025–2026 season, the model ranks avalanche days above quiet ones (AUC-ROC 0.715), and *its warnings are about seven times more precise than chance*. Total snow depth carries the most weight, followed by weak-layer burial depth, which lowers risk as it deepens. New snow and TA_{max} both raise risk, with warming weighing about as much as new-snow loading, suggesting that warm, likely rain-influenced storms are at least as important as loading in driving these events. *Detection climbs with a station's accumulated history, from 56% with one recorded avalanche to 100% with three, counting events from any season rather than only the current winter.* At a fixed threshold, false alarms do not fall along that same curve, though four of the stations already meet an operational standard of roughly one false alarm per month. This work incorporates a novel data-fusion and weak-supervision framework for avalanche forecasting, integrating NWP-driven snowpack modeling with opportunistic, publicly sourced observations. However, this framework is constrained by a limited event dataset and a model chain that lacks rigorous, component-wise validation, introducing uncertainty in both inputs and predictions. As such, the framework is best interpreted as a decision-support tool rather than a fully autonomous forecasting system. Coupled with expert validation and interpretation, it provides a scalable pathway for developing operational guidance in regions where conventional data streams are sparse or unavailable.

KEYWORDS: Avalanche Forecasting, SNOWPACK, Bayesian Modeling, Central Asia

1. INTRODUCTION

Avalanches are a severe and recurring hazard across the Hindu Kush, Pamir, Karakoram, and western Himalaya. A single 2012 avalanche cycle in Afghanistan, Pakistan, and Tajikistan killed more than 100 people (Chabot and Kaba 2016),

motivating the Aga Khan Development Network and Aga Khan Agency for Habitat (AKAH) to build a community-based avalanche warning system around village-run weather monitoring posts, coordinated evacuation planning, and trained search-and-rescue teams. When a comparably severe cycle struck the same region in February 2023, 38 people died (Chabot 2024), a fraction of the 2012 toll attributed to the network built in the intervening years. Providing operational avalanche forecasts is a central component of these pro-

^{*} Corresponding author address:

Itai Sheleg, Lake County High School,
Leadville, CO, USA,
email: itai.sheleg@gmail.com

grams, yet the observational data that underpin modern forecasting are largely unavailable across the region. Modern avalanche forecasting relies on dense observational networks, including automated weather stations, snowpack observations, and systematic avalanche records. Across much of Central Asia, this infrastructure is largely absent, and the challenge has three components. Field snowpack observations are essentially absent. Automated weather stations are rare at avalanche-relevant elevations, and where they exist, cost constraints and government restrictions on cross-border data sharing often prevent their use in regional forecasting (Chabot 2024). Systematic records of avalanche occurrence needed to establish historical baselines and evaluate model performance are similarly lacking. Each limitation would significantly constrain an operational forecasting program; together, they render conventional forecasting approaches impractical. We combine three globally available data sources: numerical weather prediction (NWP) model output, satellite imagery, and avalanche events identified from social media and local news reports to test whether these limitations can be substantially mitigated. Systematic records of avalanche events in this region exist almost entirely outside peer-reviewed literature, in technical reports, news, and social media (Acharya et al. 2023), making these sources not merely supplementary but the primary available record. This paper focuses on the NWP-driven snowpack modeling and observation components of the framework; satellite-derived inputs are part of the broader system architecture but fall outside the present evaluation. This paper presents the system architecture and a pilot deployment in Afghanistan, Pakistan, and Tajikistan. The operational network runs 1155 virtual SNOWPACK stations; the evaluation reported here uses the 30 stations matched to 69 avalanche observations compiled from AKAH's structured observation logs and social media reports. We evaluate how well the model separates avalanche days from quiet ones, whether that skill holds up in a season it has never seen, and whether the weights it learns are physically coherent, then discuss the steps required to transition the framework toward reliable operational forecasting in data-sparse regions worldwide.

