

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC2
<https://doi.org/10.5194/nhess-2022-198-AC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Oya Kalaycioğlu et al.

Author comment on "Using machine learning algorithms to identify predictors of social vulnerability in the event of a hazard: Istanbul case study" by Oya Kalaycioğlu et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2022-198-AC2>, 2022

Responses to Reviewer 2's comments

Authors' Response to Overall Comments of the Reviewer 2: Dear Dr. West, firstly we thank you for your valuable and constructive comments that will improve our manuscript. Also, thank you very much for appreciating our use of household level survey data, ML methodology and data visualization. We will revise the manuscript carefully in line with your comments. Please see below for a point-by-point response to your comments and concerns.

Specific Suggestions of the Reviewer 2 :

1. *An overarching question I would like the authors to address is whether, and how, the ML analyses might be used to improve upon the original SVI measure for Istanbul. What do the ML models add to the understanding of social vulnerability in Istanbul that was not clear previously? I think the manuscript will be strengthened if the authors can better connect these dots for readers.*

Authors' response: We thank the reviewer for this comment, giving us the chance of elaborating on this topic. In this study, we used the pre-obtained SVI score (based on the survey on 41,093 households) as the output, which the reviewer refers to as the original SVI measure. In that study (we refer to as Phase 1), the SVI score and SV category of the households were calculated based on pre-determined variables identified through a literature review. These variables are related to social, economic and demographic properties of the households as well as cultural characteristics of the community, such as the perception of and preparedness against earthquake risk. The required information to represent these predictors were obtained via survey research since such representative data of the population in Istanbul was not directly available from any institutional source before. Therefore, in this paper, we aimed to build a ML based model, that can predict the SV status of these households using publicly available databases without requiring to implementation of a household-based survey. This is valuable to reduce the time and economic burden that one may spend conducting surveys to calculate SVI and assess households with high SV. Our modeling approach could also be used by decision makers to identify and prioritize action towards target groups in the population in the interests of risk mitigation, by classifying the household with respect to their SV level. In this regard, we believe such an ML approach to identify SV is beneficial to effectively and practically interpret the social context better in the face of disasters.

We also believe that such practicality will make it possible to understand SV better as it will enable researchers to adopt this study on different spatial contexts with different variables. Eventually, this will lead to a more comprehensive understanding of the phenomenon. In the revised version of this paper, we are going to highlight this message more explicitly with proper justification.

2. I recommend building upon the discussion of the scale at which data were analyzed as a strength of this research. Social vulnerability is not often able to be analyzed at the household level, so that is a significant potential contribution of this research worth discussing.

Authors' response: We appreciate the reviewer for drawing attention to the importance of measuring social vulnerability at the household level, which is lacking in the literature due to limited data availability. In that sense, our study is one of the few studies which is based on large-scale household survey data to assess social vulnerability predictors. In the revised manuscript, we will emphasize this major contribution of our study by providing details on the methodology of the survey and its outputs.

3. The current description of the data and methods used to construct the Istanbul SVI are not yet sufficiently complete and accurate to allow their reproduction by fellow scientists. As I understand, there is a vulnerability index for Istanbul that was created as a Phase 1 of this study. However, its description is not available in English (the language of this journal), and there seems to be no peer reviewed publication associated with the SVI. In light of this, the authors need to describe the construction of the index in detail before using it as the outcome variable in the ML models. Thus, I recommend adding a section to the manuscript describing the construction of the social vulnerability index. I encourage the authors to acknowledge the limitations of various approaches to index construction, including those raised by Spielman et al. (2020) and others. I include some recommended publications on this topic at the end.

Authors' response: Unfortunately, the previous study explaining the calculation of SVI is available only in Turkish as an institutional report by Istanbul Metropolitan Municipality (IMM, 2018) and it was presented at a conference (Mentese et al., 2019). As the reviewer suggests, in the literature, there are various approaches and indicators to construct a social vulnerability index. In Phase 1 of this study, SVI was calculated following the factor analysis framework and 53 indicators, which were reduced to 7 factors. The strategy which was proposed by Cutter et al. (2003, <https://doi.org/10.1111/1540-6237.8402002>) was used. The indicators chosen for the calculation of SVI have been selected following extensive literature reviews and discussions with experts. In the revised manuscript, we will add a section which is devoted to describing the social vulnerability index, including a discussion of various approaches in the literature. We are also grateful for the publication suggestions that we believe will improve our article.

4. Building upon the explanation of the SVI construction, please also explain why the decision was made to evaluate the vulnerability index as a binary variable that refers to the top 20% of the vulnerability index. For instance, why not use the continuous index as the outcome variable? Why not use the top 25%? Finally, discuss any potential strengths, weaknesses, or consequences of defining vulnerability in this way.

Authors' response: In the previous work (phase 1), SVI's of the households were classified into four categories: Very low SVI, Low SVI, High SVI and Very High SVI. In the dataset that was used in here, instead of using four SVI categories as an outcome, we used binary outcome. Within the scope of our study, we aimed to identify and predict the households with the highest SVI which require the urgent action. Thus, we compared very high SVI group to all others, and this group corresponds to 20% of the households. Also statistically speaking, the available performance metrics for a multi-class confusion matrix

are limited compared to a binary classification problem (Markoulidakis et al., 2021, <https://doi.org/10.3390/technologies9040081>). Therefore, in accordance with our motivation and for the sake of interpretability and ease of application we used binary outcome. In the revised manuscript, in the new section that we will add regarding the construction of SVI, we will clarify this.

