Sustainable Cities and Society 138 (2026) 107182

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Infrastructure access and availability as determinants of community vulnerability: A spatial analysis of 733 districts in India

Srijith Balakrishnan [a],*, Shivam Srivastava [b], Chirag Kothari [b]

[a] Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, Netherlands
[b] Department of Civil Engineering, Indian Institute of Technology Kanpur, G67J+FG5, Nankari, Kalyanpur, Kanpur, Uttar Pradesh 208016, India

ARTICLE INFO

Dataset link: https://github.com/srijithbalakrishnan/india_resilience

Keywords:
Community resilience
Spatial modeling
Infrastructure planning
Equitable development
India
Global South

ABSTRACT

As natural disasters increase in frequency and severity, infrastructure planning must account for how existing deficits and disparities shape community vulnerability and recovery. While regional vulnerability and resilience frameworks consider infrastructure systems as core determinants, the interaction between infrastructure characteristics, such as access and availability, and their influence on community vulnerability remain underexplored, particularly in the Global South. To this end, we leverage open-source geospatial data to construct granular datasets for 733 districts in India and apply statistical methods to assess relationships between community and infrastructure-related characteristics. Specifically, we consider three dimensions of infrastructure at the district level: (a) regional critical infrastructure availability, (b) social infrastructure density, and (c) household-level essential utility access. Geographic distributions and distinct profiles of infrastructure characteristics, are identified by employing clustering algorithms on composite indicators, while spatial regression models evaluate the association between community vulnerability and infrastructure dimensions. Findings suggest that access to essential utilities is the strongest factor associated with reduced vulnerability, with a significant interaction between critical infrastructure availability and utility access, indicating potential synergistic effects. Regions with below-average infrastructure provisions are associated with higher community vulnerability and may therefore warrant greater resource investment to achieve equitable vulnerability reduction outcomes. This study provides empirical evidence of synergistic effects using district-level geospatial analysis, supporting coordinated infrastructure development at network and household levels to enhance community capabilities and resilience.

1. Introduction

As climate extremes intensify, infrastructure planners in the Global South face compounding developmental challenges: prevailing infrastructure deficits expose communities to disaster risks, while heightened community vulnerability to disasters threatens to reverse long-term socioeconomic progress (Zhang et al., 2025). A recent report by the CDRI (2023) pegged global annual losses to infrastructure from natural hazards and climate change at $732–845 billion, with low- and middle-income countries (LMICs) bearing 54% of the risk (totaling $397 billion) despite holding only 32.7% of exposed asset value. Well-designed infrastructure systems are critical for boosting productivity and reducing poverty, yielding substantial economic returns (Calderón & Servén, 2010). However, the benefits of infrastructure are not always equitably distributed, and poorly planned projects can increase social inequalities (Suharto et al., 2025) and displace communities (Aboda et al., 2019), altering the regional community vulnerability profiles. In the Global South, where infrastructure deficits are pronounced, strategic investments are essential for achieving sustainable development; yet challenges such as limited funding and governance constraints persist (Mathur, 2024). A study by Asian Infrastructure Investment Bank (Han et al., 2019) estimated that developing countries should invest 6%–10% of gross domestic product (GDP) on infrastructure development to bridge the existing deficit alone. Among the major countries in the Global South, India, with its diverse geography, rapid economic growth, and regional socioeconomic disparities, exemplifies these opportunities and challenges. For instance, large-scale infrastructure projects, such as the national highway development initiative have enhanced connectivity and spurred economic activity (Tripathy et al., 2016); however, they often lead to displacement and unequal access across regions (Yadav & Kalambe, 2022), undermining community resilience and exacerbating existing societal vulnerabilities. Therefore, empirical insights on infrastructure development patterns and

* Corresponding author.
E-mail addresses: s.balakrishnan@tudelft.nl (S. Balakrishnan), shvmsri@iitk.ac.in (S. Srivastava), ckothari@iitk.ac.in (C. Kothari).

https://doi.org/10.1016/j.scs.2026.107182
Received 1 September 2025; Received in revised form 7 November 2025; Accepted 23 January 2026
Available online 5 February 2026

their associations with community vulnerability is crucial for informing infrastructure policy and resource allocation decisions.

1.1. Defining infrastructure availability and access

When examining how infrastructure systems influence community vulnerability, three key considerations emerge:

1. Availability: Does the community have functional infrastructure networks and services?
2. Access: Can households within the community physically or virtually access these services?
3. Reliability: For those households with access, are the services reliable and of sufficient quality?

Availability, which we define as the presence of functional infrastructure, such as roads, hospitals, and communication networks, is essential because it supports a community's resilience by enabling critical services and connectivity during both everyday life and emergencies (Henshel & Ashby, 2023). However, mere presence of infrastructure networks and services in a region does not guarantee that all constituent households benefit from them and household access to such facilities can ensure individuals to cope with disasters (McGlade et al., 2019; Smit et al., 2017). Access, referred to as the ability of households to use these services, can be limited by financial barriers, sociodemographic composition, cultural factors, among others. For instance, Corscadden et al. (2018) found that income, sex, and age were the primary determinants of access to healthcare services, based on data from 11 developed countries. While service quality characteristics, such as reliability, are also vital, since inconsistent services like frequent power outages or an unreliable water supply can undermine household resilience (Guidotti et al., 2019; Verschuur & Balakrishnan, 2024), these aspects are not considered in this study.

In addition to the availability of and access to critical infrastructure and services, the spatial density of social infrastructure, such as schools and hospitals, also plays a key role in shaping community vulnerability (Chen et al., 2020). Social infrastructure density, referring to the concentration of facilities, directly affects the access and availability of public services and amenities. Higher density of social infrastructure elements can improve access and support well-being, but often comes with trade-offs. Again, a major assumption here is that all community members have access; however, whether and how they make use of the services is not considered in this study. Prioritizing critical infrastructure (e.g., transport, energy, water) may enhance economic resilience (Balakrishnan et al., 2022) while neglecting social needs, whereas focusing too much on social infrastructure may reduce capacity to handle physical disruptions (Hassan & Mahmoud, 2021). Balancing both is essential for effective, context-specific infrastructure planning. However, unless empirical evidence is available for the individual and combined effects of infrastructure investments, planning based on theoretical knowledge can result in suboptimal outcomes.

1.2. Infrastructure characteristics and community vulnerability

Infrastructure systems are pivotal in shaping community vulnerability and managing disaster risks, particularly due to their interconnectedness and potential for cascading effects on community functions (Chang, 2016). Critical infrastructure systems, such as transportation, power, and telecommunications, are highly interdependent; a disruption in one asset can propagate through the interconnections, leading to cascading failures and affect communities and businesses (Yang et al., 2018). This means that vulnerability does not always pertain to individual asset failure alone but is a systemic network property (Rinaldi et al., 2001). For instance, Anderson et al. (2025) demonstrated the concept of 'functional isolation', where communities retained physical access to essential services but could not use them due to cascading failures after the occurrence of hazard events. Quantifying these complex vulnerabilities often involves assessing cascading effects and identifying geographic hotspots (Conrad et al., 2024).

Given the risks of cascading failures in interdependent infrastructure systems, there is a growing emphasis on designing and managing them for durability, adaptability, and preparedness (Balakrishnan et al., 2024; Comfort et al., 2010). Disaster-resilient infrastructure is strategically engineered to withstand and recover from disruptions with substantial economic benefits, marking a shift from reactive response to proactive preparation (National Infrastructure Advisory Council, 2010). For instance, a recent report estimated that every dollar invested in climate-resilient infrastructure generates approximately four dollars in socioeconomic benefits (Hallegatte et al., 2019). Several studies have also explored how social infrastructure facilities, such as schools and hospitals can enhance disaster response and recovery efforts by reconfiguring their functions (Nowell et al., 2017).

Beyond its disaster mitigation role, past research has also developed both theoretical and empirical evidence for the role of infrastructure in influencing socioeconomic characteristics during stable periods. For example, Rives and Heaney (1995) found a strong positive relationship between physical infrastructure and regional economic development using a cross-sectional dataset from Iowa, United States, and emphasized the need for continued development and maintenance of infrastructure systems, such as highways, to ensure local communities have access to central places. Ascher and Krupp (2010) argued that well-developed physical infrastructure systems can efficiently maintain dense populations, support economic functions, and enhance the standard of living for communities by providing uninterrupted services and ensuring mobility. Similar findings on the positive role of infrastructure on communities are reported by Jaafari et al. (2025), Song et al. (2024). However, existing research often overlooks how these infrastructure characteristics and their related spatial inequalities work together and are associated with long-term community vulnerability, especially in diverse regions like developing countries in the Global South.

While infrastructure systems are crucial for community development and resilience, disparities in infrastructure provisions can significantly exacerbate vulnerabilities (Tu et al., 2025), particularly for disadvantaged populations (Coleman et al., 2020; Rendon et al., 2021). Similarly, evidence confirms that disadvantaged groups often have reduced access to infrastructure services (Nicoletti et al., 2023). In the literature, several methods have been proposed in the past to analyze infrastructure disparities and their effects on community characteristics. These methods can be broadly classified into six approaches: composite indicator analysis, geospatial analysis, network analysis, statistical modeling, simulation modeling, and inequality measurement.

Composite indicators, such as the Social Vulnerability Index (SoVI) (Cutter & Emrich, 2017) and the Baseline Resilience Index for Communities (BRIC) (Camacho et al., 2023), synthesize diverse socioeconomic and infrastructure factors. They provide a macro-level view to identify at-risk communities and geographic hotspots where infrastructure deficits coincide with social vulnerability (Fox et al., 2024). Geospatial analysis, using geographic information systems (GIS) and geostatistical techniques, can unravel spatial disparities and associations in infrastructure provisions across different administrative and geographic entities. For example, gridded remote sensing datasets have been used to show significant global inequalities in infrastructure provisions (Chen et al., 2024; Zhou et al., 2022) and community vulnerability (Fox et al., 2024). The third category, network analysis, is grounded in graph theory and is effective for understanding systemic risks, examining infrastructure interdependencies, and identifying potential cascading failures (Balakrishnan & Zhang, 2018; Duenas-Osorio & Vemuru, 2009). For causal inferences, statistical approaches — such as cross-sectional and longitudinal modeling — have been widely used to develop empirical relationships between community resilience and infrastructure characteristics, including overall quality (Taghizadeh-Hesary et al., 2021) and household access to essential services (Hartwig & Nguyen, 2023). Dynamic modeling techniques, such as agent-based models (Aghababaei & Koliou, 2023; Valinejad et al., 2022) and system dynamics (Feofilovs et al., 2020), simulate changes in vulnerability over time by representing infrastructure characteristics and their interactions, using mechanistic frameworks to infer causal relationships and predict future risks. Finally, inequality measurement, through economic metrics like the Gini coefficient, has been applied to study disparities in access to various infrastructure types and their consequences for urban and community outcomes (Pandey et al., 2022). For instance, a recent study by Tu et al. (2025) revealed that infrastructure access in Global South countries ranges from 50%–80% of that in Global North countries, while their associated inequality levels are 9%–44% higher.

A comparative analysis of the methods reveals trade-offs in their effectiveness in analyzing associations between infrastructure and vulnerability. Quantitative approaches (composite indicators, statistical models, inequality metrics) provide actionable insights but risk oversimplification, lacking causal depth for tailored policy development. Spatial methods (geospatial techniques, network analysis) are effective at identifying disparities and systemic risks but depend on scarce, high-quality data, especially in resource-constrained settings in the Global South. On the other hand, longitudinal modeling techniques can capture causal links and temporal dynamics; however, their applications are limited by data paucity.