2. DATA

2.1. NWP Forcing and SNOWPACK Simulations

We implemented the Weather Research and Forecasting model (WRF) (Skamarock and Klemp

2008) over Central Asia in 2022 to produce winter weather forecasts. Initial and lateral boundary conditions come from the US National Weather Service Global Forecast System (GFS). We run a nested grid, a 12-km outer domain with a 4-km inner nest, and put out 7-day forecasts twice a day. Post-processing turns that into standard parameter imagery and point forecasts on a regional website. The first winter told us the forecasts were worth building on. The following season we added SNOWPACK (Lehning et al. 2002; Morin et al. 2020), following the approach the Colorado Avalanche Information Center developed (Snook et al. 2022; Snook 2016). WRF writes meteorological output straight into SNOWPACK, which simulates layer-by-layer snowpack evolution at each grid point. The system builds profiles at every 4-km grid point above 3,500 m inside the forecast domain, updates them daily, and serves them on the same website, with full seasonal profiles available on demand. For this study, we pulled SNOWPACK output from 30 virtual stations that turned out to be nearest to at least one avalanche report across the Darvoz district of Tajikistan and northwest Pakistan. Each station carries a specific aspect (N, E, S, W, or flat) and elevation, so the network represents snowpack evolution tuned to slope exposure rather than one regional average. We ran two consecutive winters: 2024–2025 for training, 2025–2026 for evaluation. Sub-daily `.smet` files gave us the meteorological time series, and `.pro` files gave us per-layer stratigraphy at each timestep. Table 1 lists the five variables the model uses. Figure 1 shows the station network and a representative forecast window from 2025–2026.

2.2. Avalanche Observations

AKAH collects avalanche reports and forwards them to the forecaster. Almost all of them come from locals' Facebook pages in Afghanistan and Pakistan, and off Instagram in Tajikistan. Most carry video or photographs showing how far the avalanche ran, which is more than we get from most formal reporting channels in the region.

These posts give valley-floor runout locations, not start zones. Each has a date, an approximate location, and rough estimates of slope angle, aspect, and size, derived from the posts rather than measured. AKAH provided the initial notification of avalanche events, but every piece of observational data used here was pulled independently from publicly available posts. The combined set holds 69 observations: 26 from 2024–2025 and 43 from 2025–2026. We treat these as weakly labeled events. A post confirms that an avalanche

Table 1: The five model features, their SNOWPACK source, and physical role.

Feature	Source	Physical role
HS	SNOWPACK	Total snow depth: size / valley-reach potential
HN24	NWP	24-hour new snow: primary loading trigger
$TA_{\max}$	NWP	Daily max air temperature: wet vs dry mechanism
$Sn38_{\min}$	SNOWPACK	Whole-profile minimum stability index: weakest-layer strength
depth_lower_wl	SNOWPACK	Weak-layer depth: proximity to ground / basal layer

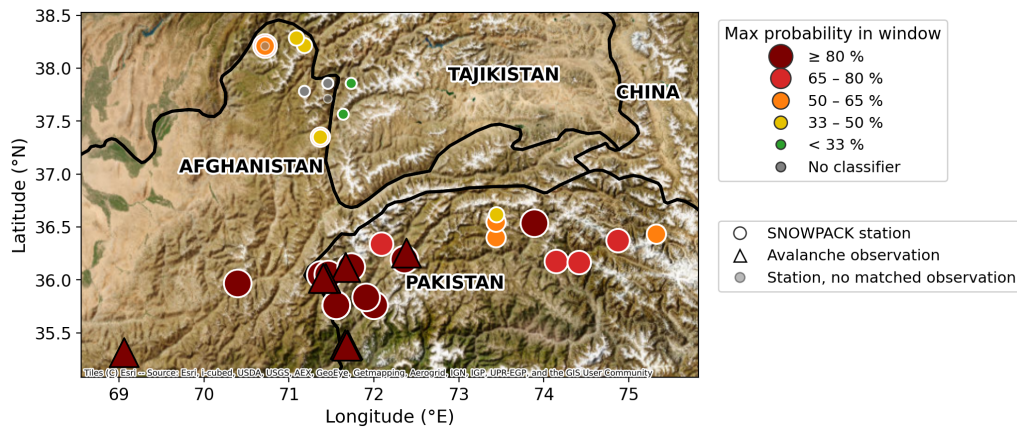


Figure 1: Study area map with the January 23, 2026 event cycle selected. Triangles mark the valley-floor avalanche reports for that date; circles are SNOWPACK virtual stations with rings colored by modeled probability in the January 22–25 forecast window. The dark-red cluster ($\geq 80\%$) coincides with the reported events, while stations without matching observations are shown faded.

reached the valley floor. It says nothing about the trigger, the release zone, or the state of the snowpack. More importantly, no post does not mean no avalanche. Our record is opportunistic and almost certainly incomplete, so some unknown number of real events sit in the data as quiet days. That is a positive-unlabeled problem (Elkan and Noto 2008), and the practical consequence works in our favor. Any skill we report is a conservative floor, because some of the quiet days the model gets penalized for flagging probably did have avalanches nobody posted about. The supplementary technical report (Sheleg et al. 2026) covers this in more detail, including why we apply no formal correction for it.

