First, we want to thank the reviewers for their insightful comments and recommendations. Following the suggestions, we have revised the manuscript thoroughly incorporating all the suggested changes.
Below is our response (italics) to each comment (regular font) from the reviewers.

<u>Review by RC1:</u>

General comments

The study has merit in its outline, as a better understanding of flood impacts at this scale is definitely needed and I agree that doing this with a generally simple approach for a general kick start to the options of AI in this domain makes sense. However, I have a number of concerns on how the study is built and executed, which I believe are at this stage too big to recommend the manuscript for publication. I detail the concerns below and would encourage the authors to rethink their strategy before moving to an eventual submission. I fully understand that this is a submission from an ECR and I want to complement you on the aim and pulling this together – definitely work that should be pursued and there is a lot of demand for outcomes of such approaches! I would have hoped to see more scrutiny here before a submission from the more experienced co-author team.

My concerns range from (a) general sloppiness of manuscript writing (many simple editing mistakes that can always happen for drafts but should not occur for a submitted manuscript over (b) the lack of appreciation of existing data and simply depicting the target region as 'data scarce' to avoid scrutiny from what is known already to (c) a lack of proper documentation of data sources on the exposure side as well at times confusing jumping between topical (what types of floods) as well as spatial (national, watershed, HMA wise) domains.

Response: *We thank the reviewer for his/her in-depth and critical review. We revised the manuscript thoroughly to correct these unnoticed mistakes and improve the writing quality. Regarding the missing references and low quality of the text and figure captions, we did a thorough review of the manuscript to correct all these errors.*

We would like to note that we defined HIMAT as 'data scarce' because of the complexity of the HMA region challenged by the scarcity of ground observations covering consistent timeframes homogeneously, as highlighted by various works in literature (Barandun et al., 2020; Dollan et al., 2024; Miles et al., 2021). We understand that some of the statements were phrased in an ambiguous way, which we have revised to address accordingly.

Regarding the lack of proper documentation, in the revised paper we provided a table (Table 1) with a detailed description of the datasets considered in this study.

We further discussed different flood disasters happening in the HMA region. Specifically, our study focuses on pluvial and fluvial flooding, which we made clearer in the introduction and throughout the manuscript.

Line 139-150: The analysis follows a multistep approach, beginning with data at both watershed and district scales. Initially, the focus was on the district scale, as socioeconomic data for Nepal, selected as primary training ground, were readily available at this level through the Nepal Disaster Risk Reduction Portal (http://drrportal.gov.np/). For this region, furthermore, there is a comprehensive coverage of high-resolution (8-meter) Digital Elevation Models (DEMs) from prior High Mountain Asia (HMA) work (High Mountain Asia 8-meter DEMs Derived from Along-track Optical Imagery, 10.5067/0MCWJJH5ABYO). Subsequently, all the information is aggregated at the watershed scale, as phenomena such as fluvial and pluvial flooding occur at this level, necessitating a dataset tailored to this scale. To transfer the demographic information from the district to the watershed scale, we performed a weighted spatial join between the watersheds and districts. For each watershed, we attributed the statistical characteristics of the intersecting districts, with weights based on the overlapping areas. Generally, the districts in Nepal are smaller in extent compared to the various watersheds.

For the work, furthermore, we provided a quality assessment of the model performance at the watershed and the district scale. The results highlighted that the watershed scale proves more accurate than the district-scale training. We clarified this in the method description.

Please consider also our response to your detailed comments in the following paragraphs.

a) General statement on 'no data'
   You make general unsupported statements on the region being data scarce on hydromet data. That is decidedly not the case. While data may often not be readily accessible, it is available and many studies have been published on this, especially for China and India and data is generally reachable from China, India, Pakistan, Afghanistan as well as Central Asian states. The data that is available you do away with as 'not trustworthy' in a single sentence. This coming from an all-US based author team is problematic and I guess you could imagine how stunned a reviewer from the US (or Europe) would be if a Chinese author would make that claim before proceeding to apply ML on all of the US or Europe. You will need to make a clearer description (with references) on what is lacking and how your approach fills that gap.

Response 1: *We have revised the manuscript thoroughly to correct these unnoticed mistakes and to improve overall the quality of the writing. Please note that when we defined HIMAT as 'data scarce' we did not do so to avoid scrutiny of the data, but rather to highlight the complexity of HMA itself, challenged by several factors, including the scarcity of ground observations covering consistent timeframes homogeneously, as highlighted by other works in literature (Barandun et al., 2020; Dollan et al., 2024; Miles et al., 2021). We understand that some of the statements were phrased in an ambiguous way, which we have revised to address accordingly. The main idea behind this paper was to provide a tool for a rapid estimate of potentially highly impacted areas, based on information accessible and updateable quickly, such as population number, rainfall intensities, and a geomorphologic index that can be derived from global DEMs (or high-resolution local ones when available).*
Please consider the revisions made from line 51-90 as follows.

Line 51-90: Accurate evaluation of the socioeconomic impacts of natural disasters is paramount to mitigate the sufferings of the affected people and rehabilitation (Cavallo & Noy, 2010; Meyer et al., 2013; Noy, 2015, 2016a). To date, available studies (Diehl et al., 2021; Mohanty & Simonovic, 2022; Pangali Sharma et al., 2019; Pervin et al., 2020; Piacentini et al., 2020; Yang & Tsai, 2000) have primarily concentrated on vulnerability mapping and risk analysis, employing case studies and descriptive event-based methodologies at a local level. Scaling up the analysis over the entire HMA region is indeed a difficult task, as it requires collecting data from several countries and multiple sources, and this poses challenges due scarcity of ground observations covering consistent timeframes homogeneously (Barandun et al., 2020; Dollan et al., 2024; Miles et al., 2021). Especially in the context of the impact of floods using socioeconomic data, the analysis involves examining the number of fatalities, injured and people otherwise affected, as well as the financial damage that natural disasters cause, and this information is generally collected at the local scale based on reported events. Significant disasters are documented in global databases like The International Disaster Database (EMDAT, www.emdat.be) or, as an example for HMA and this study, the Nepal Disaster Risk Reduction Portal (http://drrportal.gov.np/). However, these databases typically operate at a global or national level resolution, potentially overlooking minor disasters. For example, EMDAT only considers events with at least one of the following criteria: 1)10 fatalities; 2)100 affected people; 3) a declaration of state of emergency; 4) a call for international assistance. Additionally, those databases utilized to support insurance may prioritize countries with existing or potential insurance coverage (World Bank, 2012).
The integration of geomorphic properties, population data, and rainfall characteristics for assessing

socioeconomic flood impact is seldom explored comprehensively on a large scale. For HMA. this is primarily due to the inherent challenges associated with conducting on-site surveys in rugged and often inaccessible terrain. However, leveraging remote sensing data has emerged as a valuable approach for delving deeper into these dynamics and effectively quantifying flood impacts. Modern global datasets, featuring improved resolution and coverage, further enhance the utility of remote sensing in this regard (Diehl et al., 2021; Jongejan & Maaskant, 2015; Mosavi et al., 2018; Bentivoglio et al., 2022; Mazzoleni et al., 2022; Hawker et al., 2018; Kirschbaum et al., 2020; Mohanty and Simonovic, 2022; Pangali Sharma et al., 2019; Sanyal and Lu, 2004; Yang and Tsai, 2000; Zheng et al., 2018).

