ARTICLE

Check for updates

A data-driven approach to rapidly estimate recovery potential to go beyond building damage after disasters

Sabine Loos [1,2,8✉], David Lallemant [2,3], Feroz Khan[2,4], Jamie W. McCaughey [5], Robert Banick [6], Nama Budhathoki[7] & Jack W. Baker[1]

Following a disaster, crucial decisions about recovery resources often prioritize immediate damage, partly due to a lack of detailed information on who will struggle to recover in the long term. Here, we develop a data-driven approach to provide rapid estimates of non-recovery, or areas with the potential to fall behind during recovery, by relating surveyed data on recovery progress with data that would be readily available in most countries. We demonstrate this approach for one dimension of recovery—housing reconstruction—analyzing data collected five years after the 2015 Nepal earthquake to identify a range of ongoing social and environmental vulnerabilities related to non-recovery in Nepal. If such information were available in 2015, it would have exposed regional differences in recovery potential due to these vulnerabilities. More generally, moving beyond damage data by estimating non-recovery focuses attention on those most vulnerable sooner after a disaster to better support holistic and nuanced decisions.

[1] Civil & Environmental Engineering, Stanford University, Stanford, CA 94305, USA. [2] Earth Observatory of Singapore, Nanyang Technological University, Singapore, Singapore. [3] Asian School of the Environment, Nanyang Technological University, Singapore, Singapore. [4] Minority Health and Health Disparities Research Center, Dillard University, New Orleans, LA, USA. [5] Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland. [6] Poverty and Equity Global Practice, World Bank, Kathmandu, Nepal. [7] Kathmandu Living Labs, Kathmandu, Nepal. [8] Present address: U.S. Geological Survey, Golden, CO, USA. ✉email: sloos@alumni.stanford.edu

Natural hazards often cause disproportionate impacts on vulnerable populations and amplify inequality for years after an event. Among many examples, four times more women than men died in the 2004 Indian Ocean Tsunami[1], and multi-family, Hispanic, and linguistically-isolated households had inadequate access to loss-based assistance programs following the 1994 Northridge Earthquake[2]. When repeated examples of disaster-exacerbated inequality are evident over time, we must recognize that recovery policies, and the underlying information that supports them, fail to address and prevent deepening inequality.

Housing recovery policies are a powerful tool to prioritize vulnerable populations after an event[3,4]. Early decisions can shape the long-term recovery trajectory of an entire region, both positively and negatively. Currently, recovery assistance is not designed as a means of redistribution[2]. For example, aid is based on losses (or damage) incurred to pre-disaster homes, therefore prioritizing those who had assets before the disaster[2,5]. Alternative needs-based, area-targeting, or subsidiary approaches exist, where policies prioritize groups who may lack necessary resources to support their own recovery[2,4,6]. However, the implementation of sustainable recovery approaches is challenging[7], partly due to a lack of timely information that prioritizes these communities and identifies factors that will impede their recovery in the weeks following a disaster when these plans are made[8,9]. While it is clear that marginalized populations fall behind in recovery, few quantitative data exists in the hours to days after a disaster that explicitly acknowledges these inequities to better support more equitable and sustainable recovery approaches. For example, even though the the Post-Disaster Needs Assessment (PDNA) was developed as a process to collaboratively rapidly evaluate human recovery needs to support government recovery efforts[10,11], quantification of disaster impacts still largely focuses on economically quantifiable damages rather than social needs[6,11].

Advances in technology and increased data availability provide an opportunity to develop information on populations whose recovery may be impeded by factors other than damage to their home. Non-traditional post-disaster data, from remote-sensing or digital crowdsourcing, overwhelmingly focus on quantifying building damage[12,13], because it is relatively easy to quantify[14] and supports top-down recovery agendas[6,15]. While damage does represent a reduction in housing quality, the sole focus on immediate impact (Fig. 1a) is a myopic measure of long-term recovery needs (Fig. 1b). To identify communities with disproportionate needs long after a disaster, we propose focusing on those who fall behind in recovery over time, or non-recovery. We focus on non-recovery since it places attention on those who do not recover rather than delineating the characteristics of successful recovery. Importantly, to be able to regionally estimate non-recovery after future disasters, we develop a data-driven approach that uses census, remotely-sensed, and modeled data that would be readily available in the hours to days afterwards and represent a range of sociodemographic, economic, environmental, and geographic factors that are likely to affect recovery. Recognizing that recovery occurs along multiple dimensions[16–21], here we demonstrate the concept of non-recovery for the housing sector, an important dimension for recovery[17,22,23]—and overall health[24]—in which inequalities are amplified[25]. Here, we develop one specific metric of non-recovery using empirical reconstruction data from the 2015 Nepal earthquake, which is the probability that a household living in a severely damaged house will only partly rebuild or not rebuild at all within five years (Fig. 1c). We use a large-scale survey conducted in 2019 in the earthquake-affected districts by The Asia Foundation and local partner Inter-Disciplinary Analysts to assess long-term impacts and recovery patterns[26]. The developed approach combines surveyed reconstruction outcomes with the census, remotely-sensed, and modeled factors to rapidly estimate the regions that are least likely to recover by evaluating which factors are most related to surveyed recovery outcomes, which we validate with a held-out dataset and verify with Nepal-specific and broader literature on vulnerability and recovery.

Our study shifts attention beyond estimating building damage as the main source of information after an earthquake by introducing an approach to rapidly identify the obstacles that lead to the lack of household recovery progress. In Nepal, we found that the most important predictors of incomplete reconstruction fall into three categories: hazard exposure, rural accessibility and poverty, and reconstruction complexity. The relationships between these predictors and non-recovery are complex and disparate between households, though general patterns emerge amongst groups of affected households. Notably, the spatial pattern of non-recovery brings to focus regions that were not highlighted by damage alone. The approach for estimating non-recovery provides a needed alternative over prevailing methods to capture social vulnerability or resilience that have been historically developed for pre-disaster mitigation planning in developed countries. The specific model developed for Nepal can directly guide future risk reduction planning in this region. For post-disaster planning in general, evaluating and quantifying
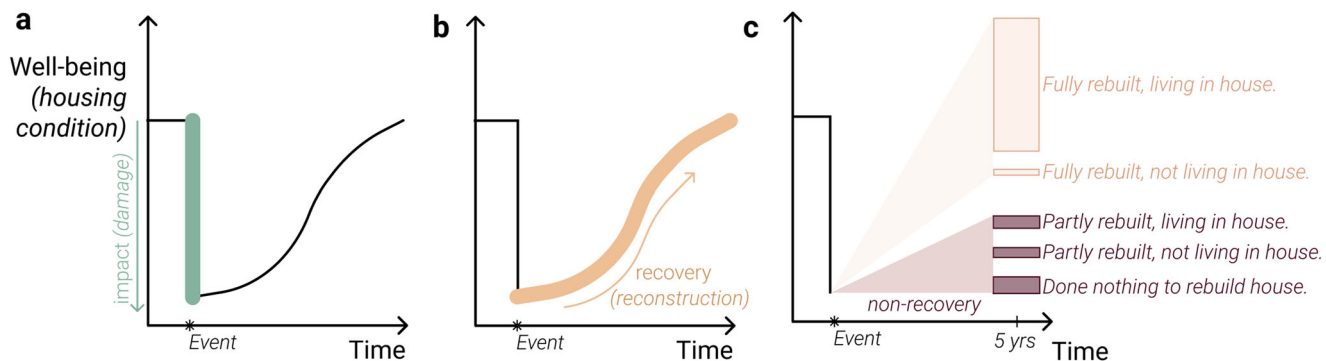

Fig. 1 Non-recovery focuses on those who are unlikely to recover in the long-term. Rapidly available post-disaster data often focuses on quantifying building damage, which captures immediate impact (**a**), as opposed to long-term recovery needs (**b**). Non-recovery identifies the impacted households who are not able to fully recover over time, as shown by the dark red recovery outcomes in (**c**). In the Nepal case study, we demonstrate non-recovery for the housing sector based on the five responses on reconstruction progress shown in (**c**), where the sizes of the rectangles represent relative proportions of each response among the survey sample ($n = 3376$). General concepts are shown, with specific metrics of impact and recovery used in this study included in parentheses.
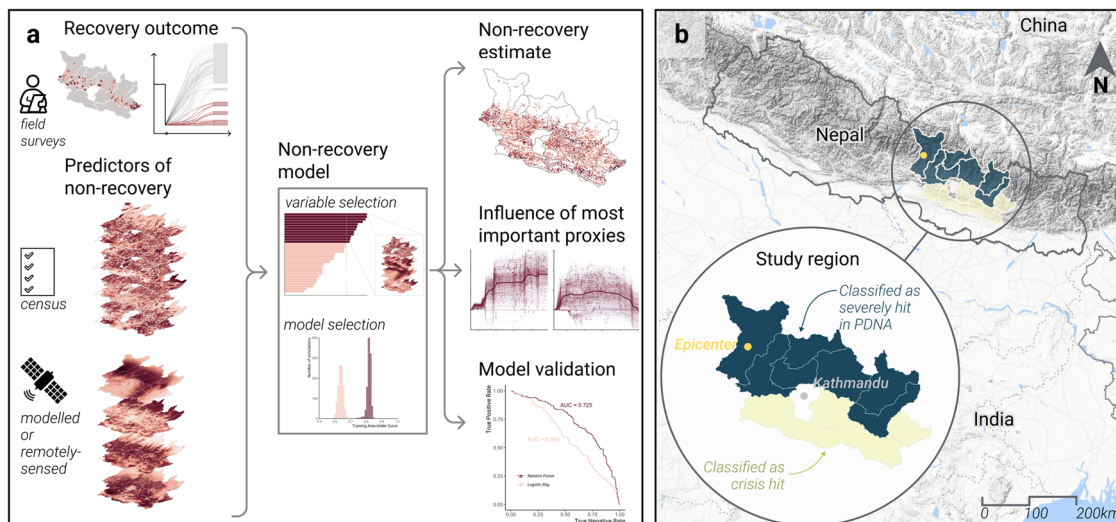
Fig. 2 Non-recovery estimation approach and study area in Nepal. a The model for non-recovery is calibrated on surveyed recovery outcomes, and uses readily available predictor variables representing sociodemographic, environmental, and geographic factors likely to influence recovery. Outputs include a spatial estimate of non-recovery, the relative influence of each variable on non-recovery, and a metric of performance by validating the model on a test set (See Methods for more information). **b** The study area considered here are the 11 most affected districts outside of Kathmandu Valley affected by the 2015 Nepal earthquake. The areas in blue were originally classified as severely hit (higher impact) and green as crisis hit (lower impact) by the Government of Nepal[28]. Map data ©2022 Google.

non-recovery can expand our data that represents disasters beyond damaged buildings and support decisions that target areas that will struggle during recovery due to, largely, pre-existing vulnerabilities.