1.3. Literature gaps and study objectives

The existing literature extensively explores the role of infrastructure provisions in community development, vulnerability, and resilience. Studies on community resilience overwhelmingly focus on short-term disaster preparedness, response, and recovery, often treating infrastructure as a resilience capability that enables communities to absorb shocks and recover faster after disasters (Bruneau et al., 2003; Koliou et al., 2018). Most empirical research examines the association between a single dimension of vulnerability (for instance, health, as in Tu et al. (2025)) and a specific infrastructure provision, such as access to electricity, drinking water, or healthcare facilities. While such sector-specific analyses are useful for targeted investments, they provide limited insight for cross-sectoral or multi-level resource allocation, where interdependencies among different infrastructure systems become critical.

Although a few studies have acknowledged the role of infrastructure services in shaping peacetime vulnerability (Fang et al., 2016), empirical evidence on how infrastructure deficits and spatial disparities jointly influence vulnerability remains limited. Furthermore, the literature rarely distinguishes between infrastructure access (e.g., household-level availability of utilities) and infrastructure availability (e.g., the regional presence of critical systems such as road networks and power grids), despite their different implications for equitable service delivery. This distinction is particularly important in developing countries, where uneven resource allocation priorities contribute to substantial subnational disparities. Additionally, limited research has examined the synergistic or compensatory relationships between these infrastructure characteristics—how one dimension may strengthen or offset another in influencing community vulnerability.

Addressing these gaps is essential for guiding infrastructure investments that reduce community vulnerability in an equitable manner. This study addresses these identified gaps by investigating the individual and synergistic associations of regional infrastructure on community vulnerability. We develop a district-level analytical framework using an empirical, spatial approach. We consider two broad types of infrastructure systems while evaluating availability and access at the district level: (a) utility-based infrastructure that provide essential services to households, including electric grids, water supply systems, and roads; and (b) social infrastructure, including schools, hospitals, and police stations. Leveraging datasets from India, a geographically and demographically diverse country, this research provides evidence for the associations between these infrastructure dimensions and community vulnerability. The specific objectives of the study are:

1. To develop a district-level composite indicator framework for assessing the relationships between household utility access, critical infrastructure availability, and social infrastructure density, and community vulnerability.
2. To compile and harmonize high-resolution, open-access geospatial and demographic datasets to construct standardized indicators across infrastructure and vulnerability dimensions.
3. To analyze regional disparities in infrastructure availability and access through spatial mapping and clustering.
4. To quantify the individual and interaction effects of infrastructure access and availability on community vulnerability, with a focus on identifying synergistic and limiting factors.

This study advances existing frameworks by explicitly integrating multiple infrastructure categories (i.e., critical- and social infrastructure systems) within a unified, spatially explicit analytical framework. By distinguishing between access and availability and modeling their combined and interaction effects, the study moves beyond sectoral or single-dimensional assessments to provide a more holistic view of regional infrastructure characteristics relate to community vulnerability. Moreover, by leveraging standardized, high-resolution datasets across all 733 districts of India, the framework operationalizes the concept of equitable infrastructure planning within a national-scale, cross-sectoral context.

The remainder of the article is structured as follows: Section 2 delineates the methodological framework. Section 3 discusses geographic disparities, limiting factors, and cluster-based findings. Section 4 summarizes policy implications of the results and potential directions for future research.

2. Methods and data

The methodology adopted in this study is illustrated in Fig. 1. It consists of four key steps: (a) data collection, processing, and categorization; (b) development of composite subnational indicators (focusing on community and infrastructure); (c) spatial analysis of subnational disparity in infrastructure and community capabilities; and (d) analysis of influence of infrastructure on community vulnerability profiles. The core of the methodology lies in the development of a comprehensive set of indicators to capture the associations between community vulnerability and infrastructure provisions. These indicators were derived from relevant features sourced from various datasets, including government data, surveys, and open-source databases. They were organized into four dimensions: regional critical infrastructure availability, social infrastructure density, household access to essential services, and community vulnerability. The indicators were then standardized and integrated into composite indicators using principal component analysis (PCA), a statistical technique commonly used to reduce dimensionality and identify underlying patterns in datasets. The framework builds upon this foundation in a series of steps. The first step involved gathering and processing the necessary datasets to develop the indicators. In the second step, PCA was applied to generate the composite indicators. In the third step, unique regional archetypes, which represent distinct combinations of the four pillars, were identified using hierarchical clustering analysis. Finally, spatial regression modeling was employed to explore the empirical relationships between community vulnerability and infrastructure provisions at the district-level. In addition, the results from the regression model were combined with clustering analysis to derive optimal infrastructure resource allocation strategies tailored to the district groups with distinct infrastructure profiles.

In the rest of the section, the above steps are discussed in detail. Since the study is based in India, the methodological descriptions are tailored to the datasets from the country.

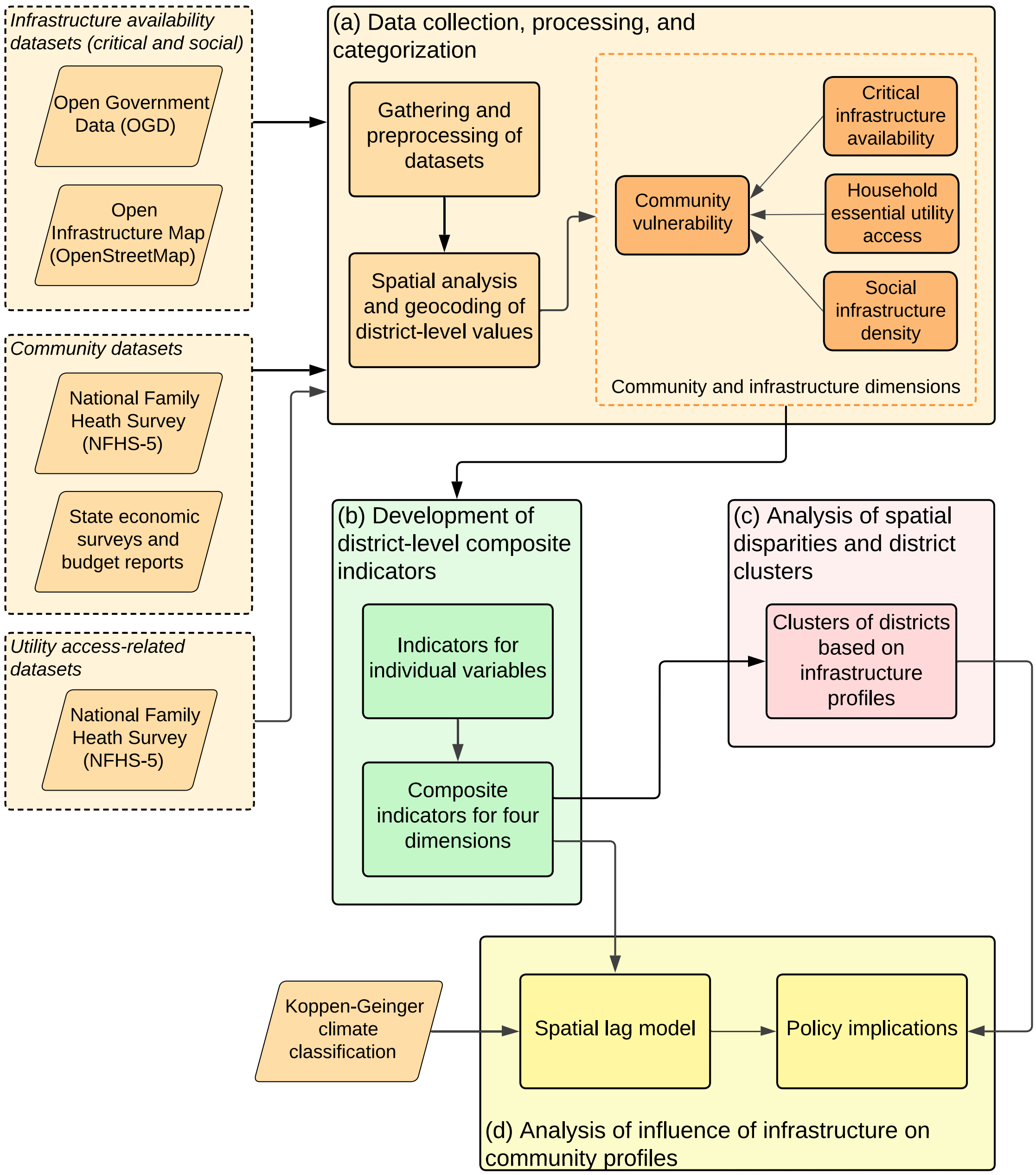

Fig. 1. Methodological framework illustrating the four steps: (a) data collection and processing, (b) composite indicator development, (c) spatial disparity analysis, and (d) analysis of influence of infrastructure on community vulnerability.

2.1. Data collection, categorization, and preprocessing

To the best of the authors' knowledge, there is no centralized database providing required data for the study. Therefore, the first step of the study was to identify relevant datasets collected in recent years pertaining to infrastructure- and community-related characteristics. Since the study aims to examine the associations between community characteristics and infrastructure provisions at the regional level, the scope of the data collection efforts was restricted to datasets providing district-level information. These datasets were then georeferenced using natural language processing tools and manual methods.

Table 1 presents the final list of features selected under each pillar for further analysis, including their sources. For community vulnerability, five features related to economic status, education, and health were selected from multiple sources, such as economic census surveys and annual budget documents, and National Family Health Survey 5 (MoHFW, 2022). Critical infrastructure availability was represented using four features related to density of highways and paved roads, electric transmission network, and extent of ground water use. Open-source spatial datasets from CGWB (2024), ESRI India (2024), OpenStreetMap (2024) were used for extracting critical infrastructure availability related features. The third dimension — household essential utility access, which represents the degree of access of services, such as water and electricity supply, was captured using four features obtained from National Family Health Survey 5 (MoHFW, 2022). Finally, social infrastructure density was measured using four features related to per capita availability of police stations, schools, and hospitals derived from ESRI India (2023), MoE (2023), MoHFW (2018, 2019).

We preprocessed the raw features in three steps: spatial harmonization and geocoding (aligning datasets to consistent district boundaries),

Table 1
Selected variables corresponding to community and infrastructure-related dimensions and their data sources.

Pillar	Feature	Unit	Original/ Derived	Effect on dimension (+/−)	Original source
Community vulnerability	Per capita income	INR ’000	Original	−	Economic surveys, budget documents
	Women (age 15-49) who are literate	%	Original	−	MoHFW (2022)
	Underweight children <5 years	%	Original	+	
	Women (age 15-49) with Low BMI	%	Original	+	
	Women (age 15-49) who are anaemic	%	Original	+	
Critical infrastructure availability	Highway density	km per 100 sq.km.	Derived	+	ESRI India (2024)
	Road density	km per 100 sq.km.	Derived	+	ESRI India (2024)
	Transmission line density	km per 100 sq.km.	Derived	+	OpenStreetMap (2024)
	Percentage level of groundwater extraction	%	Original	-	CGWB (2024)
Household essential utility access	Households with electricity	%	Original	+	MoHFW (2022)
	Households with improved sanitation facility	%	Original	+	
	Households with improved drinking water source	%	Original	+	
	Households with clean fuel for cooking	%	Original	+	
Social infrastructure density	Police stations	nos. per 1,00,000 pop	Derived	+	ESRI India (2023)
	Schools	nos. per 1,00,000 pop	Derived	+	MoE (2023)
	Primary healthcare	nos. per 1,00,000 pop	Derived	+	MoHFW (2018, 2019)
	Secondary & tertiary healthcare	nos. per 1,00,000 pop	Derived	+	

missing data handling (using ordinary-kriging), and feature standardization, to ensure their suitability for developing indicators. The technical details of the preprocessing steps are explained in Appendix A.1 (see Appendix).

2.2. Development of composite regional indicators related to community vulnerability and infrastructure provisions

In the next step, we developed district-level indicators (both feature-specific and composite). The list of features used in the study to represent the four dimensions are presented in Table 1. The following steps were performed to derive the feature-specific indicators from the pre-processed features.

