Introductory statement:

The manuscript titled "Predicting avalanche danger in Northern Norway using statistical models" is focused on applying machine learning algorithms for estimating the avalanche danger and interpreting the results for hindcasting the period from 1970 to 2023. The meteorological input and the snow cover information is provided by Norwegian reanalysis NORA3. The results were associated and discussed with climate indices.

A distinction is made between multiple classifications (danger levels 1 to 4, level 5 not considered) and a binary classification (1-2 or 3-4). Random forest classification and artificial neural networks (ANN) were used in the study. Results of the ANN were mainly provided in supplementary information.

The manuscript is an essential contribution to the natural hazard community. It describes the methods and results precisely. In the discussion, climate indicators are carefully linked to avalanche activity in northern Norway.

The notes in the main comments must be considered for publishing the manuscript.

Response:

We thank the reviewer for the careful reading of our manuscript and the for the extensive comments. We have considered all the comments and provide a point-for-point response below.

We note that in the meantime we were made aware of updates and additional material at the International Snow Science Workshop 2024 and based on this we have added some recent references and small points to the manuscript.

The new references include among others Pérez-Guillén et al. (2024a) who found in their case study in Switzerland that on days without an avalanche the average danger level was 1.9 ± 0.8 and on days with an avalanche it was 3.2 ± 0.5. We now use this as an argument for considering our binary case as a measure of avalanche activity. Another new reference is Bee et al. (2024) which appears to be the most recent contribution when it comes to connecting avalanches with climate modes (e.g., NAO) in Europe. Further references are Winkler et al. (2024) and Techel et al. (2024) who gave updates on the status of the inclusion of machine learning models in operational avalanche forecasting in Switzerland, as well as Herla et al. (2024) who updated on the implementation of the SNOWPACK model in Norway. Further references are given in the responses to the individual comments.

References:

Bee, C., Zugliani, D., and Rosatti, G.: A correlation between avalanches and teleconnection indices in the Italian Alps, International Snow Science Workshop Proceedings 2024, Tromsø, Norway, pp. 147-152, http://arc.lib.montana.edu/snow-science/item/3126, 2024.

Herla, F., Widforss, A., Binder, M., Müller, K., Horton, S., Reisecker, M., and Mitterer, C.: Establishing an operational weather & snowpack model chain in Norway to support avalanche forecasting, International Snow Science Workshop Proceedings 2024, Tromsø, Norway, pp. 168-175, http://arc.lib.montana.edu/snow-science/item/3129, 2024.

Pérez-Guillén, C., Simeon, A., Techel, F., Volpi, M., Sovilla, B., and van Herwijnen, A.: Integrating automated avalanche detections for validating and explaining avalanche forecast models, International Snow Science Workshop Proceedings 2024, Tromsø, Norway, pp. 52-57, https://doi.org/http://arc.lib.montana.edu/snow-science/item/3111, 2024a.

Techel, F., Helfenstein, A., Mayer, S., Perez-Guillen, C., Purves, R., Ruesch, M., Schmudlach, G., Soland, K., and Winkler, K.: Human vs. machine - Comparing model predictions and human forecasts of avalanche danger and snow stability in the Swiss Alps, International Snow Science Workshop Proceedings 2024, Tromsø, Norway, pp. 31‑38, http://arc.lib.montana.edu/snow-science/item/3108, 2024.

Winkler, K., Trachsel, J., Knerr, J., Niederer, U., Weiss, G., Ruesch, M., and Techel, F.: SAFE - a layer-based avalanche forecast editor for better integration of model predictions, International Snow Science Workshop Proceedings 2024, Tromsø, Norway, pp. 124‑131, http://arc.lib.montana.edu/snow-science/item/3123, 2024.

Major comments

Major comment #1

The introduction of an article should provide background information to the topic, it should explain why it is important, it should explain past attempts to solve the problem, and it should mention the specific objectives of the study. The introduction meets these criteria. However, the introduction of the manuscript additionally describes the results of the study and discusses them. This should not be part of the introduction. I suggest shortening or removing the lines 117-140.