Approach for estimating non-recovery

The approach to rapidly estimate non-recovery after a disaster captures likely inequalities in long-term recovery progress due to a range of sociodemographic, economic, environmental, and geographic factors. This approach consists of relating surveyed non-recovery outcomes to predictor variables reflecting these factors from remotely-sensed, modeled, and census-based data that would be readily available when a future earthquake occurs (Fig. 2a). We demonstrate this approach considering the case of the 2015 Nepal earthquake, which is emblematic of a major modern disaster with substantial data produced from sensors, field surveys, and digital crowdsourcing[27].

Among multiple non-recovery outcomes for the Nepal earthquake, we focus on one metric–reconstruction progress four and a half years after the event–due to the emphasis on owner-driven reconstruction in this timeframe by the Government of Nepal (GoN), international agencies, and NGO's. After Nepal's National Planning Commission led their Post-Disaster Needs Assessment (PDNA) in the first three months after the earthquake, they estimated a total of 350,540 million NPR (~$3.3 billion USD in 2015) in damages and losses to the housing sector[28]. The PDNA categorized districts by the severity of their damage as shown in Fig. 2b, largely driven by the housing sector, which comprised half of overall losses. Afterward, the Government of Nepal implemented the Earthquake Housing Reconstruction Program in affected districts, which delivered reconstruction grants to repair or rebuild severely damaged or collapsed homes[29]. This program was designed to be carried out over a five-year period with a specific focus on rural households outside of Kathmandu Valley[29,30]. Due to data availability and the focus on rural households in Nepal's Reconstruction Program, we center our study on those affected districts outside of Kathmandu Valley (Fig. 2b). While these districts are mostly rural, each district consists of several municipalities are relatively less dense than

urban Kathmandu but more dense than many other areas within the district (see Methods).

Here, reconstruction progress is measured by whether a household fully completed reconstruction four and a half years after the April 2015 earthquake (Fig. 1c). The majority of rural households in the study region owned their homes, so reconstruction progress, in this case, is informative of household recovery. To ensure that the developed model is measuring reconstruction ability rather than differences in initial damage, we consider only households with damaged or collapsed homes ($n = 3376$). Controlling for damage, we focus on the impacted households that have still not completed reconstruction in the recovery policy's timeframe despite being targeted for support. We refer to the lack of reconstruction progress as non-reconstruction when discussing Nepal's model and recovery outcome, recognizing that the overarching approach for estimating non-recovery should be applied to alternative measures of recovery outcomes in the future (see Discussion).

Initial predictor variables of non-reconstruction were selected through a combination of exploratory interviews with Nepali community leaders and recovery stakeholders; studies on impacts and recovery from the Nepal earthquake; and broader theories on sustainable development, vulnerability, and resilience (see Methods). To support generalizability of this approach beyond Nepal, we only considered variables that could be represented with commonly available data or easily developed in other countries. Due to this constraint, some variables that have been found to be important for recovery in Nepal through purely survey-based or ethnographic studies could not be fully represented. However, the resulting initial suite of 32 predictor variables capture a range of human, social, natural, economic, and physical dimensions that are expected to influence non-reconstruction and are included in Table 1. We then reduced this initial set of variables to improve model parsimony using an automatic selection technique to remove variables less predictive than random noise (see Methods). A few variables were further removed to increase model interpretability.

By relating recovery outcomes to predictor variables, we take an empirical and data-driven approach to identify the factors that

Table 1 Predictors of non-reconstruction in Nepal.

Variable	Unit	Category
Earthquake Shaking Intensity	Modified Mercalli Intensity	Hazard exposure
Tree Cover	%	Rural accessibility and poverty
Population Density	people km^{-2}	Reconstruction complexity
Remoteness	hours	Rural accessibility and poverty
Rainfall-Triggered Landslide Hazard	index (unitless)	Hazard exposure
Tap Water	%	Reconstruction complexity
Topographic Slope	°	Reconstruction complexity
Food Poverty Prevalence	%	Rural accessibility and poverty

The final eight predictors fall into three main categories, as indicated in the third column. Variables are ordered from most to least important as identified through the variable selection process.

best predict a household's likelihood (or lack thereof) to reconstruct. This empirical approach provides a needed alternative to prevailing index-based methods of estimating vulnerability or resilience[31,32], which were originally designed for pre-disaster mitigation planning in developed countries but can end up being misused or poorly adapted to post-disaster contexts in developing countries. Instead of beginning exclusively with census data that may not be readily available or vulnerability theories that may not be applicable in the affected location, we use a combination of openly-available geospatial data with local recovery surveys. We develop a data-driven model that uses the surveyed data to weight open geospatial data that represent a range of sociodemographic, economic, environmental, and geographic factors, similar to models to estimate various development indicators like mortality or literacy in the Global South[33,34]. Here, we apply a random forest model to predict the probability of non-reconstruction, which is an interpretable method that is able to capture nonlinear influences and interactions between variables, does not require distributional assumptions of predictor variables, and performed better than traditional modeling methods we also tested (see Methods). The result is a set of concrete factors that affect a household's lack of recovery and a predictive map of a tangible outcome, non-reconstruction, as opposed to a metric that provides relative values of vulnerability or resilience.

Results

Our analysis reveals that eight predictors explain the probability of a household with a destroyed home to be unable to complete reconstruction 4 and a half years after the 2015 Nepal earthquake (Table 1). We categorise these predictors into three main categories: (1) hazard exposure, (2) rural accessibility and poverty, and (3) reconstruction complexity. These categories linked with impeded reconstruction are consistent with some of those defined in other short- and long-term recovery studies in Nepal[8,9,35–38] and broader frameworks of vulnerability[39,40]. The range of predictors indicates that impeded reconstruction relates to a collection of socioeconomic, environmental, and geographic factors. This finding supports the perspective that housing recovery is a nonlinear and multifaceted process that depends both on pre-existing social vulnerabilities[25,38,41] and practical constraints[3]. While this is well-documented in natural hazards research, it is not clearly acknowledged or accounted for in rapidly available post-disaster information systems.

Influence of predictors on non-reconstruction. Figure 3 shows the marginal effect of each variable, indicating its relative influence on predicted non-reconstruction. These Individual Conditional Expectation plots allow us to see the average relationship between each variable and impeded reconstruction found in the trained model. Each variable generally has a trend where greater values

lead to higher probabilities of non-reconstruction. However, these relationships are not purely monotonic and vary from household to household. This variation demonstrates the diverse and complex reality of recovery experienced by affected households. Because random forest models capture interactions between variables, these relationships represent the influence of one variable given the inclusion of all the other variables in the model.

Hazard exposure. Hazard exposure includes variables relating to the intensity of the main earthquake or other ongoing or historical hazards that may compound the effects of the earthquake. Since Earthquake Shaking Intensity and Landslide Hazard emerged as important predictors of non-reconstruction, our model confirms hazard exposure influences reconstruction. Here, our analysis shows that areas that experienced the most intense shaking from the mainshock (at a Modified Mercalli Intensity of 8.5) are predicted to have an average of nearly 40% greater probability of impeded reconstruction independent of the level of damage to the home. Note that we have already controlled for initial damage by considering only damaged houses, so this metric likely quantifies damage to the surrounding community infrastructure, land with associated reduced access to roads, and disruption to livelihoods. Similarly, Landslide Hazard is associated with up to 20% greater predicted probability of non-reconstruction.

While we expected areas with higher mainshock Earthquake Shaking Intensity to be less likely to reconstruct due to immediate impacts to surrounding infrastructure, the inclusion of Landslide Hazard reflects the importance of compounding or more frequently occurring hazards on recovery capacity. The mainshock triggered nearly 20,000 identified landslides in Nepal[42], which already faced ongoing landslide risk due to monsoons and urban development[43,44]. Since the earthquake, many rural and remote households faced additional landslides during monsoon season[45]. The relationship we found demonstrates confirms perceived landslide risk by households[38] and how secondary risks like landslides compound pre-existing vulnerabilities of exposed Nepali communities, putting them at greater risk to immediate damage, leading to long-term displacement, and hindering regaining of livelihood[36,46].