1. In the case of community vulnerability, the composite indicator was designed such that lower values indicate lower vulnerability. To ensure consistency, all the features that were considered to reduce vulnerability were inverted to maintain consistency in directionality (for example, percentage women who are illiterate).
2. In the case of infrastructure-related dimensions, the corresponding composite indicators were designed such that higher values indicate better access and availability and were thus positively correlated with the respective dimensions. Therefore, features from these dimensions that wee negatively correlated with the dimension were inverted (e.g., percentage of groundwater-level extraction).
3. Once all features were aligned for directionality as above, ‘standard scaling’ transformation was applied and corresponding indicators were developed. Standard scaling, also known as ‘z-score normalization’, transforms data to have a mean of zero and a standard deviation of one, allowing for comparison of features with different units or scales. Standard scaling operation was performed as follows:

$$Z = \frac{X - \mu_X}{\sigma_X} \quad (1)$$

where $Z \in \mathbf{Z}$ is the transformed indicator, μ is the mean of the standardized feature, and σ is the corresponding standard deviation.

Standard scaling ensures no feature gains disproportionate importance due to its original scale. This allows decision-makers to make relative comparisons between subjects (here, districts) using standardized values.

After developing the indicators, the next task was to construct composite indicators for the four dimensions at the district-level using principal component analysis (PCA). PCA is a statistical technique for developing composite indicators by transforming correlated features into uncorrelated principal components. This approach effectively addresses multicollinearity and provides a data-driven framework for weighting and aggregating features into composite measures (Wu et al., 2022). For a detailed step-by-step derivation of the principal components, including the computation of weights, readers are referred to Appendix A.2 (see Appendix). Although standardization of input features is typically a prerequisite for PCA, this step was omitted in our study, as the indicators developed in the previous stage were already normalized using standard scaling, ensuring comparability across features.

To assess and evaluate the geographic patterns in the infrastructure composite indicators, we conducted Moran’s I tests (Moran, 1950). These tests examined whether districts with similar infrastructure profiles were geographically clustered, informing the selection of appropriate spatial models in subsequent analyses.

2.3. Analysis of infrastructure dimensions and their regional disparities

Following the development of composite indicators, we employed clustering analysis to identify groups of districts possessing similar infrastructure profiles (based on availability and access characteristics). Clustering-based infrastructure patterns could simplify the identification of infrastructure gaps and assist in fine-tuning infrastructure planning by grouping districts that share similar strengths and weaknesses.

Unlike individual composite indicators, which provide isolated district-level snapshots, clustering reveals underlying interrelationships and systemic gaps that might otherwise remain hidden.

To identify district clusters, we experimented with six distinct clustering algorithms: K-Means, Bisecting K-Means, agglomerative clustering (Ward hierarchical clustering), affinity propagation, Gaussian mixtures, and BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies). Among these, K-Means, Bisecting K-Means, agglomerative clustering, and BIRCH are distance-based algorithms. However, in the case of affinity propagation, the clusters are formed by maximizing the similarity between data points and their exemplars in each cluster. On the other hand, Gaussian Mixtures assumes that each cluster is generated from a mixture of Gaussian distributions. To ensure robust results and prevent overfitting, we implemented a 3-fold cross-validation procedure and performed appropriate hyperparameter tuning for each algorithm before selecting the final clusters. Detailed review of state-of-the-art clustering algorithms can be found in the works of Ezugwu et al. (2022), Firdaus and Uddin (2015).

To determine the optimal number of clusters, we calculated three well-known evaluation metrics: the Silhouette Score (S), the Calinski-Harabasz Index (CH), and the Within-Cluster Sum of Squares (WCSS). The details of the calculation of the metrics are discussed in Appendix A.3. After clustering, the distributions of composite indicators were derived for each of the clusters which represent the distinct regional infrastructure profiles in India.

2.4. Modeling associations between community and infrastructure characteristics

One of the key objectives of the study is to unravel how infrastructure provisions (both social and critical infrastructure systems) influence community vulnerability. While there are limitations in establishing causal relationship between community vulnerability and infrastructure provisions using the dataset, the vast spatial variations in these aspects across India can provide valuable indications about the associations that may be crucial in infrastructure planning and climate adaptation decisions.

Preliminary correlation analysis was performed using visual inspection of scatter plots and Kendall's rank correlation tests (Kendall's τ) for composite indicator pairs. Kendall's rank correlation tests are capable of investigating non-linear associations among feature pairs unlike Pearson's correlation analysis.

After confirming correlations and testing for spatial autocorrelation among composite infrastructure indicators using Moran's I test, a spatial lag model was constructed to capture the relationships between composite indicators related to community vulnerability and infrastructure provisions, while explicitly accounting for spatial dependence across state–climate zone combinations. The model, selected based on Lagrange Multiplier (LM) tests for spatial dependence, was estimated using Maximum Likelihood (ML) estimation method and specified as follows:

$$CI_{comm,d} = \rho \sum_{d' \in N_d} w_{dd'} CI_{comm,d'} + \beta_0 + \sum_{i \in I} \beta_i CI_{i,d} + \sum_{\substack{i,j \in I \\ i<j}} \beta_{ij} (CI_{i,d} \times CI_{j,d}) + \gamma_s + \epsilon_d \quad (2)$$

where $CI_{comm,d}$ is the composite indicator value corresponding to community vulnerability in district d, $I = \{crit, soc, ess\}$ denotes the three infrastructure-related dimensions (critical infrastructure availability, social infrastructure density, and essential utility access, respectively), $\sum_{d' \in N_d} w_{dd'} CI_{comm,d'}$ is the spatial lag of the dependent variable, representing the weighted average of community vulnerability in neighboring districts d' (with weights $w_{dd'}$ defined by a row-standardized Queen contiguity matrix $\mathbf{W}$), ρ is the spatial auto-regressive parameter capturing the strength of spatial dependence, γ_s represents fixed effects for state and Köppen–Geiger climate zone combinations (where district d belongs to state-climate combination s), ϵ_d is the district-level error term (assumed to be $i.i.d.$), and the β_{ij} terms capture all two-way interactions among infrastructure dimensions. The combined state and climate zone (instead of treating state and climate zone as separate variables, their unique combinations were identified and encoded as a variable) fixed effects control for regional and climatic variations, while the spatial lag term accounts for spatial spillovers in community vulnerability across neighboring districts.

The model specification presented in Eq. (2) also captures the synergistic or compensatory relationships between infrastructure dimensions through interaction terms (β_{ij}), revealing whether improvements in one dimension (e.g., critical infrastructure availability) amplify or diminish the vulnerability-reducing effects of another (e.g., essential utility access). The spatial lag term ($\rho \sum_{d' \in N_d} w_{dd'} CI_{comm,d'}$) models the influence of neighboring districts' vulnerability on local vulnerability, reflecting regional interdependencies. These associations reveal district-specific vulnerability profiles, highlighting areas where infrastructure deficiencies cluster or where strengths in one dimension (e.g., critical infrastructure) mitigate weaknesses in another (e.g., social infrastructure).

3. Results and discussion

In this section, the results from the study are presented. First, we discuss the spatial distribution of the developed composite indicators for community vulnerability and infrastructure provisions across districts in India. Later, we present the clustering results that identified the unique infrastructure profiles of the districts. Finally, we discuss the results of the spatial regression analysis.

3.1. District-level composite indicators for community vulnerability

Composite indicators for the three infrastructure dimensions and community vulnerability were developed following the procedure outlined in Section 2.2. The feature weights derived using PCA for developing the composite indicators are presented in Table A.6. Fig. 2 shows the computed composite indicator values for community vulnerability across 733 districts in India. A composite indicator value of zero represents the national average (because of the standardization step in Eq. (1)). Negative values indicate below-average vulnerability, while positive values signify above-average vulnerability.

The spatial distribution map for community vulnerability reveals considerable regional disparity. This is further confirmed by Moran's I value of 0.64 (significant at $\alpha = 0.05$), indicating presence of strong spatial autocorrelation. Several districts in southern states (e.g., Kerala, Karnataka, and Tamil Nadu), northern Himalayan states (e.g., Uttarkhand and Haryana), and northeastern states (e.g., Sikkim, Arunachal Pradesh, Mizoram) exhibit below-average community vulnerability. On the other hand, majority districts in northwestern states (e.g., Rajasthan, Gujarat), northern states (e.g., Uttar Pradesh), central (e.g., Madhya Pradesh, Chhattisgarh), and eastern states (e.g., Bihar, West Bengal, Jharkhand, Odisha) has significantly higher levels community vulnerability. The spatial distributions are in line with recent disaster vulnerability-related studies in India, such as George and Sharma (2022), Mahapatra et al. (2022).

3.2. District-level composite indicators for infrastructure-based dimensions

Fig. 3 presents the spatial distributions of the district-level composite indicators developed for the three infrastructure-related dimensions.

For critical infrastructure availability (Fig. 3a), several districts in southern and western India (parts of Karnataka, Tamil Nadu, Kerala, Maharashtra, and Gujarat) and a few parts of northern states, such as Haryana, and Uttar Pradesh, show above-average provisions (yellow to green), suggesting relatively well-developed infrastructure

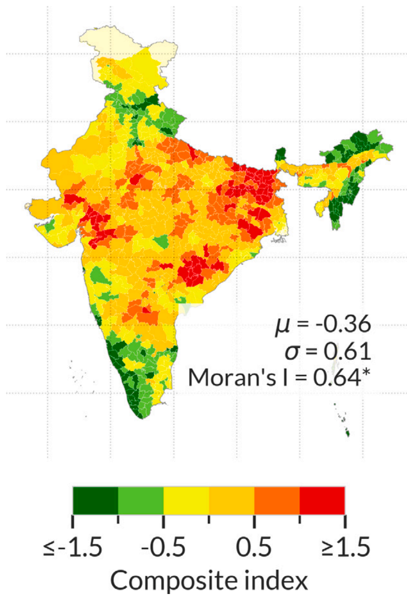

Fig. 2. District-level composite indicators of community vulnerability (higher values indicate higher vulnerability).

systems. In contrast, northwestern states, such as Rajasthan and Punjab, northeastern states, and central-eastern regions like Chhattisgarh and Odisha exhibit below-average provisions (orange to red), indicating significant gaps in infrastructure density. Regarding essential utility access (Fig. 3b), southern, western and northwestern states (e.g., Gujarat, Karnataka, Tamil Nadu, Andhra Pradesh and Kerala) demonstrate above-average resilience, while northern, central, eastern, and northeastern states, such as Uttar Pradesh, Bihar, and Assam, show below-average resilience. For social infrastructure density (Fig. 3c), northern Himalayan states (e.g., Himachal Pradesh, Uttarakhand) and parts of the northeast exhibit above-average provisions (yellow to green), indicating better access to facilities and amenities. However, districts in Uttar Pradesh, Bihar, and West Bengal show below-average performance (orange to red), underscoring an urgent need for increased resource allocation to improve per capita social infrastructure availability in these areas.

Overall, the findings suggest that infrastructure provisions exhibit strong spatial inequalities. Moran's I tests reveal significant spatial clustering for all three infrastructure dimensions (critical infrastructure availability: 0.30, essential utility access: 0.56, social infrastructure density: 0.30, all significant at $\alpha = 0.05$), indicating that districts with similar infrastructure profiles tend to be geographically concentrated. Further analysis is needed to investigate the role of regional factors, such as geography, climate, and governance, to identify the causal aspects of such spatial dependence; however, this is not addressed as they are not within the scope of the current study. These findings are aligned with the recent literature that provides empirical evidence on infrastructure inequalities and spatial associations from regions outside India (Guo et al., 2024; Pandey et al., 2025).