Furthermore, machine learning (ML) techniques have emerged as increasingly popular tools in advanced prediction systems over the past two decades. They offer more cost-effective solutions with performance that can be aggregated, surpassing the complexity and time demands associated with simulating the complex development of flood processes. Recent research (Bentivoglio et al., 2022; Deroliya et al., 2022; Mosavi et al., 2018) has showcased encouraging advancements by integrating machine learning (ML) techniques with global datasets. This contemporary approach to mapping flood vulnerability notably streamlines the computational processes associated with data-intensive simulations, enhancing flood risk management strategies. However, ML systems rely on existing data for learning. Insufficient or incomplete data coverage can hinder effective learning, leading to suboptimal performance when deployed in real-world scenarios. Therefore, ensuring robust data enrichment, encompassing both quantity and quality, is imperative.

In this study, we introduce a streamlined methodology for preliminary flood vulnerability assessment on a large scale, leveraging available global datasets. Specifically, we introduce a flood-risk assessment model designed to quantify spatially distributed socioeconomic susceptibility in flood-prone regions. We utilize this model to augment disaster understanding by integrating remotely sensed data, including climate variables and high-resolution terrain information.

Finally, we apply this model in the High Mountain Asia (HMA) regions to analyze changes in socioeconomic flood impacts spanning from 1980 to 2020.

b) Poor documentation of socio-economic data
As I detail below there is very poor documentation on where the exposure data is taken from and there is no way to make this traceable (no stable links, and also no attempt so far to make your own produced data available, see comment on Availability statement).

Response 2: *We thank the reviewer for this comment. We added a table with information on all the datasets used. Regarding the links being "not stable" – the data required for the index were accessed and we tested the links before submission. In the revised paper, we added the date of the latest access so that the data is more clearly referenced. We further added in the table the 'standard' value we adopted, with the reference for each value.*

Table 1: Parameters used to calculate LYI

Variable	Description	References
M	Mortality (number of deaths due to disaster	*Nepal Disaster Risk Reduction Portal* http://drrportal.gov.np/
Aexp	Average life expectancy at birth (by year)	WHO (https://data.who.int/countries/524)
Amed	Median age (by year)	WHO (https://data.who.int/countries/524)
e	Welfare reduction weight associated with being exposed to a disaster	set to e = 0.054 according to Noy, (2016a) , based on Mathers et al., 2013
T	Time taken by the affected person to get back to normal	Noy, (2016a)

N	Number of affected people	http://drrportal.gov.np/
c	Percent of time not used in work-related activities (.75)	Noy, (2016a)
Y	Y = Financial damage (value of destroyed/damaged infrastructure)	http://drrportal.gov.np/
PCGDP	Income per capita (by year)	http://drrportal.gov.np/

c) I also fail to see how you take census data to the watershed and how you align using Nepal government data with your approach to model at the watershed scale (which do not follow national borders).

Response 3: *We added more details in the manuscript. Please see below-*

Line 139-150: The analysis follows a multistep approach, beginning with data at both watershed and district scales. Initially, the focus was on the district scale, as socioeconomic data for Nepal, selected as primary training ground, were readily available at this level through the Nepal Disaster Risk Reduction Portal (http://drrportal.gov.np/). For this region, furthermore, there is a comprehensive coverage of high-resolution (8-meter) Digital Elevation Models (DEMs) from prior High Mountain Asia (HMA) work (High Mountain Asia 8-meter DEMs Derived from Along-track Optical Imagery, 10.5067/0MCWJJH5ABYO). Subsequently, all the information is aggregated at the watershed scale, as phenomena such as fluvial and pluvial flooding occur at this level, necessitating a dataset tailored to this scale.

To transfer the demographic information from the district to the watershed scale, we performed a weighted spatial join between the watersheds and districts. For each watershed, we attributed the statistical characteristics of the intersecting districts, with weights based on the overlapping areas. Generally, the districts in Nepal are smaller in extent compared to the various watersheds.

In general, the districts we have for Nepal are of a smaller extent than those of the various watersheds.

d) General scope and methodology
At multiple points of the manuscript I was a bit confused on the scope. There is an introduction on all types of high flow events but the methods suggest you only look at fluvial floods with exceptionally high impacts. There is a relatively rapid investigation of the methods for watersheds that do lie to some part in Nepal compared against data only from areas within Nepal and then an upscaling to all of HMA, which in turn is not clearly defined in its scope or climatologies. I would strongly suggest to maybe limit the study to areas where data is available before scaling it up, allowing you more space for methodological and data based issues.

Response 4: *We thank the reviewer for this suggestion. This paper focuses on fluvial and pluvial flooding. In the revised work, we reworked parts of the introduction to be clearer about this.*

Please refer to ***Response 3*** *for the explanation as to why we chose the watershed scale for the analysis.*

Regarding the scope and climatologies, we agree with the comment, and revised the text to provide a wider context.
Line 222 and subsequent: *The climatology in HMA is highly variable (Dollan et al. 2024). Summer monsoons drive precipitation in the Ganges-Brahmaputra basins and the Tibetan Plateau (Bookhagen and Burbank, 2010; Shamsudduha and Panda, 2019); synoptic storms dominate winter precipitation impacting areas in the northwestern Karakorum mountains (Winiger et al., 2005; Barlow et al., 2005). Overall, as well, variations in elevation gradients contribute to diverse microclimates, exemplified by Nepal's swift transition from high mountains to lowlands (Kansakar et al., 2004; Karki et al., 2016). Winter precipitation in the area is primarily influenced by the westerly weather system, with western disturbances*

originating in the Mid-Atlantic or Mediterranean Sea and traversing through northwest India to western Nepal after passing over Afghanistan and Pakistan (Kansakar et al., 2004; Hamal et al., 2020). In Nepal, which was used as the training site for the model, regional climate variations exist, mostly driven by changes in elevation, with an overall homogeneity in trends (aside from a few hotspots) and regional statistics of precipitation, in line with the variability of HMA, as highlighted by the recent study by (Khanal et al., 2023).
For this work, for the main rainfall driver of the model, we focused on daily climate concentration. As climate concentration values are mostly related to the temporal variability of the rainfall, not to the total amount or the average yearly and seasonal statistics, using this index allows to capture well various climates globally (Monjo and Martin-Vide, 2016a). The variability of climate concentration, furthermore, has been proven to be highly linked to pluvial/fluvial flooding impacts in various regions of the world, including for example Italy (both in mountainous landscapes and floodplains (Sofia et al., 2019), the US (Saki et al., 2023) [over a variety of physiographic regions], or China (Du et al., 2023). Different authors have adopted different methods to determine the temporal concentration of precipitation, and the Concentration Index (CI) (Equation 2) is one of the most used parameters (Caloiero et al., 2019; Martin-Vide, 2004; Monjo, 2016; Sangüesa et al., 2018; Serrano-Notivoli et al., 2018).

