Article

Machine learning approach for disaster risk and resilience assessment in coupled human infrastructure systems performance

Check for updates

Xiangpeng Li ✉ & Ali Mostafavi

There is a gap in the literature on data-driven analyses for post-disaster evaluation of community risk and resilience, particularly in utilizing features related to the performance of coupled human-infrastructure systems. This study developed an index and machine learning-based method for assessing community risk and resilience after a disaster. Using feature groups related to population protective actions, infrastructure/building performance, and recovery features, the study examined risk and resilience performance in communities affected by Hurricane Harvey in Harris County, Texas, in 2017. It analyzed disparities across four archetypes of risk-resilience status, and income groups, revealing how spatial areas are shaped by the performance of coupled human-infrastructure systems. The findings also highlight the complex relationship between socio-economic factors, risk exposure, and resilience. This study provides researchers and practitioners with new data-driven and machine intelligence-based methods to evaluate community risk and resilience during disasters, offering insights to inform future policies and decision-making.

Flood risk is are crucial element in vulnerability assessments[1,2], and it can generally be categorized into two key aspects: managing existing systems and disaster response[3]. The first involves mitigating flood disasters through maintenance improvement[4,5], enhanced preparedness[6], and organized evacuation plans[7] to ensure a rational and proactive approach to flood risk management. The second includes activities like rescue operations during or immediately after a flood event. Resilience represents the capacity of an individual, community, city, or nation to withstand, absorb, or recover from an event—such as an extreme flood—or to effectively adapt to adverse conditions or changes, such as climate change or economic challenges, in a timely and efficient way[8]. Risk and resilience are fundamental components in vulnerability assessments that directly influence coupled human-infrastructure systems, where the functionality and resilience of one system critically influence the other[9]. Most existing research on community resilience to disasters and crises primarily focuses on pre-disaster (before-the-event) assessment of risk and resilience, with limited attention devoted to post-disaster (after-the-event) assessment based on the actual performance of communities' coupled human-infrastructure systems. Post-disaster community risk and resilience assessment involves examining and drawing conclusions about the extent of hazard impact and how effectively the coupled human-infrastructure systems performed in coping with the impacts, providing key insights for future resilience planning and strategy development[10,11]. Yet, the current literature primarily emphasizes anticipatory, predictive, or pre-disaster resilience assessments conducted before hazard events[12–14]. Such analyses pay little attention to the resilience analysis at the post-event level, failing to capture, specify, and evaluate the actual status of community risk and resilience in an actual event context. In addition, the current approach to post-disaster assessments[15] often relies heavily on survey methods, which are valuable for grounded-level experience from affected communities. However, surveys are also time-consuming, include a time lag, and put the burden of data collection on impacted communities[16]. Another major limitation in the current approaches to post-disaster risk and resilience assessments is their focus on their limited number of coupled human-infrastructure systems features. For example, some studies[17,18] focus primarily on hazard impacts, and some focused on digital twin with graphs[19], while other studies primarily focus on infrastructure disruptions[20,21] or risk analysis[22–24]. Across these domains, traditional index-based methods, known as composite indicators, are widely employed in risk and resilience assessments by combining multiple metrics into a single numerical score[25]. These scores typically rely on predefined

Urban Resilience.AI Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, USA.
✉e-mail: xplli@tamu.edu

feature weights that are derived from expert judgment or simple statistical models, which may fail to capture the complex, non-linear interactions among diverse risk and resilience factors[26]. The emergence of large-scale data affords an unprecedented opportunity to use a broad range of features related to the coupled human-infrastructure systems' performance in post-disaster assessment of risk and resilience across different areas of communities. Despite increasing recognition of the benefits, there remains a scarcity of literature that systematically applies machine learning within the development or refinement of risk-resilience indices, which limits the ability to provide comprehensive insights into the coupled human-infrastructure systems to analyze the risk and resilience in the context of post-disaster big data analysis.

To address these limitations, machine learning-based approaches can be integrated with index-based methods to allow for data-driven feature weighting, thereby offering a more adaptive and robust framework for risk-resilience evaluations[27]. Consequently, this study aims to bridge that gap by proposing a hybrid framework that leverages machine learning to enhance the reliability and explanatory power of traditional index-based approaches. To this end, using data from the 2017 Hurricane Harvey in Harris County, Texas, this study captures three components of coupled human-infrastructure systems performance (Fig. 1). Using PCA and Hierarchical clustering method, this study classified spatial areas (census block groups (CBGs)) based on their index similarity. The clusters are categorized into four archetypes: high-risk/high-resilience (HH), low-risk/high-resilience (LH), high-risk/low-resilience (HL), and low-risk/low-resilience (LL), based on an in-depth analysis of the coupled human-infrastructure systems features in each cluster (Fig. 1). This approach unveils the risk and resilience of areas of the community based on the coupled human-infrastructure systems performance related to protective actions, infrastructure/buildings disruptions, and population recovery. Specifically, the results reveal that different spatial areas exhibit varying levels of risk and resilience, confirming the existence of four archetypes of risk and resilience status. This study addresses an important gap related to the dearth of data-driven approaches for post-disaster community risk and resilience assessment. The

contributions of this study are fourfold. First, unlike previous approaches that assessed post-disaster resilience based on infrastructure performance. This study captures heterogeneous features (Infrastructure/Building disruptions, Population protective actions, and Population activity recovery) related to the observed performance of the components of coupled human-infrastructure systems (Fig. 1). Second, the computed features related to the performance of coupled human-infrastructure systems (i.e., protective actions, infrastructure/building impacts, and population activity recovery) are based on novel data sources, enabling a data-driven approach for post-disaster rather than survey-based methods. Third, the use of a machine learning approach enables classifying risk and resilience of spatial areas based on the similarity of their coupled human-infrastructure systems performance features. Accordingly, the machine learning-based approach addresses the limitations of index-based methods that rely on subjective feature weights. Fourth, evaluation of the patterns gives insight into features that shape risk and resilience patterns, providing knowledge for informing future plans and actions. These contributions address the current limitations in methods for post-disaster risk and resilience assessment and offer new methods and insights to interdisciplinary researchers and practitioners across disaster science, urban science, and emergency management for post-disaster assessment of community risk and resilience through data-driven and machine intelligence-based methods.

In disaster contexts, population protective actions—such as evacuation, sheltering, and preparedness measures—can reduce immediate risks and significantly alter the loads placed on infrastructure networks[28]. These actions, in turn, shape both the vulnerability and the adaptive capacity of the community[29]. For example, research[9] developed a multi-agent simulation model that integrates human behaviors with infrastructure dynamics, illustrating how decision-making at the household level can ripple through infrastructure systems and ultimately affect steady-state performance. Researchers have employed a mix of quantitative methods to capture the human–infrastructure interplay, which includes agent-based modeling of evacuee decision-making and traffic flow[30], machine learning and big data analytics using location data to infer evacuation and preparedness

Fig. 1 | Characterization of community resilience based on coupled human-infrastructure systems performance. Coupled human-infrastructure systems performance is analyzed based various features: infrastructure/building disruptions, population protective action, and population activity recovery. Each feature category has sub-components for total of 11 features. Based on the coupled human-infrastructure systems performance, spatial areas are grouped into four clusters of risk and resilience archetypes: high-risk/high-resilience (HH), low-risk/high-resilience (LH), high-risk/low-resilience (HL), and low-risk/low-resilience (LL)
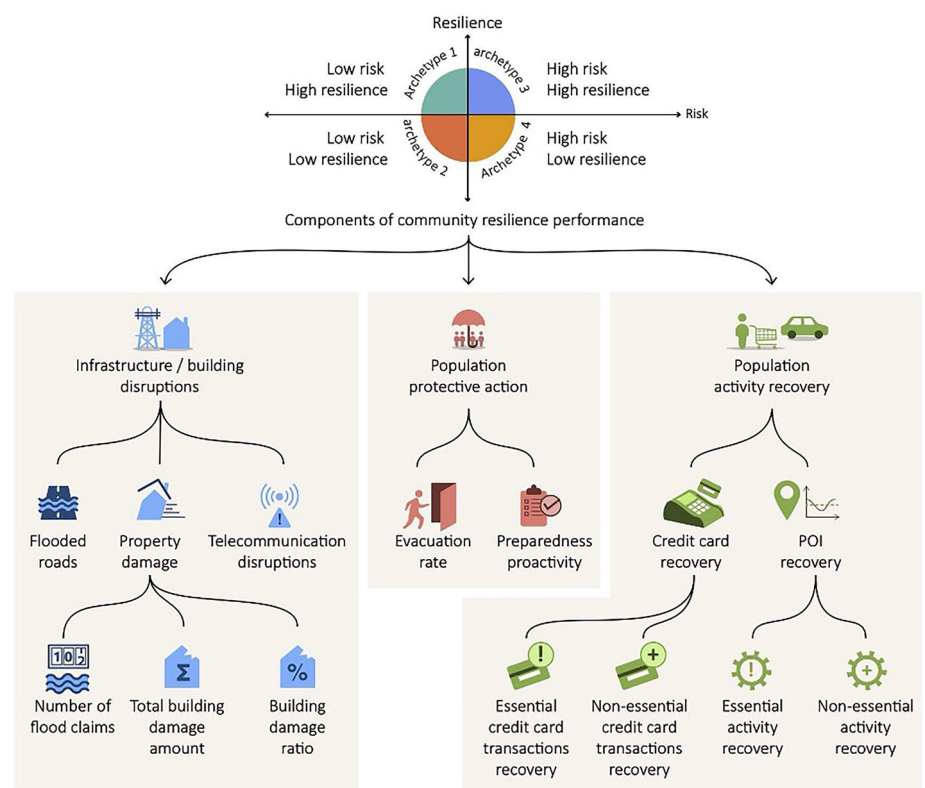
patterns[31]. Therefore, the first component of the coupled human-infrastructure system performance is population protective actions, including evacuation rate and preparedness proactivity. Preparedness proactivity measures the extent to which people take preparatory actions, such as visiting grocery stores or pharmacies before a hurricane. These behaviors provide insight into a community's readiness and awareness ahead of a disaster. Lower levels of proactive behavior may indicate greater vulnerability, which indirectly contributes to the overall risk profile in the Risk Index. Preparedness and evacuation actions moderate the extent of harm people experience and how they cope with impacts in a hazard event[32]. The effectiveness of people's preparedness actions is directly linked to their ability to stay shelter-in-place and cope with the impacts of hazards[33–35]. As part of preparedness action, people would visit critical facilities to purchase necessary items for hazard preparation, including visits to grocery stores, gas stations, and other points of interest (POIs)[36]. Preparation for evacuation, with its primary objective being to minimize the risk of loss of life or injury in the case of a disaster event, is primary protective action, serves as a crucial life-saving measure[37,38]. The evacuation rate has consistently been a key element in risk and disaster management research[39], and studying evacuation practices plays a crucial role in flood mitigation and reducing flood impacts. Evacuation rates serve as an important risk feature because they reflect the extent to which populations are exposed to hazards. For instance, low evacuation rates can lead to higher exposure and increased risk, while high evacuation rates may indicate communities responding to imminent threats[40]. However, it is critical to recognize that evacuation does not always lead to improved safety. As noted by Opdyke[41], evacuees may move to locations with higher storm surge or other disaster-related risks than their primary residences. Moreover, barriers such as lack of transportation, economic constraints, or disabilities can limit the ability of vulnerable populations to evacuate, thereby exacerbating pre-existing social inequalities[42–44]. Given these complexities, evacuation rates should not be viewed as a standalone indicator of resilience; rather, they represent one dimension of risk within a broader framework that accounts for both geophysical hazards and social vulnerabilities. Areas with low evacuation rates may be more vulnerable due to barriers preventing residents from leaving hazardous areas, whereas high evacuation rates might signal heightened exposure to threats or the potential for secondary displacement risks. Consequently, integrating evacuation rates as an element of risk assessment—rather than a direct measure of resilience—better captures the spectrum of potential outcomes and underscores the importance of context-specific analyses when evaluating community readiness and adaptive capacity[40,43].