Response:

We thank the reviewer for this suggestion. In our view, one aim with an introduction should generally also be to place the present study in the research context and point out novelties, thus requiring a brief summary of the work. However, we agree with the reviewer that our description of the results was too extensive. We have now considerably shortened lines 117-140 and only briefly point to the novelty regarding the linkage between (Norwegian) avalanches and climate modes, which is shown by our study, since this appears a topic of untapped research potential. Lines 117-140 are now replaced by:

"We optimise two different RF models: (1) for the original '4-level case' (ADL 5 has not been forecast in northern Norway) and (2) for a 'binary case', where ADLs 1 and 2 and ADLs 3 and 4 are combined. The latter allows us to perform an 'avalanche activity' hindcast for 1970-2023 and to investigate a linkage between avalanches and regional climate modes, such as the North Atlantic Oscillation (NAO) or the Arctic Oscillation (AO). Our findings have potential implications for the predictability of avalanche activity and danger, which is a salient point as only a few studies have previously investigated connections between avalanches and regional climate modes."

Major comment #2

In the section "Summary and conclusions" (lines 498-501) the results are compared to previous studies using accuracy as metric. However, the accuracy does not only depend on the algorithms which were developed in the study. The accuracy also varies for each data set (size, kind of test data, proportion between training and test data, climate region, topography, etc.). For this reason, comparisons with other studies are difficult to interpret. Add this information to the manuscript.

Response:

We have added the following in the Summary and conclusions section:

“However, these studies, including our work, differ in type and quality of data, in background climate and topography, as well as warning-region size. Thus, the comparison of accuracies between different studies should be regraded with care.”

Continuation Major comment #2

Additionally, using the previous day's value of ADL simplifies the task enormously. However, the authors only mention this in passing in section 5 "Model evaluation" (lines 380-381). A paragraph which discusses the problems when comparing metrics of different studies would help readers to interpret the results correctly.

Response:

We are a little unsure what the reviewer means here. First, we respectfully disagree that we only mentioned “in passing” the point about using the previous day’s ADL. We have considered this aspect and based on the findings of Perez-Guillen et al. (2022) we decided not to include it. However, we have now added a further argument, i.e., that our model is also intended to be applied in a hindcast setting as well as (in upcoming work) for future projections of avalanche activity, and for both these applications no previous-day ADLs exist. Second, the whole paragraph from 364-389 in the original manuscript is mainly concerned with describing the differences between accuracies in different studies and what could be the possible reason for these differences. That is, essentially it discusses why it is difficult to compare the different studies. However, we have explicitly added this point and now try to discuss more potential differences:

“Notably, our warning regions in northern Norway have an average size of about 6500 km^2 (see Table S4 in the supplementary information), while in Switzerland the average size is about 200 km^2 (Perez-Guillen et al., 2024b). The smaller warning regions potentially imply a clearer connection of avalanche danger to meteorological conditions and thus generally less noisy data, which may explain part of the higher prediction accuracies of the Swiss models. More fundamentally, the different climates and topographies of the different study regions generally complicate comparisons among studies. Much of the cited work was conducted in Central Europe (i.e., in the mid-latitudes) while our study area is in northern Norway and thus in the Arctic. The mountains in the Alps are often higher and the climate is more continental than in the fjord landscape of northern Norway. This leads to different snow and avalanche characteristics (e.g., van Herwijnen et al., 2024) and potentially implies differences in predictability, thus hampering comparability across studies.”

Note that we have added an additional figure (Fig. S6) to the supplementary information showing the warning regions in Norway as well as an additional Table (Table S4) giving the names, region codes, and areas in km^2 of the regions. Both figure and table are included here at the end of the response.

We hope these changes accommodate the concerns of the reviewer.