2.3. Observation-to-Station Matching

Valley-floor reports provide approximate runout locations rather than start zones. For each observation, the forecaster, who has regional terrain knowledge, worked out the likely start zone from valley geometry, known paths, and whatever context came with the report. We then matched each start zone to the nearest SNOWPACK station sharing the same predominant aspect (N, E, S, or W), so the simulation reflects the right solar radiation and wind loading regime. That hybrid step puts

terrain knowledge where distance alone would fail. Two stations equidistant from a reported runout but facing opposite aspects can carry completely different snowpack histories. This matching introduces forecaster subjectivity and is not automatically reproducible. We consider that a feature rather than a gap. It is the mechanism through which terrain expertise enters the framework.

3. METHODS

3.1. Feature Engineering

We aggregated sub-daily SNOWPACK output to one value per day. From the weather side we took daily maxima for total snow depth (HS), 24-hour new snow (HN24), and air temperature ($TA_{\max}$). From the modeled profile, we split each timestep into an upper zone, everything buried shallower than 40% of total depth, and a lower zone below. The upper zone is where storm slabs live; the lower zone is where persistent weak layers sit. Within each zone we took the minimum $Sn38$ stability index, plus the burial depth of the weakest layer in the lower zone. The 2024–2025 training set holds 35 positive days across the 13 stations with usable 2024–2025 SNOWPACK data. Anything larger would be modeling noise instead of signal. The five



Figure 2: A valley-floor avalanche photo of the kind local observers post to Facebook and Instagram and AKAH compiles into the observation dataset. The debris is visible, but the start zone is not. This is the gap this paper’s start-zone matching (Section 2.3.) is built to close.

variables in Table 1: HS , $HN24$, TA_{max} , $Sn38_{min}$, and $depth_lower_wl$ are physically distinct and do not track each other closely.

3.2. Labeling Under Weak Supervision

An event day gets labeled positive, and so does the day before it, accounting for the one-day reporting lag inherent in social media and field log sources. Every other day is negative. That leaves roughly 0.7% of days positive in the 2024–2025 training set, an imbalance worth remembering when reading the skill scores. Some stations had no valid zone features on positive days, usually early season with little snow down. Those fell back to weather alone (HS , $HN24$, TA_{max}), sacrificing snowpack structural information but preserving trainability.

3.3. Combining Station and Regional Data

We built one model that lets every station borrow from the rest of the region while keeping its own local signal (Gelman and Hill 2007). Each of the

13 stations with at least one recorded event holds its own baseline avalanche rate. The influence of the five variables is shared across the whole network. How much a station leans on itself versus the region is learned from the data, not set by us. A station with several observed events keeps mostly its own signal. A station with one event leans heavily on the regional average. That is the behavior you want from a sparse network: trust local history where there is enough of it, fall back on regional experience where there is not. We left stations with no recorded events out of training. In this region, an empty record more likely means nobody was watching than that nothing ran, and training on the opposite assumption teaches the model the wrong lesson. At prediction time, those stations are served by the regional pattern. This also gives us a confidence estimate on every prediction at no extra cost, which Section 4.4. describes. The model is a hierarchical, or multilevel, logistic regression, fit with PyMC (Abril-Pla et al. 2023).