Rural accessibility and poverty. Several affected communities in the study area were in rural, geographically isolated, or mountainous regions[26]. The inclusion of Remoteness, Tree Cover, and Food Poverty Prevalence reflect the particular challenges that impede reconstruction for more isolated communities. Remoteness captures the travel time to, and therefore ability to access, municipal headquarters, which host local markets, services, and government offices[47,48]. The analysis predicts that the most remote households were nearly 20% less likely to reconstruct.
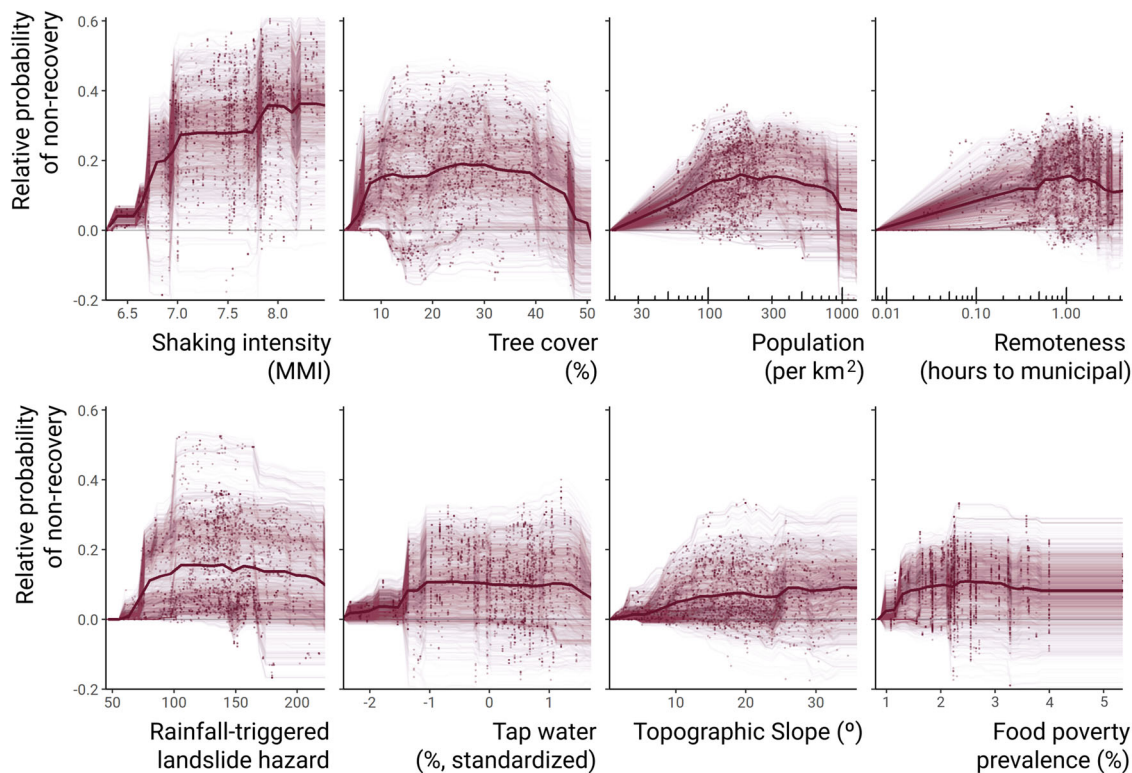
Fig. 3 The diverse, relative influence of each predictor on the probability of non-reconstruction. Each plot shows the values of each of the eight final predictors (x-axis) and the resulting probability of non-reconstruction from the analysis in Nepal (y-axis). Each point is a household used to develop the model of non-reconstruction. The thin lines running through these points show how the predicted probability of non-reconstruction would change for that household when varying the value of the predictor on the x-axis from least to greatest, while keeping the other characteristics of the household fixed. The dark line shows the average relationship among all households. All results are scaled to the predicted probability of the minimum value of each predictor, therefore relative probabilities are shown. The right-most 1% of data was truncated for tree cover and tap water for sample representation.

Other studies have found that remoteness complicated the economics of household reconstruction: remote households struggled to attract or afford wage labor in highly competitive post-disaster labor markets; construction materials were much costlier to transport where vehicles could not reach[37,38,49,50]; and the lenders, non-profits, and governmental actors supporting reconstruction tended to neglect difficult-to-reach populations[51]. The challenge in reconstruction for remote communities further compounds previous inequities in immediate damage from the earthquake[8,9]. The relationship found from Tree Cover and non-reconstruction adds complexity to the positive relationship between remoteness and non-reconstruction. Here, we find once an area reaches a certain level of Tree cover (> 40%), households were more likely to reconstruct. This result mirrors findings on short-term recovery along multiple dimensions in Nepal, including physical reconstruction, that natural capital serves as a resource in areas where there is low accessibility, since communities can rely on local materials to rebuild[35]. By providing evidence of a relationship between reconstruction ability and Tree Cover, a proxy for natural capital, this result supplements the current gaps in our understanding of the benefits of natural capital for disaster risk reduction, which has largely focused on other hazards, risk, and short-term recovery[35,52,53], rather than long-term reconstruction. The underlying mechanisms for this relationship require further research, but it is likely the model underestimates one benefit of tree cover on reducing landslide risk through slope stabilization[53], as it is directly captured in the Rainfall-triggered landslide hazard variable.

Additionally, areas with greater prevalence of pre-existing food poverty were less likely to reconstruct. The importance of food security–which is a component of human capital or human skills, knowledge, and health[54]–has been evaluated for shorter term income recovery in rural Nepal[55]. This relationship provides further evidence that human capital is important for other dimensions of recovery, here, longer term household reconstruction. Rural, mountain communities, who are often already experiencing economic and food insecurity[9], will face additional challenges during reconstruction. Overall, the negative impact of food poverty on reconstruction found here reflects the intertwined relationship between food security, building damage, and reconstruction[45,56], and highlights the complex risk arbitration these households will need to make in balancing food security and safe shelter after the earthquake.

Reconstruction complexity. The significance of population density, percentage of houses with tap water, and topographic slope likely reflect the logistical complexity of reconstructing. In the case of population density, our model predicts that households in denser areas are less likely to reconstruct. Urban areas in Nepal had unique challenges with reconstruction, such as shared landownership[51] and strict rebuilding requirements for settlements in heritage sites[57], resulting in slower reconstruction progress[26]. Our model captures the differential ability to reconstruct between urban and rural areas in terms of a continuous gradient of population density, and further reflects this urban/rural difference in reconstruction, both in Nepal and globally[3]. While some factors affecting reconstruction, like land ownership, have been found to be more pronounced in more urban areas[51], the results of the non-reconstruction model in Nepal point to the
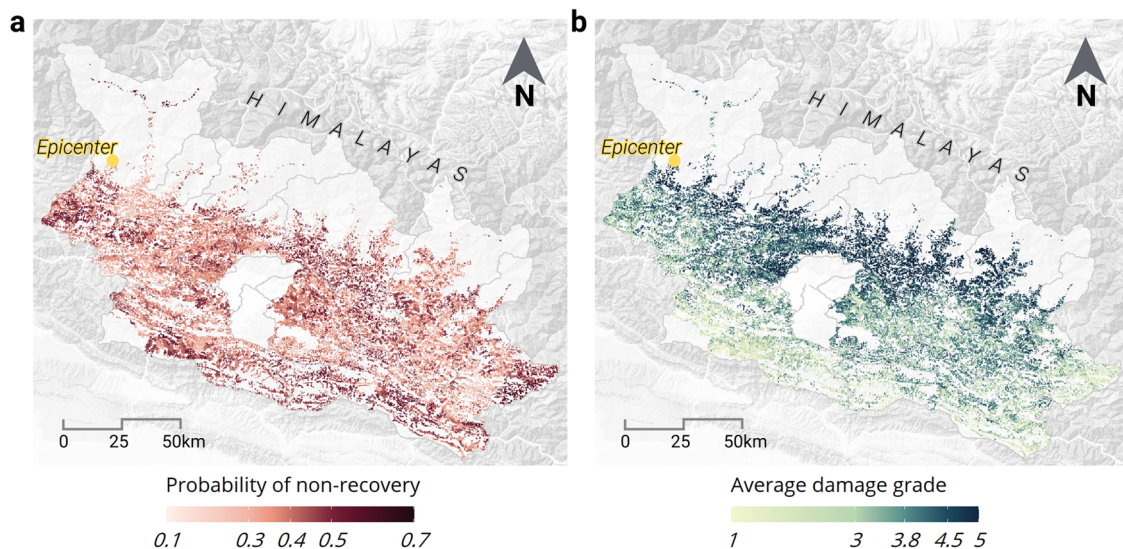
Fig. 4 The regions predicted least likely to recover were not necessarily those most damaged from the 2015 Nepal earthquake. a The spatial distribution of non-reconstruction using data from the 2015 earthquake shows areas likely to have impeded reconstruction scattered throughout the center, west, east, and south of the study area. Dark red areas correspond to high probabilities of non-reconstruction, or top of the y-axis in Fig. 3. **b** The pattern of damage was largely concentrated in the north, near the Himalayas. Damage data is from the Central Bureau of Statistics Nepal[117]. Dark blue areas correspond to areas of high damage. Both maps only show locations with buildings and are colored by quantiles of the distribution. Map data ©2022 Google.

need for more detailed comparisons on the reasons behind slowed reconstruction between dense urban areas, like Kathmandu Valley, and peri-urban municipalities, like those in the districts included in these studies. The inclusion of the percentage of households with tap water per ward exhibits a similar relationship —greater prevalence of tap water in a region is associated with higher probability of non-reconstruction. Here, the percentage of households with access to tap water per ward was initially chosen as a proxy for quality of life and infrastructure access[35,36], those who have tap water spend less time obtaining clean drinking water. While greater access to tap water could reflect infrastructure access, a factor that can be viewed as promoting resilience, the relationship here indicates that access to tap water is instead related to slowed reconstruction, considering all other variables. This perhaps indicates that, at the household level, reconstruction can be slowed in areas with greater access to infrastructure (often in more population dense, connected villages) due to the more complex logistics of reconstructing these houses, which often require external laborers or have permitting requirements[51,58]. This result illuminates how this model can capture unexpected relationships between variables and non-recovery, though the reasons behind these interactions require further research.

Topographic slope shows an influence on non-reconstruction beyond its link to hazard and accessibility. It can be explained by the difficulty of finding rebuilding sites and reconstructing on steep slopes or increased costs associated with retaining walls necessary in hillside communities[38,59], though studies have also found that marginalized populations in Nepal have settled on steep slopes as well[8,38].