New references added here:

Perez-Guillen, C., Techel, F., Volpi, M., and van Herwijnen, A.: Assessing the performance and explainability of an avalanche danger forecast model, EGUSphere [preprint], https://doi.org/10.5194/egusphere-2024-2374, 2024b.

van Herwijnen, A., Muccioli, M.,Wever, N., Saiet, E., Mayer, S., and Pugno, N.: Is Arctic snow different from alpine snow? Delving into the complexities of snow cover properties and snow instability, International Snow Science Workshop Proceedings 2024, Tromso, Norway, pp. 401‑408, http://arc.lib.montana.edu/snow-science/item/3165, 2024.

Major comment #3

The study of Dekanová et al. (2018) is very similar to an earlier study published by Stephens et al. (1994). Add this information to the appropriate places in the manuscript.

Response:

We thank the reviewer for pointing out this connection. We were aware of this study, but since it is different from our study in that it is based on avalanche occurrence instead of avalanche danger (similar to some of the studies suggested in Major comment #4), we had decided not to discuss it further. However, we have now added the required information where we first refer to Dekanová et al. (2018).

We note that previously we were convinced that Dekanová et al. (2018) used avalanche danger levels from their regional avalanche bulletin to train their ANN. However, upon further reading and comparing the two studies, we admit to being somewhat confused regarding the training data in Dekanová et al. (2018). They initially state in section *A.* that the output of their neural network is "avalanche danger in range between zero and one". However, subsequently they state in section *B.* that the output of the neural network is "avalanche danger degree in international danger level scale." Also, in this section they first state that the knowledge base of the neural network is past weather and *determined avalanche danger*, while further below they also mention *past known avalanches* (i.e., avalanche occurrence). Thus, we are unsure if the model is trained on avalanche danger or on avalanche activity/occurrence. Despite this confusion, we have decided to keep the parts of the discussion about Dekanová et al. (2018) unchanged (except for the reference to Stephens at el., 1994).

Stephens, J., Adams, E., Huo, X., Dent, J. and J. Hicks (1994). Use of neural networks in avalanche hazard forecasting. In proceedings of the International Snow Science Workshop 1994, Snowbird, UT: 327-340.

Major comment #4

The authors have carefully cited earlier studies. However, references from the Asian region are missing. The following references are examples (incomplete list).

Joshi, J. C., Kumar, T., Srivastava, S., Sachdeva, D., & Ganju, A. (2018). Application of Hidden Markov Model for avalanche danger simulations for road sectors in North-West Himalaya. Natural Hazards, 93(3), 1127–1143. https://doi.org/10.1007/s11069-018-3343-7

Joshi, J. C., Kaur, P., Kumar, B., Singh, A., & Satyawali, P. K. (2020). HIM-STRAT: a neural network-based model for snow cover simulation and avalanche hazard prediction over North-West Himalaya. Natural Hazards, 103(1), 1239–1260. https://doi.org/10.1007/s11069-020-04032-6

Yousefi, S., Pourghasemi, H.R., Emami, S.N. et al. A machine learning framework for multi-hazards modeling and mapping in a mountainous area. Sci Rep 10, 12144 (2020). https://doi.org/10.1038/s41598-020-69233-2

Yariyan, P., Omidvar, E., Minaei, F., Abbaspour, R. A., & Tiefenbacher, J. P. (2021). An optimization on machine learning algorithms for mapping snow avalanche susceptibility. Natural Hazards, 111(1), 79–114. https://doi.org/10.1007/s11069-021-05045-5

Response:

We thank the reviewer for these references of which we were mostly unaware.