3.4. Evaluation

We tested the model three ways:

1. **Rolling per-station test.** For each station, we put its observed events in time order and asked whether the model warned before each one, using only what it knew beforehand. Each test trains on everything that came earlier at that station, pulling from either season, and tests on the next event. Training only looks backward, so no future information leaks in. This measures how much accumulated local history helps. It mixes seasons on purpose, so it is not a test of a brand-new year. We counted an event as detected if the highest predicted probability in the three days up to and including the event reached 50%.
2. **Cross-season holdout.** We trained on 2024–2025 only and evaluated entirely on 2025–2026, so nothing from the test season ever reached training. This is the honest test of whether the model holds up in a year it has never seen. We report two scores. AUC-ROC measures how well the model ranks dangerous days above quiet ones, where 1.0 is perfect, and 0.5 is a coin flip. Average precision is the more useful of the two when real events are rare, and they are: about 2% of station-days carry a positive label in the 2025–2026 evaluation season, versus 0.7% in 2024–2025 training. That gap is season coverage, not a change in reporting; avalanches were recorded at all 30

matched stations in 2025–2026 but only 15 of 30 in 2024–2025, so more of the pooled 2025–2026 days sit near a reported event (Saito and Rehmsmeier 2015).

3. **Event-level scoring.** For each test event, we took the model's highest probability in the three days before it as the warning score, then compared that against equal-length windows on quiet days at the same stations. This asks the question a forecaster cares about, whether a timely warning went out, rather than whether every single day got labeled correctly.

4. RESULTS

4.1. Per-Station Performance and Learning Curve

We asked the practical question at each station: using only what the model knew before an avalanche ran, did it flag the day? Sixteen stations had enough history to test. Eight caught every event. Three caught about half. Five missed completely, and all five had exactly one prior event to learn from. The pattern inside that failure is the most useful thing we found (Table 2 and Figure 3). *With one logged event at a station, whichever season it came from, the model caught 56% of the next ones. With two, 86%. With three, it caught all of them.* Those counts are cumulative across the record rather than within a single winter, so an event logged in an earlier season still carries weight. Every added observation delivers a large, consistent gain, and the dataset sits on the steep part of that curve. The limiting factor here is not the model but the number of recorded avalanches. Three events at a station is roughly where we would start trusting it. Detection is only half the ledger: false alarms at a fixed threshold do not improve along the same curve, and the full accounting (threshold sweeps and per-station false alarm rates) is in the supplementary technical report (Sheleg et al. 2026).

4.2. Cross-Season Performance

The harder test is whether the model holds up in a season it has never seen. We trained on 2024–2025 and evaluated on the 2025–2026 season's 37 unique station-day events (43 raw reports that season, six of which shared a station and date with another report and so collapsed into one event), with nothing from that winter in the training data. It ranked dangerous days above quiet ones with an AUC-ROC of 0.715, *and its warnings landed about seven times more precisely than random guessing.* We also tried simpler setups: one

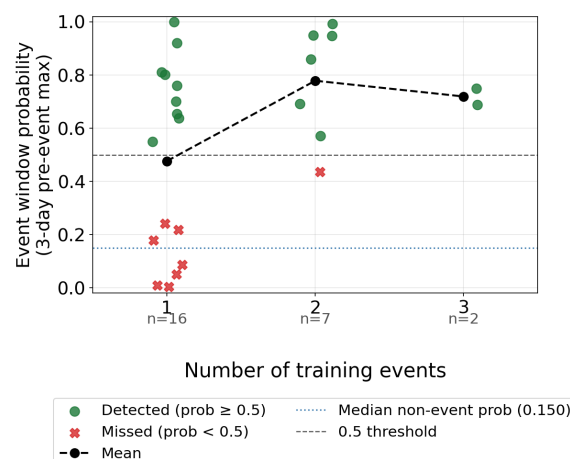


Figure 3: Learning curve: detection rate and mean warning probability as a function of the number of prior observed events at a station (any season).

model per station, one model for the whole region, and a blend of the two. The shared, hierarchical model beat all of them. A station with few logged events borrows the pattern from better-observed stations nearby while still keeping whatever local history it has. The full model comparison and statistical detail are available in the supplementary technical report (Sheleg et al. 2026).

4.3. What the Model Learned

The weights the model settled on are foundational rules every forecaster already knows (Figure 4). Total snow depth dominates: a deep snowpack carries the volume and energy needed to reach the valley floor, and thin years do not produce these events. Weak-layer burial depth comes second and pulls the other way. The deeper a weak layer sits, the harder it is to trigger, so shallow weak layers drive the modeled risk up. New snow and warm temperatures both raise risk, and warming carries about as much weight as loading. The *Sn38* stability index adds a smaller, secondary signal. These patterns hold across the region and align with what experienced forecasters expect from the terrain and climate in Central Asia.