Spatial distribution of non-reconstruction given damage. The model can be used to map the projections of non-reconstruction. Figure 4a shows the probability of a household with a damaged house having not reconstructed within four and a half years. It can be used in addition to the map of building damage in Fig. 4b (from an auxiliary eligibility survey by the Government of Nepal). Comparing these maps shows that areas that would have been

predicted to face the greatest and most persistent recovery needs are not necessarily those that were most damaged from the 2015 Nepal earthquake. The building damage caused by the earthquake was lowest in the southwest Hill districts of our study area and increased moving north towards the Mountain Districts near the Himalayas, closer to the epicenter and adjacent districts (Fig. 4b). This pattern of damage is largely dictated by the high shaking intensity and prevalence of vulnerable construction types in the mountains. In contrast, Fig. 4a shows that non-reconstruction is predicted to be likely scattered throughout the center, west, east, and south of the study region. This shows a pattern of non-reconstruction dictated by the spatial pattern of the social, geographic, and environmental predictors included (Supplementary Fig. 5). The map is shown in high-resolution to show the resolution used to develop the model, but should be aggregated depended on the scale required by users. Overall, the resulting map of non-reconstruction points to areas that were not originally estimated as the most impacted in influential documents like the PDNA (Fig. 2b), but that would require support during their recovery due to both physical and non-physical characteristics.

Discussion

To shift the focus from damaged buildings to vulnerable communities, we propose emphasizing and quantifying non-recovery, which focuses on those households who will not recover years after an event and due to obstacles to long-term recovery. The proposed approach employs data that is readily available after an earthquake representing a range of factors influencing recovery to identify communities likely to be unable to recover. The variables that predict non-reconstruction in Nepal fall in the categories of hazard exposure, rural accessibility and poverty, and reconstruction complexity. Some variables and their relationships with non-reconstruction could generalize to other countries outside of Nepal, like earthquake shaking intensity. Whereas other variables, importantly, are characteristics of some of the most vulnerable communities in Nepal, like remoteness and food poverty prevalence[8,60]. The empirical approach for estimating non-reconstruction can therefore inherently capture both

broadly-applicable and context-specific, in this case Nepal-specific, factors of vulnerability in technical information that can be used as a basis to guide recovery planning and policies soon after a disaster occurs.

Having information on ongoing social and environmental vulnerability and risks after a disaster, in addition to building damage, is necessary to better inform recovery policy. Developing rapid data on non-recovery emphasizes a broad range of potential factors to consider during recovery, offering a useful supplement to the myriad of building damage data produced after a disaster[12,13,15,61] and one solution towards providing the necessary high-resolution and rich data required during response and early recovery[8,9]. For example, ongoing risks from landslides and food insecurity were identified as important concerns before the 2015 earthquake[44], but were largely not considered in Nepal's post-earthquake housing recovery program, partly because the least resilient households were physically disconnected, exposed to higher hazards, and were the least visible in post-earthquake datasets[8]. Because the model of non-recovery explicitly incorporates factors on ongoing risks, the resulting estimate of non-recovery emphasizes that these risks are indeed associated with recovery capacity, and should therefore be incorporated into post-earthquake planning. While this specific model of non-recovery would not have been available after the 2015 event, the model's results could have informed the eligibility survey for the household reconstruction grant, which inadequately addressed landslide risks[14,21]. Consequently, Nepal's recovery policy primarily focused on in-situ reconstruction using more earthquake-resistant housing designs irrespective of whether that household experienced chronic landslide hazards. In addition, much of the nongovernmental organization (NGO) activity supporting reconstruction inundated high damage areas, or near the epicenter in Gorkha[62], though work was also required in areas with chronic social vulnerability that received less media attention[8,63]. Rapid non-recovery information would have helped to mitigate spatial inequalities caused by the prioritization of recovery efforts[37], by highlighting the importance of considering both areas that experienced high damage and were unlikely to recover, with their respective underlying factors. While the examples listed here demonstrate how a variety of post-disaster data would be useful for earthquakes in Nepal, multifaceted post-disaster data is also needed, and have been desired[15], for other hazards in other regions.

Methodologically, we chose to model non-recovery because, after an earthquake, many organizations understand the most vulnerable to be those groups who have trouble reconstructing[49]. Compared to index-based methods for mapping vulnerability or resilience, our modeling approach provides a direct measure of a recovery outcome of interest—in this case, whether a house will finish reconstruction—rather than a unitless aggregate of factors of vulnerability that is challenging to validate[64–66]. Index-based methods also rely on factors that are context-specific, like minority population, and therefore require several place-based studies of vulnerability that may not be readily available in most countries, limiting the implementation of indices after an earthquake to mostly high-income countries where they have been heavily studied and developed. Because estimating non-recovery takes a data-driven approach, a country's resulting model can empirically identify those most important socioeconomic, environmental, and geographic factors that are important within that context and their relationship with the recovery outcome of interest. While these identified factors could be used to develop an index, the approach presented here instead uses these factors to directly estimate a more actionable metric, non-recovery. Importantly, the non-recovery model can capture the complex relationship between those factors and the recovery outcome of interest, and

therefore maintain a level of interpretability that is often impossible with index-based approaches[66]. In addition, validation is built into the model building process, providing a sense of the accuracy of the model, a feature that is often lacking in index-based methods[67].

The main components necessary to develop a non-recovery model are surveyed recovery outcomes from a previous disaster in the region to calibrate the importance of each predictor variable. In this study, we demonstrate the concept of non-recovery with one recovery outcome of interest, which practitioners in the housing sector viewed as important in Nepal. Therefore, the non-reconstruction model we develop here is relevant to reconstruction to the housing sector in Nepal, though the overall approach of relating recovery outcomes to commonly available data can and should be applied to alternative measures of recovery at varying timescales and in other contexts. For example, in Nepal, several recovery outcomes of interest have been studied at varying time scales ranging from months to years after the earthquake, including income and livelihood recovery[55,68], reconstruction quality[69], time to reach a recovery outcome[35], or as a combination of multiple outcomes[21,35,37,38]. Future approaches to estimate non-recovery could consider these alternative recovery outcomes at different stages of progress as well. The choice of recovery outcome to apply would ultimately depend on the needs of future users, but several metrics for recovery exist outside of Nepal as well, including population displacement[70], food security[56,71], population health[18], perceived recovery[72], and many others[23,73].

Certainly, issues can arise when overly relying on technical disaster information[15,74]—hazards researchers should use data-driven models responsibly when representing complex processes[8,75–77]. The area under the Receiver Operating Characteristic (ROC) curve shown in Supplementary Fig. 3a demonstrates that the model of non-recovery provides an informative prediction with known uncertainty. With new data becoming increasingly available, it is expected that the uncertainty could be further reduced in the future. Nonetheless, the model fills an important gap in existing information that is developed rapidly after a disaster in that it incorporates a range of factors in the projection of long-term and multifaceted recovery needs instead of only considering immediate physical impacts. More generally, including a holistic and reflective set of initial variables is essential to modeling non-recovery. There were basic factors that are known to influence recovery that we would have liked to include at a more granular level, such as gender and economic poverty. Unfortunately, little openly-available data exists on many important social characteristics of recovery, pointing to the need for global high-resolution social, economic, and mental well-being data. This model instead identifies large-scale spatial patterns of recovery, which serves an initial step to more detailed and diverse assessments of recovery needs on the ground.

Many agencies and nongovernmental organizations are focusing on harnessing non-traditional data and methods to estimate damage after a disaster[78,79]. There is great potential to harness this data with data-driven methods to develop early estimates on the likely inequalities in recovery, which are just as important to understand when developing long-term plans. The approach presented in this paper extends beyond current practices of estimating immediate damage to instead estimating the obstacles to long-term recovery after an earthquake. Estimations of non-recovery can be used to identify early after an event the unexpected but relevant factors that are important during recovery which should then be further evaluated on the ground. Having quantitative data on how to support those least likely to recover can frame recovery actions like how to invest in recovery capacity[35], how best to handle reconstruction versus

resettlement[36,80,81], or how to consider the community rather than just reconstructing the building[6]. Many policymakers and international agencies are moving towards developing data-driven evidence to support disaster decision-making[15,82]. Non-recovery is one crucial mechanism to focus our attention on quantifying metrics that prioritize those most vulnerable and support more nuanced recovery planning sooner after a disaster.

Methods

Study area. The April 25th Nepal earthquake caused extensive damage and loss of life. The location of the earthquake's epicenter meant it affected not only the major cities in Kathmandu Valley, including the capital, but also the surrounding and relatively more rural areas. After the earthquake, the Government of Nepal established the Earthquake Housing Reconstruction Program (EHRP) to support households, mostly rural and outside of Kathmandu Valley, with reconstructing more safely[29]. This program provided grants of three lakh ($3000 USD) to houses that experienced severe damage or collapsed (damage grades at or above three using the European Macroseismic Scale, EMS-98, damage scale) based on a detailed eligibility survey conducted in the 18 months after the earthquake. The EHRP was an owner-driven reconstruction program, given the high rates of ownership in these districts. Due to the government's program focus and data availability, we center our study on these highly impacted and relatively more rural districts outside of Kathmandu Valley. These districts include Gorkha, Dhading, Rasuwa, Nuwakkot, Sindhupalchowk, Dolakha, Okhaldunga, Ramechhap, Sindhuli, Kavrepalanchowk, and Makawanpur.

Outside of earthquakes, Nepal faces frequently recurring and ongoing hazards, where the greatest loss of life over the 20 years prior to the 2015 earthquake was caused by landslides and flooding[83]. In the years after the 2015 earthquake, affected communities faced multiple aftershocks, yearly flooding during the monsoon months, and landslides. These ongoing risks mean that Nepal is in a constant cycle of recovery from previous disasters[46].

In addition to Nepal's multihazard risk, the country's geography and changing political landscape make its recovery unique. Rural households face varying levels of remoteness to the nearest municipality, primarily due to the Himalayas' rugged terrain and the inability to access roads. While the districts included in this study are less densely populated than Kathmandu Valley, each district contains several municipalities and rural municipalities that gained increased responsibility to support the recovery of households throughout the district. After decades of a monarchist government (which transitioned to a multiparty democracy in the 1990s)[84,85], Nepal underwent a decentralization and devolution process in 2015-2017 that transferred governing power from the central to local governments located in these municipality headquarters throughout the country[86]. Therefore, the importance of local governments for reconstruction increased throughout the recovery period, especially in supporting more rural regions throughout the district[87,88].