Infrastructure and building disruptions represent the second critical component of the coupled human-infrastructure systems, as failures or outages in transportation, utilities, and other essential services directly impede recovery processes[45,46]. Restoring infrastructure functionality is fundamental for enabling population mobility, public services, and economic activities—key factors in long-term community resilience. The National Disaster Recovery Framework by FEMA[45] underscores that rebuilding infrastructure and redeveloping damaged areas can take years or even decades, highlighting the profound influence that infrastructure recovery exerts on overall resilience. This study captured features related to flooded roads, telecommunication disruptions, and property damage to capture infrastructure/building disruptions. The rationale for focusing on these features was the extensive disruptions to roads and buildings[47] caused by Hurricane Harvey flooding, as well as disruptions in telecommunications services. Notably, Hurricane Harvey had a profound impact on transportation, such as road inundation and accessibility, with its devastating effects lingering for weeks after the storm[48,49]. Flooded roads highlight infrastructural vulnerabilities[50], as road flooding disrupts mobility, hampers emergency responses, and affects evacuation efforts, exacerbating the impact of disasters. Spatial analyses indicate that hazards can cause internet disruptions even in unaffected areas, highlighting the complex relationship between hazard severity and internet service continuity[51]. The number of flood claims serves as an indicator of the scale and frequency of flood

damage, providing a quantitative measure of risk exposure and the financial burden on affected communities. The damage building ratio and total building damage amount reflect the severity of structural damage from the flood claims caused by disasters, offering insight into the physical and economic vulnerabilities of a region. Other infrastructure disruptions (such as power outages) were not considered since Harvey did not cause extensive power outages in Harris County.

The third component of the coupled human-infrastructure system performance was population activity recovery. Population activity recovery serves as a vital indicator of resilience by reflecting how quickly and comprehensively a community returns to normal or adaptive functioning[52]. Large-scale mobility data, for instance, has been used to track the resumption of activities at points of interest (POIs) and to evaluate both the speed and extent of post-disaster recovery[53,54]. Yabe[55] specifically illustrates how analyzing changes in human mobility patterns within interdependent urban socio-physical systems reveals both the progress and bottlenecks in the recovery trajectory. Moreover, fluctuations in visits to POIs can provide granular insights into the socio-economic dimensions of resilience, helping researchers and policymakers identify which sectors or demographics might be lagging behind[56]. The speed at which affected populations resume their normal life activities has been shown in prior studies to provide an important indicator for community recovery[57]. Population activities are considered to recover when people settle into a pre-disaster lifestyle after coping with impacts, disrupted infrastructure is restored, and businesses resume operations. To capture population activity recovery features, this study compared variations in visits to points of interest (POIs) after the hazard event with normal period trends, using this comparison to measure the duration of recovery. Prior research[58] has shown the effectiveness of evaluating fluctuations in visits to POIs for measuring and quantifying population activity recovery.

The components of the coupled human-infrastructure systems performance discussed above were captured using features computed from various datasets. Community-scale big datasets allow us to observe the dynamics of coupled human-infrastructure systems in the risk and resilience index of communities after hazard events. Harnessing community-scale big data is instrumental in the process of enhancing predictive flood risk monitoring, quick impact assessment, and situational awareness[59,60]. Multiple aspects of coupled human-infrastructure systems performance can be captured from community-scale big data and used to evaluate community risk and resilience status after the fact. By synthesizing these three domains within a coupled human-infrastructure systems framework, this study captured the reciprocal influence of human behaviors and infrastructure functionality. This perspective not only aligns with established community resilience models[29] but also enables a more holistic evaluation of risk and adaptive capacity.

Results
Risk and resilience archetypes
Principal Component Analysis (PCA) loading refers to the correlation coefficients that represent the contribution of each original variable to the principal components derived from PCA[61,62]. From Fig. 2, the PCA loadings tell which specific features had the greatest influence on each index. For the Risk Index, "Number of Flood Claims" and "Total Building Damage Amount" are prominent, with loadings of 0.572 and 0.555, respectively. These features highlight the substantial role of physical damage and historical flood impact in shaping the overall risk profile of a region. Other features in the Risk Index, such as "Damage Building Ratio" (loading = 0.526) and "Evacuation Rate" (loading = 0.278), also show significant influence, although to a lesser extent. The moderate loading of "Evacuation Rate" suggests that behavioral responses to potential hazards contribute to risk but do not dominate the index. Additionally, features like "Flooded Roads" (loading = 0.082) and "Preparedness Proactivity" (loading = 0.048) have relatively lower weights, indicating their limited direct impact on the aggregated risk measure. "Telecommunication Disruptions" has a negative loading of −0.044, reflecting a slight inverse relationship, where increased
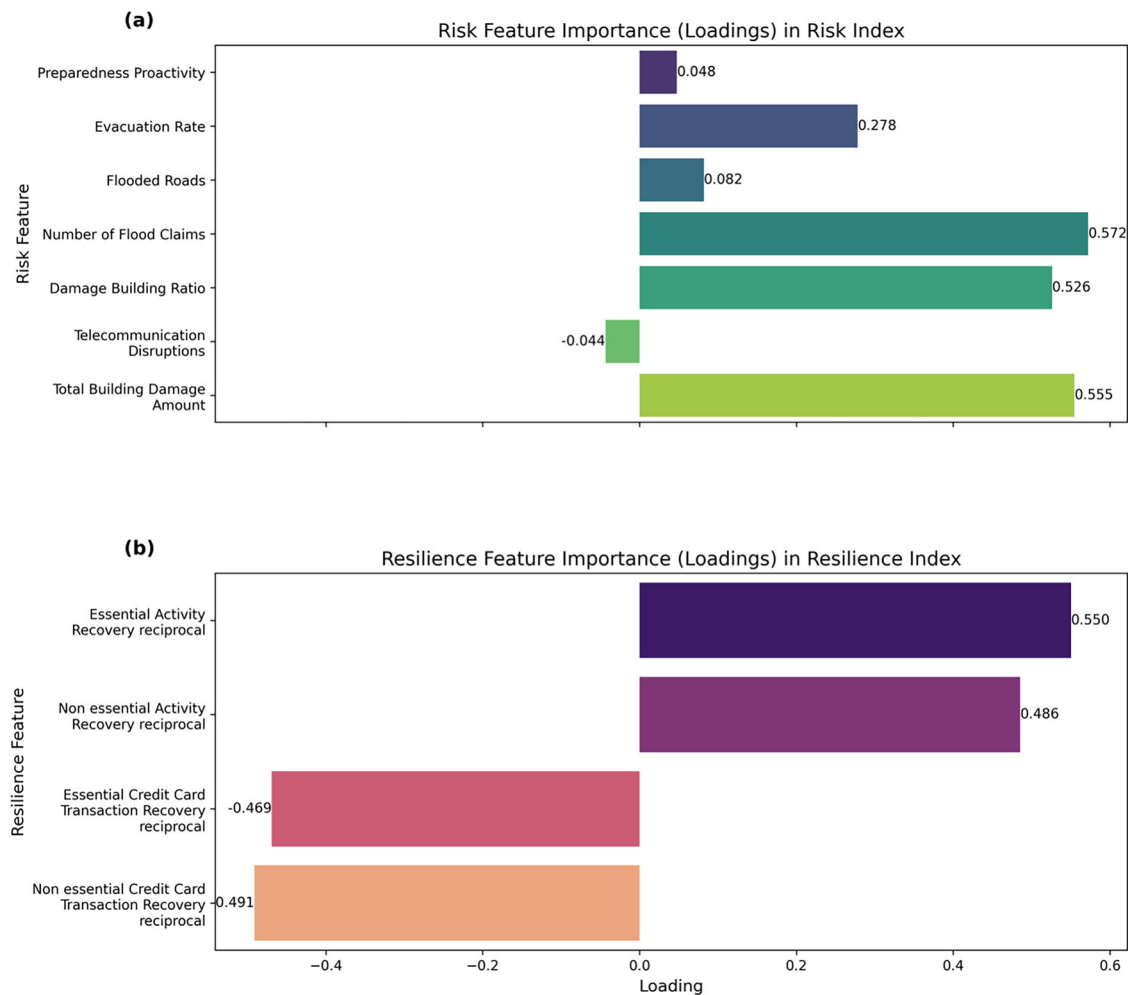
Fig. 2 | Feature importance to index. a Risk features importance in risk index; **b** Resilience features importance in resilience index.

telecommunication stability might correlate with higher perceived risk regions. In constructing the Resilience Index, we focused on features that indicate recovery efficiency across different sectors. The loading plot for resilience demonstrates that the reciprocal of "Essential Activity Recovery" duration has the highest influence, with a loading of 0.550, underscoring its critical role in the resilience framework. This feature signifies how quickly essential activities return to baseline functionality, highlighting regions with robust recovery mechanisms. Similarly, the reciprocal of "Non-essential Activity Recovery" carries a substantial loading of 0.486, pointing to the importance of restoring non-essential services in evaluating resilience. The negative loadings observed for the reciprocal of "Essential Credit Card Transaction Recovery" (-0.469) and the reciprocal of "Non-essential Credit Card Transaction Recovery" (−0.491) reflect the impact of economic activity recovery patterns; regions where economic transactions resume slowly exhibit lower resilience scores.

Then the index values are used into hierarchical cluster, with setting the cluster numbers ranging from 2 to 9. The silhouette result revealed that a four clusters solution maximized the silhouette score at around 0.46, which, although moderate, is acceptable in the context of complex social and environmental data where overlapping clusters are common. To visualize different archetypes, we calculate the mean value of risk and resilience index of all areas. The scatter plot in Fig. 3 categorizes regions into four distinct risk and resilience archetypes, each representing different combinations of risk and resilience characteristics within the framework. The plot is organized with a horizontal dashed line marking the mean resilience value and a vertical dashed line representing the mean risk value, which helps to visually separate each cluster based on its mean risk and resilience index. The

presence of confidence ellipses also highlights the internal heterogeneity of clusters. This segmentation reveals four archetypes that classify LL as low-risk and low-resilience, HL as high-risk and low-resilience, LH as low-risk and high-resilience, and HH as high-risk and high-resilience. We used a Chi-square test to examine the relationship between clusters and their distribution across four quadrants of risk and resilience, based on the overall mean values of the Risk and Resilience indices. The Chi-square test statistic is 1646.16, with a corresponding p value of 0.0, indicating a statistically significant association between clusters and quadrants.