4.4. Knowing What the Model Doesn't Know

A key advantage of the shared, hierarchical approach is that it reports not just a probability, but how confident it is in that probability. Where a station has several logged events, the estimate is tight. Where it has none, the model falls back on the regional average and flags uncertainty in the prediction. A high probability at a station with history behind it is a strong signal with confidence.

Table 2: Detection rate improves steeply with each additional observed event.

Prior events at station (any season)	Detection rate	Mean warning probability
1	56%	0.48
2	86%	0.78
3	100%	0.72

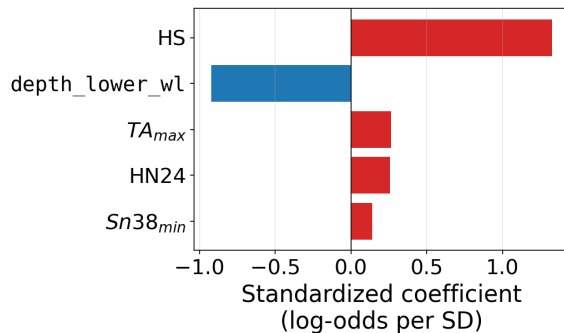


Figure 4: Feature influence on predicted avalanche risk, with standardized model coefficients as a horizontal bar chart. Red bars increase risk; blue bars decrease risk; longer bars mean stronger influence.

The same number at a station with nothing behind it is a reason to prompt extra caution and expert judgment rather than automated action.

4.5. Operational Readiness

We set the bar at roughly one false alarm a month, no more than 10% of non-event days, since a forecast nobody believes is worse than no forecast at all. Four stations are operationally ready today. One threw zero false alarms across the entire 2025–2026 season; two more flagged fewer than three days all winter. Three stations sit in a marginal band. Those are usable if a forecaster reviews every alert before it goes out, which is what we currently do. The rest are not ready, almost always because they have too few logged events to set a threshold on. Another season of reports will move stations from not-ready to marginal or ready.

4.6. Case Study: Model Output for a Known Event

Figure 5 shows the best station through the 2025–2026 season. The probability spikes right before the January 23 avalanche cycle, then drops back within days. A narrow, time-specific signal like this is something a forecaster can act on, versus a persistent high-risk background that gets ignored. This chart is what field staff and forecasters actually look at.

5. DISCUSSION

5.1. Strong Signal from Sparse Data

The system pulls a lot of signal out of very little data. We held it to five common variables: snow depth, new snow, air temperature, weak-layer strength, and burial depth. That keeps it tied to real avalanche triggers instead of chasing noise. It also means a forecaster can see the reasoning: the logistic-regression coefficients show exactly which variable pushes the probability up on a given day, so a forecaster can weigh that against their own expertise rather than just trusting the number. The model also improves as it accumulates local observations. It already performs well on the small training pool available today, and every additional observation makes a measurable improvement.

5.2. A Chain of Unverified Assumptions

Building a forecasting tool for a data-sparse region forces a chain of assumptions, and each link carries risk that compounds down the chain. With no weather stations at avalanche elevation, we feed SNOWPACK forecast weather instead of measured weather. Errors in that forecast land in the simulated snowpack, where they stack up layer by layer into a stability estimate that can drift a long way from what is actually happening. We have no surface stations to catch that drift and no pits to check the profile against. The modeled stratigraphy is a hypothesis, not an observation.

Reading the warning signal has its own catch. Our virtual stations sit higher and colder than the actual start zones, so we cannot see precipitation type where an avalanche actually releases. Checking the model's warm-flagged events against traced start-zone elevation and a standard lapse rate backs up a rain-influenced release for most of them, but a bare station-level threshold on its own got at least one of them wrong. Thus, a physically sensible coefficient still needs terrain-aware interpretation before it becomes an operational claim. The full methodology and per-event results are in the supplementary technical report (Sheleg et al. 2026).