3.3. Clustering based on infrastructure provisions shows unique district patterns

Observing the spatial patterns of infrastructure characteristics across districts from the above analysis, the next step was to apply clustering algorithms to identify unique district patterns. As discussed in the methodology section, six clustering algorithms were applied: K-Means, Bisecting K-Means, Agglomerative (Hierarchical) Clustering, Affinity Propagation, Gaussian Mixtures, and BIRCH. The `scikit-learn` Python package was used to develop the clustering models. The performance metrics of the algorithms are presented in Table 2. The performance of the clustering algorithms was evaluated using three metrics: Silhouette score, Calinski-Harabasz score, and within-cluster sum of squares. After hyperparameter tuning and 3-fold cross-validation, the results showed that Agglomerative Clustering outperformed other algorithms based on the three metrics. It achieved relatively higher average Silhouette ($S = 0.28 \pm 0.02$) and Calinski-Harabasz ($CH = 99.59 \pm 12.93$) scores, indicating better-defined clusters with less overlap and greater separation. Additionally, it minimized the within-cluster sum of squares ($WCSS = 95.12$). The cluster evaluation metrics suggested that Agglomerative Clustering provided moderate clustering accuracy for the study region and was a suitable approach for resource allocation decision-making and prioritization. Based on the metrics, the optimal number of clusters was fixed at six, and the respective groups of districts were identified. Agglomerative Clustering was selected over BIRCH, which was excluded due to its higher WCSS (indicating substantial within-cluster variability), and Gaussian Mixtures, which was rejected due to its lower Silhouette score.

Fig. 4 presents the results from the clustering subtask. Specifically, Fig. 4a shows the cluster to which each district belongs, while Fig. 4b shows the distributions of composite infrastructure indicators across the different clusters. For brevity, the plots depict only the mean values and 95% confidence intervals of the composite indicators. If the lower and upper bounds of the confidence interval are entirely below zero, the district-level capability corresponding to an infrastructure dimension is considered below the national average; if both are above zero, it is deemed above the national average. If only the lower bound falls below zero, the infrastructure capability is regarded as at par with the national average.

Clustering analysis revealed the unique district-level infrastructure profiles across six distinct clusters, each characterized by unique combinations of critical infrastructure availability, social infrastructure density, and household essential utility access. Since the composite indicators are standardized, each unit represents one standard deviation. Understanding these cluster-specific profiles is fundamental for developing targeted and effective resource allocation strategies for infrastructure development with a focus on community well-being and vulnerability.

Cluster 0, comprising 61 districts, exhibits a moderately positive composite index for social infrastructure density, while critical infrastructure availability and household essential utility access show minimal deviation from the national average. Cluster 1, with 259 districts, has slightly below-average social infrastructure density and slightly positive values for the other two dimensions, indicating infrastructure provisions comparable to the national average across all three. In contrast, Cluster 2, encompassing 229 districts, shows slightly below-average values for infrastructure provisions. Cluster 3, with 28 districts, stands out with a strongly positive composite index for critical infrastructure availability and household essential utility access, while social infrastructure density remains at the national average level. Cluster 4, consisting of 108 districts, presents a mixed profile with slightly below-average critical infrastructure availability but slightly above-average provisions for social infrastructure density and household essential utility access. Lastly, Cluster 5, consisting of 48 districts, is characterized by significantly below-average household essential utility access. However, critical infrastructure availability and social infrastructure density values are comparable to the national average, with slightly lower values.

Comparative analysis of the cluster properties highlighted several similarities and differences across these clusters as follows:

- Critical infrastructure availability: Cluster 3 shows a strong positive performance in critical infrastructure availability, standing in contrast to Clusters 1, 2, 4, and 5, which are all below or around the national average.

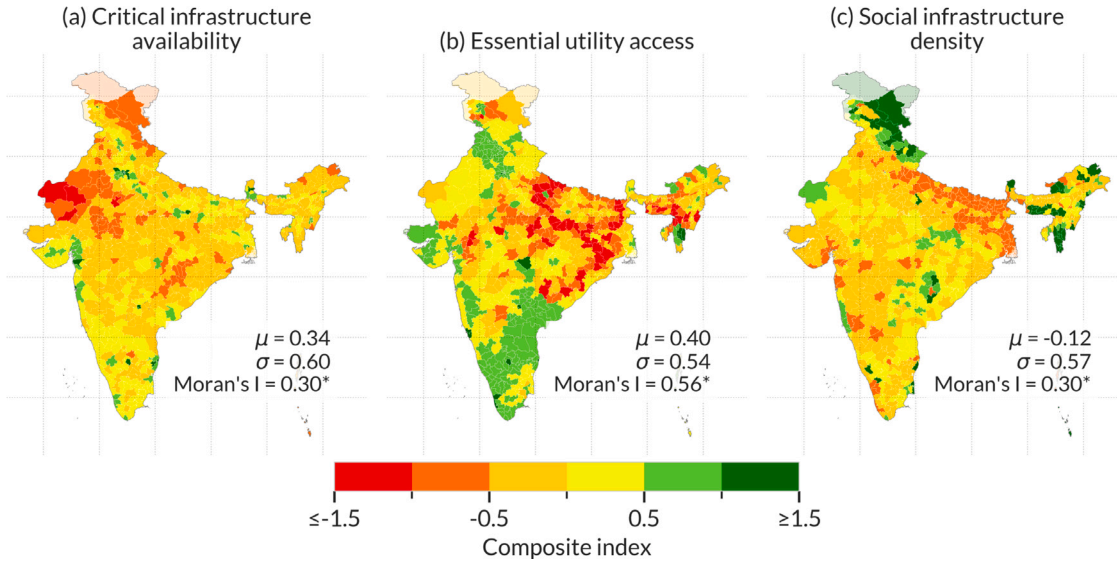

Fig. 3. District-level composite indicators of infrastructure provisions (higher value indicate better infrastructure characteristics.

Table 2
Performance metrics of various clustering algorithms.

Clustering algorithm	Number of clusters	Silhouette score	Calinski-Harabasz score	Within-cluster sum of squares
K-Means	6	0.27 ± 0.01	107.51 ± 3.89	180.16
Bisecting K-Means	6	0.26 ± 0.02	85.15 ± 1.02	213.01
Agglomerative	6	0.28 ± 0.02	99.59 ± 12.93	95.12
Affinity propagation	9	0.26 ± 0.02	93.46 ± 0.88	NA
Gaussian Mixtures	6	0.24 ± 0.02	86.73 ± 11.85	2.24
BIRCH	6	0.30 ± 0.03	57.49 ± 8.05	518.17

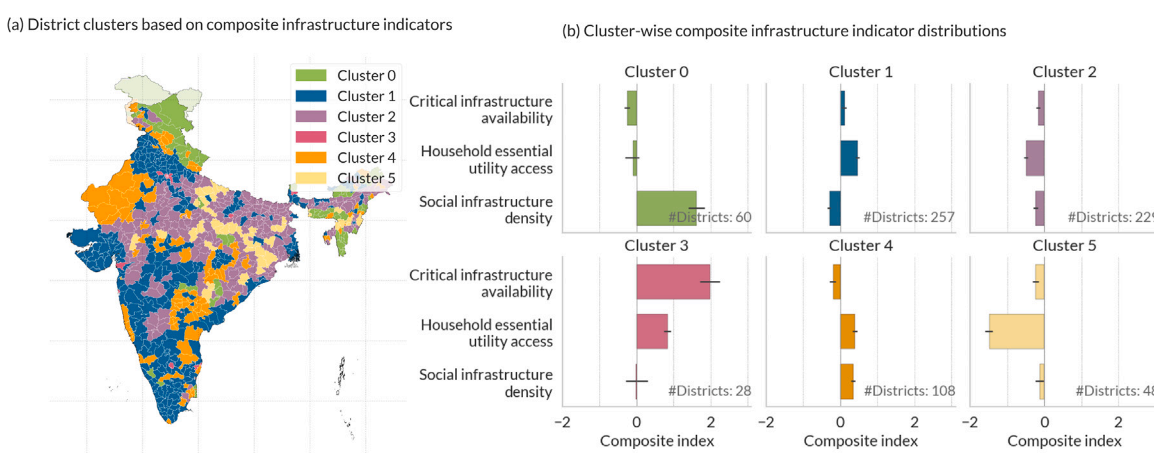

Fig. 4. Agglomerative clustering results and corresponding infrastructure patterns at district-level.

- Social infrastructure density: Cluster 0 stands out with significantly above-average social infrastructure density, contrasting with other five clusters which are all around the national average.
- Household essential utility access: Cluster 5 exhibits a dramatic deficit in household essential utility access, far below the national average. This contrasts with Cluster 3, which shows very high level of utility access. Cluster 1 and 4 shows moderately high levels whereas, cluster 2 suggest a moderately below national average capability for utility access.

Some interesting regional patterns also emerged from the analysis. Fig. 4(a) indicate the presence of several spatial subclusters within each identified cluster. At the same time, we observe that subclusters from different clusters exist within individual states, highlighting significant regional disparities at the sub-state level. This pattern is evident in most states, with a few exceptions, such as Gujarat, Kerala, and Punjab. The results, therefore, also highlight the need for region-specific and dimension-specific infrastructure interventions, addressing the unique challenges and strengths of each area to ensure equitable and sustainable community development. However, this requires understanding how different infrastructure provisions, individually and collectively, contribute to or are associated with community capabilities and vulnerability reduction.

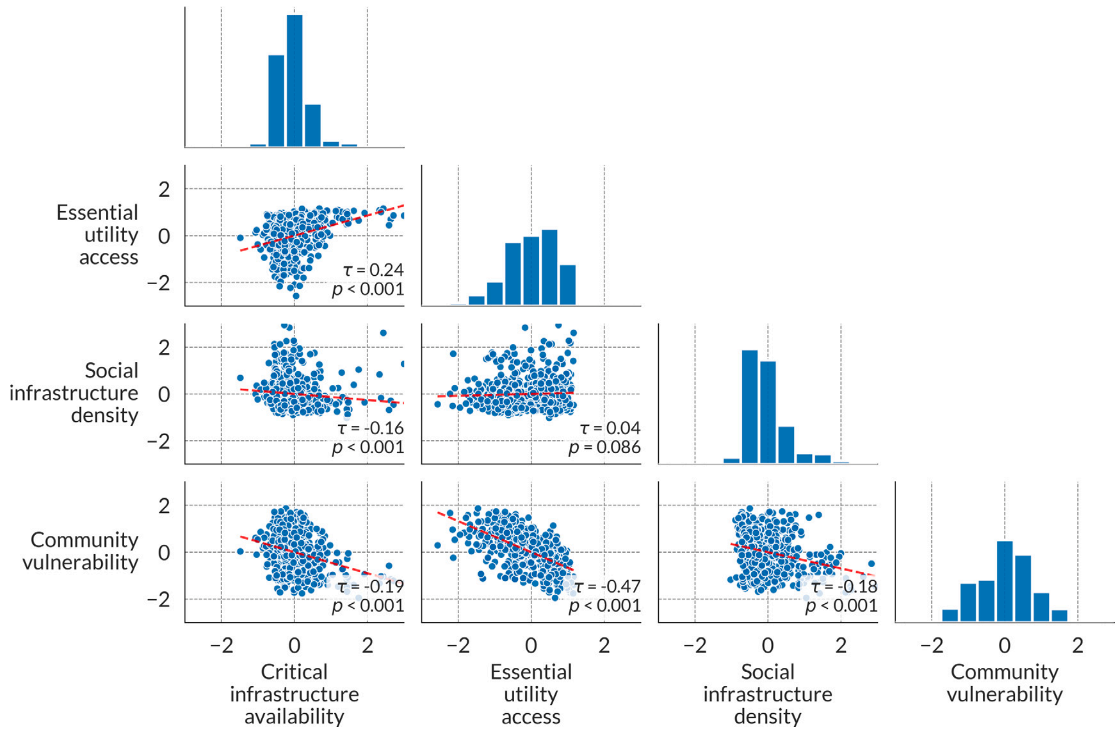

Fig. 5. Pair-plot of composite indicators along with Kendall's rank correlation coefficients (τ).