Survey data. The field survey data used in this study were collected by The Asia Foundation (TAF) and local partner Inter Disciplinary Analysts as part of their Independent Impact and Recovery Monitoring (IRM) project[26], funded by UKAid. This survey was part of a series of five surveys meant to monitor longer-term impacts, observe recovery patterns, and track the evolving needs of people affected by the earthquake in Nepal. For a detailed analysis of the fifth round of surveys used in this study, refer to[26]. Here, we only used data from their fifth round of surveys ($n = 5857$), conducted between September-October 2019, or four and a half years after the April 2015 earthquake. For this round of the survey, TAF sampled households using a stratified random sampling technique, representative at the district level. Eleven districts were surveyed, five of which were classified as "Severely-Hit" in the Post-disaster Needs Assessment, three as "Crisis-Hit," two as "Hit with Heavy Losses," and one "Hit," in order of most affected to least affected.

In this study, we only considered households from the six most impacted and rural districts classified as severely-hit and crisis-hit, which overlapped with our study area ($n = 3484$) after removing all non applicable responses to the two questions we considered below ($n = 83$). These districts include Ramechhap, Gorkha, Sindhupalchowk, Nuwakot, Dhading, and Okhaldhunga. The survey question we used as a metric of non-reconstruction asks "If your house was damaged or completely destroyed by the earthquake, have you done any of the following?". Respondents designated whether they have done nothing to reconstruct their house, have started rebuilding, or have finished rebuilding. Even though this question is conditioned on severe damage, we further ensured this condition by only including households that stated in a separate response that their house was partially or fully damaged ($n = 3376$). Conditioning on damage controlled for the differences in reconstruction rates between damage states and the EHRP reconstruction grant that was geared towards only severely damaged homes. By only including those households fit the eligibility criteria for the grant (by being damaged) we mostly address the effect of the reconstruction grant. We controlled for the reconstruction grant through damage statistics rather than receipt of the aid itself, because receiving each tranche of the grant removed another subset of vulnerable households. For example, requirements to receiving each tranche

included understanding the grant rules, having proper documentation, and bank accounts, all of which are barriers for vulnerable households[26,38]. Once we prepared the dataset, we evaluated the distribution of household demographics aligned with those of the original sample collected by The Asia Foundation[26].

We used these responses as a binary variable for our probability classification model. Households that did not complete reconstruction by the time of the survey were classified as one ($n = 727$), and all other households were classified as zero ($n = 2649$). By classifying the survey data in this way, our model predicts the probability of a household not completing reconstruction four and a half years after an earthquake. A map of this survey data is shown in Supplementary Fig. 1.

Predictor data. We represented factors expected to influence non-reconstruction with a set of 32 variables, which come from openly available census, remote-sensing, or modeled datasets. The initial selection of these variables were based on a combination of exploratory interviews with Nepali community leaders and recovery stakeholders; studies on impacts and recovery from the Nepal earthquake; and broader theories on sustainable development, vulnerability, and resilience. The inclusion of a variable was constrained by whether data would be available to represent that variable if the model is in the weeks after a future earthquake in Nepal to predict areas of non-reconstruction. Therefore, predictor data that would be accessible after an event was used, rather than additional questions from the survey data, because survey data would take years to collect after an earthquake and therefore predict spatial patterns of non-reconstruction. Many of the predictor datasets employ the version that was most recently available for that variable after the 2015 earthquake. For example, the modeled population density was developed in 2015 after earthquake, but census variables were taken from the most recent census data in 2011. Though, if these datasets are updated and measured in the same way (for example, if a new census is carried out), those updated variables can be used if the model is implemented in the future in Nepal.

All variables that were considered are listed in Table 1. This table includes only those variables that had reliable and readily available data. Columns describe specifics of the data (Units, Name, Source) as well as information on how that variable was selected (Related Factor(s), Dimension(s)/Capital(s)). Here, we describe the initial selection process for all variables and fully describe the predictor datasets for those eight variables that were selected as most important for predicting non-reconstruction through the variable selection process that is further detailed below.

Initial selection of predictor variables. Initial selection of predictor variables to include based on research on the recovery in Nepal in conjunction with broader studies of vulnerability, resilience, and sustainable development. Through exploratory interviews with ward leaders and organizations active in supporting housing reconstruction ($n = 14$), we identified potential variables that broadly influenced recovery, ranging from additional hazards (like monsoon flooding) to economic capacity (like savings and remittances). The Government of Nepal also instated four criteria for vulnerable populations who require additional assistance during recovery, including single women, the elderly, children, and persons with disabilities[89]. Potential variables from the interviews corroborated with and supplemented variables that were broadly recognized as important for recovery in Nepal from larger-scale studies conducted in affected districts prior to the fifth round of survey data above[35,36,63,90-94]. For example, a longitudinal study on recovery conducted in the first two years after the event similarly found that additional environmental hazards (like flooding) led to lower likelihoods of recovering pre-earthquake livelihood, while broader economic disruptions (like price increases) led to poorer diets and reduced likelihood to reconstruct[90]. The combination of informal interviews and a review of impact and recovery studies led to an initial set of Nepali-specific variables that were expected to influence reconstruction, which we completed with variables that are aggregated into broader indexes of risk and vulnerability that have been aimed at country-level analyses[95,96] or developed countries[31]. These related factors are highlighted in Table 1.

An important constraint on the inclusion of these initial variables in the variable selection process was ensuring data exists to reasonably represent these variables at a high enough spatial resolution to develop rapid estimates of non-recovery after future disasters in Nepal. While several datasets exist to provide a reliable representation of environmental variables at a high resolution, fewer datasets exist for socioeconomic and demographic variables. For example, precipitation models were chosen to represent flood hazard, of which there were several alternative estimates with varying quality and resolutions. However, the most reliable poverty data were from wealth indices, because higher resolution estimates of poverty are largely based on nighttime lights data that were biased towards Kathmandu Valley, an area outside of our study. Because of the greater availability of data to represent environmental variables, we evaluated whether our initial set of predictor data was relatively evenly distributed across multiple dimensions by employing the five capitals necessary for sustainable livelihoods[54,72,97] as a guide to evaluate the range of dimensions of the variables in Table 1. While we would have liked to represent several dimensions in more detail that have been found to be important for reconstruction in Nepal, for example the bridging or bonding dimensions of social capital[90] or detailed remittance and migration information[68,94], more detailed measures would not be readily available in Nepal or similar regions soon after a future earthquake would occur.

Final predictor data. The eight variables that were selected through the variable selection process represented several factors, as described in the results. Histograms with descriptive statistics of these predictors are included in Supplementary Fig. 4 and maps in Supplementary Fig. 5. Shaking intensity consists of the Modified Mercalli Intensity in the United States Geological Survey's Shakemap developed for the main earthquake on April 25, 2015[98]. The seismic landslide hazard map developed by the British Geological Survey provides an index of relative landslide hazard triggered by extreme 24 h rainfall[99]. The remoteness variable estimates the time to travel to the nearest municipality headquarters, accounting for walking and driving time if roads are accessible, through a model developed by the World Bank Poverty and Equity Global Practice[47,48]. While we only calculated remoteness to municipalities, this variable is highly correlated to remoteness to other landmarks (e.g., district headquarters, roads, financial institutions). Original tree cover is derived from Landsat data and shows the per-pixel percentage estimate of tree canopy cover in 2010[100]. To capture the tree cover in the surrounding vicinity of each point in our study area, we took the average percentage within a 30-minute walking distance. Food poverty prevalence is the proportion of individuals living in an local government unit (LGU) who are in households that have a per capita food expenditure that is below the food poverty line. LGUs are a sub-district administrative unit in Nepal that is a collection of multiple villages, similar to a county. Food poverty prevalence per LGU is a small area estimation derived from a statistical model combining surveys and auxiliary data[101]. Population density is the estimate of population per 100 square meters from WorldPop, which we converted to people per square kilometer[102]. Slope was derived in R from the digital elevation model developed by CGIAR[103]. The tap water percentage is from Nepal's 2011 census[104].

Data preparation. Each predictor variable was originally produced or aggregated to different spatial scales (cells, wards, and LGUs), noted in Table 1. To merge with the survey data, we extracted the value of each predictor at the coordinates of each household location. Once merged, we split the combined dataset into six folds using stratified random sampling to ensure each fold had roughly the same proportion of households that are reconstructed and not reconstructed as the full dataset. We also visually inspected whether each fold covered the same spatial distribution of the study area as the full dataset. We used five folds (84%) as the training set to build the model of non-reconstruction and one fold (16%) as the test set for evaluating how the model would perform on a future dataset. For the spatial prediction of non-reconstruction over the study region (Fig. 4a), we converted each proxy to the same resolution of 300m by 300m by resampling raster data using nearest neighbor methods to preserve original data or converting ward and LGU data to cells through extracting the first data value (usually of a vector of one) in a cell.

Models to predict probability of non-reconstruction. We developed a statistical relationship between the surveyed response of non-reconstruction (Y) and the suite of proxies ($\mathbf{X}$) using the training set. Our goal was to predict the probability that a damaged household has not completed reconstruction given its proxy values ($P(Y = 1|\mathbf{X} = \mathbf{x})$). We used a random forest, which is a non-parametric statistical model that averages the results of many individual, decorrelated decision trees[105]. Here we extended the typical random forest to predict probabilities of each household belonging to each reconstruction outcome (1 = not reconstructed, 0 = reconstructed)[106]. A bootstrapped sample of the training dataset is recursively split into distinct subsets for growing one tree in the random forest. Each split divides the data at that split, or parent node, into two child nodes. The parent node is split using a proxy variable that minimizes the mean squared error over all of a set of randomly selected features ($mtry$). For probability estimation, we continued to grow the tree until we reach the minimum nodesize of 10% of the bootstrapped sample. The probability of each node was the proportion of $Y = 1$'s. This process was repeated for a designated number of trees ($ntree$). For our model, we tuned hyperparameters using a grid search and minimized the mean squared error.

