

Response to the Reviewers

Reviewer #1

To allow a better communication of our replies, we have divided them into three categories: **Agree/Clarification/Disagree**.

1.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
This manuscript investigates the influence of different data-scaling strategies on SVM-based fault classification using synthetic and borehole-derived TIN datasets. The topic is relevant to machine-learning-assisted geological modelling, particularly because preprocessing choices are often treated as a technical detail despite their potentially important influence on spatial predictions. The manuscript therefore addresses a useful methodological problem. However, in its present form, substantial revisions are required before the main conclusions can be fully supported.	We believe that the paragraph provided by Reviewer #1 summarizes our manuscript very well.	None.

2.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
---	-------------------	--------------------------

I provide the following major comments. 1. Definition of spatial additivity and spatial continuity The manuscript should distinguish more clearly between spatial additivity, prediction invariance, spatial continuity, and geological accuracy. The current subset experiment mainly evaluates whether predictions for the same observations change when those observations are processed within different spatial extents. This is a measure of prediction or spatial additivity, but it does not directly demonstrate that the resulting fault traces are geometrically or topologically continuous. I recommend introducing a formal mathematical definition of spatial additivity in the methods and using separate metrics for fault continuity.	Agree/Clarification We agree that the definition of spatial additivity should be in Methods. We believe that it is too early to formally introduce the metric “fault continuity”. Therefore, only “spatial additivity” is formalized.	We’ve moved the definition of “spatial additivity” to Methods (section 3.6). We’ve removed the occurrences of “spatial continuity”, as we believe that it is too early to provide a strict definition and to differentiate between “spatial additivity” and “spatial continuity”. We’ve corrected the indices (replaced “n” with “k” for the iterated subset or classification because “n” should stand only for the last in the sequence.
---	---	---

3.

Suggestion, Question, or Comment from the Reviewer #1	Author’s Response	Change in the Manuscript
---	-------------------	--------------------------

2. Dataset description and reproducibility Although the Study Area section provides enough geological information about the 810 boreholes, the actual dataset used in the study for the machine learning remains insufficiently described. A dedicated subsection should explain the spatial distribution and characteristics of the boreholes, the specific stratigraphic horizon extracted from each borehole, how these observations are converted to a TIN, the resulting number of vertices and triangles, and exactly what constitutes one observation supplied to the supplier.	Clarification We agree that more information could be useful. However, the bullet points provided by the Reviewer appear to contain both data-specific and methodological information. Therefore, we believe that the components dispersed across different sections of the manuscript should be corrected instead of a dedicated subsection. Additionally, we would argue that spatial distribution of boreholes is best conveyed through a figure rather than text. We believe that some of the requested information was available in the submitted manuscript such as the stratigraphic horizon marked in Fig. 6. However, we've attempted now to make it even more clear. 1. the spatial distribution and characteristics of the boreholes, 2. the specific stratigraphic horizon extracted from each borehole, 3. how these observations are converted to a TIN, 4. the resulting number of vertices and triangles, 5. and exactly what constitutes one observation supplied to the supplier.	We have corrected Fig. 6 by adding borehole points. We have corrected Fig. 6 by adding borehole points. We added a sentence to the caption of Fig. 6 to explain that the studied horizon is marked using a red arrow. We have provided more information about creating the TIN model in the first paragraph of Methods (section 3.1). We have added more information about the number of vertices (810) and triangles (1603) in section 5.1.
--	--	---

		We have added the information about features of an observation in section 5.1.
--	--	--

4.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
The manuscript should also clearly separate the synthetic and real datasets. The number of synthetic surfaces, parameter ranges used for regional dip, dip direction, fault throw, noise, and other variables, the distributions from which these parameters are sampled, and the class proportions should be reported. A table summarizing these parameters would greatly improve reproducibility.	Agree We agree that a table including the following information should improve reproducibility: 1. The number of synthetic surfaces, 2. parameter ranges used for regional dip, dip direction, 3. fault throw, 4. noise, and other variables, 5. the distributions from which these parameters are sampled, 6. and the class proportions	• We've created a new section in Results (5.1 Features used for classification and parameters) • We've added a new Table which presents the parameters requested by the Reviewer