We decided to use Nepal as the training site as it represents rapid variations in climatologies over a smaller extent, for which we had complete coverage of good-quality data. For Nepal, as we showed in the paper, we have a gradient of CI values, and as ML models learn from the data they ingest, we believe the system can work across various regions from the climatic point of view. We added comments on this in the paper, highlighting the strengths and weaknesses of the approach.

Response to Specific comments:

1. L31: Be careful in your framing – population growth does not increasing likelihood of flooding, it increases flood risk!

Response 5: *We rephrased this in the revised manuscript. The sentence was meant to refer to flood impacts.*

2. Also, in the abstract and your general analysis, you focus on precipitation as a flood driver but here then passingly mention glacial melt as well – those are very different flood drivers and would be crucial to be clear what kind of flooding you wish to tackle here.

Response 6: *We tried to be clear in the revised manuscript on the fact that this work focuses on pluvial/fluvial flooding. We added background information on flood hazards in general in the region, as we believe some overview of this is needed to frame the work correctly in the context of the various possible flood hazards in the region.*

3. L54: '*HMA does not have enough hydrological stations for region-wide flood monitoring*' is a huge statement to make without a citation – what is an appropriate number? Also most countries in HMA, especially China, India, Pakistan and Nepal have large and dense network of hydro(-met) monitoring, which they also use for forecasting. That is not as open as in the US, but the statement that there is 'not enough' needs to be qualified. You then claim '*Moreover, the available meteorological datasets may not be sufficiently trustworthy.*', which again lacks any qualification. Imagine me making that statement for a European or North American country, that would be thrown out. The region has a large amount of met data (see e.g. the overview figure in (Nepal et al. 2023)) and if you do not trust the data you need to justify why.

Response 7: *As mentioned earlier, we revised this part of the manuscript and rephrased our statements. It was not our intention to underrepresent available datasets. Gathering data for the scale of the HMA* *region requires collecting data from several countries. This is time-consuming and to some extent, it can be nearly impossible due to political constraints and limits on accessing the data for some regions. As a matter of fact, the goal of this study is to overcome the data and time constraints and provide a quick tool for disaster management. Also, please refer to the response regarding the datasets available throughout the project's scope.*

4. L61: '*The use of remote sensing technology for disaster studies in HMA is comparatively new*' – I also do not quite agree. Remote sensing itself isn't very old and it has been used in HMA for many studies already (which maybe anyway would need some acknowledgement here).

Response 8:
We want to reiterate our intention was not to undermine the remote sensing data used in HMA. Instead, we want to emphasize that Remote Sensing data usage particularly in disaster studies in HMA is comparatively modern. We rephrased this in the manuscript as follows,
Line 68: The integration of geomorphic properties, population data, and rainfall characteristics for assessing socioeconomic flood impact is seldom explored comprehensively on a large scale. For HMA. this is primarily due to the inherent challenges associated with conducting on-site surveys in rugged and often inaccessible terrain. However, leveraging remote sensing data has emerged as a valuable approach for delving deeper into these dynamics and effectively quantifying flood impacts. Modern global datasets, featuring improved resolution and coverage, further enhance the utility of remote sensing in this regard (Diehl et al., 2021; Jongejan & Maaskant, 2015; Mosavi et al., 2018; Bentivoglio et al., 2022; Mazzoleni et al., 2022; Hawker et al., 2018; Kirschbaum et al., 2020; Mohanty and Simonovic, 2022; Pangali Sharma et al., 2019; Sanyal and Lu, 2004; Yang and Tsai, 2000; Zheng et al., 2018).

5. L87: You focus here a lot on monsoon changes with intense precipitation – but if you actually focus on HMA (rather than just the Hindukush Himalaya) there are a lot of other processes – Westerlies in Central Asia, Eastern Monsoons in the Upper Yangtze etc. Maybe it is required to reconsider the total spatial scope of the study here?

Response 9: *We wanted to draw the focus on the fluvial/pluvial flooding caused by the precipitation. In the revised manuscript we provided a more detailed discussion of the climatology of the area, as mentioned in our previous response. Please see* ***Response 4*** *for the exact changes we made.*

6. L92: You now finally get to actual numbers of potential affected, but leave it to the reader to get the data from EMDAT.It would be prudent to explain here (or rather in the introduction) what the actual numbers are and for what types of hazards, to then narrow down and which ones you actually focus.

Response 10: *The EMDAT link was used as a reference to the statement, rather than as an actual dataset. We added the numbers calculated using EMDAT data. This was not provided to direct the reader to get the data from EMDAT.*

6. Figure 1: Up to this point there was no clear description how the watersheds are selected, i.e. what boundary you used for HMA. This needs to be provided to give context to why so many watersheds outside HMA are also included.

Response 11: *We have added the following to the manuscript.*
Line 108-114: This study considers approximately 6,000 watersheds across HMA as main target area (Figure 1): the watershed were selected to be consistent with the HMA domain and all the datasets produced throughout the different phases of the NASA-funded HiMAT project (https://himat.org/). The analysis initially centered on training and testing a machine learning model specifically for Nepal. To achieve this, we collected fine-resolution topographic data along with district-scale socioeconomic

information pertaining to population characteristics and documented flood impacts for this region. Subsequently, leveraging the insights gained from this initial phase, we extended the application of the trained model to predict socioeconomic impacts across all watersheds in HMA.

7. L116: At this point you mention that you will predict impacts of 'floods', i.e. all of them? The way you describe your research you are narrowing this down on pluvial floods, as glacial lake outburst floods or debris flows etc need very different driver analysis. Can you be precise here? In L185 you then suddenly just focus on 'fluvial flooding', so is it just that you focus on?

Response 12: *Thank you for this comment. We tried to be consistent with the terminology. And we have used only fluvial and pluvial flooding in the analysis.*

8. L120: This part is crucial as you present the socioeconomic data and how you treat it. However there are a few issues that would need to be addressed with respect to traceability and presentation of data used.

- You refer to data sources that are questionable, the knoema.com page is not stable and it is unclear from where their data is sourced or where it is known needs to be documented here.
- You refer to general government and Worldbank websites (like http://drrportal.gov.np/) that exist but what data you took from there at what point in time remains unclear. Copernicus Journals subscribe to FAIR practices, that includes the documentation of third party data used in a publication.
- You introduce a lot of data as well as parameters from literature (like T and e) without any questioning of their accuracy, uncertainty etc. This would propagate and need to be addressed, especially as you seem to upscale from this approach with a few numbers on Nepal government websites to all of HMA.
-

Response 13: *We thank the reviewer for this comment. Please note that in the revised paper we added a table with the data sources used for the work and a reference to all the literature from where we took any 'standard value' used in the calculation. We further checked the links and added a reference date for access. Please note that in the revised manuscript, given the comment regarding some of the parameters, we confirmed their value through other sources (WHO).*
In general, some standard parameters in the LYI formula (like T and e) are suggested by (Noy, 2016) and they were applied in the referenced work to calculate LYI for Nepal as a whole, so we used them consistently. We added the following table to the manuscript where we clarified the data sources considered for the analysis.