Because the random forest model is non-parametric, it does not require assumptions of the distribution of the data or specification of interaction terms. This is attractive for predicting non-reconstruction if a sufficient amount of training data is available because it allows for nonlinear relationships between the predictor variables and reconstruction outcome and for unexpected interactions to occur. We found the random forest outperformed (explained below) the standard probability prediction model, the logistic regression, both on the training and test sets (Supplementary Fig. 3).

Variable selection. To prevent overfitting and for practicality, we reduced the number of variables used in the non-reconstruction model. Prior to model development, we ensured that none of the predictor variables are highly collinear by manually removing all but one variable with a Pearson correlation coefficient greater than 0.75 over the entire study region. These variables tended to be a variation of the same class of predictors (e.g., remoteness to municipality versus remoteness to financial institutions).

During model development, we applied an automatic variable selection technique to identify those predictor variables that are consistently predictive of

non-reconstruction in Nepal, irrespective of training data sample. For the random forest, a simulated noise variable was inserted as an additional predictor and all the predictor variables with a greater Gini importance[107] than the noise variable were selected. To account for variation in the variable selection due to sample location, we repeated the model building process 1000 times using a bootstrapped sample of the training data[108]. Through this automatic selection, we narrowed down the 32 original predictor variables to 12 that occurred more than 75% of the time in the 1000 models, shown in Supplementary Fig. 2, and retrained a new random forest model using these variables.

Once we reduced the variables through this automatic selection, we then manually inspected the partial dependence plots (described below) of the remaining 12 variables to evaluate whether the model found influential relationships between those variables and impeded reconstruction. We found four of the twelve variables to have negligible relationships, as shown in Supplementary Fig. 6. As shown in Supplementary Fig. 6, both Thatch Roof percentage and Dalit Caste Group percentage were selected due to households in the top 1% of the sample–too small of a sample to draw conclusions from. The differences in percentage during the monsoon and dry season also shows a negligible average relationship and were therefore also removed. This manual inspection stage is necessary to ensure that final predictor variables add predictive ability to the model and are therefore useful to collect in the future. A summary of the variables selected and removed are shown in Table 1.

Recovery outcome–predictor variable relationships. The partial dependence plots shown in Fig. 3 provide insight into so-called "black-box" statistical methods, like the random forest[105]. The dark red line is the average marginal effect of a proxy of interest, X_s, on the random forest function, $f(\mathbf{X})$, when all other complementary proxies, $\mathbf{X}_c$, vary over the training data used to build the model of non-reconstruction. The resulting partial dependence function on X_s can be estimated with:

$$\hat{f}_{X_s}(X_s) = \frac{1}{N}\sum_{i=1}^{N}\hat{f}(X_S, \mathbf{X}_{Ci}), \qquad (1)$$

where $\mathbf{X}_{Ci}$ are the values of the proxy variables in the training data of size N. Here, we show these relationships for the training data, as indicated by the light red lines, which is the partial dependence function $\hat{f}_{X_s}(X_s)^{(i)}$ disaggregated for each household and centered to the minimum value of $X_S^{(i)}$[109]. These plots can be interpreted as the average prediction of the model when varying a proxy of interest—it shows what is happening inside the model. However, it does not indicate causal mechanisms in the real world.

Validation. To evaluate the logistic regression and random forest models' performance, we calculated the area under the receiver operating characteristic (ROC) curve[110]. This curve assesses the trade-off between the rate of true positives versus false positives of our trained model of non-reconstruction when varying the cutoff used to classify its predictions as reconstructed or not. The closer the area under the ROC curve (AUC) is to one, the better the model is classifying an outcome. Here, we found an average training AUC of 0.817 for the random forest and 0.636 for the logistic regression (Supplementary Fig. 3a). The AUC of the test set (Supplementary Fig. 3b), which indicates performance on a hypothetical future dataset, was 0.725 for the random forest and 0.592 for the logistic regression. Thus, the random forest model's prediction performs better than the logistic regression and was used for our final model. We also compared the spatial prediction of non-recovery in Fig. 4a with the surveyed reconstruction data in Supplementary Fig. 1 and found the two to visually align.

Model limitations. add here about these being developed for a very specific user Several limitations of this model should be noted. The first is the interaction between reconstruction and aid in Nepal. The model attempts to predict differences in reconstruction trends that can not be explained by building damage. All damaged and collapsed homes were part of the EHRP's standardized assistance program, but we were unable to control for external, non-governmental assistance that households may have received and were influential for several communities[62]. The second is the transferability in time of the model to future earthquakes. Since we learn from each disaster, it remains to be seen whether this specific model can be applied to future earthquakes that may occur in Nepal, though it is likely that many of the identified risk factors will continue to be relevant in the future. Furthermore, predictions of non-recovery cannot capture household level differences in non-recovery due to limitations in the resolution of the predictor data. Finally, the model in Nepal does not include several Nepali-specific sources of vulnerability mentioned in previous vulnerability studies[111,112], such as gender or caste. This does not mean they are unimportant; rather, the final selected variables had a more representative sample or were more predictive for completing reconstruction in Nepal.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

References

1. MacDonald, R. How women were affected by the tsunami: A perspective from Oxfam. *PLoS Med.* **2**, 0474–0475 (2005).
2. Kamel, N. M. & Loukaitou-Sideris, A. Residential assistance and recovery following the Northridge earthquake. *Urban Stud.* **41**, 533–562 (2004).
3. Jha, A. K., Duyne Barenstein, J., Phelps, P., Pittet, D. & Sena, S. *Safer Homes, Stronger Communities: A Handbook for Reconstructing after Natural Disasters*. (The International Bank for Reconstruction and Development / The World Bank, Washington, DC, 2010).
4. Tafti, M. T. & Tomlinson, R. Theorizing distributive justice and the practice of post-disaster housing recovery. *Environ. Hazards* **18**, 7–25 (2019).
5. Howell, J. & Elliott, J. R. Damages done: The longitudinal impacts of natural hazards on wealth inequality in the united states. *Soc. Probl.* **66**, 448–467 (2019).
6. Mukherji, A. Post-disaster housing recovery. *Oxf. Res. Encyclopedia, Nat. Hazard Sci.* **1**, 1–35 (2017).
7. Daly, P. et al. Rethinking relief, reconstruction and development: Evaluating the effectiveness and sustainability of post-disaster livelihood aid. *Int. J. Disaster Risk Reduc.* **49**, 101650 (2020).
8. Mulder, F. Humanitarian data justice: A structural data justice lens on civic technologies in post-earthquake Nepal. *J. Contingencies Crisis Manag.* **28**, 432–445 (2020).
9. Klein, J. A. et al. An integrated community and ecosystem-based approach to disaster risk reduction in mountain systems. *Environ. Sci. Policy* **94**, 143–152 (2019).
10. The European Commission, The United Nations Development Group & The World Bank. Post-Disaster Needs Assessments Volume A Guidelines. Tech. Rep. A, European Union, GFDRR, The United Nations Development Group (2013).
11. Jeggle, T. & Boggero, M. Post-Disaster Needs Assessment: Lessons from a Decade of Experience. Tech. Rep., World Bank, Washington, DC (2018).
12. Loos, S. et al. G-DIF: A geospatial data integration framework to rapidly estimate post-earthquake damage. *Earthq. Spectra* **36**, 1695–1718 (2020).
13. Loos, S., Levitt, J., Tomozawa, K., Baker, J. & Lallemant, D. Efficacy of Damage Data Integration: A Comparative Analysis of Four Major Earthquakes. *Nat. Hazards Rev.* **23**, 04022026 (2022).
14. Soden, R. & Lord, A. Mapping silences, reconfiguring loss: Practices of damage assessment & repair in post-earthquake Nepal. Proceedings of the ACM on Human-Computer Interaction 2 (2018).
15. Liboiron, M. Disaster data, data activism: Grassroots responses to representing Superstorm Sandy. In Leyda, J. & Negra, D. (eds.) Extreme Weather and Global Media, chap. 6, 144-162 (Routledge, New York and London, 2015).
16. Tierney, K. & Oliver-Smith, A. Social dimensions of disaster recovery. *Int. J. Mass Emergencies Disasters* **30**, 123–146 (2012).
17. Comerio, M. C. Disaster recovery and community renewal: Housing approaches. *Cityscape: J. Policy Development Res* **16**, 51–68 (2014).
18. Abramson, D. M., Stehling-Ariza, T., Park, Y. S., Walsh, L. & Culp, D. Measuring Individual Disaster Recovery: A Socioecological Framework. *Disaster Med. Public Health Preparedness* **4**, S46–S54 (2010).
19. Johnson, L. A. & Hayashi, H. Synthesis Efforts in Disaster Recovery Research - International Journal of Mass Emergencies and Disasters. *Int. J. Mass Emergencies Disasters* **30**, 212–238 (2012).
20. Miles, S. B. & Chang, S. E. Modeling community recovery from earthquakes. *Earthq. Spectra* **22**, 439–458 (2006).
21. Spoon, J., Gerkey, D., Chhetri, R. B., Rai, A. & Basnet, U. Navigating multidimensional household recoveries following the 2015 Nepal earthquakes. *World Development* **135**, 105041 (2020).
22. Peacock, W. G., Dash, N. & Zhang, Y. Sheltering and Housing Recovery Following Disaster. In Rodríguez, H., Quarantelli, E. L. & Dynes, R. R. (eds.) Handbook of Disaster Research, 258-274 (Springer New York, New York, NY, 2007).
23. Hamideh, S., Peacock, W. G. & Zandt, S. V. Housing type matters for pace of recovery: Evidence from Hurricane Ike. *Int. J Disaster Risk Reduc.* **57**, 102149 (2021).
24. Capasso, L. & D'Alessandro, D. Housing and Health: Here We Go Again. *Int. J. Environ. Res. Public Health* **18**, 12060 (2021).
25. Peacock, W. G., Van Zandt, S., Zhang, Y. & Highfield, W. E. Inequities in long-term housing recovery after disasters. *J. American Planning Assoc.* **80**, 356–371 (2014).
26. The Asia Foundation. Aid and Recovery in Post-Earthquake Nepal: Quantitative Survey November 2019. Quantitative Survey Phase 5, The Asia Foundation, San Francisco, CA (2021). https://asiafoundation.org/publication/aid-and-recovery-in-post-earthquake-nepal-quantitative-survey-november-2019/.
27. Dennison, L. & Rana, P. Nepal's emerging data revolution background paper. Tech. Rep. April, Development Initiatives (2017).
28. Government of Nepal National Planning Commission. Post Disaster Needs Assessment, Nepal Earthquake 2015. Tech. Rep. B, National Planning Commission, Kathmandu (2015). https://www.npc.gov.np/images/category/PDNA_volume_BFinalVersion.pdf.
29. Nepal Earthquake Housing Reconstruction Multi-Donor Trust Fund. Nepal earthquake housing reconstruction annual report. Tech. Rep., Nepal Earthquake Housing Reconstruction Multi-Donor Trust Fund, Kathmandu, Nepal (2016).
30. Government of Nepal National Reconstruction Authority. Nepal earthquake 2015: Post Disaster Recovery Framework 2016-2020. Tech. Rep., National Reconstruction Authority, Kathmandu, Nepal (2016).
31. Cutter, S. L., Boruff, B. J. & Shirley, W. L. Social vulnerability to environmental hazards. *Soc. Sci. Quarterly* **84**, 242–261 (2003).
32. Cutter, S. L., Burton, C. G. & Emrich, C. T. Disaster resilience indicators for benchmarking baseline conditions. *J. Homeland Secur. Emerg. Manag.* **7**, 14 (2010).
33. Golding, N. et al. Mapping under-5 and neonatal mortality in Africa, 2000–15: A baseline analysis for the Sustainable Development Goals. *Lancet* **390**, 2171–2182 (2017).
34. Bosco, C. et al. Exploring the high-resolution mapping of gender-disaggregated development indicators. *J. Royal Soc Interface* **14**, 20160825 (2017).
35. Mishra, A. et al. Building ex ante resilience of disaster-exposed mountain communities: Drawing insights from the Nepal earthquake recovery. *Int. J. Disaster Risk Reduction* **22**, 167–178 (2017).
36. He, L., Aitchison, J. C., Hussey, K., Wei, Y. & Lo, A. Accumulation of vulnerabilities in the aftermath of the 2015 Nepal earthquate : Household displacement, livelihood changes and recovery challenges. *Int J. Disaster Risk Reduc.* **31**, 68–75 (2018).
37. Spoon, J. et al. Anatomy of disaster recoveries: Tangible and intangible short-term recovery dynamics following the 2015 Nepal earthquakes. *Int. J. Disaster Risk Reduc.* **51**, 101879 (2020).
38. Spoon, J. et al. Understanding short-term household recoveries from the 2015 Nepal earthquakes. Lessons learned and recommendations. *Prog. Disaster Sci.* **10**, 100169 (2021).
39. Turner, B. L. et al. A framework for vulnerability analysis in sustainability science. *Proc. Natl. Acad. Sci. USA* **100**, 8074–8079 (2003).
40. Birkmann, J. Measuring vulnerability to promote disaster resilient societies: Conceptual frameworks and definitions. In Measuring Vulnerability to