5.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
3. Relationship between synthetic and real data The method assumes that the synthetic training surfaces represent the geometric characteristics of the real-borehole derived surface. This is a fundamental assumption and should be evaluated rather than only stated. I recommend comparing the distributions of the principal geometric attributes between synthetic and real datasets. Otherwise, differences in classification may partly result from synthetic-to-real domain shift rather than preprocessing alone.	Agree We agree that comparing the distributions of the principal geometric attributes between synthetic and real datasets is a reasonable idea. The main conclusion from the comparison is that synthetic surfaces are smoother than real data (lower variance of features). Therefore, we recommend adding a greater level of noise in future studies.	We've provided a comparison (chapter 6.7). We've conducted an analysis for means and standard deviations (Table B2). We've created QQ plots to better present differences (Figs. B1-B2).

6.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript

4. Mathematical description of the three scaling methods Since scaling architecture is the central methodological variable, the three approaches should be defined mathematically. This would make the distinction among independent subset scaling, per-surface local scaling, and training-derived global scaling much clearer than the current predominantly descriptive presentation. The workflow figures should also be revised. Software-specific terms such as FIT_TRANSFORM and TRANSFORM are less informative for the general geological readership than explicitly showing where the mean and standard deviation are estimated and where those same parameters are subsequently reused. A single three-panel comparative figure may be more effective than separate figures.	Clarification Initially, we were requested by the Editor to adjust the abstract to justify why our paper is relevant to Solid Earth which is mostly read by geologists. As such, we believe that too much mathematical formalism has the potential of obscuring the idea between scaling methods. We believe that the current presentation of scaling methods allows better reproducibility because it directly illustrates data organization and order of tasks. We agree that the software-specific terms could be avoided. We would like to respectfully disagree with the suggestion to merge the three figures into one figure. The problem is that there is a small font applied for some components of Figure 3 (describing Method 2) and the smallest characters would likely be unreadable.	We have removed software-specific terms such as “fit_transform” and “transform” and replaced it with “learn&scale” and “scale”, respectively. To better explain where the mean and standard deviation are estimated and reused, we improved caption to Fig. 4 using the proposed vocabulary.
---	---	--

7.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
---	-------------------	--------------------------

6. Data-leakage discussion The Data Leakage subsection is useful but relies heavily on a limited number of sources (Kumar and Narayan) and focuses mainly on conventional ML leakage. The discussion should be expanded to include spatial dependence and spatially correlated train/test observations, which are directly relevant to TIN-based geological data.	Agree/Clarification The Reviewer raised a very interesting and relevant limitation of our study which is often discussed in ecological/Remote sensing studies. We agree that when spatially dependent observations are randomly partitioned between the training and test sets, spatial autocorrelation can lead to overly optimistic performance evaluation. This result was observed in our study in a limited (largely theoretical) way. One would expect that the most optimistic results would be for Methods 1 and 3 because they allow triangles from one surface to be in training and test sets due to random split. However, the most optimistic results were observed for Method 2 in which surfaces are grouped together (first and last surfaces are in the training and test sets, respectively). We agree to add more sources about spatial dependence.	We've added three references (see below) to the manuscript which explains the issue of spatial autocorrelation for geological/remote sensing/ecological studies. To address the Reviewer's concern about possible autocorrelation for TIN data, we've expanded section 6.2.
--	--	---

References (spatial autocorrelation):

Pawley, S. M., Atkinson, L., Utting, D. J., Hartman, G. M. D., and Atkinson, N.: Evaluating spatially enabled machine learning approaches to depth to bedrock mapping, Alberta, Canada, PLoS One, 19, <https://doi.org/10.1371/journal.pone.0296881>, 2024.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., and Pélissier, R.: Spatial validation reveals poor predictive performance of large-scale ecological mapping models, Nat. Commun., 11, <https://doi.org/10.1038/s41467-020-18321-y>, 2020.