In Fig. 3, LL represented in purple and positioned in the lower-left quadrant. With a mean Risk Index of −0.56 and a Resilience Index of -0.72, this archetype suggests areas that may not face severe risks but are also not good at recovery. Although they currently experience low risk, these areas remain vulnerable due to their limited recovery capacity. HH, shown in orange and located in the upper-right quadrant, represents regions that face high risk but have high resilience. With a high-risk Index of 7.74 and a high Resilience Index of 0.31, HH represents a high-risk, high-resilience group. This archetype identifies areas that, despite significant exposure to risk, have relatively strong recovery capabilities. The large confidence ellipse around HH's centroid suggests a broad variation within this group, indicating that while some areas might have moderate resilience, the overall archetype represents a segment that is predominantly vulnerable in the face of high risks. LH, depicted in blue and situated in the upper-left quadrant, stands out as a low-risk, high-resilience group. With a mean Risk Index of -0.36 and a high Resilience Index of 1.18, this archetype represents regions that are well-prepared for recovery even though their risk exposure is relatively low. The tighter ellipse around LH's centroid indicates less variation within the

Fig. 3 | Risk and resilience archetype for each archetype. There are four resilience and risk archetypes: LL is categorized under the low-risk/low-resilience (LL) quadrant; HH is positioned in the high-risk/high-resilience (HH) quadrant; LH falls into the low-risk/high-resilience (LH) quadrant; HL is situated in the high-risk/low-resilience (HL) quadrant.
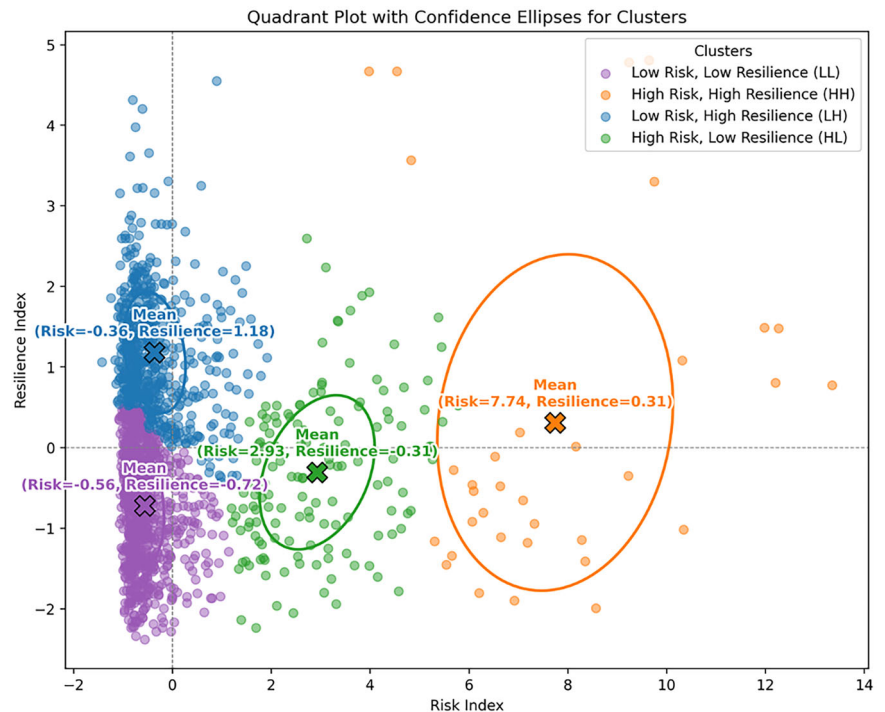
archetype. Finally, HL, represented in green and positioned in the lower-right quadrant, combines high risk with low resilience. With a mean Risk Index of 2.93 and a Resilience Index of -0.31, this archetype represents areas that are highly exposed to risk but exhibit limited recovery capacity.

Figure 4, the map of Harris County[63], TX, illustrates the spatial distribution of the four identified archetypes, each represented by a unique color. LL (purple) occupies a significant portion of Harris County, particularly in the outer regions and sparsely populated areas. This archetype corresponds to regions that are less urbanized and have limited infrastructure support for rapid recovery. The dominance of LL in less densely developed areas suggests insufficient disaster preparedness and recovery resources. One possible reason for this pattern is the poor drainage systems and low flood depth thresholds in these areas, which can lead to stagnation when extreme events occur[64]. Additionally, peripheral or suburban regions are more likely to fall into the LL category due to fewer emergency resources, lower levels of public infrastructure, and weaker community networks[65]. Despite a lower immediate flood risk, these communities remain vulnerable because of limited preparedness and response capacity, particularly in extreme events[66]. HH (orange) has a more restricted spatial distribution compared to other archetypes. However, these regions demonstrate stronger infrastructure, adaptive strategies, and proactive flood management, which contribute to their resilience despite facing significant flood risks[67]. LH (blue) is scattered across Harris County but is more prevalent in moderately urbanized zones. These areas maintain high resilience despite lower natural hazard exposure due to greater preparedness, stakeholder engagement, and strong governance[65,67]. The combination of proactive flood management and community-driven disaster preparedness plays a crucial role in their ability to recover effectively from hazards. HL (green) is found in various dispersed areas throughout the county. These areas experience frequent and severe flooding but lack adequate resilience measures, such as emergency response systems, flood defenses, or well-maintained drainage infrastructure[64]. HL regions are particularly vulnerable due to their above-average flood risk and below-average access to emergency shelters, making them susceptible to prolonged recovery periods. Furthermore, these areas are often home to historically marginalized or low-income communities, where infrastructure investments have not kept pace with increasing hazard exposure[66].

Features' contribution to index in different archetypes

To visualize the features' contribution to their own indexes for each archetype, Clustered Feature Aggregation was applied, and examined the median values of various resilience and risk-related features across four distinct archetypes (Fig. 5). The resilience features are transformed using the reciprocal function (smaller values after transformation indicate better resilience). Median values of risk and resilience features are calculated within each archetype, and this step enables comparison across archetypes, revealing patterns in how different archetypes vary in their resilience and risk profiles. After aggregating the features within each archetype, Min-Max scaling is applied separately to the median values to bring all features into a comparable range (typically 0 to 1). This normalization facilitates a more interpretable comparison across archetypes, regardless of the original feature scales.

In LL, we observe minimal exposure to most risk factors, as evidenced by the relatively small, stacked values in the positive risk features. Telecommunication Disruptions is the primary risk feature for this archetype, but it is negative value in risk index, suggesting that telecommunication disruptions do not lead to high risk. In terms of resilience, LL shows low values for essential and non-essential activity recover, but high values in non-essential card transactions. The study shows that card transactions is negatively related to resilience, so LL has lowest resilience compared to other archetypes. HH displays substantial values across several risk indicators, including high scores for "Total Building Damage Amount" and "Damage Building Ratio". These high values contribute significantly to the archetype's Risk Index, indicating that these areas are particularly vulnerable to physical and infrastructural risks. Despite this vulnerability, HH exhibits strong resilience, as seen in the high median values for essential and non-essential activity recovery. The resilience plot shows that activity recovery speed plays a crucial role in mitigating the impact of disruptions. Thus, although HH faces considerable risk, its recovery capabilities allow it to effectively bounce back, particularly in critical services. LH shows low values across risk features, which like LL. Resilience in LH is particularly strong, as shown by highest values in essential and non-essential activity recovery, highlighting a robust capacity for restoring critical services after disruptions. LH also shows minimal reliance on credit card recovery activities, yet maintains a
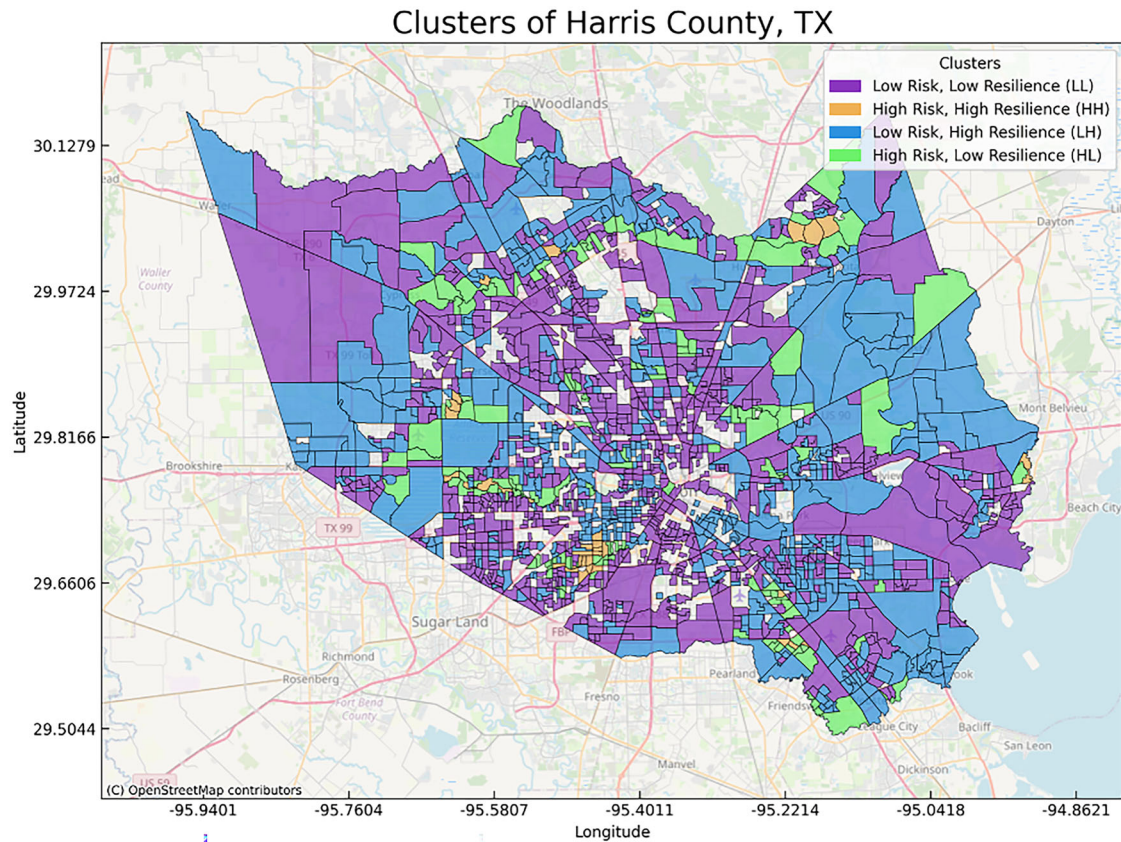
Fig. 4 | Spatial distribution of archetype in Harris County, TX. Harris County is divided into four archetypes based on the Hierarchical Clustering analysis: LL (863 CBGs), representing the low-risk, low-resilience archetype, is shaded purple; HH (34 CBGs), representing the high-risk, high-resilience archetype, is orange; LH (557 CBGs), representing the low-risk, high-resilience archetype, is blue; and HL (144 CBGs), representing the high-risk, low-resilience archetype, is green. Areas with missing data are shaded gray.

high Resilience Index due to strong activity recovery. HL reveals significant exposure to risk factors, with high values for "Number of Flood Claims," "Building Damage Ratio," and "Total Building Damage Amount." These prominent risk factors contribute to its high-risk Index, indicating that this archetype is highly susceptible to disruptions. In terms of resilience, HL shows lower scores for essential and non-essential activity recovery and negative scores for card transaction activities, reflecting limited recovery capabilities. Comparing HH and HL, we observe that HH has a higher "Total Damage Amount" than HL, leading to a higher risk index. This suggests that the damage amount serves as a significant indicator of an area's risk level. LH has highest "Essential Activity Recovery" speed and "Non-essential Activity Recovery" speed; also, it does not have significant credit card activity, and this allows LH to have the highest resilience.

Overall, there is the significant role that the building system plays in determining the risk index. Features such as the building damage amount and building damage ratio are particularly influential, as higher values for these features correspond to an increased level of risk. This underscores the importance of infrastructure vulnerability in shaping the overall risk profile of a region. At the same time, the resilience index is strongly associated with the recovery of essential and non-essential activities. Higher performance in these activities—indicating a quicker return to normalcy after a disruptive event—captures greater resilience. This highlights the critical role of economic and social activity recovery in mitigating the long-term impacts of disasters and improving community adaptability. By linking the building system to risk and activity recovery to resilience, our study emphasizes the interconnected nature of physical infrastructure and social dynamics in disaster management.

Disparities in risk and resilience archetypes across income groups

Both descriptive statistics and statistical tests were used to explore how income varies across the archetypes. From Fig. 6, we observe that LL (colored in purple) has a median income of approximately $54,618 and contains the largest number of areas, with a count of 845. This archetype, with a relatively lower median income, represents areas with significant dispersion and several low-income outliers, indicating economic diversity. HH (in orange) has the highest median income of about $113,380, albeit with a much smaller sample size of 34 areas. LH (blue) and HL (green) have similar median incomes, around $61,924 and $62,962, respectively, with 552 and 142 areas in each archetype.