Table 1: Parameters used to calculate LYI

Variable	Description	References
M	Mortality (number of deaths due to disaster	*Nepal Disaster Risk Reduction Portal* http://drrportal.gov.np/
Aexp	Average life expectancy at birth (by year)	WHO (https://data.who.int/countries/524)
Amed	Median age (by year)	WHO (https://data.who.int/countries/524)
e	Welfare reduction weight associated with being exposed to a disaster	set to e = 0.054 according to Noy, 2016a, based on Mathers et al., 2013
T	Time taken by the affected person to get back to normal	Noy, (2016a)
N	Number of affected people	http://drrportal.gov.np/
c	Percent of time not used in work-related activities (.75)	Noy, (2016a)
Y	Y = Financial damage (value of destroyed/damaged	http://drrportal.gov.np/

	infrastructure)	
PCGDP	Income per capita (by year)	http://drrportal.gov.np/

9. You calculate these values for Nepal as a whole but then work on the watershed scale – how is this compatible?

Response 14: *Please note that we did not calculate the values for Nepal as a whole, but rather the LYI is calculated for each flood event, starting from the information provided by the Nepal Disaster Risk Reduction portal. The LYI total value is then summarized for each district as an overall number of LYI for all the events geolocated within that district. To scale it up to the watershed level, we did a weighted spatial join.*

There are indeed a few parameters that go in the LYI equation that are standard, but this is consistent with the calculation of the index as it was defined by Noy et al. 2016. Most of the other parameters, such as median age, or life expectancy at birth may vary from country to country. As we 'train' the model considering a temporal variability of this parameter, we have enough variability of this parameter to represent a good statistical sample of values. For example, Figure R1 reported below shows the life expectancy over time for Nepal, to show that we do have a variability in time for the index. We will clarify this in the manuscript.

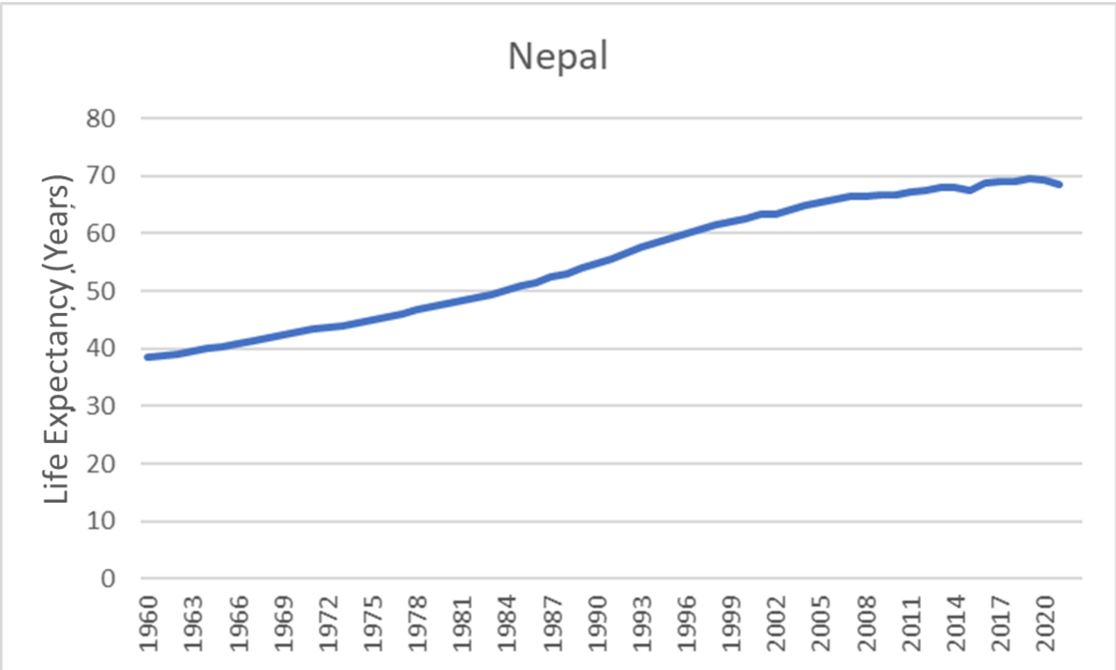

Figure R1: Life expectancy at birth for Nepal over time (Source: https://data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2021&locations=NP&start=1960&view=chart&year=2021)

For Nepal, we considered ***a weighted spatial join between the watersheds and the districts.*** *To each watershed, we attributed the statistics of the district intersecting it, weighted by the overlapping areas. In general, the districts we have for Nepal are of a smaller extent than those of the various watersheds.*

10. Figure 3: I am not sure whether these are now LYl only due to floods or all disasters. Considering that there are no jumps for earthquake events like 2015, I assume this has been calculated for floods only? Then this needs to be made very clear in the caption rather than just calling it 'disasters'.

Response 15: *This is for fluvial and pluvial floods only. We revised the manuscript to be clear and consistent.*

11. L176 + Figure 4: What is HAND in Figure 4. Is this from (Delalay et al. 2018)? The publication is not open access and only limited to Sindupalchowk, how does it go to all of Nepal? What does it actually map?

Response 16: *We thank the reviewer for this comment, we clarified this in the manuscript.*

Line 205 and subsequent: We identified flood-prone areas by grouping them into six classes by their FGP index. For each watershed, we then considered the areas covered by the classes with FGP greater than 4, which, when compared to published data, proved to correspond realistically with areas subject to floods of about 100-year depth. Figure 4b compares the Flood Geomorphic Potential (FGP) automatic classes derived for select rivers in Nepal, with baseline inundation scenarios evaluated using standard inundation depths associated with critical flood events and their return periods provided in the work of Delalay et al. (2018). This visual comparison serves to highlight the efficacy of flood inundation mapping facilitated by the FGP.
It's worth noting that the FGP methodology has been previously published and applied in various contexts (Samela et al., 2017). While testing the quality of the FGP lies beyond the scope of this work, its effectiveness for flood mapping has been well-established in previous studies (Manfreda et al., 2011, 2014; Manfreda & Samela, 2019; Samela et al., 2016, 2018), which have demonstrated the utility of the methodology, particularly in ungauged conditions, for preliminary identification of flooded areas in regions where conducting expensive and time-consuming hydrologic-hydraulic simulations may not be feasible.