Persistent disparities in infrastructure characteristics across districts reflect historical patterns of uneven regional development, where post-independence industrial (Chakraborty, 2024) and infrastructure investment priorities (Saxena et al., 2018) followed differing pathways across regions, leading to varied levels of growth. Geographic constraints and variations in governance capacity (Mundle et al., 2016) further may have shaped the reach and quality of infrastructure networks, particularly in rural and interior regions. At the household scale, utility access is strongly influenced by local economic conditions, as income levels, employment opportunities, and fiscal capacity determine both the demand for and provision of basic services (Kumar, 2015). The broader analysis of factors influencing the spatial variability of infrastructure access and availability remains beyond the scope of this study.

3.4. Modeling of associations between infrastructure and community characteristics

To analyze how infrastructure provisions and community vulnerability in districts are associated, we applied spatial regression analysis. The correlation among the infrastructure and community vulnerability dimensions were evaluated using Kendall's rank correlation (Puka, 2011). Fig. 5 shows the results from the correlation analyses along with the corresponding scatter plots. Overall, the infrastructure dimensions are negatively correlated with community vulnerability dimension, indicating that districts with better infrastructure provisions are associated with lower community vulnerability characteristics. The following are the prominent observations that emerged in the analysis.

1. A strong negative correlation is observed between essential utility access and community vulnerability (Kendall's $\tau = -0.47$, $p < 0.001$). This suggests that improved access to essential utilities (e.g., water, electricity) may significantly reduce community vulnerability by supporting health, safety, and economic resilience.
2. A weak yet statistically significant negative correlation is seen between critical infrastructure availability and community vulnerability ($\tau = -0.19$, $p < 0.001$), indicating that the presence of critical infrastructure networks is also an important determinant that could influence local vulnerability profiles.
3. A weak negative correlation is found between social infrastructure density and community vulnerability ($\tau = -0.18$, $p < 0.001$), suggesting a modest association between access to higher social infrastructure (e.g., schools, health centers) and reduced community vulnerability.
4. A moderate positive correlation exists between critical infrastructure availability and essential utility access ($\tau = 0.24$, $p < 0.001$), highlighting how physical infrastructure underpins access to basic services. However, the modest strength of this relationship implies that other socioeconomic or governance factors, and local infrastructure options may also influence household access to utility services.
5. Social infrastructure density shows a weak negative correlation with critical infrastructure availability ($\tau = -0.16$, $p < 0.001$), possibly pointing to regional disparities where areas prioritized for physical infrastructure differ from those prioritized for social services. This may indicate trade-offs in planning or uneven development.

Next, we developed a spatial lag model with fixed effects for state-climate zone combinations (Eq. (2)) to capture the influence of various infrastructure-related characteristics on community vulnerability (CI_{comm}), while accounting for spatial spillovers across districts. The model results are summarized in Table 3. The spatial lag model also captures the indirect spillover effects of infrastructure dimensions on the community vulnerability of neighboring districts. The indirect effects are presented in Table 4.

The model examines the infrastructure-related determinants of community vulnerability using district-level data from 729 observations (four out of 733 districts are isolated islands and therefore removed for spatial weights calculation), accounting for spatial spillovers through the spatial lag term. The model demonstrates strong explanatory power,

Table 3
Spatial lag model results.

Variable	Coefficient	Std. error	z-stat	p-value
Intercept (Reference state-climate zone category)	0.1150	0.0777	1.480	0.139
CI_{crit}	−0.0727	0.0316	−2.304	0.021*
CI_{ess}	−0.3850	0.0238	−16.171	<0.001***
CI_{soc}	−0.0452	0.0215	−2.105	0.035*
Interaction effects				
$CI_{crit} \times CI_{ess}$	−0.1092	0.0380	−2.875	0.004**
$CI_{crit} \times CI_{soc}$	0.0425	0.0279	1.523	0.128
$CI_{ess} \times CI_{soc}$	0.0026	0.0280	0.092	0.927
$S\sum w_{dd'} CI_{comm,d'}$	0.5500	0.0288	19.112	<0.001***
Model Statistics				
Observations (districts)				729
Pseudo R^2				0.8841
Spatial pseudo R^2				0.8398
Log-Likelihood				−57.45
AIC				222.90
Moran's I (residuals)				0.0329 (p = 0.070)

- Notes: Standard errors are computed using Maximum Likelihood estimation.
- Coefficients of combined state-climate zone variables not shown.
- $CI_{comm}, CI_{ess}, CI_{crit}, CI_{soc}$ indicate composite indicators for community vulnerability, household essential utility access, critical infrastructure availability, and social infrastructure density. $S\sum w_{dd'} CI_{comm,d'}$ denotes the spatial lag of CI_{comm}.
* Significance level: $p < 0.05$.
** Significance level: $p < 0.01$.
*** Significance level: $p < 0.001$.

with a $pseudo - R^2$ of 0.8841 indicating that 88.41% of variance in the composite community vulnerability indicator is explained by the infrastructure-related composite indicators. The Moran's I test on model residuals (I = 0.0329, p = 0.0.70) indicates that the model effectively accounts for spatial autocorrelation and the spatial autoregressive parameter ρ captures most of the spatial dependendence effects.

All the three infrastructure-related variables shows statistially significant influence on community vulnerability. Household level essential utility access (CI_{ess}) showed the strongest direct impact (β = −0.3850, $p < 0.001$), where each standard deviation of improvement is associated with approximately 0.39 standard deviations of reduction in community vulnerability (CI_{comm}), with an additional indirect effect (β = −0.4988) through neighboring districts, totaling a 0.89 standard deviation reduction. This finding underscores the fundamental role of reliable utility access in community capacities, amplified by regional spillover effects. Critical infrastructure availability (CI_{crit}) also demonstrated a moderate direct effect (β = −0.0727, p = 0.021), suggesting that spatial density of infrastructure system and services, such as road networks, groundwater availability, and energy systems are associated with reduced community vulnerability, with an indirect effect (β = −0.0992) enhancing the total impact to 0.18 standard deviations. Social infrastructure density (CI_{soc}) exhibits a modest direct impact (β = −0.0452, p = 0.035), implying that per capita density of schools, hospitals, and health facilities are associated with lower community vulnerability, although the relationship appears relatively modest. This is intuitive because social infrastructure density may not have a direct and immediate effect on vulnerability and may manifest through higher-order impact pathways in the long term via transformation of community well-being.

The analysis of interaction effects revealed how the infrastructure dimensions operate together in localized contexts. A significant negative interaction effect of critical infrastructure availability and essential utility access (β = −0.1092, p = 0.004) indicates synergistic associations, where communities with strong performance in both areas show compounded reductions in predicted vulnerability (total effect = −0.2347) beyond what either factor is associated with independently. Achieving a unit reduction in community vulnerability (CI_{comm}) requires more resources in regions where both critical infrastructure availability (CI_{crit}) and essential utility access (CI_{ess}) are relatively lower, due to the lower baseline direct effects (−0.0727 for CI_{crit} and −0.3850 for CI_{ess})

Table 4
Average direct, indirect, and total effects based on the spatial lag model.

Variable	Direct	Indirect	Total
CI_{crit}	−0.0767	−0.0992	−0.1759
CI_{ess}	−0.3857	−0.4988	−0.8845
CI_{soc}	−0.0529	−0.0683	−0.1212
$CI_{crit} \times CI_{ess}$	−0.1023	−0.1323	−0.2347
$CI_{crit} \times CI_{soc}$	0.0545	0.0704	0.1249
$CI_{ess} \times CI_{soc}$	−0.0022	−0.0029	−0.0051

combined with a significant negative interaction (−0.1092), necessitating substantial simultaneous improvements. In contrast, regions where both CI_{crit} and CI_{ess} are above average show synergistic associations, exhibiting lower predicted vulnerability as the combined direct and indirect effects correspond to stronger relationships than those of individual effects. Regions with universal access to essential utility services backed up by a reliable critical infrastructure network can lead to more economic activities and opportunities for social development, and ultimately result in low vulnerability of populations. The findings suggest that there should be a balance between the two dimensions while allocating resources for optimal reduction in community vulnerability. In summary, the results suggest that regions with lower levels of infrastructure provisions may need more resources to achieve the same level of community vulnerability reduction compared to those with well-developed infrastructure networks and accessibility.

Additionally the significant spatial lag term ($\rho = 0.5500$, $p < 0.001$), and the indirect effects presented in Table 4 indicate that community vulnerability in a district is strongly influenced by the vulnerability in neighboring districts, emphasizing regional interdependencies. Indirect effects show that improvements in essential utility access (indirect effect = −0.4988) and critical infrastructure (indirect effect = −0.0992) in one district is associated with reduced vulnerability in neighboring districts, highlighting regional spillovers. These effects underscore the importance of coordinated infrastructure investments across districts , as such coordination is associated with lower overall community vulnerability.

3.5. Identifying the best resource allocation strategy for each district cluster

To guide targeted resource allocation, we integrate the spatial regression results with our clustering framework, which groups the 729

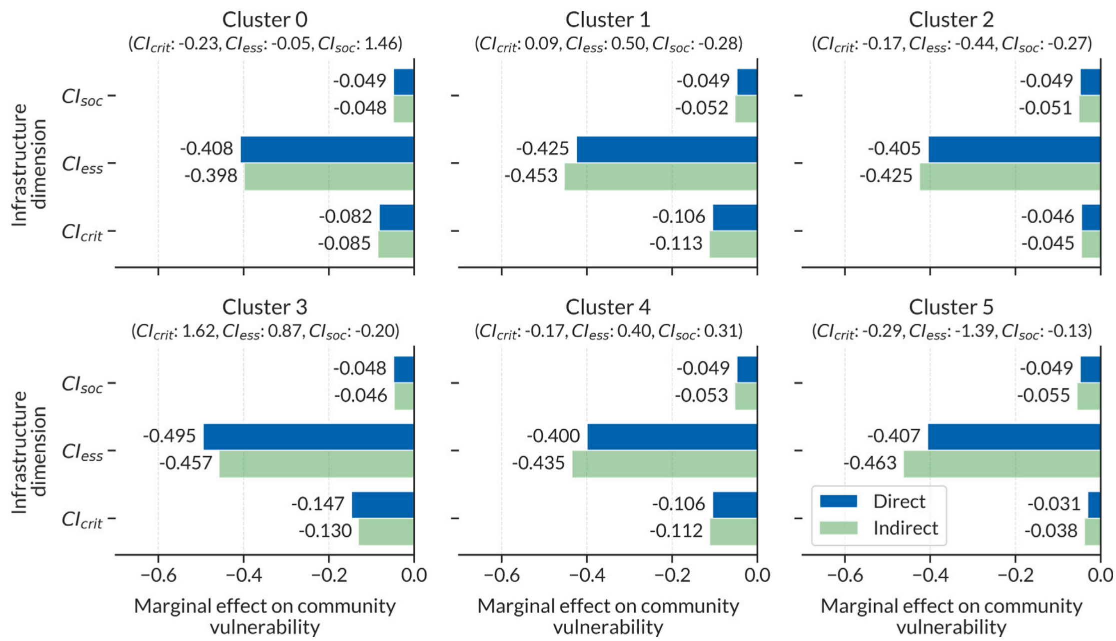

Fig. 6. Marginal influence on community vulnerability due to enhanced infrastructure provisions. The marginal effects are calculated based on the updated spatial lag model after removing the insignificant interaction terms $CI_{crit} \times CI_{soc}$ and $CI_{ess} \times CI_{soc}$.

districts into six distinct clusters based on infrastructure characteristics. District-level marginal effects, capturing direct (local) and indirect (spillover) impacts, are extracted from the spatial multiplier $(\mathbf{I}-\rho\mathbf{W})^{-1}$, where $\mathbf{W}$ is a queen contiguity spatial weights matrix. The marginal effects for each district are defined as:

$$\frac{\partial CI_{\text{comm},d}}{\partial CI_{\text{crit},d}} = (\mathbf{I}-\rho\mathbf{W})^{-1}(\beta_{\text{crit}} + \beta_{\text{crit,ess}} \cdot CI_{\text{ess},d}),$$
$$\frac{\partial CI_{\text{comm},d}}{\partial CI_{\text{ess},d}} = (\mathbf{I}-\rho\mathbf{W})^{-1}(\beta_{\text{ess}} + \beta_{\text{crit,ess}} \cdot CI_{\text{crit},d}), \tag{3}$$
$$\frac{\partial CI_{\text{comm},d}}{\partial CI_{\text{soc},d}} = (\mathbf{I}-\rho\mathbf{W})^{-1}\beta_{\text{soc}}.$$

The district-level marginal effects computed using Eq. (3) were subsequently aggregated by clusters developed in Section 3.3 to facilitate comparison across groups of districts with similar infrastructure characteristics. The mean direct and indirect marginal effects for each cluster were then used to interpret the relative importance of infrastructure dimensions and to identify priority areas for targeted resource allocation.