To capture the compound effect of these variables, we constructed an interaction term, "Risk-Income Interaction (RII)," calculated as the product of Risk Index and Median Income (Risk Index × Median Income). The OLS regression analysis (Fig. 7) examines how the interaction between risk index and median income influences resilience index across four distinct community archetypes.

For the negative interaction pattern (LL and LH archetypes), which is also low-income areas, both Low Risk, Low Resilience (LL: -3.72e-06, $p < 0.001$) and Low Risk, High Resilience (LH: -3.15e-06, $p < 0.001$) archetypes exhibit significant negative interaction effects. This pattern indicates that within these contexts, higher income does not effectively buffer against increased risk exposure. The negative interaction suggests that as both risk and income increase simultaneously, resilience actually decreases. The findings here imply that in low-income areas, resources may not be effectively allocated to resilience-enhancing activities, or that socioeconomic factors limit the resilience benefits of higher income. The similarity in coefficients across LL and LH suggests this limitation transcends
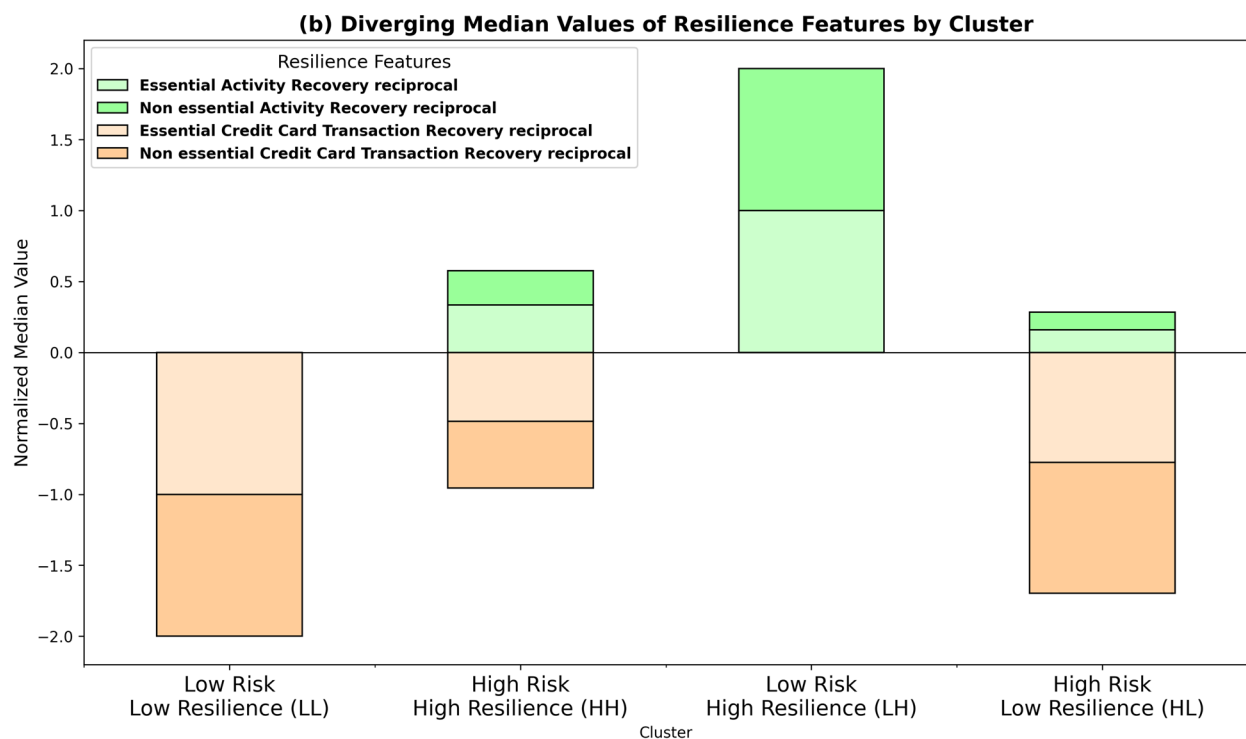
(a)
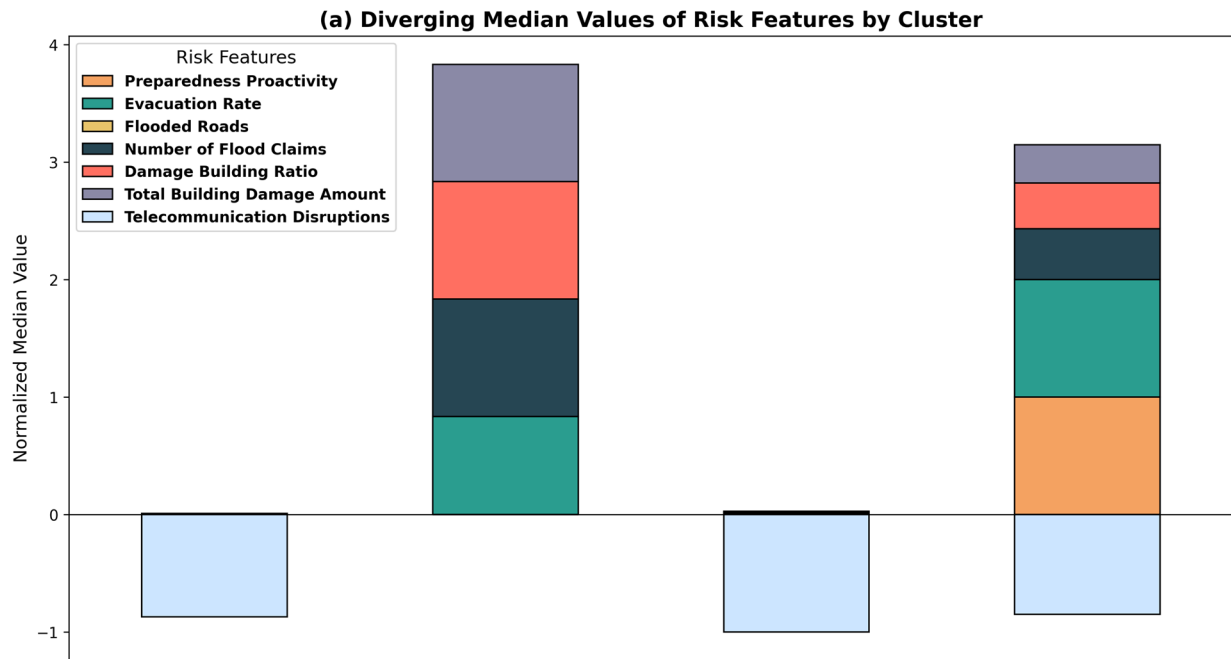
(b)

Fig. 5 | Risk and resilience contributions. a Weighted median value of risk features by archetype; **b** Weighted median value of resilience features by archetype. Positive and negative contributions are visualized separately, with color-coding that enhances interpretability: positive resilience features are shaded in green, negative resilience features in orange, positive risk features are assigned random colors, and negative risk contributors, such as "Telecommunication Disruptions," appear in blue.

baseline resilience levels and may be tied to broader socioeconomic or institutional constraints.

For the positive interaction pattern (HH and HL archetypes), which is also high-income archetype, both High Risk, High Resilience (HH: 1.58e-06, $p < 0.05$) and High Risk, Low Resilience (HL: 9.62e-07, $p < 0.05$) archetypes demonstrate positive interaction effects. This

pattern indicates that income effectively moderates the risk-resilience relationship, with higher income providing protective benefits in high-risk contexts. While HH shows a stronger effect than HL, likely due to its higher median income, both archetypes demonstrate the capacity to leverage financial resources for resilience enhancement under risk exposure. This suggests that income contributes

Article

meaningfully to resilience in this archetype, likely due to enhanced capacity for investments in infrastructure, insurance, and community resources that can buffer against risk. These high-income areas are thus better equipped to withstand and recover from disruptions, underscoring the role of financial resources in risk management and community resilience.
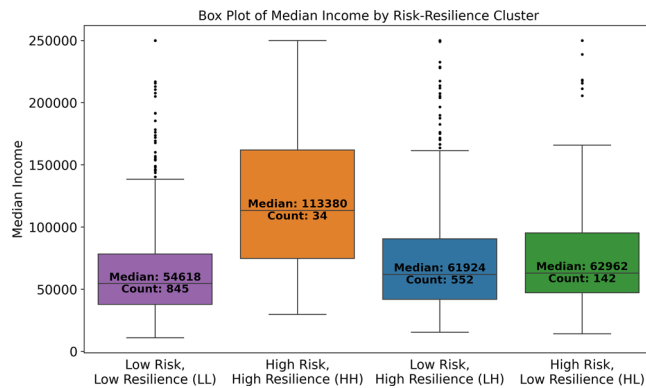

Fig. 6 | Box plot of median income by archetype.

The findings from each archetype emphasize that the impact of the interaction between risk and median income on resilience is context-dependent. Archetypes with higher baseline income levels, such as HH and HL, demonstrate a positive association between income and resilience in high-risk areas, likely due to more substantial investments in resilience resources. Conversely, lower-income archetypes like LL and LH exhibit a negative interaction effect, indicating that higher income alone does not suffice to offset the challenges posed by high risk. Overall, this analysis reveals that income has a complex and context-specific influence on resilience in high-risk environments. Higher-income areas have a more protective response to risk, while lower-income areas may struggle to achieve resilience despite high income levels. This underscores the need for tailored strategies that consider both income and risk levels to enhance resilience effectively across diverse communities.

Discussion

This study proposed an index and machine learning-based approach for post-disaster assessment of community risk and resilience based on coupled human-infrastructure systems performance. Despite significant advancements in the interdisciplinary field of community resilience, most existing studies focus primarily on pre-disaster assessments, and relatively less attention has been paid to post-disaster assessments. Additionally, many post-disaster assessments are based primarily on disaster reconnaissance
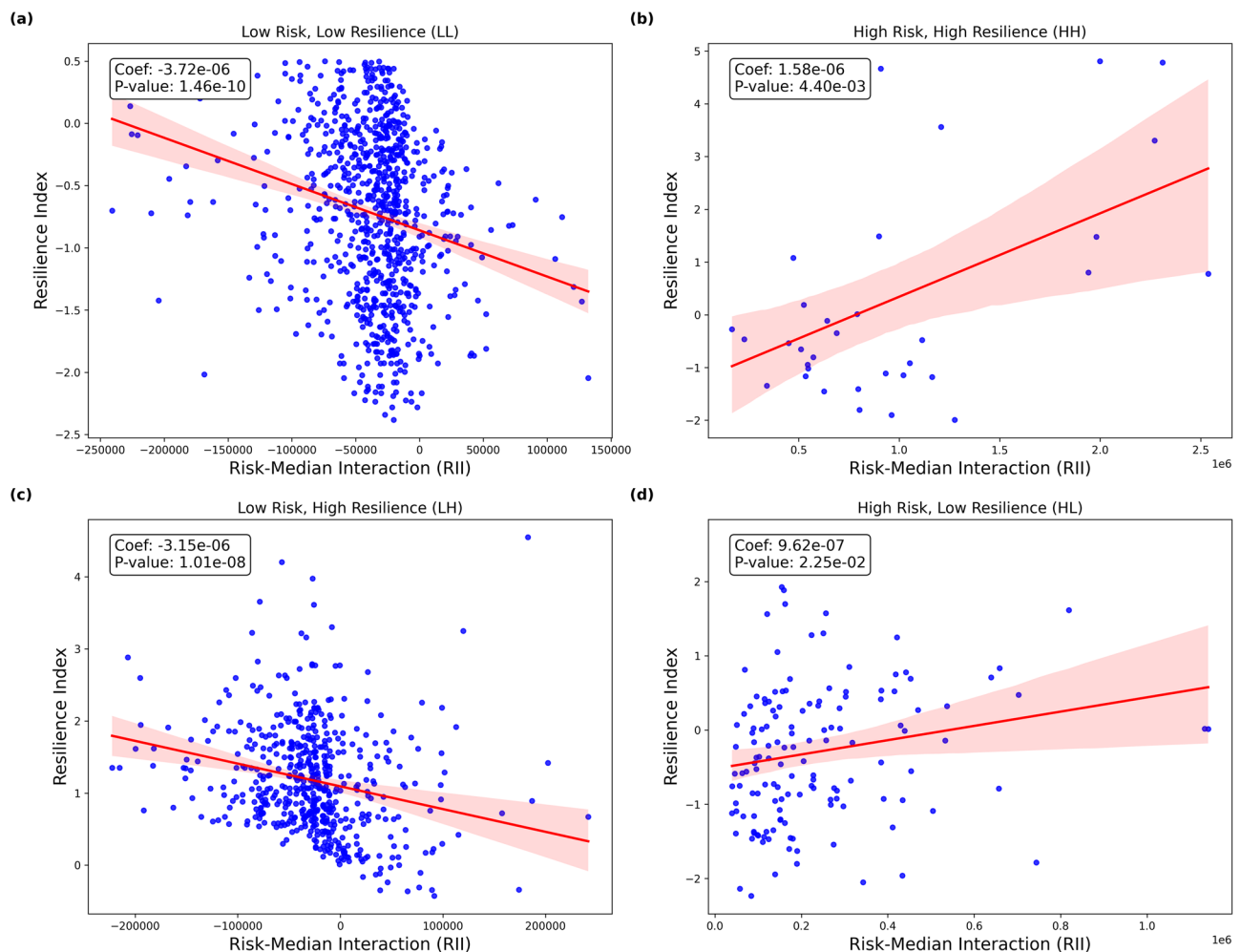