12. L194: While I understand that it would be well beyond the scope of this study to evaluate the suitability of ERA5 data for flood simulations (let alone in a mountain context where precipitation products are of poor quality) but it would be crucial to address this and dispel concerns from the get go by referring to discussions of this data in mountain regions as well as for flood mapping.

Response 17: *This study is a part of the HiMAT project. There are several research groups working on different aspects of HMA and the comparison of various rainfall dataset. Speficically for ERA5, for example, other related works from the HiMAT team, such as Maggioni & Massari, 2018; Maina et al., 2023, Dollan et al. 2024 have analyzed different precipitation products. Dollan et al. 2024 stated that ERA5 although overestimates the monthly precipitation, can capture extreme events quite accurately compared to other products. We added some comments on this in the manuscript.*

13. L210: As for the other socioeconomic data above, the description of population data here remains lacking. For Nepal you only refer to the Census Bureau, which does not report distributed data or data by watershed (so how was that brought in line with inundation maps) and you also do not specify where on the general page you retrieved the data from. You then refer to the GHSL but do not provide a citation or link where this data was retrieved. Distributed data in Asia is generally of problematic and definitely not homeogenous quality, hence a discussion of how this was dealt with need a much more thorough description than the short paragraph here without any references.

Response 18: *We thank the reviewer for this comment. Indeed, obtaining population data at a detailed scale depends on local authorities, and for Nepal, we relied on the official source of the census bureau, which provides this information by the district.*

Line 262: To extend the model to the whole HMA, we computed the population for each watershed *across the region from the Gridded Population of the World (GPW), v4 | SEDAC, 2024) dataset by the Center for International Earth Science Information Network. This dataset provides spatially explicit estimates of population density for the years 2000, 2005, 2010, 2015, and 2020, based on counts consistent with national censuses and population registers, as raster data to facilitate data integration. We used a simple linear regression to retrieve data for the missing years.*

14. A detail but you also call it LYI (capital I) here while it should be Lyl (lower case L)!

Response 19: *We have decided to call the life year index (LYI) as an abbreviation for Life years lost. We checked for any inconsistency in the manuscript and corrected it.*

15. L227: You discuss your first results here on the F score and model performance discussion – this should come under Results and Discussion respectively, not Methods! Figure 6 as well as Table 1 also lacks a description of variables and results presented. Unclear how this should be interpreted.

Response 20: *Thank you for this comment. We moved this part to the results and discussion section. Also, we will add more explanations to make it clear for the reader.*

16. L248: Apart from the Brakenridge citation not having a date nor being present anywhere in the references, and agreeing that in principle such a dataset would be an interesting set for validation, the fact that the whole dataset only has 46 events from Nepal since 2021 and <10 with the 1000 deaths plus displaced criterium you introduce below makes its use questionable considering this is the area you run your model in. Wouldn't data from Nepal (like https://bipadportal.gov.np/) be much more appropriate then?

Response 21: *We thank the reviewer for this comment. We understand that some information might not have been clear from the submitted manuscript, which we improved during revision.*
Please note that we trained and validated our model considering information for NEPAL from the Nepalese government through their Nepal Disaster Risk Reduction Portal, which includes flood/heavy rain/flash flood events for all districts in Nepal from different sources. This database includes more than 46 events reported in the DFO. Please note that indeed the Bipadportal you suggested, reports information sourced from the DRR Portal, which is what we use to train and validate our model.

We provided the analysis of both Nepal and HMA from 1980-2020. To the best of our knowledge, the datasets from EMDAT and DFO have the longest and most detailed series of point datasets for different events for the period we are interested in when scaling up at the HMA level Therefore, we considered these two, with their limits, to highlight how our model could help target priority areas over HMA as a whole, and we showcased that areas highlighted by our model as potentially high-risk areas where indeed indeed affected by high-impact events, as highlighted by floods reported in these two independent datasets.

17. Also this database captures lowland floods, rather than mountain floods, making me wonder whether the aim to characterize 'High Mountain Asia' floods is really the right scope here. Also the DFO reports single coordinates, are you then simply assuming the watershed that matches the coordinate is the only one affected? Likely the reported numbers refer to much larger areas, as the size of the watershed you chose is rather small (guessing from the Figure, it's not actually described anywhere!)

Response 22: *We added the following to our paper.*
Line 284 and following: To verify our findings, we compared the predictions at the HMA level with flood events reported in the Dartmouth Flood Observatory's (DFO) Global Active Archive of Large Flood Events,

1985–Present. This comprehensive database compiles information on major floods sourced from diverse channels such as news reports, governmental records, ground observations, and remote sensing data. Notably, the DFO dataset encompasses various flood types, including lowland floods and mountainous river floods characterized as fluvial and pluvial floods.
The dataset provides point locations, representing the centroids of affected areas during floods. While acknowledging that flood centroids may oversimplify the complexities driving flood events, we utilized this dataset to showcase our model's capability to target high-risk locations historically impacted by floods within the specified timeframe. Identifying high-risk areas with recorded flood occurrences centered around these locations underscores the robustness of the model beyond the confines of its training and validation site in Nepal.

L261: You include a crucial boundary condition of your model here, i.e. '1000 deaths plus displaced'. Does this mean your model will only be useful in this domain? It would be crucial to report how many such events have actually happened in your domain then. Also how is the adding up of 'dead and displaced' justified? These are quite 'different' responses to a flood.

Response 23: *In the LYI calculation, the formula refers to the number of affected people, without differentiating deaths from displaced. Hence, we considered the number of people from the DFO as a proxy of socioeconomic impacts. If there were fine-scale datasets with this complete information, we would have validated the model outside Nepal for those datasets, but unfortunately, this information is not available at a fine scale, but only at the country scale from Noy et al., or it would be possible to recalculate the index by disaster, using EMDAT data (For example). Please note that the paper by Noy, 2016, where the life year is calculated globally country-wise, considers the EMDAT source for the analysis, so numerically it reports the life year lost based on the available numbers, with the limitation expressed in the revised paper regarding EMDAt. Regarding the ML model we propose, the model itself is bounded by the data it ingested. The training set contains a variability of LY lost, that is in the order of magnitude of 1 up > 1000 life year lost, consistent with the examples provided in Noy et al. 2016 globally.*

18. Figure 9: Panel a is elevation not rainfall as your legend suggests!

Response 24: *There was a mistake in the legend. The rainfall in the yellow contour lines overlayed on the elevation. The yellow line is not showing up in the legend. We have corrected the legend and added more information in the caption.*

19. L265: To be honest I am not entirely sure how I should interpret Figure 7 – doesn't it just confirm that people live close to wide river channels? Then there is really no link to atmospheric characteristics as you claim in L270. There is a lot of discussion already as well on convection patterns all stemming from other literature and not really relevant to what I read in the Figure.