We recognise that our review of previous research did not cover all previous work or all regions, but we in fact did include references from Asia. Blagovechshinskiy et al. (2023) investigated a region in Kazakhstan, and we extensively referred to this study and compared it with our and other previous studies. Furthermore, we twice referred to a study from the Tianshan Mountains in China (Hao et al., 2023). However, since this study investigated avalanche activity instead of danger, and is hence to some degree out of the scope of our study, we did not consider it in more detail. Some of the studies suggested by the reviewer (Yousefi et al., 2020; Yariyan et al., 2022) appear to diverge from our methodology in that they train machine learning models on observed avalanche occurrences and afterwards classify the resulting probability into something akin to avalanche danger levels (e.g., Yariyan et al., 2022, p. 103, and similarly, Yousefi et al., 2020, Fig. 1). Thus, these studies appear out of the scope in the context of our study, and we would like to refrain from adding them as references. On the other hand, Joshi et al. (2020) is close to our work since they train their model on danger levels from an avalanche bulletin. We have added this study in our discussion of previous work.

Minor comments

line 8: reorder words

replace "… optimized and trained …"

with "… trained and optimized …"

Response: Done.

line 10-11

The second part of the sentence is unclear. Does the "confusion" relate to (i) the model or to (ii) the underlying observational data?

Response:

We are unsure what exactly the reviewer means here. The applied statistical model is a random forest that at the basis classifies the data by building decision rules with thresholds (in decision trees). It is nearly impossible to say if the reason for the misclassification is fundamentally due to the data (for instance, they are just too noisy to be classified more clearly) or due to insufficiency of our model. To hopefully increase clarity and precision, we have changed the second part of this sentence to: "…, which is due to the latter model often misclassifying ADL 1 as ADL 2 and ADL 4 as ADL 3."

Continuation line 10-11

In general, avoid using the word "confusing" in a scientific context. Explain the reasons for misclassification of 1-2 and 3-4

(i) Is it easier for the model to decide between two classes compared to four classes? Why? Would larger data sets solve the problem?

(ii) Is the source of uncertainty a human factor? Or is the regional scale (and simplifications) the origin of the uncertainty? Are the weather prediction models uncertain?

Response:

See the response above. Also, we do not think there is space for further speculation/explanation of the misclassification in the abstract. However, we had briefly covered this in section 5 (lines 394-395 and 404-409 of the original manuscript). We have now changed the part in lines 394-395 slightly to: "Thus, a large part of the misclassification in the 4-level case is due to the confounding of levels 1 and 2 and levels 3 and 4. This explains the higher accuracy achieved in the binary case, where these levels are combined to one class, respectively."

We believe that the large size of the warning regions and the implied noisy relation between danger level and weather data is one of the main reasons for the misclassification and have accordingly added the following to the end of section 5:

"More fundamentally, we again point to the large warning regions in Norway. Various meteorological conditions may simultaneously be prevalent within a given region, implying a noisy relationship between the weather data and the ADLs, likely contributing to the high rates of misclassification. A decrease of warning region size may be necessary for a clearer relationship between weather data and ADLs to substantially reduce misclassification and increase prediction accuracy."

Note that "again" here refers to earlier parts of the text we added to accommodate Major comment #2, which are documented in our response to this comment above.

line 28: reformulate sentence

replace "… for industry, farming, and fishery and are thus strongly important for the planning…"

with "… for industry, farming, and fishery are important for the planning…"

Response:

The sentence reads: "Other environmental indicators, such as nutrient concentration may be related more to the conditions for industry, farming, and fishery and are thus strongly important for the planning of these industries and their infrastructure (…)"

With the suggested change the sentence becomes intelligible. Thus, we would like to retain the original formulation, except that we will remove the word "strongly".

line 34

replace "… meteorological weather data."

with "… meteorological data."

Response: Done.

line 34-35

Remove or simplify the sentence "Thus, this data …", because it is obvious that meteorological data affect snow avalanches.

Response:

Please note that in this part of the Introduction we were still talking in general about environmental parameters or indicators that may be statistically inferred from meteorological data and we were not yet specifically referring to snow avalanches. We turned to snow avalanches only in the following paragraph (lines 38-41). Furthermore, our point here was that one of the reasons for using meteorological data to infer environmental indicators is that they are so widely modelled. Thus, we would like to retain this sentence as is.

line 37

replace "... on the modelled future changes in weather conditions."

with "… on climate scenarios."