The reports carry their own bias. We only know about avalanches somebody saw. A path that runs

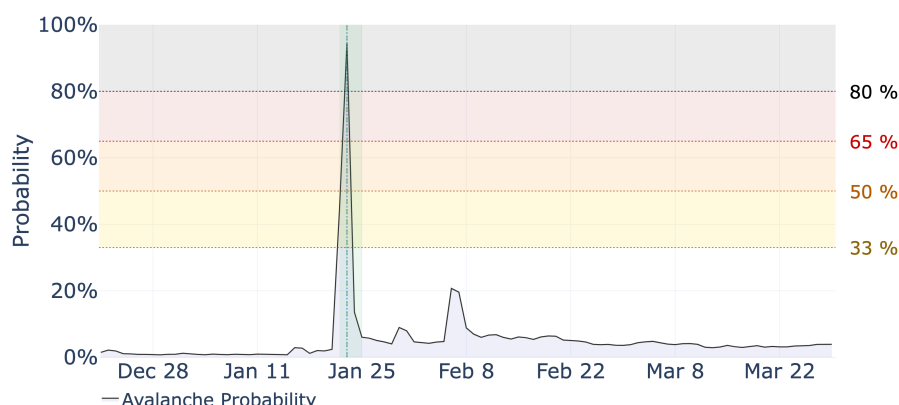


Figure 5: Per-station avalanche probability for station 160942_res (east-facing station) over the 2025–2026 season, with operational threshold bands. The model assigns probability ~ 1.00 in the highlighted three-day window around the observed January 23 event, with markedly lower probabilities on surrounding non-event days.

in an empty valley leaves no record, and the model logs that day as quiet. The record leans toward valleys with people and roads in them, and under-samples remote terrain.

5.3. Decision Support, Not a Replacement

These gaps set a hard limit on how forecasters should use the tool. It informs a person's decision; it does not make it. We built it as a second opinion for a forecaster, not a replacement for one. Expert judgment is not an afterthought bolted onto the output: it is what turns a valley-floor report into a start zone in the first place, the step every observation passes through before it ever reaches the model. The model scales that judgment; it does not replace it.

5.4. Future Work

Four things would close the biggest gaps, and the first one costs the least.

A standing avalanche log. The single highest-value improvement available to this framework is not an algorithm or an instrument. It is a written record of avalanches that AKAH staff already witness. Our learning curve puts a number on it: every station that accumulated three logged events detected all of its subsequent avalanches, while every station that missed completely had exactly one event to learn from. Moving a station from one record to three bought more skill than any modeling change we tested, and it requires only a date, an approximate location, and rough size, aspect, and elevation, captured on a form as staff encounter avalanches during their normal work.

No new sensors, no field campaign, and no additional budget line stand between AKAH and that gain. We therefore recommend that AKAH and its partner organizations treat routine avalanche logging as operational infrastructure and sustain it season after season. Each additional winter of records lifts more stations past the three-event threshold, and the benefit compounds, because every event logged this year sharpens each forecast the model issues in the years that follow.

Satellite detection. The observation bias is the worst weakness we have, and remote sensing is the direct fix. Synthetic-aperture radar (SAR) sees avalanche debris through cloud and darkness, so it catches the paths that run in remote, unpopulated drainages that no observer would ever report. Capturing these detections into the training corpus would enlarge the record and lessen the bias. Just as important, it would let us hold model output against activity somebody actually measured rather than activity we simulated.

A regional picture. Point probabilities at a handful of stations do not tell you what the snowpack is doing across the whole range. We want to overlay satellite detections and human reports onto the terrain so a forecaster can see at a glance where things have been failing regionally, then read each station's probability in that wider context.

Runout modeling. Right now we match a report back to a start zone and stop there. Coupling the stability output to a flow and runout model, building directly on the terrain-tracing already in the pipeline, would let the system reason about individual paths: which ones can reach the road, which ones can reach the village, and under what

loading. That turns a per-station probability into a spatially explicit hazard estimate.

6. CONCLUSIONS

We ran NWP-forced SNOWPACK and matched the output against avalanche reports locals posted to Facebook and Instagram. In countries with no weather stations at start-zone elevation and nobody digging pits, that gives a forecaster something to work with. Trained on 2024–2025 and tested on 2025–2026, *the model put dangerous days above quiet ones and flagged them seven times more precisely than chance*. Four stations already meet an operational false-alarm standard.

What it learned makes physical sense. Deep snow drives valley-floor runouts, shallow weak layers raise the odds of triggering an avalanche, and warming carries about the same weight as new snow. Since our stations sit higher and colder, we can't see precipitation type at the actual start zones. Since warm-signal events needed less total snowfall to trigger an avalanche than the cold ones did, we assume that these events are consistent with rain-on-snow, and since rain loads the slab and weakens it at the same time, it takes less of it to go. None of this is checked against actual snowpack observations, so every pattern here is a physically sensible hypothesis, not a proven one. But a model with no physical grounding wouldn't land on the mechanisms forecasters already trust by chance.