Response 25:
Line 310-335: From this analysis, we can see how the variability of CI is complex. If expectedly, the variability of the index is related to atmospheric characteristics (Sangüesa et al., 2018), the index varies also due to geographical factors influencing climate (Tuladhar et al., 2020). In their study based on Nepal, Karki et al., 2017 highlighted the difference in the spatial pattern of high-intensity storm events from that of annual and monsoon events. The rapid rate at which physical processes (e.g., convection) take place regulates the high temporal concentration of precipitation in the regions where the sea surface and ground are highly affected by warmer temperatures (Monjo & Martin-Vide, 2016b). On the other hand, the low temporal concentration of rainfall is characterized as a normal pattern caused by cyclical weather events (Monjo & Martin-Vide, 2016). Watersheds with lesser floodplain extents (that is, less areas with high FGP) are related to higher and steeper mountains, with complex orography. Research has shown that low areas in Nepal are susceptible to receiving high-intensity storm events even though they have fewer wet days (Karki et al., 2017). The authors of the same study also observed that

the low-intensity events (annual and monsoonal precipitation) were mostly predominant over Nepal's western middle mountains and central high mountains. In another study, however, Subba et al., 2019 stated that the frequency of extreme events had decreased significantly over the past two decades in the eastern part of Nepal. For our case, areas having the larger physical potential to floods (high FGP), appear to be areas showing the largest variation in CI, with values ranging from low (0.2) as well as very high (0.75), indicating a potential compound effect of highly torrential rains (CI=0.7) in locations where much of the landscape is potentially floodable. Readers should consider that higher FGP values do not imply locations having wider channels, but rather they indicate how the landscape is potentially more flood-prone as highlighted by (Samela et al., 2017; Manfreda & Samela, 2019; Samela et al., 2016, 2018). In Figure 6, we showcase how for our landscape, areas where we have higher variability of CI (>0.60) correspond to locations with high physical flood potential, as well as larger exposure in population.

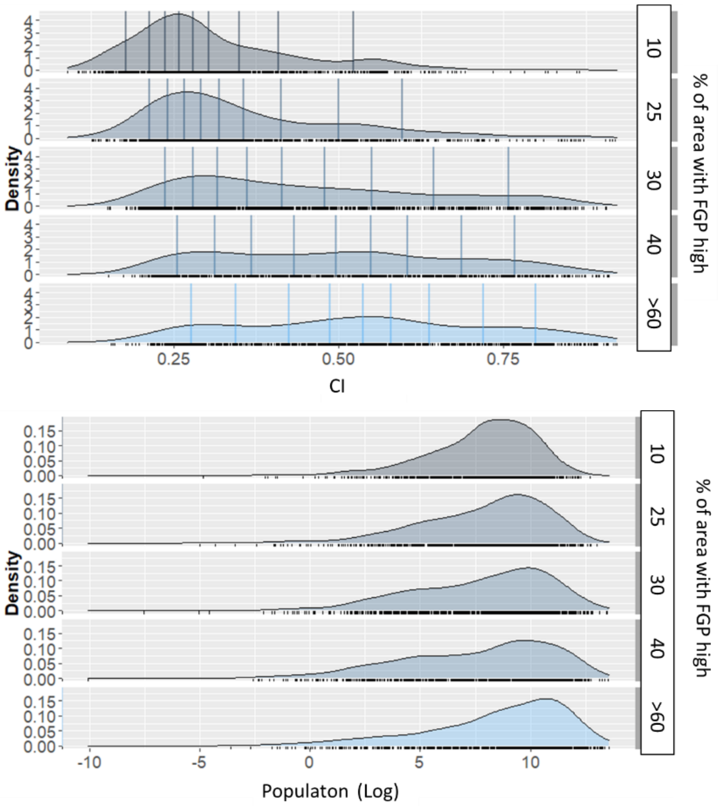

Figure 6: Average variability of the CI (top) and population (bottom) compared to FGP from 1980-2020

20. L295: A main concern I have here is that I am still not very clear on where the observed events come from you compare this to. I am also wondering if your Figure 8 simply only confirms one thing – that there are many people (an input to your model) where there are many people (a validation of your model). How does your model compare on actually coming up with an observed flood from the input 'ERA5 rain'? This concern then propagates into the result for the whole region, where you 'predict' the biggest impacts with the highest population densities. That isn't quite so surprising and it is unclear to me how I can see the power of ML in these results. To be provocative, would the results have been different if you would have just distributed rainfall across the watersheds without a model in between?

Response 26: *We thank the reviewer for his comment. We revised the manuscript thoroughly and we believe it is now clearer. Please note that we trained the model and validated it only using the data for Nepal, at the district scale and then at the watershed scale. Overall, we opted for a 90-10 approach, for which 90% of the Nepal data were used for training and 10% for validation.*

Line 360 and following: Comparing predicted Lifeyears Index (LYI) flood impacts with observed data

showed good correspondence between high-risk areas identified by the ML method and historical flood locations in Nepal. This suggests that the proposed approach effectively delineates flood risk on a national scale. Figure 8 illustrates this comparison, showcasing observed (empirically evaluated) and ML-predicted LYI values at both watershed (upper row) and district (lower row) levels.
The 'observed' LYI values were empirically calculated from observational data (Table 1) and categorized into three groups: 'low', 'medium', or 'high', with basins/districts labeled as 'high' for LYI values exceeding 1000 years, 'medium' between 100 and 1000 years, and 'low' below 10 years. The 'predicted' values represent the outputs from the machine learning model.
In Nepal, we achieved an overall training accuracy of 97% and a test accuracy of 63%. Notably, training the model at the watershed level yielded higher accuracy compared to the district level. This is attributed to watersheds being hydrologic units that integrate geomorphological and climatic properties, thus providing a more accurate representation of flood dynamics compared to administrative district boundaries.

At the watershed level, nearly all year ranges exhibited a 100% match with observed impacts. In instances where the model's accuracy fell below 100% (e.g., 1985–90 and 1990–95), the LYI values in the affected watersheds were low, indicating that the predictors considered were more indicative of major flooding events.
The superior accuracy achieved at the watershed level underscores the value of implementing the model at this scale when scaling up the system.