Response: Changed.

lines 38-41: comment

From my perspective, "Hazard" is often used in long term context (e.g. hazard mapping) and the term "danger" is related to the present situation (e.g. danger level). McClung's (2002ab) articles can probably help to understand the differences.

McClung, D.M. The Elements of Applied Avalanche Forecasting, Part I: The Human Issues. Natural Hazards 26, 111–129 (2002a). https://doi.org/10.1023/A:1015665432221

McClung, D.M. The Elements of Applied Avalanche Forecasting, Part II: The Physical Issues and the Rules of Applied Avalanche Forecasting. Natural Hazards 26, 131–146 (2002b). https://doi.org/10.1023/A:1015604600361

Response: We thank the reviewer for pointing out this informative pair of articles.

line 46

replace "note" with "noted"

Response: Done.

line 56

replace "… avalanche occurrence and danger…"

with "… avalanche occurrence and hazard mapping…"

Response:

The part of the sentence where this formulation occurs reads: "… as climate change likely impacts avalanche occurrence and danger (…)." We would like to retain this formulation as we think it is more consistent that climate change impacts avalanche danger itself rather than the mapping of avalanche hazard.

line 58

remove “exact”

Response: Done.

line 76

reorder “Lehning et al., 2002b, a” to Lehning et al., 2002a, b”

Response:

We were aware of this issue, but we were hoping this would be fixed later in copy editing since it appears to require changes to the copernicus.bst file which controls the bibliography style. However, we have now fixed this by manually changing the .bbl file.

lines 117-140

these lines describe what the authors did. And this should not be content of the introduction section. Some components belong to the section methods and others are assigned to the results. Remove, shorten or move these lines.

Response: Please see the response to major comment #1.

Figure 1

The rectangle in the inset map is rotated. But this is wrong! This is indicated by the grid lines of the main figure which are not rotated. The lower left corner of the main figure is located between islands, but the inset map shows this point in the sea.

The boundary lines do not match.

Response:

We recognised the ca. 45° rotation in the rectangle in the inset unfortunately only after manuscript was published as a preprint. We thank the reviewer for pointing it out. This is now corrected.

line 174

replace “discuss” with “discussed”

Response: Done.

line 175

replace “find” with “found”

Response: Done.

lines 216-219

Reformulate these sentences, because they are unclear and avoid phrases like “work horse”.

Response:

Please not that “work horse” is the term used in Morin et al. (2020), which is why we quoted it here. However, we have removed this term (see the updated passage below).

We have rewritten these lines which now read:

“The NORA3 reanalysis provides no data on the snow conditions at the surface. Thus, in order to obtain information about, e.g., the snow depth and density and snow water equivalent (SWE), we run the snow model seNorge (Saloranta, 2012) version 1.1.1 (Saloranta, 2014, 2016) using NORA3 daily 2-m temperature and total precipitation amount as input. Due to a lack of both in-situ and satellite observational data on snow, seNorge is the main tool used to provide information on snow for the avalanche warning system in Norway (Saloranta, 2012; Morin et al., 2020).”

The remainder of the paragraph is left as is.

Table 1

replace “400 m” with “400 m a.s.l.” and replace “900 m” with “900 m a.s.l.”

Response: Done.

lines 226-232

Remove this paragraph, because neither SNOWPACK nor CROCUS were used in context with the present study.

Response: We have removed this paragraph.

line 263

reformulate "... little impact ...". The effect was so low that it was not considered below.

Response: Changed to “… no impact …”

line 285

A dot is missing at the end of the sentence.

Response: Added.

line 331

replace "... similarly find ..."

with " similarly found ..."

Response: Done.

line 338

replace "... wind ..." with "... wind speed ..."

Response:

The part of the sentence where this word appears reads "..., since both new snow and wind, especially associated with storms (e.g., Davis et al., 1999), are prominent avalanche triggers..."

We would like to retain the word "wind" here, as it appears to us more sensible to call the wind itself the avalanche trigger instead of the wind speed.