Fig. 7 | Relationship between RII and resilience index. a Relation of RII and resilience index in LL; **b** Relation of RII and resilience index in HH; **c** Relation of RII and resilience index in LH; **d** Relation of RII and resilience index in HL.

approaches focusing on measuring impacts and damage. Furthermore, post-disaster assessments rely primarily on surveys for capturing impacts and damage, resulting in delays in data collection and assessment. Finally, the current approaches to post-disaster assessment of community resilience focus mainly on single systems (e.g., transportation, housing, or businesses) and do not consider various aspects of coupled human-infrastructure performance. These limitations are addressed in this study based on evaluating various features related to coupled-human infrastructure performance computed from various data sources and by adopting a machine learning-based approach to classify spatial areas of a community (census block groups) based on the intertwined features to unveil various risk and resilience performance archetypes.

This study contributes to disaster science and resilience planning through several key innovations. We developed a novel approach to post-disaster assessment of community resilience using an index and machine intelligence-based approach that provides interdisciplinary researchers with a new tool for evaluation. Our methodology integrates diverse features related to coupled human-infrastructure systems performance, including infrastructure service disruption and restoration, protective actions, property damage, and life activity recovery of populations. This comprehensive framework offers advantages over existing approaches that focus primarily on human or infrastructure systems separately. By utilizing observational data from various sources, our data-driven approach enables comparison of coupled human-infrastructure systems performance across different events and regions, while guiding response and recovery efforts to fairly effect resource allocation and prioritization. The identification of four distinct community risk-resilience archetypes according to their coupled human-infrastructure systems performance provides essential insights for future risk reduction and resilience improvement strategies. For example, strong resilience, particularly in essential and non-essential activity recovery, suggests robust systems in place to resume key activities after disruptions. Additionally, features like total building damage amount consistently correlate with increased risk exposure across different datasets, highlighting specific vulnerability patterns. These findings can inform targeted interventions such as building retrofitting, enhanced construction standards, and damage mitigation strategies to reduce risk in vulnerable areas. In addition, the analysis of Harris County during Hurricane Harvey revealed important disparities in risk-resilience profiles across different median income communities. While higher baseline income levels can enhance resilience in high-risk areas through greater investment in resilience resources, this positive association is not universal. Lower-income areas, even those with relatively high-income levels within their context, may still struggle to achieve resilience when facing significant risk factors. This demonstrates that income alone does not uniformly confer resilience, particularly in communities exposed to elevated risk levels. Therefore, effective resilience enhancement requires tailored strategies that consider not only income levels but also the unique risk profiles of each community. These findings can inform targeted interventions such as building retrofitting, enhanced construction standards, and damage mitigation strategies to reduce risk in vulnerable areas.

While this study provides valuable insights into community resilience, it is important to acknowledge several limitations. First, data availability constraints may affect the quality and comprehensiveness of the analysis. Certain critical datasets on human-infrastructure interactions (e.g., building codes, real-time infrastructure usage, demographic distributions) may be inaccessible or unavailable, restricting the scope of the assessment. Besides, NFIP data has important limitations, as many residents are unaware they live in floodplains, and cannot afford flood insurance. Hence, it does not capture uninsured losses. Future studies can include other datasets that can complement NFIP data to measure flood impact and losses in communities. Second, our analysis relied on certain assumptions (such as income as a key indicator) and generalizations (including current features) when interpreting risk and resilience patterns. These simplifications may not fully capture the complex socio-economic, cultural, and governance factors that influence how different communities experience disasters[31]. This limitation underscores the importance of incorporating local data and fostering cross-

sector collaboration to develop more effective and contextually appropriate resilience strategies. Despite these constraints, the overarching framework—encompassing population protective actions, infrastructure disruptions, and activity recovery—can be adapted to various disaster contexts. For instance, in earthquake-prone areas, data on building codes, seismic retrofits, and population evacuation behavior could replace storm surge or flood-relevant variables. In wildfire scenarios, evacuation routes, air-quality metrics, and infrastructure vulnerabilities (e.g., power grid stability) might be emphasized to reflect region-specific risks[68]. As more hazard- or region-specific data becomes available, researchers can integrate additional indicators tailored to local contexts, thereby refining the robustness of the coupled human-infrastructure system analysis. Building on these insights, future studies should pursue at least two avenues of investigation. First, they can incorporate new data sources—such as high-resolution remote sensing, crowd-sourced surveys, or real-time IoT sensors—to enhance the granularity of resilience metrics. Second, cross-regional and cross-event analyses applying this framework to different disaster types will enable more consistent comparisons of resilience patterns and the identification of universally effective (or context-specific) interventions. Third, other socio-demographic features can be investigated to have more comprehensive understanding of community vulnerability and resilience to disasters, as factors such as educational attainment, age distribution, housing characteristics, and occupational diversity may provide additional insights into the complex socio-economic dynamics that influence disaster impacts and recovery trajectories. Through this broader application, key factors that strengthen resilience can be validated and translated into evidence-based policy recommendations, thereby broadening the framework's utility for diverse hazard contexts.

Methods
Study context
Beginning as a tropical storm in early August 2017, Hurricane Harvey rapidly escalated to a Category 4 hurricane. By August 23, it had intensified, necessitating mandatory evacuations in several Texas counties as forecasts predict the storm to escalate into a major hurricane[69]. Making landfall about 50 kilometers (about 30 miles) northeast of Corpus Christi, Texas, on August 26. Harvey then moved towards Houston, causing severe and widespread flooding in southeastern Texas, with particularly devastating impacts occurring on August 27. These conditions persisted until August 30, 2017, and Harvey was officially declared over on August 31, leading to widespread recovery and reconstruction efforts across numerous Texas and Louisiana counties[70].

From August 25 through 30, 2017, Hurricane Harvey drenched Harris County with more than 130 centimeters (50 inches) of rain, causing unprecedented flooding across vast areas of the city[71]. Even though the hurricane's center bypassed Houston to the south, it brought significant rains and floods to the area, particularly in the northeast, due to a stationary front around the storm[72]. Harris County, encompassing the densely populated Houston area, experienced more than 100 centimeters (40 inches) of rain, resulting in considerable flooding and damage[73]. Harvey's ferocious winds, peaking at 150 miles per hour, caused localized damage and necessitated emergency actions to avert dam failures, further aggravating the flooding and infrastructure destruction[72,74]. The significance of impacts, as well as the breadth of geographic area that was impacted, makes this study context a suitable setting for post-disaster assessment of community risk and resilience in this study.

General approach
Figure 8 presents a detailed illustration of our analysis steps. This study examines 11 specific features related to coupled human-infrastructure systems performance during flood events, which we categorized into two primary indices: Risk Index and Resilience Index. Risk Index captures the vulnerability and exposure aspects through seven features: flooded roads, telecommunication disruptions, total building damage amount, building damage ratio, number of flood claims, preparedness proactivity, and evacuation rate. Resilience Index measures recovery capacity through four

Fig. 8 | The overview of analysis steps. Total of eleven features are classified into two indexes and then classified into four clusters using the PAC indexed Hierarchical algorithm. These clusters are then labeled as risk and resilience archetypes based on their index values. Accordingly, disparities in risk-resilience archetypes of CBGs with varying median income levels are examined.
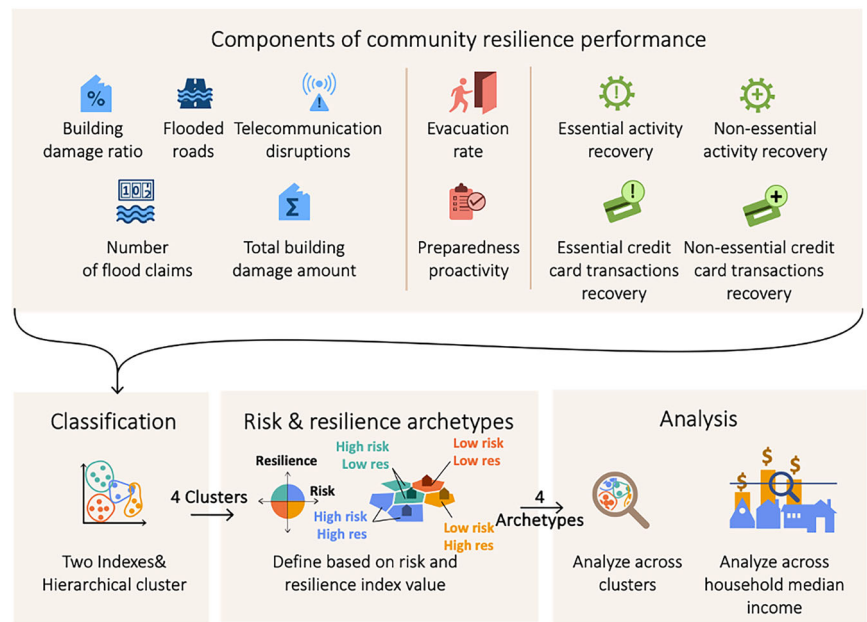
features: essential activity recovery duration, non-essential activity recovery duration, essential credit card transaction recovery duration, and non-essential credit card transaction recovery duration. The first step employed Principal Component Analysis (PCA) to generate two comprehensive indices—risk index and resilience index—based on eleven distinct features that characterize community response to flooding events. Then the study applied a Hierarchical Clustering method to organize communities into four distinct risk-resilience archetypes based on their index values. The findings reveal four distinct risk-resilience archetypes distributed across Census Block Groups (CBGs). The analysis examines how individual features contribute to both the Risk Index and Resilience Index within each archetype, and how specific features contribute to risk and resilience in different areas. Additionally, this study investigated the relationship between household income levels and archetype classification to understand socioeconomic influences on community risk and resilience patterns during flood events.