Schratz, P., Muenchow, J., Iturrity, E., Richter, J., and Brenning, A.: Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data, *Ecol. Modell.*, 406, 109–120, <https://doi.org/10.1016/j.ecolmodel.2019.06.002>, 2019.

8.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
The authors should also distinguish classical data leakage from inconsistent preprocessing. Fitting a scaler using combined training and test data constitutes direct leakage, whereas independently fitting a transformation on the test or production dataset represents a different methodological problem. These situations should not be treated as conceptually identical.	Agree We agree that these are different methodological problems and should be distinguished.	For clarification, we have expanded section 2.2 (the last sentences just before Methods).

9.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
7. Quantitative assessment of spatial additivity The Results section currently relies too heavily on visual inspection. The number of changed classifications should be treated as a primary quantitative result rather than appearing mainly in the Discussion.	Agree We agree to move Tab. 1 (changes classification) from the Discussion to Results section. We've calculated the fault-to-homocline and homocline-to-fault transitions.	We moved the table about the changed classification to Results (now Table 2 in section 5.4). We've added the calculated fault-to-homocline and homocline-to-fault transitions to the table.

For each method, the authors should report at least the number and percentage of changed classifications, together with separate counts of fault-to-homocline and homocline-to-fault transitions. A formal disagreement metric could be defined as the proportion of observations within a subset whose predictions differ between complete-domain and subset processing.	We agree that the formal disagreement metric proposed by the Reviewer makes sense. However, because showing the subset serves only to illustrate the effect of violating spatial additivity, we believe that the current data in the table are sufficient to achieve this objective.	We haven't introduced the disagreement metric.
---	--	--

10.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
The current analysis also appears to use only one selected rectangular subset. The result may depend strongly on subset size, position, point density, and local feature distribution. I therefore recommend repeating the experiment using multiple spatial subsets of different sizes and locations and reporting the mean, variability, and range of the disagreement metric for each preprocessing method.	Clarification We acknowledge that the selected subset and its properties (position, size, density, local distribution) may impact the "changed classification" metrics and the magnitude of the disagreement. However, we believe that one arbitrary subset is sufficient to illustrate the effect of violating spatial additivity, while measuring the magnitude of the disagreement is a separate topic which could be followed in future studies.	We've added the information to the caption of the table with changed classification: "We note that the magnitude of the observed effects depends on the selected subset (red rectangle marked in Fig. 8)."

11.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
Some of the tables appear in the Appendix instead of the Results, where they should be properly described.	Disagree/Clarification Please note that Appendix A contains classification results for synthetic data which is somewhat repetitive to what already has been published in Michalak et al. (2025 - a reference given below). As such, we didn't want to put at risk the originality of the work and to obscure the main concept - maintaining spatial additivity of classification. Therefore, we would prefer to leave the content in Appendix A. Please note that the content of Appendix B was discussed in section "6.3 Statistical considerations and model generalization". Appendix B provides a statistical basis for explanation why Methods 1 and 3 provide similar results despite different scaling architecture - this is not the core of the paper. After the requested changes, Appendix B also includes comparison between synthetic and real data.	None.

References:

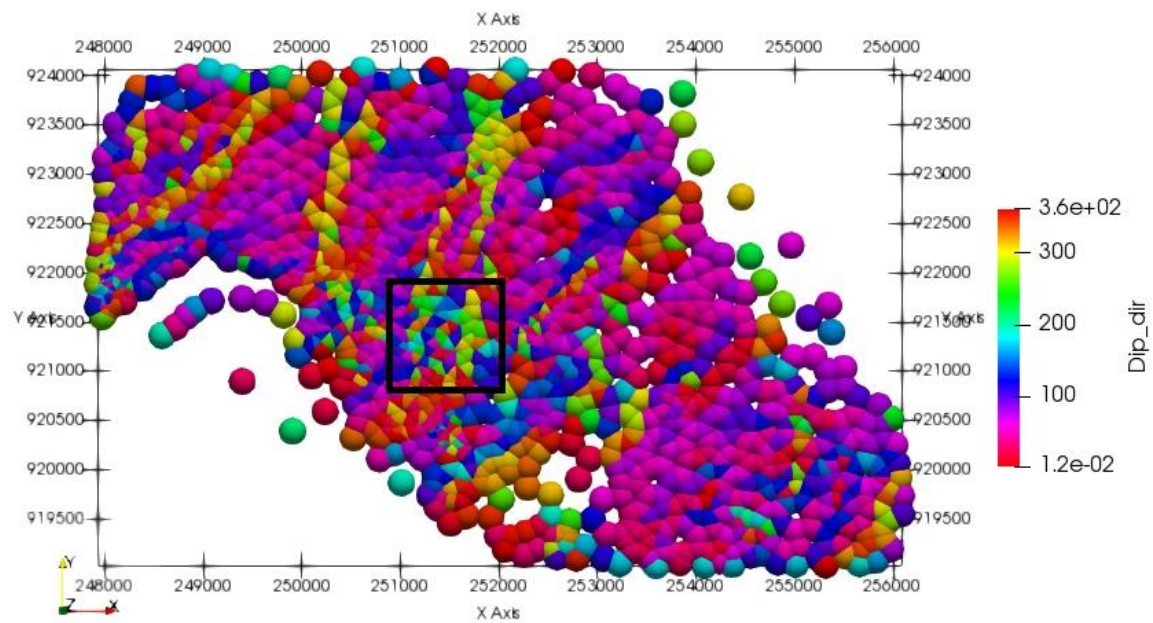
Michalak, M. P., Gerhards, C., and Menzel, P.: SubsurfaceBreaks v. 1.0: a supervised detection of fault-related structures on triangulated models of subsurface homoclinal interfaces, Geosci. Model Dev., 18, 4469–4481, <https://doi.org/10.5194/gmd-18-4469-2025>, 2025.

12.

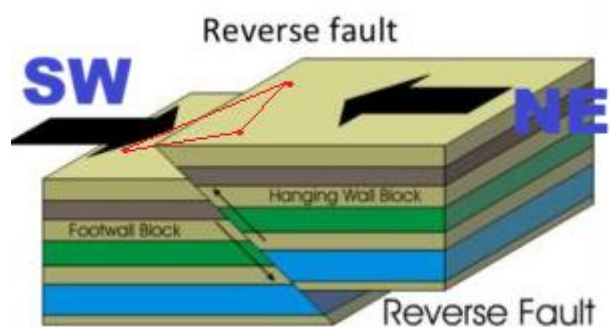
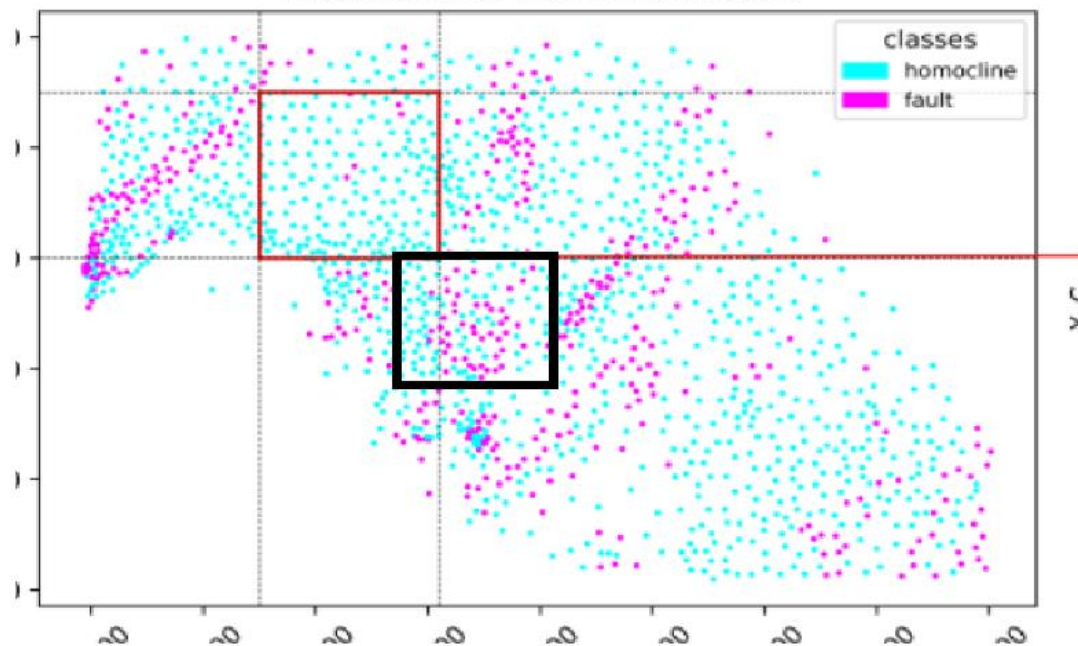
Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
8. Discussion The Discussion requires substantial strengthening. Several parts currently repeat the Results, whereas greater emphasis should be placed on explaining the mechanism causing the observed behaviour, the conditions under which local scaling becomes problematic, the influence of subset size and feature distributions, and the geological implications of the classification patterns.	Clarification We thank the Reviewer for this comment because it helped discover interesting directional patterns which can explain why the algorithm proposed faults in some locations. We propose the following points to strengthen the Discussion section: 1. Regarding geological implications: We've checked why the classification algorithm suggested faults which are perpendicular to the regional dip, which can be best observed for Method 2 (see the attached figure below). We can see that in that area (marked using a black rectangle) the triangles have a dip direction which is opposite (to SW) to the	We've added a geological interpretation of SW-oriented anomalies (section 6.4).