Table 4 typing error

replace "hypperparameter"

with "hyperparameter"

Response: Done.

line 340

remove "Interestingly,". The subsequent sentence already starts with "This is remarkable ...".

Response: Done.

line 351

replace "... snow cover." with "... snowpack."

Response: Done.

line 470

replace "We choose ..." with "We decided ..."

Response: Done.

Supplementary information: line 21

The threshold for categorization is 0.5. However, the case equal 0.5 is undefined. Use either "lower than or equal to" (<=0.5) or "larger than or equal to" (>=0.5).

Response:

We thank the reviewer for pointing out this oversight. We used the “lower than or equal to” case and have changed this accordingly in the text.

Added Tables and Figures:

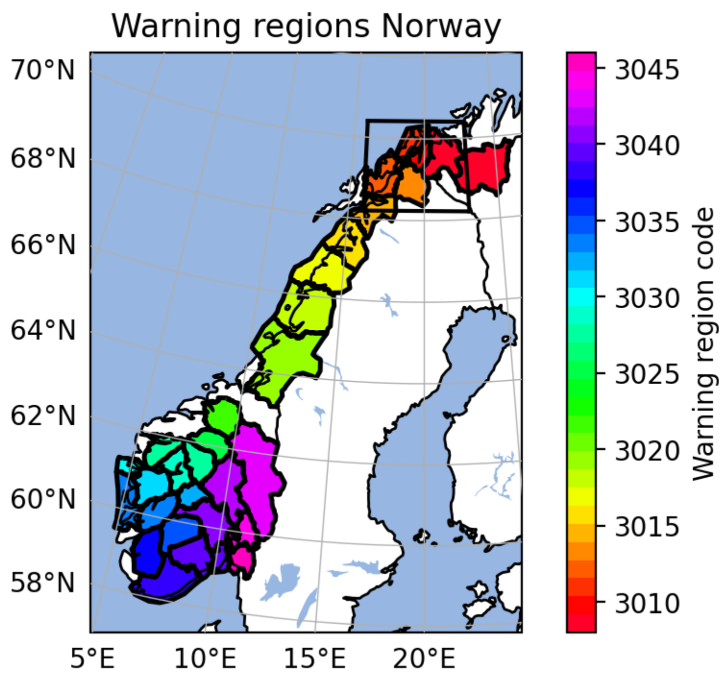

Figure S6: Map of Norwegian avalanche warning regions. The black rectangle indicates our study region in northern Norway. The colour corresponds to the region code (see Table S4).

Table S4. List of the Norwegian avalanche warning regions. The area is given in km^2. See Fig. S6 for a map of the individual regions.

region code	region name	area
3008	Finnmarksvidda	15062.0
3009	Nord-Troms	9392.1
3010	Lyngen	2842.5
3011	Tromsø	4722.1
3012	Sør-Troms	8202.5
3013	Indre Troms	7371.7
3015	Ofoten	7579.8
3016	Salten	10279.1
3017	Svartisen	13718.3
3018	Helgeland	18110.7
3019	Nord-Trøndelag	27303.6
3022	Trollheimen	8098.8
3025	Nord-Gudbrandsdalen	6035.8
3027	Indre Fjordane	5715.9
3028	Jotunheimen	6586.0
3029	Indre Sogn	6678.7
3030	Ytre Sogn	2703.0
3031	Voss	7778.6
3032	Hallingdal	4240.9
3033	Hordalandskysten	8160.7
3034	Hardanger	7134.4
3035	Vest-Telemark	8083.1
3037	Heiane	8636.3
3038	Agder sør	13343.3
3039	Telemark sør	8600.3
3040	Vestfold	3132.0
3041	Buskerud sør	10534.7
3042	Oppland sør	13077.7
3043	Hedmark	26999.0
3044	Akershus	4988.7
3045	Oslo	474.1
3046	Østfold	4801.8
	average northern Norway	6506.2
	average Norway	9074.6