As discussed earlier, the direct effects represent the influence of infrastructure investments on community vulnerability (CI_{comm}) within a district, while the indirect effects capture spillovers to neighboring districts, amplified by the spatial autoregressive parameter. These effects are aggregated by cluster, with the mean marginal effects presented in Fig. 6. The reported values are based on the updated spatial lag model, obtained after removing the insignificant interaction variables from the model presented in Table 3. The updated model summary and coefficients are provided in Table A.5. For essential utility access (CI_{ess}), direct effects range from −0.495 (Cluster 3) to −0.400 (Cluster 4), with indirect effects ranging from −0.463 (Cluster 5) to −0.398 (Cluster 0). For critical infrastructure availability (CI_{crit}), direct effects range from −0.147 (Cluster 3) to −0.031 (Cluster 5), with indirect effects from −0.130 (Cluster 3) to −0.038 (Cluster 5). The effect of social infrastructure density (CI_{soc}) is uniform across clusters, with direct effects of approximately −0.049 and indirect effects of −0.051, reflecting its consistent but modest contribution.

For districts within each clusters, these results inform resource allocation strategies by identifying priority areas for investment. Cluster 3, consisting of mostly highly urbanized districts and characterized by high critical infrastructure availability and essential utility access, exhibits the largest total marginal effects for CI_{ess} ((−0.457)+(−0.495) = −0.952) and CI_{crit} ((−0.130) + (−0.147) − 0.277), indicating that higher levels of these dimensions are associated with substantially lower vulnerability, both locally and regionally through spatial spillovers. Cluster 4, with strong total CI_{ess} effects (−0.835), is characterized by above-average household utility access and social infrastructure density but relatively lower critical infrastructure availability. In this cluster, enhancing household-level utility access could generate significant regional benefits as well, while improvements in critical infrastructure availability also show notable marginal associations with reduced vulnerability. In contrast, Cluster 5, which includes districts mostly located in central and eastern India, is characterized by severe deficits in household access to basic utilities, while levels of social infrastructure density and critical infrastructure availability are only slightly below the national average. Consequently, improvements in CI_{ess} yield the largest marginal effect in community vulnerability, whereas the total marginal effect of CI_{crit} (−0.069) is the smallest among all clusters, due to the interaction between average critical infrastructure availability and extremely low household-level access.

The uniform and relatively larger marginal effect of CI_{soc} underscores its consistent, though secondary, association with community vulnerability across all clusters.

The results also highlight critical insights for equitable infrastructure planning at the national level. As shown in Fig. 6, Clusters 1 and 4, where critical infrastructure availability (CI_{crit}) and essential utility access (CI_{ess}) exceed average levels (i.e., composite indicator values above zero, representing infrastructure characteristics exceeding the national mean), exhibit the highest marginal effects for CI_{crit} and CI_{soc}, reflecting synergistic interactions that correspond to lower predicted vulnerability. Prioritizing resource allocation to these clusters over infrastructure-deficient clusters like Clusters 0, 2, and 5 could further enhance infrastructure characteristics and result in reduced vulnerability in Clusters 1 and 4, but could widen regional disparities in both infrastructure access and community vulnerability. A similar pattern emerges for CI_{soc} investments in Clusters 2 and 5, where lower baseline infrastructure are linked to limited returns. Therefore, priority should be given to regions with low infrastructure provisions in infrastructure-deficient clusters which could lead to reduced spatial inequality and promote equitable development, even though the marginal reduction in vulnerability is lower compared to that in regions with improved infrastructure characteristics.

3.6. Policy implications

The results highlight the need for spatially differentiated infrastructure investment strategies to effectively reduce community vulnerability. The strong association with essential utility access underscores the importance of ensuring universal access to water, electricity, and sanitation, particularly in regions facing severe infrastructure deficits. For instance, expanding reliable household electricity supply in agrarian regions where power supply deficits are highest (for example, some parts of northern and eastern India) could simultaneously strengthen health and livelihood resilience by enabling irrigation and digital connectivity. Targeted investments in these areas can generate substantial local and regional benefits through spatial spillovers.

Moderate yet significant effects of critical infrastructure availability suggest that enhancing the reliability and spatial density of transport, energy, and communication systems can amplify the benefits of household-level service improvements. The observed interaction between regional infrastructure availability and household utility access further indicates that co-investments across these sectors may yield synergistic reductions in vulnerability, underscoring the importance of integrated and multi-scalar infrastructure planning across administrative boundaries. Infrastructure initiatives, such as improving all-weather rural roads under the *Pradhan Mantri Gram Sadak Yojana* (Wagale et al., 2020) and expanding safe water supply through the *Jal Jeevan Mission* (Department of Drinking Water and Sanitation, 2022), illustrate efforts that simultaneously strengthen regional connectivity and ensure household-level access. Such programs demonstrate the potential of coordinated investments that bridge network expansion with household-level access.

While the marginal returns on infrastructure investments may be lower in districts with severe deficits compared to more developed areas, such investments remain vital for promoting equitable community and regional development. Critical infrastructure networks also strengthen household-level utility reliability and indirectly support the functionality of social infrastructure and amenities. For instance, stable electricity and road access enhance the effectiveness of health and education facilities, improving long-term adaptive capacity. The relatively smaller but consistent effects of social infrastructure highlight the continued importance of investments in community facilities, such as schools and health centers, for long-term resilience building, even if their immediate impact on vulnerability is relatively modest. Finally, the strong spatial dependence observed in the model suggests that siloed, district- or state-level infrastructure interventions could be less effective. Therefore, coordinated multi-scalar planning, spanning local, district, and state levels, can maximize community-centric gains and ensure equitable development outcomes.

4. Conclusions

In the context of escalating climate risks and uneven development trajectories, this study presented an analysis of district-level community vulnerability in India, emphasizing the pivotal role of infrastructure access and availability. Leveraging open-source geospatial and socioeconomic data, we constructed composite indicators to assess how critical infrastructure availability, household utility access, and social infrastructure density interact and influence community vulnerability outcomes.

The results revealed that household access to essential services — such as electricity, drinking water, clean cooking fuel, and sanitation — had the most substantial relationship with reduced community vulnerability among the three infrastructure dimensions. This dimension alone was associated with the largest marginal reduction in community risk, underscoring the importance of ensuring reliable essential service delivery at the household level in shaping resilience. Availability of critical infrastructure networks also exhibited significant associations, particularly when combined with strong household utility access. The interaction between these two dimensions suggests that physical infrastructure systems and services do not operate in isolation; rather, their combined presence corresponds to universal, equitable household access. Social infrastructure, while comparatively less influential in the short term, remains an important correlate of community capabilities and lower vulnerability through its support for health, education, and public safety systems. However, resource allocation strategies must also consider cost-effectiveness and spatial disparities of infrastructure interventions (availability- or access-related) considering the local needs and deficits.

The distinct infrastructure profiles identified through clustering highlight the limitations of centralized, uniform infrastructure planning. These profiles emphasize the need for context-specific strategies that address the spatially varied and context-specific nature of community vulnerability. Within several states, districts with contrasting infrastructure characteristics reveal complex development outcomes, likely shaped by governance, geography, and historical investment patterns.

Availability and access of infrastructure systems not only provide capacity to absorb shocks and accelerate recovery during hazard events but, as the study suggests, also correspond to lower community vulnerability and greater resilience in peacetime. Collectively, the findings advocate tailoring infrastructure policy to local contexts to address deficits that may shape vulnerability to climate shocks and stresses. Enhanced household access to essential utilities emerges as an important correlate of resilience-building efforts, particularly in districts with persistent access gaps despite regional infrastructure availability. However, such improvements are most effective when coordinated with investments in critical infrastructure systems, such as transportation, energy, and water networks that underpin service delivery. Additionally, the results suggest that long-term resilience is linked to higher levels of social infrastructure density.

This study also demonstrated the feasibility and value of integrating diverse open datasets into a spatially disaggregated framework for assessing community vulnerability and resilience. As climate adaptation planning becomes more urgent, such data-driven approaches can inform targeted, equitable, and contextually appropriate infrastructure strategies that address the multidimensional nature of vulnerability.

Summarizing, this study provided valuable insights into the role of infrastructure access and availability in shaping community vulnerability, leveraging diverse datasets to inform resilience strategies. However, the results are derived from available datasets, which may not fully capture all dimensions of vulnerability. The following limitations contextualize the findings without diminishing their contributions:

1. Certain factors (e.g., employment, level of education, etc.) influencing district-level vulnerability may not be fully reflected in the clusters due to the unavailability of recent datasets.
2. The agglomerative clustering algorithm introduces a degree of randomness, such as through cross-validation, which may influence cluster assignments and composite indicator distributions. Nonetheless, the identified patterns remain robust for informing targeted policies, provided resilience features and composite indicators are analyzed in conjunction with clustering results.
3. Along with availability and access, reliability also play a significant role in community vulnerability. This aspect needs to be investigated to get deeper insights on how quality and reliability of infrastructure services affect households and its role in determining vulnerability under shocks and stresses.
4. The absence of panel data constrained the ability to establish causal relationships between infrastructure dimensions and community vulnerability. However, the observed associations provide a strong foundation for policy development and future longitudinal studies.

Future research could address these limitations by incorporating more recent and granular data, as well as exploring the integration of additional dimensions such as environmental and economic resilience. Nevertheless, the framework offers actionable insights for policymakers and urban planners, enabling them to identify and address specific infrastructural gaps.

CRediT authorship contribution statement

Srijith Balakrishnan: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Shivam Srivastava:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Chirag Kothari:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Jay Mestry who worked as an intern at Indian Institute of Technology Kanpur and partly contributed to the data collection and preprocessing efforts in this study.

Appendix. Supporting descriptions and results

A.1. Preprocessing of resilience and vulnerability features

The raw datasets were subjected to three sequential preprocessing steps to ensure suitability for developing infrastructure and vulnerability indicators. These steps were as follows:

1. District-based spatial referencing: Most datasets contained location information in non-standardized textual formats and were manually converted to spatially geocoded datasets. This process reconciled administrative boundary changes, particularly for districts bifurcated or unified in recent years, through manual matching prior to geocoding.
2. Handling missing data and outliers: Missing data were estimated using ordinary kriging, a spatial interpolation method that derived values based on spatial relationships among known observations. Ordinary kriging–based imputation was selected instead of simple mean or median methods, as it leverages spatial correlation between neighboring data points to generate statistically optimal estimates that maintain the spatial continuity of infrastructure and vulnerability variables. Estimated values replaced missing entries, while district values exceeding the 99th percentile for each feature were clipped to mitigate the influence of extreme outliers on the indicators.
3. Feature standardization: To enable comparisons across districts with varying sizes and populations, raw features expressed as absolute counts were normalized. Normalization divided values by district population or area, depending on the feature, yielding per capita or density measures that reflected relative intensities across districts.