Natural Hazards; towards Disaster Resilient Societies, vol. 01, 9-54 (UNU-Press, Tokyo, Japan, 2006).

41. Oliver-Smith, A. Post-disaster housing reconstruction and social inequality: A challenge to policy and practice. *Disasters* **14**, 7–19 (1990).

42. Gnyawali, K. R. & Adhikari, B. R. Spatial Relations of Earthquake Induced Landslides Triggered by 2015 Gorkha Earthquake Mw = 7.8. In Mikoš, M., Casagli, N., Yin, Y. & Sassa, K. (eds.) Advancing Culture of Living with Landslides, 85-93 (Springer International Publishing, Cham, 2017).

43. McAdoo, B. G. et al. Roads and landslides in Nepal: How development affects environmental risk. *Nat. Hazards Earth Sys Sci.* **18**, 3203–3210 (2018).

44. Sudmeier-Rieux, K., Jaquet, S., Derron, M. H., Jaboyedoff, M. & Devkota, S. A case study of coping strategies and landslides in two villages of Central-Eastern Nepal. *Appl. Geogr.* **32**, 680–690 (2012).

45. The Asia Foundation. Aid and recovery in post-earthquake nepal independent impacts and recovery monitoring phase 3 quantitative survey: September 2016. Tech. Rep. Phase 3, The Asia Foundation (2016).

46. He, L. Identifying local needs for post-disaster recovery in Nepal. *World Development* **118**, 52–62 (2019).

47. Banick, R. & Kawasoe, Y. Measuring inequality of access: Modeling physical remoteness in nepal. Tech. Rep. Report 140353, The World Bank, Washington, DC (2019). http://documents.worldbank.org/curated/en/605991565195559324/Measuring-Inequality-of-Access-Modeling-Physical-Remoteness-in-Nepal.

48. Banick, R., Heyns, A. M. & Regmi, S. Evaluation of rural roads construction alternatives according to seasonal service accessibility improvement using a novel multi-modal cost-time model: A study in Nepal's remote and mountainous Karnali province. *J. Transp. Geogr.* **93**, 103057 (2021).

49. Coyle, D. Modelling vulnerability & shelter: Hamro Ghar approach. Tech. Rep., People in Need, Scott Wilson Nepal, Build Change, PHASE Nepal, Kathmandu, Nepal (2019).

50. Practical Action. Supply chain of construction materials in earthquake aff ected districts: An assessment in nuwakot and rasuwa. Tech. Rep., Practical Action, Nepal (2016).

51. Limbu, B., Rawal, N., Suji, M., Subedi, P. C. & Baniya, J. Reconstructing nepal: Post-earthquake experiences from bhaktapur, dhading and sindhupalchowk. Tech. Rep., Social Science Baha, Nepal (2019).

52. Sudmeier-Rieux, K. et al. Scientific evidence for ecosystem-based disaster risk reduction. *Nat. Sustainability* **4**, 803–810 (2021).

53. Moos, C. et al. Ecosystem-based disaster risk reduction in mountains. *Earth-Sci. Rev.* **177**, 497–513 (2018).

54. De Haan, L. J. Globalization, Localization and Sustainable Livelihood. *Sociologia Ruralis* **40**, 339–365 (2000).

55. Liu, W., Gerber, E., Jung, S. & Agrawal, A. The role of human and social capital in earthquake recovery in Nepal. *Nat. Sustainability* **5**, 167–173 (2022).

56. Thorne-Lyman, A. L. et al. Nutritional resilience in Nepal following the earthquake of 2015. *PLoS ONE* **13**, 1–14 (2018).

57. Daly, P., Ninglekhu, S., Hollenbach, P., Duyne Barenstein, J. & Nguyen, D. Situating local stakeholders within national disaster governance structures: Rebuilding urban neighbourhoods following the 2015 Nepal earthquake. *Environ. Urbanization* **29**, 403–424 (2017).

58. Daly, P., Barenstein, J. D., Hollenbach, P. & Ninglekhu, S. Post-disaster housing reconstruction in urban areas in Nepal: Aid governance and local rebuilding initiatives. Tech. Rep., International Institute for Environment and Development, London (2017).

59. Rodgers, J. et al. Resilience Challenges in South and Southeast Asian Hill Cities. In Proceedings of the 11th National Conference in Earthquake Engineering (Earthquake Engineering Research Institute, Los Angeles, CA, 2018).

60. Amnesty International. Building inequality: The failure of the nepali government to protect the marginalised in post-earthquake reconstruction efforts. Tech. Rep. ASA 31/6071/2017, Amnesty International, London (2017).

61. Lallemant, D. et al. Post-Disaster Damage Assessments as Catalysts for Recovery: A Look at Assessments Conducted in the Wake of the 2015 Gorkha, Nepal, Earthquake. *Earthq. Spectra* **33**, 435–451 (2017).

62. Fehr, T. Nepal's Post-Earthquake Development Surge: The Unintended Local Impacts of Reconstruction. *Sociol. Development* **8**, 272–293 (2022).

63. Lord, A. & Murton, G. Becoming Rasuwa Relief: Practices of Multiple Engagement in Post-Earthquake Nepal. HIMALAYA, the Journal of the Association for Nepal and Himalayan Studies 37 (2017). https://digitalcommons.macalester.edu/himalaya/vol37/iss2/12.

64. Rufat, S., Tate, E., Emrich, C. T. & Antolini, F. How valid are social vulnerability models? *Ann. American Assoc. Geographers* **109**, 1131–1153 (2019).

65. Fekete, A. Social vulnerability (re-)assessment in context to natural hazards: Review of the usefulness of the spatial indicator approach and investigations of validation demands. *Int. J. Disaster Risk Sci.* **10**, 220–232 (2019).

66. Spielman, S. E. et al. Evaluating social vulnerability indicators: Criteria and their application to the Social Vulnerability Index. *Nat. Hazards* **100**, 417–436 (2020).

67. Cai, H. et al. A synthesis of disaster resilience measurement methods and indices. *Int. J. Disaster Risk Reduc.* **31**, 844–855 (2018).

68. Raut, N. K. An assessment of livelihood recovery status of earthquake-affected households in Nepal: A study of coping strategies and their effectiveness. *Prog. Disaster Sci.* **9**, 100147 (2021).

69. Hendriks, E. & Opdyke, A. Adoption of seismic-resistant techniques in reconstructed housing in the aftermath of Nepal's 2015 Gorkha earthquake. *Earthq. Spectra* **37**, 2662–2686 (2021).