	regional dip of strata (to NE). One candidate hypothesis to explain this effect would be to have a reverse fault with a hanging wall and footwall with a relative location to the NE and SW, respectively and the triangles oriented to the SW (see the attached figure below). 2. Regarding the observed mechanism: The lack of spatial activity can be explained by a general rule, that the distributions of features (normal vectors, distances with neighbours) cannot be assumed equal for the entire data set and the extracted subsets. 3. An inherent property of Method 2 is the isolation of the scaling process for individual	We've added a sentence which explains the observed mechanism to section 6.2. We've added an explanation of the isolating effect of Method 2 to section 6.2.
--	---	---

	synthetic files: the standardization parameters for each file are determined solely based on its own observations; consequently, feature values from one file do not influence the mean and standard deviation used to scale the other files. This isolating effect limits the influence of anomalous surfaces on the calculated transformation parameters.	
--	--	--



Predictions for the entire dataset



13.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
The suggestion that independently scaling real data could alleviate poor representation of reality by synthetic models should be treated cautiously. Independent standardization does not generally solve synthetic-to-real domain shift.	Agree We agree with the Reviewer that the suggestion about potential advantages of independent scaling should be treated cautiously.	We have added the following sentence: "However, we advise interpreting this suggestion as tentative, as independent standardization does not resolve the synthetic-to-real domain shift." (section 6.1)

14.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
Similarly, the suggestion that Method 2 may be preferable because it produces visually smoother patterns is not sufficiently supported, particularly because the manuscript also states that this method fails to identify some known faults. Visual smoothness should not be equated with geological accuracy.	Agree We agree with the Reviewer that the suggestions about Method's 2 smoothness were not sufficiently supported and it should not be equated with geological accuracy. See also our response to Comment #12 which mentioned the isolating effect, inherent to Method 2.	We added the important limitation raised by the Reviewer to section 6.2 (second paragraph).

--	--	--

15.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
The formal definition of spatial additivity currently introduced in the Discussion should instead appear in the Methods.	Agree Done. See our reply to comment #2.	Done. See our reply to comment #2.