A.2. Principal component analysis (PCA) procedure

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms correlated variables into a set of uncorrelated principal components while preserving as much variance as possible. PCA is then applied to these standardized indicators, transforming them into uncorrelated principal components that capture the maximum variance in the data. PCA is applied to standardized indicators to transform them into uncorrelated principal components that capture the maximum variance in the data as follows:

$$\mathbf{Y} = \mathbf{Z}\mathbf{v} \tag{A.1}$$

where $\mathbf{Y}$ consists of the principal components and $\mathbf{v}$ is the eigenvector matrix. The eigenvalues (λ) and the component loadings ($\ell_{ij} = v_{ij}\sqrt{\lambda_j}$) are also computed in the same step. We then selected first k components that explain at least 95% of the total variance (proportion of variance explained) for further analysis and modeling.

Now, the weights for each original indicator (Z) are derived from the variance-weighted loadings of the principal component contributions as follows:

$$W_i = \frac{\sum_{j=1}^{k} |\ell_{ij}| \cdot \lambda_j}{\sum_{i=1}^{p}\sum_{j=1}^{k} |\ell_{ij}| \cdot \lambda_j} \tag{A.2}$$

where W_i is the weight for the ith indicator, ℓ_{ij} is the loading of the ith indicator on the kth component, λ_k is the explained variance ratio of the kth component, and p is the total number of indicators assigned to a resilience pillar.

Finally, composite indicators (CI) for each of the resilience pillars (dimensions) is computed as the weighted average of the constituent indicator values.

$$CI = \sum_{i=1}^{p} W_i Z_i \tag{A.3}$$

A.3. Cluster evaluation metrics

To determine the optimal number of clusters, we calculated three key metrics: the Silhouette Score (S), the Calinski-Harabasz Index (CH), and the Within-Cluster Sum of Squares (WCSS). The Silhouette Score quantifies how similar an object is to its own cluster compared to other clusters and is defined as:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{A.4}$$

where $a(i)$ is the average distance between i and all other points in the same cluster, and $b(i)$ is the average distance between i and all points in the nearest cluster. The overall silhouette score is computed as:

$$S_{\text{avg}} = \frac{1}{n}\sum_{i=1}^{n} S(i) \tag{A.5}$$

where n is the total number of points. The Calinski-Harabasz Index measures cluster quality by comparing variance between clusters to variance within clusters:

$$CH = \frac{B_k/(k-1)}{W_k/(n-k)} \tag{A.6}$$

where B_k is the between-cluster dispersion, W_k is the within-cluster dispersion, k is the number of clusters, and n is the total number of observations.

Additionally, we applied the elbow method to determine an appropriate number of clusters by calculating WCSS for varying values of k. WCSS measures how well data points are clustered around their respective centroids and is defined as:

$$\text{WCSS} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \text{distance}(x_j^{(i)}, c_i)^2 \tag{A.7}$$

where $\text{distance}(x_j^{(i)}, c_i)$ represents the distance between the jth data point in cluster i and its centroid c_i. We plotted WCSS against different values of k, identifying an "elbow" point where adding more clusters yields diminishing returns in explained variance.

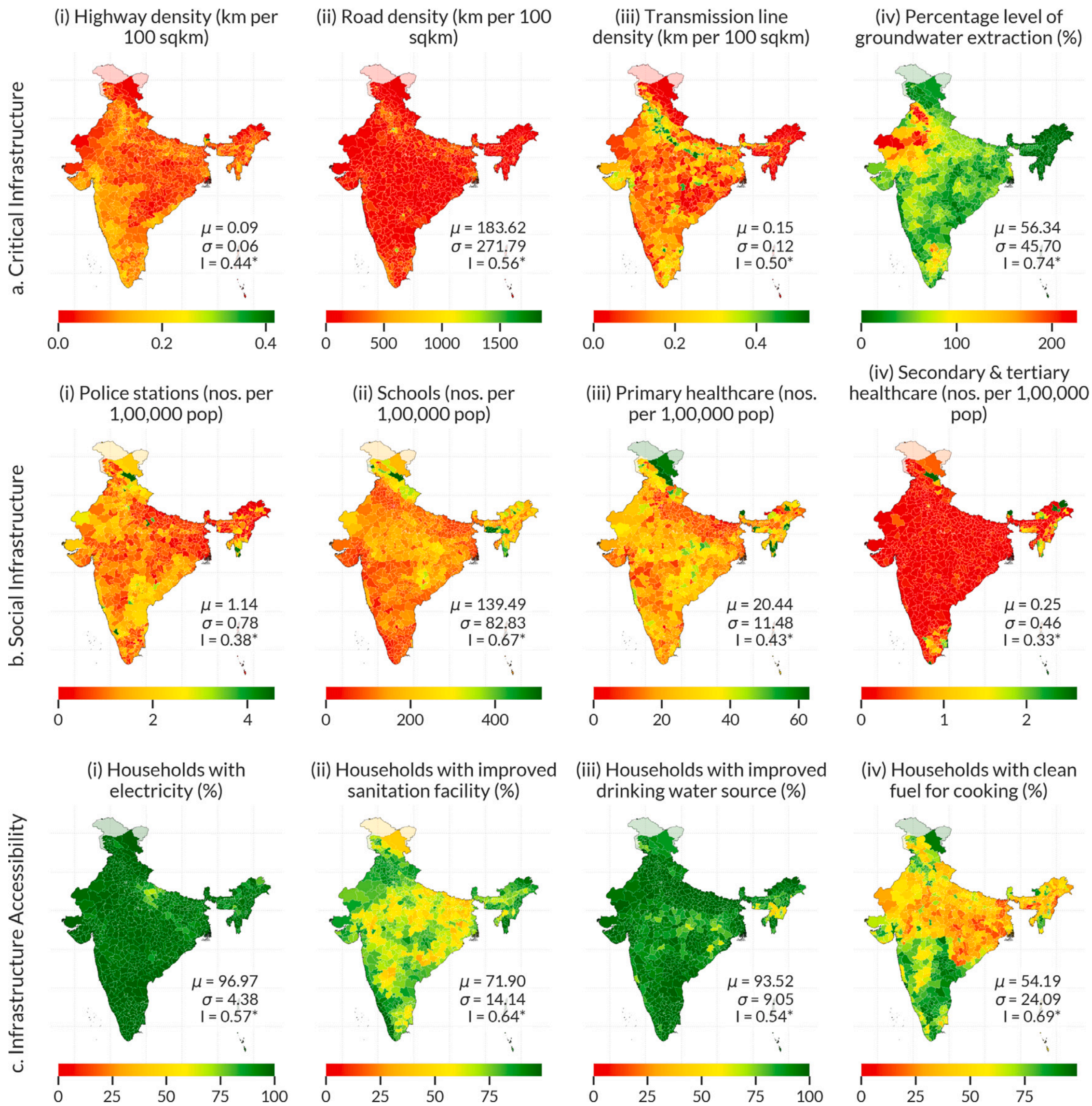

Fig. A.7. Spatial distribution of various infrastructure characteristics and community vulnerability. μ and σ are the mean and standard deviation, I is the Moran's I coefficient, and the asterisk sign (*) indicates its statistical significance at $\alpha = 0.05$).

A.4. District-level indicators for vulnerability and infrastructure provisions

Fig. A.7 shows the spatial distributions of features which are used to define infrastructure provisions and community vulnerability.

The maps also provide descriptive statistics, such as the mean (μ) and standard deviation (σ) of the features. Additionally, I represents Moran's I, a commonly used measure of spatial autocorrelation, which indicates whether there are spatial patterns in the feature distribution (Moran, 1950).

For evaluating the critical infrastructure dimension, four features are considered (Fig. A.7a): highway density ($\mu = 0.09$, $\sigma = 0.06$), road density ($\mu = 183$, $\sigma = 273$), power transmission line density ($\mu = 0.15$, $\sigma = 0.12$), all measured in $\mathrm{km}/100\ \mathrm{m}^2$, and percentage of groundwater extraction ($\mu = 56.34$, $\sigma = 45.70$). The maps indicate that both features have higher values in more urbanized districts (denoted by yellow or green colors). The spatial autocorrelation (Moran's I) values exceed 0.50, indicating that the both critical infrastructure variables exhibit moderately high spatial clustering.

Social infrastructure resilience is quantified using four density variables (all measured per 1,00,000 population; Fig. A.7b) related to social infrastructure systems: police stations ($\mu = 1.14; \sigma = 0.78$), schools ($\mu = 139.54; \sigma = 83.08$), primary healthcare facilities ($\mu = 20.44; \sigma = 11.49$), and secondary and tertiary healthcare facilities ($\mu = 0.25; \sigma = 0.46$). The maps reveal significant spatial disparities, with higher infrastructure density observed in more urbanized regions (denoted by yellow and green colors). Spatial autocorrelation (Moran's I) values for all features range between 0.32 and 0.67, indicating that social infrastructure variables exhibit moderate to high spatial clustering.

The infrastructure access dimension is captured through four variables (Fig. A.7d), all measured as percentages of households: households with electricity ($\mu = 96.97\%; \sigma = 4.38\%$), households with improved sanitation facilities ($\mu = 71.90\%; \sigma = 14.14\%$), households with improved drinking water sources ($\mu = 93.52\%; \sigma = 9.05\%$), and households with clean fuel for cooking ($\mu = 54.19\%; \sigma = 24.09\%$). The strong spatial autocorrelations (Moran's I values ranging from 0.54 to 0.69) across features indicate that infrastructure access patterns are also driven by regional factors.

The community dimension includes both resilience and vulnerability features, captured through seven variables (Fig. A.7c). Resilience features include per capita income (μ = INR 144, 820; σ = INR 87, 880), the proportion of women (aged 15–49) who are literate ($\mu = 74.37\%; \sigma = 12.09\%$), fully vaccinated children aged 12–23 months

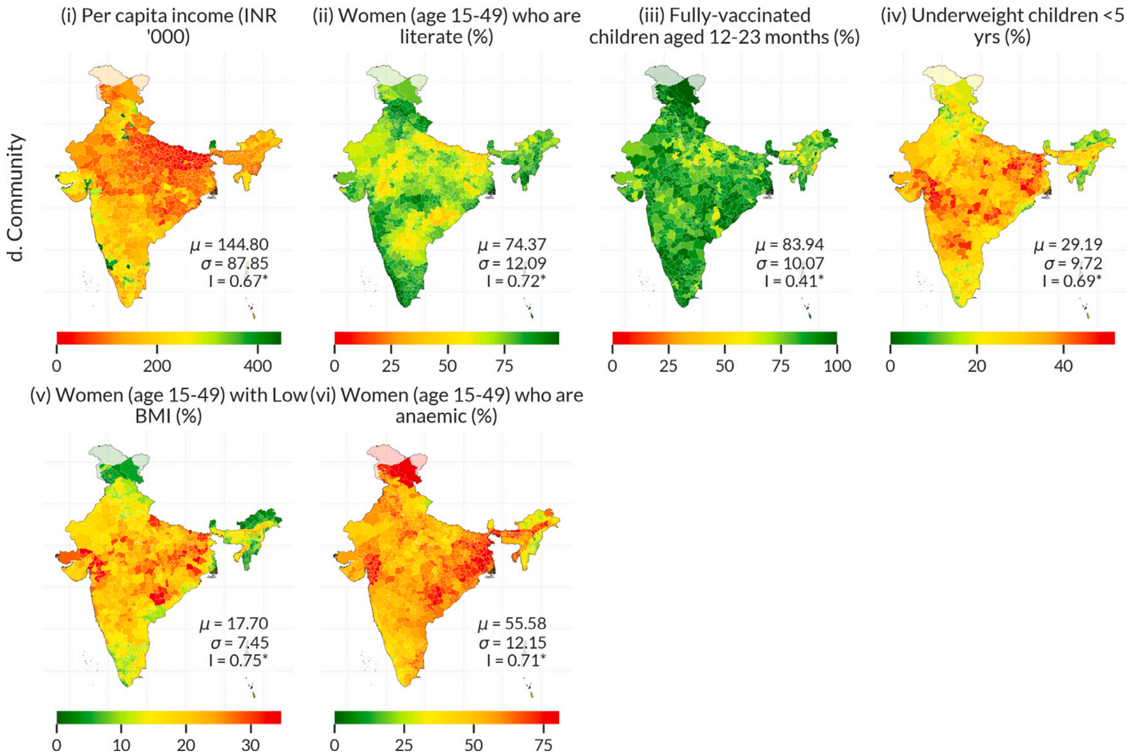

Fig. A.7. (*continued*).