70. van de Lindt, J. W., Peacock, W. G. & Mitrani-Reiser, J. (eds.) Community ResilienceFocused Technical Investigation of the 2016 Lumberton, North Carolina Flood MultiDisciplinary Approach. No. 1230 in NIST Special Publication (National Institute of Standards and Technology, 2018).

71. Clay, L. A., Papas, M. A., Gill, K. B. & Abramson, D. M. Factors Associated with Continued Food Insecurity among Households Recovering from Hurricane Katrina. *Int. J. Environ. Res. Public Health* **15**, 1647 (2018).

72. Han, Z., link will open in a new window Link to external site, t, Wang, L. & Wei, J. The correlations between livelihood capitals and perceived recovery: A longitudinal study in China after the 2008 Wenchuan earthquake. *Disaster Prevention Manag.* **30**, 194–208 (2021).

73. Jordan, E. & Javernick-Will, A. Indicators of Community Recovery: Content Analysis and Delphi Approach. *Nat. Hazards Rev.* **14**, 21–28 (2013).

74. Shrestha, K. K., Bhattarai, B., Ojha, H. R. & Bajracharya, A. Disaster justice in Nepal's earthquake recovery. *Int. J. Disaster Risk Reduc.* **33**, 207–216 (2019).

75. Blumenstock, J. Don't forget people in the use of big data for development. *Nature* **561**, 170–172 (2018).

76. Gevaert, C. M., Carman, M., Rosman, B., Georgiadou, Y. & Soden, R. Fairness and accountability of AI in disaster risk management: Opportunities and challenges. *Patterns* **2**, 100363 (2021).

77. Soden, R., Wagenaar, D., Luo, D. & Tijssen, A. Taking ethics, fairness, and bias seriously in machine learning for disaster risk management. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2-5 (Long Beach, CA, USA, 2019).

78. Gupta, R. et al. Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 10-17 (Computer Vision Foundation, 2019).

79. Yun, S.-h et al. Rapid damage mapping for the 2015 Mw 7.8 Gorkha earthquake using synthetic aperture radar data from COSMO – SkyMed and ALOS-2 satellites. *Seismol. Res. Lett.* **86**, 1549–1556 (2015).

80. McCaughey, J. W., Daly, P., Mundir, I., Mahdi, S. & Patt, A. Socio-economic consequences of post-disaster reconstruction in hazard-exposed areas. *Nat. Sustainability* **1**, 38–43 (2018).

81. Ingram, J. C., Franco, G., Rio, C. R.-d & Khazai, B. Post-disaster recovery dilemmas: Challenges in balancing short-term and long-term needs for vulnerability reduction. *Environ. Sci. Policy* **9**, 607–613 (2006).

82. Global Facility for Disaster Reduction and Recovery (GFDRR). Machine learning for disaster risk management. Tech. Rep., GDRR, Washington, DC (2018).

83. Pokhrel, D., Bhandari, B. S. & Viraraghavan, T. Natural hazards and environmental implications in Nepal. *Disaster Prevention Manag: Int. J.* **18**, 478–489 (2009).

84. Thapa, M.Forget Kathmandu: An Elegy for Democracy (Penguin, Viking, 2005), illustrate edn.

85. Sharma, K. The political economy of civil war in Nepal. *World Develop.* **34**, 1237–1253 (2006).

86. Chaudhary, D. The decentralization, devolution and local governance practices in nepal: The emerging challenges and concerns. *J. Political Sci.* **19**, 43–64 (2019).

87. Pokharel, T. et al. Political economy analysis of post-earthquake reconstruction in nepal: An assessment of emerging role of local governments. Tech. Rep. September, Kathmandu: Nepal Administrative Staff College and The Asia Foundation (2018).

88. Democracy Resource Center Nepal. The roles of local governments in disaster management and earthquake reconstruction. Tech. Rep. April, The Asia Foundation, Kathmandu (2019).

89. Multi-Donor Trust Fund. Nepal Earthquake Housing Reconstruction: Progress Report: August 2015-May 2018. Progress Report, World Bank, Kathmandu, Nepal (2018). www.nepalhousingreconstruction.org.

90. Mercy Corps. What matters for households' recovery trajectories following the gorkha earthquake? A two-year panel study. Tech. Rep., Mercy Corps, Nepal (2018). https://www.mercycorps.org/research-resources/gorkha-earthquake-recovery.

91. The Asia Foundation. Independent Impacts and Recovery Monitoring Phase 1: Synthesis Report: July 2015. Synthesis Report 1, The Asia Foundation, San Francisco, CA (2015).

92. The Asia Foundation. Independent Impacts and Recovery Monitoring Phase 2: Synthesis Report: February-March 2016. Synthesis Report 2, The Asia Foundation, San Francisco, CA (2016).

93. The Asia Foundation. Independent Impacts and Recovery Monitoring Phase 3: Synthesis Report: September 2016. Synthesis Report 3, The Asia Foundation, San Francisco, CA (2017). https://asiafoundation.org/publication/aid-recovery-post-earthquake-nepal-synthesis-report-phase-3/.

94. The Asia Foundation. Independent Impacts and Recovery Monitoring Phase 4: Synthesis Report: April 2017. Synthesis Report 4, The Asia Foundation, San Francisco, CA (2017). https://asiafoundation.org/publication/aid-recovery-post-earthquake-nepal-synthesis-report-phase-4/.

95. De Groeve, T., Vernaccini, L. & Poljansek, K. Index for Risk Management - INFORM. Concept and Methodology Version 2016. Tech. Rep. EUR 27521, Publications Office of the European Union, Luxembourg (2015). https://publications.jrc.ec.europa.eu/repository/handle/JRC98090.

96. Cardona, O. Indicators of disaster risk and risk management. Technical Notes IDB-TN-169, Inter-American Development Bank, Manizales, Colombia (2010). https://publications.iadb.org/en/publication/11419/indicators-disaster-risk-and-risk-management-program-latin-america-and-caribbean.

97. Ashley, C. & Carney, D.Sustainable Livelihoods: Lessons from Early Experience, vol. 7 (Department for International Development London, 1999).

98. United States Geological Survey. M 7.8 - 36km E of Khudi, Nepal (2015). https://earthquake.usgs.gov/earthquakes/eventpage/us20002926%5C#shakemap?source=uscode=us20002926.

99. British Geological Survey. Modelling exposure through earth observation routines (METEOR) (2021). https://meteor-project.org/.

100. Hansen, M. C. et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853 (2013).

101. Haslett, S., Jones, G., Isidro, M. & Sefton, A. Small area estimation of food insecurity and undernutrition in Nepal. Tech. Rep., Central Bureau of Statistics, National Planning Commissions Secretariat, World Food Programme, UNICEF, and World Bank, Kathmandu, Nepal (2014).

102. WorldPop. WorldPop (2020). https://hub.worldpop.org/project/categories?id=3.

103. Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. Hole-filled seamless SRTM data V4 (2008). http://srtm.csi.cgiar.org/.

104. Central Bureau of Statistics - National Planning Commission Secretariat, GoN. Nepal - National Population and Housing Census 2011, Marks 100 years of census taking in Nepal (2011). https://nada.cbs.gov.np/index.php/catalog/54.

105. Breiman, L. Random forests. Machine Learning 45, 5–32 (2001).

106. Malley, J., Kruppa, J., Dasgupta, A., Malley, K. & Ziegler, A. Probability machines: Consistent probability estimation using nonparametric learning machines. Methods Inf. Med. 51, 1–7 (2012).

107. Nembrini, S., König, I. R. & Wright, M. N. The revival of the Gini importance? Bioinformatics (Oxford, England) 34, 3711–3718 (2018).

108. Degenhardt, F., Seifert, S. & Szymczak, S. Evaluation of variable selection methods for random forests and omics data sets. Brief. Bioinform. 20, 492–503 (2019).

109. Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. J. Comput. Graphical Stat. 24, 44–65 (2015).

110. Hastie, T. J., Tibshirani, R. J. & Friedman, J. J. H.The Elements of Statistical Learning (Springer, 2009).

111. Aksha, S. K., Juran, L., Resler, L. M. & Zhang, Y. An analysis of social vulnerability to natural hazards in nepal using a modified social vulnerability index. Int. J. Disaster Risk Sci. 10, 103–116 (2019).

112. I. Sudmeier, K., Jaboyedoff, M. & Jaquet, S. Operationalizing "resilience" for disaster risk reduction in mountainous Nepal. Disaster Prevention Manag. 22, 366–377 (2013).

113. Loos, S.sabineloos/nonrecovery-nepal: Submission release (2021). https://doi.org/10.5281/zenodo.4560937.

114. Worden, C. B., Thompson, E. M., Hearne, M. G. & Wald, D. J. ShakeMap Manual Online: Technical Manual, User's Guide, and Software Guide (2020).

115. Climate Hazards Center, USB & USAID. CHIRPS: Rainfall estimates from rain gauge and satellite observations (2021). https://www.chc.ucsb.edu/data/chirps.

116. Pratesi, M., Giusti, C. & Marchetti, S. Small Area Estimation of Poverty Indicators. In Davino, C. & Fabbris, L. (eds.) Survey Data Collection and Integration, 89-101 (Springer, Berlin, Heidelberg, 2013).

117. Government of Nepal Central Bureau of Statistics. 2015 Nepal earthquake: Open data portal (2015).

Acknowledgements

Author contributions

S.L., D.L., F.K., J.M. and N.B. conceptualized model. S.L., N.B and J.M. conducted exploratory interviews. S.L., D.L., J.B., F.K. and J.M. developed methodology. S.L. and R.B. collected and curated data. S.L. conducted formal analysis and developed software. S.L., D.L. and J.B., acquired funding. S.L. developed visuals. S.L., F.K. and R.B. wrote original draft. D.L. and J.B. provided supervision. All authors revised and edited manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-023-00699-4.

Correspondence and requests for materials should be addressed to Sabine Loos.

Peer review information Communications Earth & Environment thanks Milan Shrestha and Hitomu Kotani for their contribution to the peer review of this work. Primary Handling Editor: Joe Aslin. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.