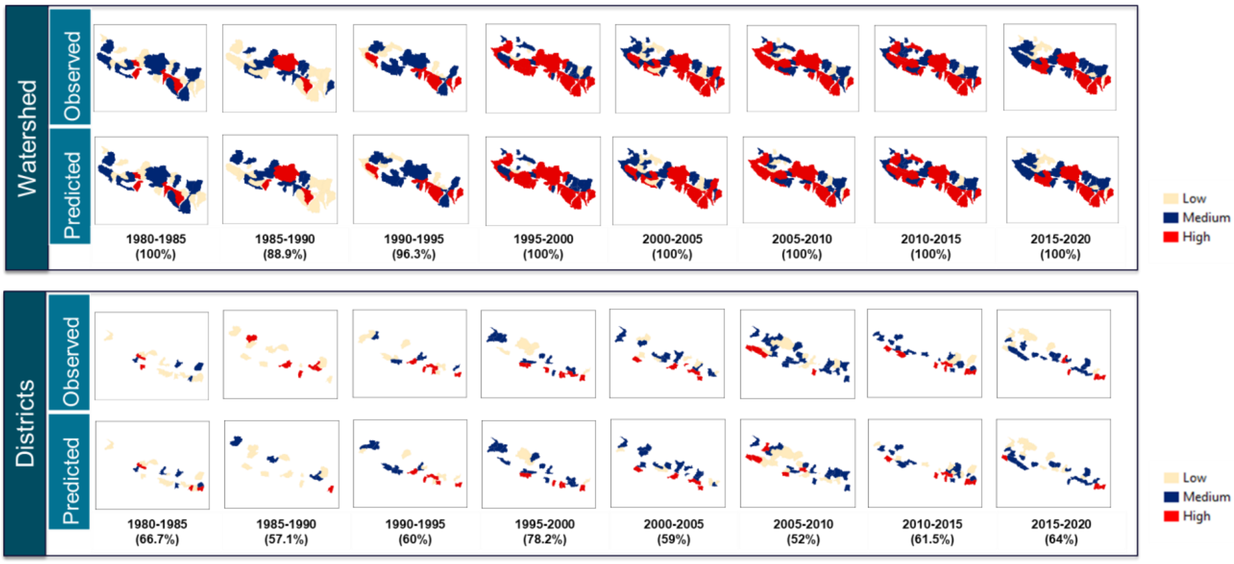

Figure 8: Comparison of prediction with actual socioeconomic impact for watersheds and districts in Nepal. Basin/districts are marked as "high" for LYI over 1000 years. Medium is between 100 and 1000, and low is less than 10. Numbers in parentheses represent accuracy.

21. L349: The figures you note here do not show what is described in the text.

Response 27: *We corrected this mistake.*

22. L356: I lack some context here - <10% of watersheds see an increase, are all other stable or see a decrease?

Response 28: *Here in section 3.6 we showcase the changes in impact [low to medium (LtoM); medium to high (MtoH); and low to high (LtoH)] increase in impact. Some watersheds have not changed, and some have decreased impact. However, we are concerned and discussed the ones that will threaten the people's future socioeconomic balance.*

23. How can you differentiate here between hazard (rain) and exposure (population) as a driver of

change?

***Response 29:** We do not attempt to differentiate between hazard and exposure. Rather we use them together to find out the impact. We simply tried to analyze the results and connect the dots by revisiting past occurrences (such as population boom, extreme events, or both).*

24. How do you explain that increase has slowed after 2010 significantly? And how is it possible that in the 1995-2010 jump the number of increasing watersheds is similar to the just 5 year jump between 1990 – 1995? Isn't that completely counterintuitive?

***Response 30:** It may be counterintuitive, however, there may have been many events that have caused more damage in that 5-year window than the longer span.*

25. L406: While in general 'an intention to make data available' shouldn't be followed, for a journal like NHESS this is definitely not acceptable. Data availability needs to be clearly described (or arhued why this is not the case).

***Response 31:** The FGP dataset produced in this study is available online. We have revised it in the document.*

26. *Technical corrections (Minor issues):*

***Response 32:** We carefully revised the entire manuscript to correct all the minor issues.*

References:

Barandun, M., Fiddes, J., Scherler, M., Mathys, T., Saks, T., Petrakov, D., and Hoelzle, M.: The state and future of the cryosphere in Central Asia, Water Secur, 11, 100072, https://doi.org/10.1016/J.WASEC.2020.100072, 2020.

Dollan, I. J., Maina, F. Z., Kumar, S. V., Nikolopoulos, E. I., and Maggioni, V.: An assessment of gridded precipitation products over High Mountain Asia, J Hydrol Reg Stud, 52, 101675, https://doi.org/10.1016/J.EJRH.2024.101675, 2024.

Du, M., Huang, S., Leng, G., Huang, Q., Guo, Y., and Jiang, J.: Multi-timescale-based precipitation concentration dynamics and their asymmetric impacts on dry and wet conditions in a changing environment, Atmos Res, 291, 106821, https://doi.org/10.1016/J.ATMOSRES.2023.106821, 2023.

Maggioni, V. and Massari, C.: On the performance of satellite precipitation products in riverine flood modeling: A review, J Hydrol (Amst), 558, 214–224, https://doi.org/10.1016/J.JHYDROL.2018.01.039, 2018.

Maina, F. Z., Kumar, S. V., Getirana, A., Forman, B., Zaitchik, B. F., Loomis, B., Maggioni, V., Xue, Y., McLarty, S., and Zhou, Y.: Development of a Multidecadal Land Reanalysis over High Mountain Asia, https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/415850, 11 January 2023.

Miles, E., McCarthy, M., Dehecq, A., Kneib, M., Fugger, S., and Pellicciotti, F.: Health and sustainability of glaciers in High Mountain Asia, Nature Communications 2021 12:1, 12, 1–10, https://doi.org/10.1038/s41467-021-23073-4, 2021.

Delalay, M., Ziegler, A. D., Shrestha, M. S., Wasson, R. J., Sudmeier-Rieux, K., McAdoo, B. G., and Kochhar, I.: Towards improved flood disaster governance in Nepal: A case study in Sindhupalchok District, International Journal of Disaster Risk Reduction, 31, 354–366, https://doi.org/10.1016/j.ijdrr.2018.05.025, 2018.

Monjo, R., & Martin-Vide, J. (2016). Daily precipitation concentration around the world according to several indices. International Journal of Climatology, 36(11), 3828–3838. https://doi.org/10.1002/JOC.4596

Noy, I.: A Global Comprehensive Measure of the Impact of Natural Hazards and Disasters, Glob Policy, 7, 56–65, https://doi.org/10.1111/1758-5899.12272, 2016.

Saki, S. A., Sofia, G., & Anagnostou, E. N. (2023). Characterizing CONUS-wide spatio-temporal changes in *daily precipitation, flow, and variability of extremes. Journal of Hydrology, 626, 130336. https://doi.org/10.1016/J.JHYDROL.2023.130336*

Sofia, G. and Nikolopoulos, E. I.: Floods and rivers: a circular causality perspective, Sci Rep, 10, https://doi.org/10.1038/s41598-020-61533-x, 2020.

Sofia, G., Ragazzi, F., Giandon, P., Dalla Fontana, G., & Tarolli, P. (2019). On the linkage between runoff generation, land drainage, soil properties, and temporal patterns of precipitation in agricultural floodplains. Advances in Water Resources, 124, 120–138. https://doi.org/10.1016/j.advwatres.2018.12.003

Sofia, G., Tarolli, P., Cazorzi, F., and Dalla Fontana, G.: Downstream hydraulic geometry relationships: Gathering reference reach-scale width values from LiDAR, Geomorphology, 250, 236–248, https://doi.org/10.1016/j.geomorph.2015.09.002, 2015.

Manfreda, S., Di Leo, M., and Sole, A.: Detection of flood-prone areas using digital elevation models, Journal of Hydrologic Engineering, 16(10), 781–790, http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000367, 2011.