Datasets descriptions

This study focused on three components of the coupled human-infrastructure systems performance that shape community risk and resilience (Fig. 1): population protective actions, infrastructure/building disruptions, and population activity recovery. These features are heterogeneous as they are derived from multiple data sources—including location-based services, claims, telecommunication networks, and financial transactions—each representing a distinct dimension of post-disaster recovery. By integrating these diverse factors, our study provides a multi-criteria perspective on how risk and resilience manifest across different aspects of community functioning. The data sources utilized in this study are summarized in Table 1. Our analysis focused on computing the coupled human-infrastructure systems features at the CBG level. As shown in Table 1, the components of the coupled human-infrastructure systems performance we examined include three main feature groups: population protective action, infrastructure/building disruptions, and population activity recovery. Population protective action, encompassing preparedness proactivity and evacuation rate, is calculated based on location-based data. Infrastructure/building disruptions include flooded roads, property damage, and telecommunications disruptions. Population activity recovery durations were computed based on location-based data (POI data) and credit card transactions. In total, the analysis examined 11 features related to the coupled human-infrastructure systems performance: flooded roads, number of flood claims, total building damage amount, building damage ratio, telecommunication

disruptions, preparedness proactivity, evacuation rate, essential activity recovery, non-essential activity recovery, essential credit card transaction recovery, and non-essential credit card transaction recovery. There is one socio-demographic characteristic feature used to analyze the research results. A detailed description of each feature and the datasets, and methods used in computing the features are presented below.

Median income. Our socio-demographic data utilizes median income, which represents the household median income in 2017 as reported by the US Census Bureau table data. Median income serves as a critical indicator of a region's vulnerability and resilience to hazards, directly influencing a community's capacity to withstand and recover from adverse events[75,76]. This economic metric reflects both the level of development and economic inequality within communities, significantly affecting how populations prepare for, respond to, and recover from disaster impacts[77].

Preparedness proactivity. Method of preparedness proactivity is adopted from Li and Mostafavi[31] to capture the earliest maximum POI change percentage (Eq. 1) at CBG level using location-based data from Spectus[78]. Spectus acquires its data through partnerships with smartphone applications, collecting information from devices where users have consented to the collection of their location data. Multiple independent studies have found that Spectus mobile location data is broadly representative of population mobility patterns. The spatial distribution of Cuebiq's opted-in users strongly correlates with actual census-based population distributions, with correlations around 0.7 at the zip code level[79], and 0.91 at the census tract level[80] and exceeding 0.93 at the county level[81]. Its demographic representativeness is further supported by a ~0.8 correlation with local census populations and matching age and household size distributions[82]. Mobile phone data offer detailed insights into human movement patterns on a large scale at an unprecedented spatio-temporal granularity and scale[83]. By collaborating with app developers, Spectus harnesses a variety of signals, including Bluetooth, GPS, Wi-Fi, and IoT, to compile a high-resolution geo-behavioral dataset.

In this study, we defined the preparedness period as August 20 through August 25, 2017. This five-day timeframe was selected to capture the critical preparation activities that occurred immediately before Hurricane Harvey's landfall, reflecting the typically short duration of the acute preparedness phase in disaster management. To assess preparedness proactivity, we

Table 1 | Coupled human-infrastructure features and datasets

Feature group	Feature	Data description	Data sources
Socio-demographic characteristic	Median income	Household median income in 2017	U.S. Census Bureau table data[99]
Protective action features	Preparedness proactivity	Maximum POI visits during preparedness period relative to normal baseline in pharmacy	Safegraph, Inc; Spectus; Microsoft building footprint[78,90,94]
		Maximum POI visits during preparedness period relative to normal baseline in gas station	Safegraph, Inc; Spectus; Microsoft building footprint[78,90,94]
	Evacuation rate	Maximum evacuation rate relative to normal baseline at CBG level	Spectus[78]
Infrastructure and building disruptions	Flooded roads	Total distance that has null traffic speed	INRIX[100]
	Number of flood claims	The total count of flood claims within each census block group area	Federal Emergency Management Agency National Flood Insurance Program[89]
	Total building damage amount	The actual cash value of the damage that the main property sustained is shown in terms of whole dollars.	Federal Emergency Management Agency National Flood Insurance Program[89]
	Building Damage Ratio	The building damage ratio was determined by dividing the total number of claims by the total number of buildings in a given area.	Federal Emergency Management Agency National Flood Insurance Program, and Microsoft Building Footprints[89,90]
	Telecommunication data	Maximum download kbps changing rate during hazard relative to normal baseline	Ookla[91]
Recovery features	Essential activity recovery	Essential human mobility recovery data (essential POI)	Spectus[78]
	Non-essential Activity Recovery	Non-essential human mobility recovery data (non-essential POI)	Spectus[78]
	Essential credit card transaction recovery	Essential credit card activity	Safegraph[94]
	Non-essential credit card transaction recovery	Non-essential credit card transactions recovery	Safegraph[94]

identified the peak visitation percentage change date for each CBG relative to the baseline and then determined the interval between this peak and the hurricane's landfall. Greater preparedness and proactivity mean people in the community prepare earlier for the natural hazard. Zero proactivity indicates people reach maximum POI change percentage on the day of the hurricane landfall. To determine the changes in POI visits due to hurricane preparations, we established baselines using POI visitation patterns related to the first two weeks in August. This period's visitation numbers capture the residents' normal POI visits with no disturbances. The preparedness rate is calculated based on visitations to pharmacies and gas stations, determined using datasets of POIs visits. Equation 1 describes the percent change of the POI visits in calculating the preparedness index of each CBG:

$$PC_{i,d,t} = \frac{V_{i,d,t} - B_{i,d,t}}{B_{i,d,t}} \quad (1)$$

where, $PC_{i,d,t}$ is the percentage change of visits to one category of POI (t) from home CBG (i) in date (d), $V_{i,d,t}$ is the number of visits to one category of POI (t) from home CBG (i) on date (d). $B_{i,d,t}$ is the calculated baseline value corresponding to the date. The maximum POI visitation change percentage of certain POI is computed, and the date is recorded. The preparedness proactivity is calculated based on the difference in days between the recorded date and August 25, 2017. For example, if an area reached its maximum POI visitation change percentage on August 24, 2017, its preparedness proactivity is one day. After compiling all preparedness proactivity measures related to gas stations and pharmacies, we determine their mean value as the overall preparedness proactivity for a CBG.

Evacuation rate. The evacuation rate was calculated for each CBG by comparing the rates before and after Hurricane Harvey (August 20 through September 10, 2017) against a pre-event baseline period (July 9 through August 5, 2017), and the maximum evacuation rate was identified. This extended timeframe for evacuation data was chosen based on

findings from the Texas A&M Hazard Reduction & Recovery Center survey, which documented ongoing evacuation activities through late August[84]. The extended timeframe captures both immediate and delayed evacuation responses. Then the study chose the maximum percentage change for each CBG as our feature value, following the methodology of previous study[38], and using location-based data from Spectus. Only users with a minimum of 240 min of daily location information were included to ensure data accuracy and minimize bias. Individuals who left their home CBGs and CBGs and stayed in another CBG for at least 24 h during Hurricane Harvey were considered to have evacuated. Here Eq. 2 describes the percent change of the evacuation rate (ER):

$$ER\ Change\ rate = \frac{ER_t - ER_B}{ER_B} \quad (2)$$

ER_t represents the evacuation rate on day t, and ER_B is the baseline. The maximum evacuation rate change percentage is the feature for evacuation in this study.

Flooded roads. Flooded road networks limit access to emergency services[59,85,86]. Harris County Road segment traffic data from August 20 through September 11, 2017, obtained from INRIX, provided average traffic speeds for each road segment at 5-minute intervals, along with historical average speeds for comparison. Following the approaches[87,88], road segments displaying a null value for average traffic speed were categorized as flooded during Hurricane Harvey. We represented flooded roads as Line String geometries based on their recorded start and end coordinates. To assess the impact at the community level, we conducted spatial intersections between flooded road segments and GEOID-defined boundaries, allowing us to quantify the total length of flooded roads within each CBG area. These lengths represent the total null distance in each CBG polygon, indicating the extent of flooded roads within those areas.

Number of flood claims. The insurance claim data utilized in this study was sourced from the FEMA National Flood Insurance Program[89] (NFIP) (FEMA, 2023). FEMA manages the NFIP by overseeing the proper processing of insurance applications, determining accurate insurance premiums, and handling the renewal, modification, and cancellation of insurance policies. This data set provides details on NFIP claims transactions, and the number of claims in a given area was determined by calculating the total count of insurance claims filed by policyholders in each CBG.

Total building damage amount. The total building damage amount, sourced directly from the National Flood Insurance Program[89], represents the actual cash value of damage incurred by the main property, expressed in whole dollars. We aggregated this data to the total building damage amount for each CBG area by summing all the reported damage amounts.

Building damage ratio. The building damage ratio was determined by dividing the total number of claims[89] by the total number of buildings in a given area. The number of claims has been mentioned before, and the data on the total number of buildings was from Microsoft Building Footprints. The Bing Maps and Microsoft Maps and Geospatial teams have extracted a comprehensive dataset consisting of 129,591,852 building footprints across the United States[90].

Telecommunication disruptions. Ookla data encompasses to cellular internet speeds, including specific metrics like upload and download rates from mobile devices, latency, and location-based information about both the device and the server[91]. This research focused on using the download kbps changing rate as the telecommunication data from August 10 to September 8, 2017. The download kbps changing rate was assumed to be the maximum changing rate during this period. This study leveraged data provided by Ookla, which offers comprehensive insights into internet capabilities, including cellular internet speeds. This dataset includes vital metrics such as upload and download rates from mobile devices, latency, and geolocation information about the devices and servers involved[51].

In order to record the maximum telecommunication disruptions, this study concentrates on the maximum changing rate of download speeds in kilobits per second (kbps) from August 10 through September 8, 2017. To establish a baseline for normal telecommunication operations, we calculated the average download speed during the first two weeks of August. This period is chosen to reflect a typical pattern of telecommunication use without the influence of an impending natural disaster. The measure of the maximum telecommunication disruption is then determined by Eq. (3).

$$Maximum\ Download\ Speed\ Change\ Rate = \frac{Speed_B - \min(Speed_t)}{Speed_B}$$

(3)

where $\min(Speed_t)$ represents the minimum download speed on day t from August 24th and September 3rd, 2017, and $Speed_B$ is the baseline.

Essential and non-essential activity recovery and credit card transaction recovery. To create smart, resilient, and sustainable urban environments, urban planners need an understanding of how extreme weather events affect human activities[92,93]. The dataset comprising points of interest visitation patterns was obtained from Spectus[78]; the dataset related to credit card transactions was sourced from SafeGraph[94]. In line with methodologies used in prior research[57,58,95], this study uses essential activity recovery, non-essential activity recovery, essential credit card transaction recovery, and non-essential credit card transaction recovery. Essential activities represent services that are critical for daily life and survival, such as healthcare, grocery shopping, gas stations, and utilities. In contrast, non-essential activities, including retail, restaurants, entertainment, and self-care, contribute to enhancing quality of life but are not

immediately critical for survival. We utilized data from the period before landfall (pre-August 25, 2017), the hurricane impact phase (August 25–31, 2017), and the immediate recovery period (September 1–30, 2017) to assess post-disaster recovery. To establish the baseline, POI visits during the three weeks leading up to Hurricane Harvey (August 1–August 21, 2017) were analyzed. Recovery weeks were calculated using a seven-day moving average of daily visits to various POIs, with this baseline serving as a reference for comparing pre- and post-disaster visitation patterns. In Eq. 4, the daily value refers to the actual observed visits to POIs on a given day, while the base value acts as the reference point for the normal activity period before the disaster. The change percentage is then calculated to reflect variations in the daily visit numbers for each category, relative to the baseline, using a seven-day rolling average to smooth the data over seven-day intervals:

$$7 - dayRolling_{avg} = \sum_{d=1}^{7} \frac{Daily}{} value - Base\ valueBase\ Value100\%$$

(4)

Resilience curves for POIs were derived by plotting seven-day rolling averages. In this study, the term duration of recovery and is defined as the time at which the seven-day rolling average of visits to POIs attains 90% of the baseline values established prior to the disruption. This calculation enabled us to determine the average time needed for each area to achieve recovery.