16.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
Likewise, much of the literature discussion presented under "Bridging the research gap" would be more appropriate in the Introduction, where the novelty of the study is established.	Agree/Clarification We can address the Reviewer's comment in a limited way because prior to the submission, we were requested by the Editor to expand the Discussion section. We received the following feedback from the Editor: <i>"Before I can accept this into review, I would need you to 1) rewrite the abstract to make it clearer how this is relevant to</i>	We've moved the initial paragraph of "Bridging the research gap" to Introduction". We have rewritten the "Bridging research gap" to additionally take into consideration: 1) research which discussed leakage due to spatial autocorrelation.

	<i>our readership and what the method brings, 2) <u>revise the discussion, to emphasize more strongly how your approach advances over other, existing approaches, and</u> 3) <u>change the manuscript type to "method".</u></i>	2) Extending the spatial additivity concept to regression tasks.
--	---	--

17.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
Figure 1 appears two times one time after the abstract and one time in the methodology. I never understood why it appears two times.	Clarification This was done according to the journal requirements. Hopefully, we interpreted it correctly: “For preprints, since these have no key figure, graphical abstracts must always be included in the paper as unnumbered graphic placed directly after the text abstract. “ https://www.solid-earth.net/submission.html	None.

18.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript

In addition, in the Introduction, 45-50 is not properly described for a non-expert reader.	Agree We agree to use a simpler language.	We have replaced 45-50 with a simpler explanation.
--	---	--

19.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
Also, in sentence 285 real data file word is not properly quoted; the commas used are wrong.	Agree/Clarification We agree that the quasi-quotation could be confusing. Instead, we propose to explain the process without any quotations and in a shorter way by connecting two sentences.	We have changed the paragraph, as follows: "In this experiment, the selected subset was processed in the same way as the original data set, which means that it received a new independent scaler "

20.

Suggestion, Question, or Comment from the Reviewer #1	Author's Response	Change in the Manuscript
Appendix A should go to the results and must be properly described and discussed. In addition, Appendix B.	Clarification/Disagree Please note that Appendix A contains classification results for synthetic data which is somewhat repetitive to what already has been published in Michalak et al. (2025 - a reference given below). As such, we didn't want to put at risk the originality of the work and to obscure the main concept - maintaining spatial additivity of classification for	None.

	updating of geological models. Therefore, we would prefer to leave the content in Appendix A. Please note that the content of Appendix B was discussed in section “6.3 Statistical considerations and model generalization”. Appendix B provides a statistical basis for explanation why Methods 1 and 3 provide similar results despite different scaling architecture - this is not the core of the paper. After the requested changes, Appendix B also includes comparison between synthetic and real data.	
--	--	--

References:

Michalak, M. P., Gerhards, C., and Menzel, P.: SubsurfaceBreaks v. 1.0: a supervised detection of fault-related structures on triangulated models of subsurface homoclinal interfaces, *Geosci. Model Dev.*, 18, 4469–4481, <https://doi.org/10.5194/gmd-18-4469-2025>, 2025.

Our own changes:

Section 6.6 (Bridging the research gap) - We’ve removed some strict claims from Discussion such as “lack of discussion”. Instead, we propose to replace it with “limited discussion”, since the Reviewer #1 provided a context (spatial autocorrelation) which allowed us to find more papers dealing with leakage but mostly in soil and hydrological science.

Nevertheless, we believe that it may be true that the discussion about leakage is very limited in seismic ML modeling because in the referenced review paper the term leakage has not been mentioned: “A systematic review of deep learning for structural geological interpretation”.

Fernandes, G. L., Figueiredo, F., Hatushika, R. S., Leao, M. L., Mariano, B. A., Monteiro, B. A. A., ... & Zavam, F. (2025). A systematic review of deep learning for structural geological interpretation: GL Fernandes et al. *Data Mining and Knowledge Discovery*, 39(1), 3.

We've corrected some typos and mistakes about best hyperparameters/numbers in tables (A1-A6). We've also performed a more comprehensive revision of language using ChatGPT (grammar, redundant expressions, awkward phrasing). The monitoring of all changes is possible using the marked version of the manuscript.