Manfreda, S., Nardi, F., Samela, C., Grimaldi, S., Taramasso, A. C., Roth, G., and Sole, A.: Investigation on the use of geomorphic approaches for the delineation of flood prone areas, Journal of Hydrology, 517, 863–876, https://doi.org/10.1016/j.jhydrol.2014.06.009, 2014.

Manfreda, S. and Samela, C.: A digital elevation model-based method for a rapid estimation of flood inundation depth, Journal of Flood Risk Management, 12(Suppl. 1), e12541 1–10, https://doi.org/10.1111/jfr3.12541, 2019.

Samela, C., Albano, R., Sole, A., and Manfreda, S.: A GIS tool for cost-effective delineation of flood-prone areas. Computers, Environment and Urban Systems, 70, 43–52, A GIS tool for cost-effective delineation of flood-prone areas. Computers, Environment and Urban Systems, 2018.

Samela, C., Manfreda, S., De Paola, F., Giugni, M., Sole, A., and Fiorentino, M.: DEM-based approaches for the delineation of flood-prone areas in an ungauged basin in Africa, Journal of Hydrologic Engineering, 21(2), 06015010, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001272, 2016., Journal of Hydrologic Engineering, 21(2), 06015010, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001272, 2016.

Samela, C., Troy, T. J., & Manfreda, S.: Geomorphic classifiers for flood-prone areas delineation for data-scarce environments, Advances in Water Resources, 102, 13–28, https://doi.org/10.1016/j.advwatres.2017.01.007, 2017., Advances in Water Resources, 102, 13–28, https://doi.org/10.1016/j.advwatres.2017.01.007, 2017.

Review by RC2

1. Clarity and Structure: The abstract is well-structured, presenting the problem, the proposed solution, and a few findings. However, some sentences are complex (starting from the title), and more concise wording could enhance clarity,

Response: *We thank the reviewer for this comment. We revised the complex sentences to enhance their readability.*

2. Methodology: The use of the Lifeyears Index (LYI) as a measure for socioeconomic flood impact is well explained. It would be beneficial to provide a brief explanation of how the geomorphologically guided machine learning approach works, even if it is in a bit summary.

Response: *We revised the methodology to be clearer on the geomorphologically guided machine learning approach. A notable advantage of the proposed approach lies in its reliance on automatic techniques leveraging globally available datasets, thereby facilitating its applicability across diverse geographical regions to forecast socioeconomic flood impacts. The framework also benefits from leveraging on geomorphologically-driven information, to have an improved characterization of the* *different aspects of the underlying physical processes shaping the landscape and possibly impacting flood characteristics. By incorporating such domain knowledge into the ML model, the framework can better generalize across different regions and conditions, improving robustness and reliability for risk mapping in diverse environments and facilitating informed decision-making for flood management and mitigation strategies.*

3. Data: The abstract mentions training the model with over 6000 flood events from 1980 to 2020, but it is mentioned that the model shows variability from 1980 to 2022 as temporal Coverage. So, /what's the correct timeline?

Response: *There is a mistake and both of them should be the same. It should be 1980 to 2020. We corrected it.*

4. Conclusion: A brief conclusion summarizing the main contributions and implications of the study would be beneficial.

Response: *Thank you for the suggestions. We revised the conclusion and summarized the main contributions and implications of the study.*

Review by CC1

The manuscript "Predictive understanding of socioeconomic flood impact in data scarce regions based on channel properties and storm characteristics: Application in High Mountain Asia(HMA)" by Khanam et al. used LYI and ML methods to evaluate and predict the flood impacts and risk due to precipitation in HMA. This work is first time to evaluate socioeconomic impacts of flood hazards in data scarce region. However, it is not good writing. The structure is not reasonable. And the XGboosting tools is not clear to solve what? Thus, I would suggest it should be major revision.

General comments:

- In HMA there are also GLOF which risk the human being and infrastructure. If possible, please include evaluating the socioeconomic flood impact.

Response: *We thank the reviewer for this suggestion. GLOFs in general are triggered by glacial melt but here we focus on a climatic driver of the flooding that is related to rainfall. Addressing the damages due to GLOF is separate from the scope of this study. In general, this paper focuses on fluvial and pluvial flooding, and we made this clearer in the introduction.*

- Data-Scarce regions should be clear (which data or which type of data). In HMA, population is scarce. And Socio activity is also low.

Response: *We have revised the manuscript thoroughly to correct these unnoticed mistakes and to improve overall the quality of the writing. Please note that when we defined HIMAT as 'data scarce' we did not do so to avoid scrutiny of the data, but rather to highlight the complexity of HMA itself, challenged by several factors, including the scarcity of ground observations covering consistent timeframes homogeneously, as highlighted by other works in literature (Barandun et al., 2020; Dollan et al., 2024; Miles et al., 2021). We understand that some of the statements were phrased in an ambiguous way, which we have revised to address accordingly. In the revised work, we explained that-*

Line 53-58: Gathering data for the scale of the HMA region is a difficult task, as it requires collecting data from several countries and multiple sources, and this poses challenges due scarcity of ground observations covering consistent timeframes homogeneously, as highlighted by various works in the literature (Barandun et al., 2020; Dollan et al., 2024; Miles et al., 2021). Especially in the context of the impact of floods using socioeconomic data, the analysis involves examining the number of fatalities, *injured and people otherwise affected, as well as the financial damage that natural disasters cause, and this information is generally not always available, or it is collected at the local scale based on reported events. Major disasters are reported in global databases such as The International Disaster Database (EMDAT, www.emdat.be), or, for Nepal, the Nepal Disaster Risk Reduction Portal (http://drrportal.gov.np/) which we used in our paper, but these datasets are also not complete. For example, EMDAT only considers events with at least one of the following criteria: 1)10 fatalities; 2)100 affected people; 3) a declaration of state of emergency; 4) a call for international assistance.*

The main idea behind this paper was to provide a tool for a rapid estimate of potentially highly impacted areas, based on information accessible and updateable quickly, such as population number, rainfall intensities, and a geomorphologic index that can be derived from global DEMs (or high-resolution local ones when available).

Specific comments

1.Section 2.2 methods. This title is not reasonable. Is section 2.3(Machine learning model) methods? In addition, the dataset and methods in this section should be divided, for example, 2.2.4 exposure(population) is datasets.

Response: *Thank you for the suggestions. We reorganized the section and added a separate section for the dataset.*

We kept the 2.2 section as "Methods". This section explains the methods we used in this study and it is based on ML thus it proceeds to explain the model as well as some of the data processing.

2.Line 135 Why is it classified by LYI values(<2,2-3, and >3).

Response: *We have classified the LYI as a basis for comparison across all the watersheds and periods. The three groups correspond to <100, between 100-1000, and over 1000.*

3.Line 217 While XGBoosting is …, this sentence is incomplete.

Line 217 this section (machine learning model) is a little difficult to understand the role that is plays.

Response: *We revised it.*