Statistical methods

This study combines index-based and clustering methods by applying hierarchical clustering to the index values obtained from the first PCA. Principal Component Analysis (PCA) approach allows to reduce a set of features into two composite indices—a Risk Index and a Resilience Index—based on a set of features related to risk and resilience ability. The Risk Index encompasses features including Evacuation Rate, Flooded Roads, Preparedness Proactivity, Number of Flood Claims, Damage Building Ratio, Total Building Damage Amount, and Telecommunication Disruptions. These features represent multiple dimensions of risk exposure and vulnerability within a region. The Resilience Index, in contrast, focuses on features that reflect the speed of recovery, calculated as the reciprocal of recovery days for various activities. This index includes Essential Activity Recovery, Non-essential Activity Recovery, Essential Credit Card Transaction Recovery, and Non-essential Credit Card Transaction Recovery.

Principal Component Analysis (PCA) was applied to the scaled risk and resilience features to extract a single index. Each index value was computed as a linear combination of the original features, where the weights correspond to the eigenvector associated with the largest eigenvalue of the feature covariance matrix. Mathematically, the first principal component is given by:

$$PC_1 = X\omega_1$$

(5)

where PC_1 represents the first principal component, and X is the matrix of standardized features, and ω_1 is the eigenvector associated with the first largest eigenvalue. For each of the two indices, we selected only the first principal component, as it explains the largest proportion of variance in the data. The Risk Index and Resilience Index are then defined as:

$$Risk\ Index = X_{risk}\omega_{risk}$$

(6)

$$Resilience\ Index = X_{resilience}\omega_{resilience}$$

(7)

where X_{risk} and $X_{resilience}$ represent the standardized feature matrices for the selected risk and resilience features, respectively, and ω_{risk} and $\omega_{resilience}$ are the first eigenvectors obtained from PCA on each feature set.

Hierarchical clustering[96,97] is to identify patterns of risk and resilience index across regions, which is a technique that builds a multilevel hierarchy of clusters by either merging or splitting clusters iteratively.

Specifically, this study used the Agglomerative Clustering method, which is a "bottom-up" approach. Each data point starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Using the Ward linkage method, the agglomerative clustering merges clusters that minimize the within-cluster variance. This is expressed by calculating the variance criterion for each potential merge and choosing the pair of clusters that minimizes the increase in total within-cluster variance. The distance between clusters A and B in Ward's method can be formalized as:

$$D(A, B) = \frac{|A||B|}{|A| + |B|} ||\bar{x}_A - \bar{x}_B||^2 \qquad (8)$$

where $|A|$ and $|B|$ are the sizes of clusters A and B, respectively, and $\bar{x}_A$ and $\bar{x}_B$ are the centroids of clusters A and B. This criterion tends to produce compact clusters with low variance, which aligns with the goal of identifying regions with distinct risk and resilience profiles.

In addition, this study also conducted Ordinary Least Squares (OLS) regression analysis[98], and the regression equation can be generally represented as follows:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i \qquad (9)$$

where y_i is the dependent variable for observation i, $X_{i1}, X_{i2}, \ldots, X_{ip}$ are the independent variables (predictors) for observation i, and ε is the error term. The coefficient $\beta_1, \beta_2, \ldots, \beta_p$ are the regression coefficients corresponding to each predictor.

Data availability
The data that support the findings of this study are available from Ookla, Safegraph, INRIX, Microsoft Building Footprint, and Spectus, Inc., but restrictions apply to the availability of data from Spectus and INRIX, which were used under license for the current study. Other data we use in this study are all publicly available.

Code availability
The code that supports the findings of this study is available from the corresponding author upon request.

References
1. Birkholz, S., Muro, M., Jeffrey, P. & Smith, H. M. Rethinking the relationship between flood risk perception and flood management. *Sci. Total Environ.* **478**, 12–20 (2014).
2. Ekmekcioğlu, Ö., Koc, K. & Özger, M. Towards flood risk mapping based on multi-tiered decision making in a densely urbanized metropolitan city of Istanbul. *Sustain. Cities Soc.* **80**, 103759 (2022).
3. Plate, E. J. Flood risk and flood management. *J. Hydrol.* **267**, 2–11 (2002).
4. Assaf, G. & Assaad, R. H. Optimal preventive maintenance, repair, and replacement program for catch basins to reduce urban flooding: integrating agent-based modeling and monte carlo simulation. *Sustainability* **15**, 8527 (2023).
5. Ansari, A. H., Mejia, A. & Cibin, R. Flood teleconnections from levees undermine disaster resilience. *Npj Nat. Hazards* **1**, 2 (2024).
6. Savari, M., Jafari, A. & Sheheytavi, A. Determining factors affecting flood risk perception among local communities in Iran. *Sci. Rep.* **15**, 4076 (2025).
7. Borowska-Stefańska, M. et al. Emergency management of self-evacuation from flood hazard areas in Poland. *Transp. Res. Part Transp. Environ.* **107**, 103307 (2022).
8. Sayers, P. et al. *Flood Risk Management: A Strategic Approach*. (2013).
9. Rasoulkhani, K. & Mostafavi, A. Resilience as an emergent property of human-infrastructure dynamics: a multi-agent simulation model for characterizing regime shifts and tipping point behaviors in infrastructure systems. *PLOS ONE* **13**, e0207674 (2018).
10. Molinari, D. et al. Ex post damage assessment: an Italian experience. *Nat. Hazards Earth Syst. Sci.* **14**, 901–916 (2014).
11. Kameshwar, S. et al. Probabilistic decision-support framework for community resilience: Incorporating multi-hazards, infrastructure interdependencies, and resilience goals in a Bayesian network. *Reliab. Eng. Syst. Saf.* **191**, 106568 (2019).
12. Beguería, S. Validation and evaluation of predictive models in hazard assessment and risk management. *Nat. Hazards* **37**, 315–329 (2006).
13. Wang, Z. et al. Flood hazard risk assessment model based on random forest. *J. Hydrol.* **527**, 1130–1141 (2015).
14. Yuan, F. et al. Smart flood resilience: harnessing community-scale big data for predictive flood risk monitoring, rapid impact assessment, and situational awareness. *Environ. Res. Infrastruct. Sustain.* **2**, 025006 (2022).
15. Brown, D. et al. Indicators for Measuring, Monitoring, And Evaluating Post-disaster Recovery. In *Proc. International Workshop on Remote Sensing for Disaster Applications* (Pavia, 2008).
16. Martín, Y., Cutter, S. L. & Li, Z. Bridging Twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma. *Nat. Hazards Rev.* **21**, 04020003 (2020).
17. Vamvakeridou-Lyroudia, L. S. et al. Assessing and visualising hazard impacts to enhance the resilience of Critical Infrastructures to urban flooding. *Sci. Total Environ.* **707**, 136078 (2020).
18. Chen, A. S. et al. From hazard to impact: flood damage assessment tools for megacities. *Nat. Hazards* **82**, 857–890 (2016).
19. Hofmeister, M. et al. Cross-domain flood risk assessment for smart cities using dynamic knowledge graphs. *Sustain. Cities Soc.* **101**, 105113 (2024).
20. Pant, R., Thacker, S., Hall, J. W., Alderson, D. & Barr, S. Critical infrastructure impact assessment due to flood exposure. *J. Flood Risk Manag.* **11**, 22–33 (2018).
21. Kasmalkar, I. G. et al. When floods hit the road: resilience to flood-related traffic disruption in the San Francisco Bay Area and beyond. *Sci. Adv.* **6**, eaba2423 (2020).
22. Najafi, M. R., Zhang, Y. & Martyn, N. A flood risk assessment framework for interdependent infrastructure systems in coastal environments. *Sustain. Cities Soc.* **64**, 102516 (2021).
23. Han, F., Yu, J., Zhou, G., Li, S. & Sun, T. Projected urban flood risk assessment under climate change and urbanization based on an optimized multi-scale geographically weighted regression. *Sustain. Cities Soc.* **112**, 105642 (2024).
24. Zheng, Q., Shen, S.-L., Zhou, A. & Lyu, H.-M. Inundation risk assessment based on G-DEMATEL-AHP and its application to Zhengzhou flooding disaster. *Sustain. Cities Soc.* **86**, 104138 (2022).
25. Gliksman, D. et al. Review article: a European perspective on wind and storm damage – from the meteorological background to index-based approaches to assess impacts. *Nat. Hazards Earth Syst. Sci.* **23**, 2171–2201 (2023).
26. Satta, A., Snoussi, M., Puddu, M., Flayou, L. & Hout, R. An index-based method to assess risks of climate-related hazards in coastal zones: the case of Tetouan. *Estuar. Coast. Shelf Sci.* **175**, 93–105 (2016).
27. Ma, X., Hao, Y., Li, X., Liu, J. & Qi, J. Evaluating global intelligence innovation: an index based on machine learning methods. *Technol. Forecast. Soc. Change* **194**, 122736 (2023).
28. Esmalian, A., Coleman, N. & Yuan, F. et al. Characterizing equitable access to grocery stores during disasters using location-based data. *Sci Rep* **12**, 20203 (2022).
29. Cutter, S. L. et al. A place-based model for understanding community resilience to natural disasters. *Glob. Environ. Change* **18**, 598–606 (2008).

30. Qin, H., Liang, Q., Chen, H. & De Silva, V. A high-performance coupled human and natural systems (CHANS) model for flood risk assessment and reduction. *Water Resour. Res.* **60**, e2023WR036269 (2024).

31. Li, B. & Mostafavi, A. Location intelligence reveals the extent, timing, and spatial variation of hurricane preparedness. *Sci. Rep.* **12**, 16121 (2022).

32. Dong, S., Wang, H., Mostafavi, A. & Gao, J. Robust component: a robustness measure that incorporates access to critical facilities under disruptions. *J. R. Soc. Interface* **16**, 20190149 (2019).

33. Bullock, J., George, H. & Damon, C. *Introduction to Emergency Management* (Butterworth-Heinemann, 2017).

34. Lindell, M., Perry, R., Prater, C. & Nicholson, W. *Fundamentals of Emergency Management* (FEMA, 2006).

35. Bronfman, N. C., Cisternas, P. C., Repetto, P. B. & Castañeda, J. V. Natural disaster preparedness in a multi-hazard environment: Characterizing the sociodemographic profile of those better (worse) prepared. *PLoS ONE* **14**, e0214249 (2019).

36. Dargin, J. S., Li, Q., Jawer, G., Xiao, X. & Mostafavi, A. Compound hazards: an examination of how hurricane protective actions could increase transmission risk of COVID-19. *Int. J. Disaster Risk Reduct.* **65**, 102560 (2021).

37. Cova, T. J., Drews, F. A., Siebeneck, L. K. & Musters, A. Protective actions in wildfires: evacuate or shelter-in-place? *Nat. Hazards Rev.* **10**, 151–162 (2009).

38. Lee, C.-C., Chou, C. & Mostafavi, A. Specifying evacuation return and home-switch stability during short-term disaster recovery using location-based data. *Sci. Rep.* **12**, 15987 (2022).

39. Sadiq, A.-A., Tyler, J. & Noonan, D. S. A review of community flood risk management studies in the United States. *Int. J. Disaster Risk Reduct.* **41**, 101327 (2019).