The fastest way to improve isn't a different algorithm. It is more logged events. *Detection went from just over half to nearly certain between one and three historical avalanche events at a station*. Every season AKAH keeps logging avalanches is the cheapest upgrade this framework gets, cheaper than any sensor, any satellite tally, and any change of algorithm, and it is the only one that costs nothing to start. And because the model reports its own confidence, a high probability at a station with several logged events is a real signal ready to alert a forecaster; the same number at a station with no local history is a reason to lean on experience and terrain judgment instead.

We view this framework as a decision-support layer whose reach grows with every observation added, and as a template for avalanche forecasting in data-sparse mountain regions worldwide.

ACKNOWLEDGEMENTS

We want to thank the American Avalanche Association for supporting this work through the ISSW

2026 Young Professional Scholarship. We are also indebted to the dedicated work of AKAH country-level staff.

REFERENCES

- Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., and Zinkov, R. (2023). PyMC: a modern, and comprehensive probabilistic programming framework in Python. In: *PeerJ Computer Science* 9, e1516. <https://doi.org/10.7717/peerj-cs.1516>.
- Acharya, A., Steiner, J. F., Walizada, K. M., Ali, S., Zakir, Z. H., Caiserman, A., and Watanabe, T. (2023). Review article: snow and ice avalanches in high mountain Asia – scientific, local and indigenous knowledge. In: *Natural Hazards and Earth System Sciences* 23, pp. 2569–2592. <https://doi.org/10.5194/nhess-23-2569-2023>.
- Chabot, D. (2024). "Snow, rain, and an earthquake: a massive avalanche cycle in Tajikistan and Afghanistan". In: *Proceedings of the International Snow Science Workshop*. Tromsø, Norway.
- Chabot, D. and Kaba, D. (2016). "Avalanche forecasting in the central Asian countries of Afghanistan, Pakistan and Tajikistan". In: *Proceedings of the International Snow Science Workshop*. Breckenridge, CO, USA.
- Elkan, C. and Noto, K. (2008). "Learning classifiers from only positive and unlabeled data". In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 213–220. <https://doi.org/10.1145/1401890.1401920>.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Lehning, M., Bartelt, P., Brown, B., Fierz, C., and Satyawali, P. (2002). A physical SNOWPACK model for the Swiss avalanche warning services. Part II: snow microstructure. In: *Cold Regions Science and Technology* 35.3, pp. 147–167.
- Morin, S., Horton, S., Techel, F., Bavay, M., Coléou, C., Fierz, C., Gobiet, A., Hagenmuller, P., Lafaysse, M., Ližar, M., Mitterer, C., Monti, F., Müller, K., Olefs, M., Snook, J. S., Herwijnen, A. van, and Vionnet, V. (2020). Application of physical snowpack models in support of operational avalanche hazard forecasting: a status report on current implementations and prospects for the future. In: *Cold Regions Science and Technology* 170, p. 102190. <https://doi.org/10.1016/j.coldregions.2019.102910>.
- Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. In: *PLOS ONE* 10.3, e0118432. <https://doi.org/10.1371/journal.pone.0118432>.
- Sheleg, I., Chabot, D., Snook, J., Peters, J., Steinkogler, W., and Simenhois, R. (2026). *Methods and Results: Avalanche Forecasting in Data-Sparse Central Asia (Supplementary Technical Report)*. URL: https://github.com/ronimos/caw-avalanche-mapping/blob/main/docs/methods_and_results.md (visited on 07/26/2026).
- Skamarock, W. C. and Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for research and NWP applications. In: *Journal of Computational Physics* 227, pp. 3465–3485.
- Snook, J. S. (2016). "Weather forecast model grid spacing – is smaller better?" In: *Proceedings of the International Snow Science Workshop*. Breckenridge, CO, USA.
- Snook, J. S., Cooperstein, M., and Greene, E. (2022). "Snowpack modeling efforts at the Colorado Avalanche Information Center". In: *Proceedings of the International Snow Science Workshop*. Bend, OR, USA.