5. *It was also my impression that the variables in the Istanbul vulnerability index are not specific to earthquake vulnerability in particular. Instead, they seem to refer to social vulnerability more generally and not in relation to any single hazard. This is not necessarily a problem with the data. However, the claim to understand earthquake vulnerability in particular needs to be more fully substantiated, or removed, because the SVI data used in this study do not appear to refer to earthquake-related vulnerability, even if the original household survey did focus on earthquakes.*

Authors' response: The outcome of the study - which is the dichotomized SVI - was derived based on the Cutter's scheme social vulnerability in relation to earthquake hazards. The interview questions used to derive SVI includes questions related to respondents' preparedness for earthquake, their risk of perception of earthquakes and their experience with earthquake incidence. We agree that the predictors assessed in our current study using ML, can also be used for social vulnerability assessment for other disaster types. However, we do not intend to claim that selected variables are universal for all types of disasters as our outcome is based on earthquake-related vulnerability. We believe that, in the new section explaining the derivation of SVI, how the outcome SVI relates to earthquake will be clarified.

6. *I am curious to know how the "risk of job loss" was assessed, and I would be interested to see more explanation of why job loss would be specific to a post-earthquake context, or whether this is a general measure of job insecurity. If possible, please describe briefly how that question was asked on the original survey and whether it was a self-assessment of potential job loss. This will hopefully help readers better understand that variable.*

Authors' response: By risk of job loss, we meant to refer the risk of employment loss following an earthquake. In the original survey, the participants were asked to assess their and other household members' risk of losing job following an earthquake. As a SV predictor, we considered risk of job loss mainly related to informal employment which may be either in the form of casual, seasonal employment or self-employment, where social security and social insurance registrations are not provided by the employers. Various authors discussed that economic losses and an increase in the number of unemployed in a community will lower the coping capacities and contribute to a slower recovery from the disaster (Cutter et al., 2003, <https://doi.org/10.1111/1540-6237.8402002>; Chen et al., 2013, <https://doi.org/10.1007/s13753-013-0018-6>; Llorente-Marrón et al., 2020, <https://doi.org/10.3390/su12093574>). In accordance with the literature, we will change the terminology to "Employment Loss" and describe it in relation to social vulnerability context.

7. *The last sentence of the abstract suggests that "The machine learning methodology and the findings that we present in this paper can serve as a guidance for decision makers..." I would like to know more about how specifically the machine learning methodology could be used by decision makers. I do understand how the SVI data can be a tool for decision makers, but it is not yet clear to me how the ML component could practically be used by decision makers to reduce vulnerability. Would you argue that this ML analysis can be used to improve the SVI as a decision making tool? If so, explaining how and providing an example of a use case might be helpful.*

Authors' response: We thank the reviewer for giving us a chance to clarify this important aspect of our study. Our motivation was to assess the variables which

contribute most to the social vulnerability of the households, given a pre-developed SVI as an outcome. By using our proposed ML model, one can input these variables – which are available in data bases of metropolitan / district municipalities, city governorship and public authorities, etc. – to our proposed final ANN model and obtain the status of social vulnerability of the household. Hence our model proposal has potential to reduce the time and economic burden that one may spend to conduct surveys to calculate household level SVI and assess households with high SV. We agree in the sense that our results are not directed towards defining disaster risk reduction policies but based on the social vulnerability characterization, these households can be prioritized by decision makers to develop new social policies targeted to disaster vulnerability. Such targeted policies are particularly missing in the Turkish context as well as in many of the low-mid income countries. We will clarify better the contribution of using ML methodology in relation to predict SV in the revised manuscript.

Grammatical/proofreading suggestions of the Reviewer 2:

1. *In the abstract, please define what the outcome variable is when mentioning it.*

Authors' response: We will define the outcome in the abstract in the revised manuscript.

2. *Change "dept" to "debt" in the last line of Table 1*

Authors' response: It is a typo. We will correct this typo as "Debt".

3. *I recommend avoiding the term "natural disasters" and instead simply using "disasters" or being more specific. (line 63, 103, 453, 529)*

Authors' response: We will use "disasters" in the revised manuscript instead of "natural disasters".

4. *Page 4, Line 121: Avoid using the word "intrinsic" to describe social vulnerability because social science research specifies that vulnerability is not intrinsic to people; it is instead a condition borne by some people under certain conditions. People are also not vulnerable at all times or in all contexts. In other words, social vulnerability does not emerge from the characteristics in the SVI. Rather these variables can sometimes indicate or signal who is more likely to bear vulnerability created by structural forces and power imbalances. It is important that the language used is clear about this.*

Authors' response: We totally agree and we will substitute "intrinsic" with "specific dynamics leading to social vulnerability/ies".

5. *Page 25, Line 497: I recommend rephrasing the sentence, "We have found that socially, economically, and environmentally vulnerable communities are more likely to suffer disproportionately from disasters," because this is not a finding of the current study as it does not evaluate impacts of disasters. Instead, this is a finding of many previous studies that you could perhaps cite here. Simply removing the words "we have found that..." may be sufficient.*

Authors' response: We agree that we made a strong statement here. We will revise this sentence accordingly.

6. *I find the phrase "social vulnerability risk" to be a bit confusing and inconsistent with most literature on this topic. Social vulnerability is typically considered a sub-component of risk, so these phrases should not be combined. Instead, use the phrase "social vulnerability" without the word "risk." For similar reasons, the phrase "social risk" should be replaced with either "social vulnerability" or "disaster risk," as appropriate.*

Authors' response: We thank the reviewer for this suggestion. By "social vulnerability risk" we intended to refer was "status of social vulnerability". We will revise "social vulnerability risk" phrase accordingly. Also, there is typo in the sentence related to "social risk" referenced to Cutter (1996). It has to be "social vulnerability" as the reviewer has suggested, not "social risk". The social risk terminology issue and the sentence will be fixed in the revised manuscript.