40. Chakraborty, J., Tobin, G. A. & Montz, B. E. Population evacuation: assessing spatial variability in geophysical risk and social vulnerability to natural hazards. *Nat. Hazards Rev.* **6**, 23–33 (2005).

41. Opdyke, A., Bodo, D. D. & Smyth, J. Higher ground or into harm's way? Household storm surge sheltering and evacuation plans. *Int. J. Disaster Risk Reduct.* **106**, 104452 (2024).

42. Hellman, J. Living with floods and coping with vulnerability. *Disaster Prev. Manag. Int. J.* **24**, 468–483 (2015).

43. Chen, X., Kwan, M.-P., Li, Q. & Chen, J. A model for evacuation risk assessment with consideration of pre- and post-disaster factors. *Comput. Environ. Urban Syst.* **36**, 207–217 (2012).

44. Renne, J. L. Emergency evacuation planning policy for carless and vulnerable populations in the United States and United Kingdom. *Int. J. Disaster Risk Reduct.* **31**, 1254–1261 (2018).

45. FEMA. *National Disaster Recovery Framework*. https://www.fema. gov/pdf/recoveryframework/ndrf.pdf (2011).

46. Bruneau, M. et al. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthq. Spectra* **19**, 733–752 (2003).

47. Rosenheim, N., Guidotti, R., Gardoni, P. & Peacock, W. G. Integration of detailed household and housing unit characteristic data with critical infrastructure for post-hazard resilience modeling. *Sustain. Resilient Infrastruct.* **6**, 385–401 (2021).

48. Rajput, A. A., Nayak, S., Dong, S. & Mostafavi, A. Anatomy of perturbed traffic networks during urban flooding. *Sustain. Cities Soc* **97**, 104693 (2023).

49. Coleman, N., Esmalian, A. & Mostafavi, A. Anatomy of susceptibility for shelter-in-place households facing infrastructure service disruptions caused by natural hazards. *Int. J. Disaster Risk Reduct.* **50**, 101875 (2020).

50. Singh, P., Sinha, V. S. P., Vijhani, A. & Pahuja, N. Vulnerability assessment of urban road network from urban flood. *Int. J. Disaster Risk Reduct.* **28**, 237–250 (2018).

51. Gupta, Y., Liu, Z. & Mostafavi, A. *Digital Divide in Disasters: Investigating Spatial and Socioeconomic Disparities in Internet Service Disruptions During Extreme Weather Events*. https://www. ssrn.com/abstract=4672649, https://doi.org/10.2139/ssrn. 4672649 (2023).

52. Yabe, T., Tsubouchi, K., Fujiwara, N., Sekimoto, Y. & Ukkusuri, S. V. Understanding post-disaster population recovery patterns. *J. R. Soc. Interface* **17**, 20190532 (2020).

53. Liu, J. et al. Complexity of coupled human and natural systems. *Science* **317**, 1513–1516 (2007).

54. Donner, W. & Rodríguez, H. Population composition, migration and inequality: the influence of demographic changes on disaster risk and vulnerability. *Soc. Forces* **87**, 1089–1114 (2008).

55. Yabe, T., Rao, P. S. C. & Ukkusuri, S. V. Resilience of interdependent urban socio-physical systems using large-scale mobility data: modeling recovery dynamics. *Sustain. Cities Soc.* **75**, 103237 (2021).

56. Sun, W., Bocchini, P. & Davison, B. D. Applications of artificial intelligence for disaster management. *Nat. Hazards* **103**, 2631–2689 (2020).

57. Jiang, Y. et al. Data-driven tracking of the bounce-back path after disasters: critical milestones of population activity recovery and their spatial inequality. *Int. J. Disaster Risk Reduct* **92**, 103693 (2023).

58. Podesta, C., Coleman, N., Esmalian, A., Yuan, F. & Mostafavi, A. Quantifying community resilience based on fluctuations in visits to points-of-interest derived from digital trace data. *J. R. Soc. Interface* **18**, 20210158 (2021).

59. Yuan, F., Xu, Y., Li, Q. & Mostafavi, A. Spatio-temporal graph convolutional networks for road network inundation status prediction during urban flooding. *Comput. Environ. Urban Syst.* **97**, 101870 (2022).

60. Praharaj, S., Chen, T. D., Zahura, F. T., Behl, M. & Goodall, J. L. Estimating impacts of recurring flooding on roadway networks: a Norfolk, Virginia case study. *Nat. Hazards* **107**, 2363–2387 (2021).

61. Guan, Q. et al. Hyperspectral technique for detection of peanut leaf spot disease based on improved PCA loading. *Agronomy* **13**, 1153 (2023).

62. De Winter, J. C. F. & Dodou, D. Common factor analysis versus principal component analysis: a comparison of loadings by means of simulations. *Commun. Stat. Simul. Comput.* **45**, 299–321 (2016).

63. OpenStreetMap contributors. Planet dump retrieved from https:// planet.osm.org (2024).

64. Wang, Y., Zhang, C., Chen, A. S., Wang, G. & Fu, G. Exploring the relationship between urban flood risk and resilience at a high-resolution grid cell scale. *Sci. Total Environ.* **893**, 164852 (2023).

65. Lee, E. S. Assessing climate vulnerability for resilient urban planning: a multidiagnosis approach. *Sens. Mater.* **35**, 3479 (2023).

66. Ermagun, A., Smith, V. & Janatabadi, F. High urban flood risk and no shelter access disproportionally impacts vulnerable communities in the USA. *Commun. Earth Environ.* **5**, 2 (2024).

67. Sharma, A. P., Fu, X. & Kattel, G. R. Shannon entropy-based quantitative method for measuring Risk-integrated Resilience (RiR) index on flood disaster in West Rapti basin of Nepal Himalaya. *Nat. Hazards* **120**, 477–510 (2024).

68. Keeley, J. E. & Syphard, A. D. Twenty-first century California, USA, wildfires: fuel-dominated vs. wind-dominated fires. *Fire Ecol.* **15**, 24 (2019).

69. Mirbabaie, M., Bunker, D., Stieglitz, S., Marx, J. & Ehnis, C. Social media in times of crisis: learning from Hurricane Harvey for the coronavirus disease 2019 pandemic response. *J. Inf. Technol.* **35**, 195–213 (2020).

70. Doubleday, A., Choe, Y., Miles, S. & Errett, N. A. Daily bicycle and pedestrian activity as an indicator of disaster recovery: a hurricane harvey case study. *Int. J. Environ. Res. Public. Health* **16**, 2836 (2019).

71. Emanuel, K. Assessing the present and future probability of Hurricane Harvey's rainfall. *Proc. Natl. Acad. Sci. USA* **114**, 12681–12684 (2017).

72. Blake, E. & Zelinsky, D. *National Hurricane Center Tropical Cyclone Report* (2018).
73. Qin, R., Khakzad, N. & Zhu, J. An overview of the impact of Hurricane Harvey on chemical and process facilities in Texas. *Int. J. Disaster Risk Reduct.* **45**, 101453 (2020).
74. Kiaghadi, A. & Rifai, H. S. Physical, chemical, and microbial quality of floodwaters in Houston following Hurricane Harvey. *Environ. Sci. Technol.* **53**, 4832–4840 (2019).
75. Greenberg, M. R. & Schneider, D. Income disparities and risk: geographical manifestations of extreme inequities in the United States. *Risk Anal*. Risa.14349 https://doi.org/10.1111/risa.14349 (2024).
76. Modica, M. & Zoboli, R. Vulnerability, resilience, hazard, risk, damage, and loss: a socio-ecological framework for natural disaster analysis. *Web Ecol.* **16**, 59–62 (2016).
77. Noy, I. & Yonson, R. Economic vulnerability and resilience to natural hazards: a survey of concepts and measurements. *Sustainability* **10**, 2850 (2018).
78. Spectus. Spectus mobility data. (2023).
79. Thommes, E. W. et al. U.S. state-level COVID-19 transmission insights from a mechanistic mobility-incidence model. Preprint at https://doi.org/10.1101/2022.06.21.22276712 (2022).
80. Ban, X. (Jeff), Chen, C., Wang, F., Wang, J. & Zhang, Y. Promises of data from emerging technologies for transportation applications: puget sound region case study. https://doi.org/10.13140/RG.2.2.20005.55522 (2018).
81. Deng, H., Du, J., Gao, J. & Wang, Q. Network percolation reveals adaptive bridges of the mobility network response to COVID-19. *PLoS ONE* **16**, e0258868 (2021).
82. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav*. **4**, 964–971 (2020).
83. Yabe, T., Jones, N. K. W., Rao, P. S. C., Gonzalez, M. C. & Ukkusuri, S. V. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* **94**, 101777 (2022).
84. Bierling, D. H. et al. Coastal bend hurricane evacuation study: hurricane harvey evacuation behavior survey outcomes and findings. https://hdl.handle.net/1969.1/188203.
85. Dong, S., Esmalian, A., Farahmand, H. & Mostafavi, A. An integrated physical-social analysis of disrupted access to critical facilities and community service-loss tolerance in urban flooding. *Comput. Environ. Urban Syst.* **80**, 101443 (2020).
86. Bhavathrathan, B. K. & Patil, G. R. Capacity uncertainty on urban road networks: a critical state and its applicability in resilience quantification. *Comput. Environ. Urban Syst.* **54**, 108–118 (2015).
87. Fan, C., Jiang, X. & Mostafavi, A. A network percolation-based contagion model of flood propagation and recession in urban road networks. *Sci. Rep.* **10**, 13481 (2020).
88. Yuan, F., Yang, Y., Li, Q. & Mostafavi, A. Unraveling the temporal importance of community-scale human activity features for rapid assessment of flood impacts. *IEEE Access* **10**, 1138–1150 (2022).
89. FEMA. National Flood Insurance Program (NFIP) redacted claims dataset (Version 2). (2023).
90. Microsoft. Microsoft Building Footprints dataset. (2023).
91. Ookla. Ookla speed test and network performance data. (2023).
92. Chen, Z., Gong, Z., Yang, S., Ma, Q. & Kan, C. Impact of extreme weather events on urban human flow: a perspective from location-based service data. *Comput. Environ. Urban Syst.* **83**, 101520 (2020).
93. Zhang, W. et al. Structural changes in intercity mobility networks of China during the COVID-19 outbreak: a weighted stochastic block modeling analysis. *Comput. Environ. Urban Syst.* **96**, 101846 (2022).
94. SafeGraph, Inc. SafeGraph Places data. (2023).
95. Li, X., Jiang, Y. & Mostafavi, A. Hazard exposure heterophily in socio-spatial networks contributes to post-disaster recovery in low-income populations. *Int. J. Disaster Risk Reduct.* **108**, 104537 (2024).
96. Murtagh, F. & Contreras, P. Algorithms for hierarchical clustering: an overview. *WIREs Data Min. Knowl. Discov.* **2**, 86–97 (2012).
97. Nielsen, F. Hierarchical Clustering. in *Introduction to HPC with MPI for Data Science* 195–211 (Springer International Publishing, 2016). https://doi.org/10.1007/978-3-319-21903-5_8.
98. National Hurricane Center. Tropical Cyclone Report: Hurricane Harvey (AL092017). National Oceanicand Atmospheric Administration (2018). https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf.
99. U.S. Census Bureau. American Community Survey 5-Year Estimates, Table B19013: Median Household Income in the Past 12 Months (in 2017 Inflation-Adjusted Dollars). (2017).
100. INRIX. INRIX traffic and mobility data (2023).

Author contributions
Both authors contributed to the final manuscript and approved its submission. X.L. is the first author and was responsible for data collection, conducting the primary analysis, interpreting key findings, and drafting the manuscript. A.M. is the faculty advisor, guiding the project's conceptual development, overseeing manuscript revisions, and providing critical feedback on the research.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Xiangpeng Li.

Reprints and permissions 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.