Table A.5
Updated spatial lag model (SLM) and coefficients.

Variable	Coefficient	Std. Error	z-stat	p-value
Intercept (Reference state–climate zone category)	0.1149	0.0778	1.477	0.140
CI_{crit}	−0.0760	0.0315	−2.416	0.016*
CI_{ess}	−0.3846	0.0238	−16.157	<0.001***
CI_{soc}	−0.0452	0.0215	−2.096	0.036*
Interaction effects				
$CI_{crit} \times CI_{ess}$	−0.1040	0.0379	−2.740	0.006**
$S\sum w_{dd'} CI_{comm,d'}$	0.5497	0.0288	19.116	<0.001***
Model Statistics				
Observations (districts)				729
Pseudo R^2				0.8836
Spatial pseudo R^2				0.8393
Log-Likelihood				−58.85
AIC				221.70
Moran's I (residuals)				0.0333 (p = 0.074)

- Notes: Standard errors are computed using Maximum Likelihood estimation.
- Coefficients of combined state–climate zone variables not shown.
- $CI_{comm}, CI_{ess}, CI_{crit}, CI_{soc}$ denote composite indicators for community vulnerability, household essential utility access, critical infrastructure availability, and social infrastructure density, respectively. $S\sum w_{dd'} CI_{comm,d'}$ represents the spatial lag of CI_{comm}.

* Significance level: $p < 0.05$.
** Significance level: $p < 0.01$.
*** Significance level: $p < 0.001$.

($\mu = 83.94\%; \sigma = 10.07\%$), and households with health insurance or financial schemes ($\mu = 40.25\%; \sigma = 22.80\%$). Vulnerability features include the proportion of the population aged under 15 years ($\mu = 26.32\%; \sigma = 5.13\%$), underweight children aged under 5 years ($\mu = 29.19\%; \sigma = 9.72\%$), and women (aged 15–49) with low BMI ($\mu = 17.70\%; \sigma = 7.45\%$). The results show that all features, except 'fully vaccinated children,' exhibit strong spatial autocorrelation (Moran's I coefficient ranging between 0.69 and 0.75), indicating that regional factors play a significant role in shaping the resilience and vulnerability profiles.

A.5. Correlation among resilience and vulnerability variables (Kendall's τ)

Kendall's τ is a non-parametric statistical test used to measure the ordinal association between two variables. It quantifies the similarity in the ordering of data points, indicating how consistently two variables move together. The Kendall's τ test is robust against the presence of outliers and non-normally distributed data. Fig. A.8 shows the Kendall's correlation coefficients for all features used in the construction of the composite indicators. A τ value closer to 1 indicates that the ordinal relationship between the ranks of two features is monotonically positive whereas a value close to −1 indicates a monotonically

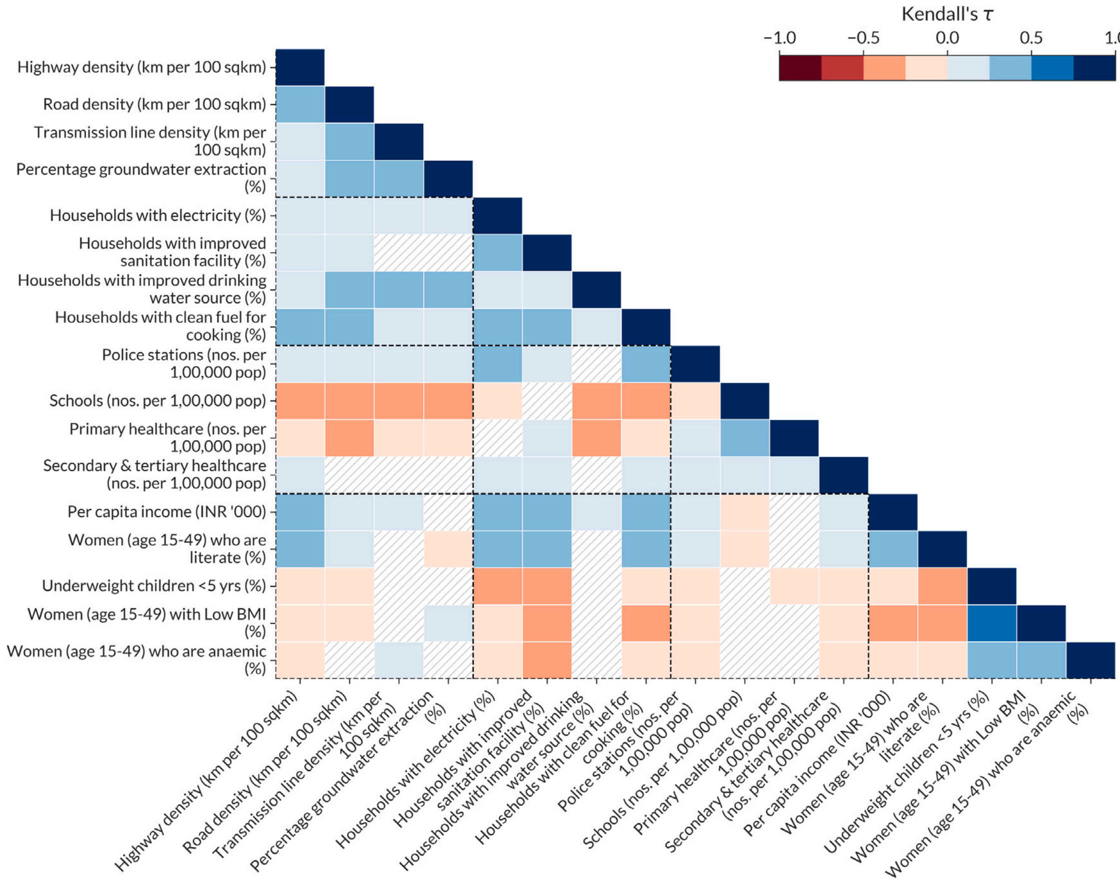

Fig. A.8. Kendall's τ for various infrastructure characteristics and community vulnerability feature pairs.

Table A.6
Feature weights derived using PCA for developing composite indicators.

Composite indicator	Feature	Weight
Critical infrastructure availability	Highway density (km per 100 sq.km.)	0.258
	Road density (km per 100 sq.km.)	0.241
	Transmission line density (km per 100 sq.km.)	0.254
	Percentage level of groundwater extraction (%)	0.247
Social infrastructure density	Police stations (nos. per 100,000 pop)	0.236
	Schools (nos. per 100,000 pop)	0.269
	Primary healthcare (nos. per 100,000 pop)	0.240
	Secondary & tertiary healthcare (nos. per 100,000 pop)	0.255
Household essential utility access	Households with electricity (%)	0.261
	Households with improved sanitation facility (%)	0.277
	Households with improved drinking water source (%)	0.230
	Households with clean fuel for cooking (%)	0.233
Community vulnerability	Per capita income (INR '000)	0.206
	Women (age 15-49) who are literate (%)	0.198
	Underweight Children <5 yrs (%)	0.199
	Women (age 15-49) with Low BMI (%)	0.193
	Women (age 15-49) who are anaemic (%)	0.204

negative relationship. A value of 0 indicates that there exists no clear relationship.

A.6. Updated spatial lag model with only significant terms

Table A.5 presents the updated results of the spatial lag model (SLM) used to assess the association between community vulnerability and infrastructure provisions. The model retains only the significant interaction between critical infrastructure availability (CI_{crit}) and essential utility access (CI_{ess}), confirming that districts with jointly higher levels of these provisions exhibit substantially lower vulnerability. All main effects remain significant and negative, indicating that greater access to critical, essential, and social infrastructure is consistently associated with reduced vulnerability. The spatial lag coefficient is positive and highly significant, suggesting that community vulnerability exhibits moderate spatial dependence—districts tend to resemble neighboring areas in their vulnerability levels. The model achieves a high pseudo R^2 (0.88), and the Moran's I of residuals indicates minimal remaining spatial autocorrelation, supporting the robustness of the specification.

A.7. Feature weights used for composite indicators using PCA

The PCA weights used to construct the composite community vulnerability and resilience indicators are presented in Table A.6.

Data availability

All datasets and reproducible codes (analyses) can be accessed from the following GitHub repository: https://github.com/srijithbalakrishnan/india_resilience. Raw data will be made available on request.

References

Aboda, C., Mugagga, F., Byakagaba, P., & Nabanoga, G. (2019). Development induced displacement; a review of risks faced by communities in developing countries. *Sociology and Anthropology*, *7*(2), 100–110. http://dx.doi.org/10.13189/sa.2019.070205.

Aghababaei, M., & Koliou, M. (2023). Community resilience assessment via agent-based modeling approach. *Computer-Aided Civil and Infrastructure Engineering*, *38*(7), 920–939.

Anderson, M. J., Brunner, L., & Logan, T. (2025). Functional isolation: The compounding burden on community capacity from cascading infrastructure failures and access disruption. *International Journal of Disaster Risk Reduction*, *118*, Article 105190.

Ascher, W., & Krupp, C. (2010). Rethinking physical infrastructure development. In *Physical infrastructure development: balancing the growth, equity, and environmental imperatives* (pp. 1–33). Springer.

Balakrishnan, S., Jin, L., Cassottana, B., Costa, A., & Sansavini, G. (2024). Developing resilience pathways for interdependent infrastructure networks: A simulation-based approach with consideration to risk preferences of decision-makers. *Sustainable Cities and Society*, *115*, Article 105795.

Balakrishnan, S., Lim, T., & Zhang, Z. (2022). A methodology for evaluating the economic risks of hurricane-related disruptions to port operations. *Transportation Research Part A: Policy and Practice*, *162*, 58–79.

Balakrishnan, S., & Zhang, Z. (2018). Developing priority index for managing utility disruptions in urban areas with focus on cascading and interdependent effects. *Transportation Research Record*, *2672*(1), 101–112.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., & von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra*, *19*(4), 733–752.

Calderón, C., & Servén, L. (2010). Infrastructure and economic development in sub-saharan africa. *Journal of African Economies*, *19*(Suppl. 1), i13–i87.

Camacho, C., Bower, P., Webb, R. T., & Munford, L. (2023). Measurement of community resilience using the baseline resilience indicator for communities (BRIC) framework: A systematic review. *International Journal of Disaster Risk Reduction*, *95*, Article 103870.

CDRI (2023). *Global infrastructure resilience: Capturing the resilience dividend - A biennial report from the coalition for disaster resilient infrastructure*: *Technical report*, New Delhi, India: Coalition for Disaster Resilient Infrastructure, URL https://doi.org/10.59375/biennialreport.ed1. (Accessed 29 September 2025).

CGWB (2024). *National compilation on dynamic ground water resources of India, 2024*: *Technical report*, Faridabad, India: Central Ground Water Board, Department of Water Resources, River Development & Ganga Rejuvenation, Ministry of Jal Shakti, Government of India, URL https://cdnbbsr.s3waas.gov.in/s3a70dc40477bc2adceef4d2c90f47eb82/uploads/2024/12/20241231588319401.pdf. (Accessed 25 June 2025).

Chakraborty, M. (2024). Industrial clustering and location in India: Sectoral patterns of investments and employments. *Regional Science Policy & Practice*, *16*(6), Article 100041.

Chang, S. E. (2016). Socioeconomic impacts of infrastructure disruptions. In *Oxford research encyclopedia of natural hazard science*.

Chen, L., Chen, M., Zhang, X., & Xian, Y. (2024). Evaluating inequality divides in urban development intensity between the global north and south. *Land Use Policy*, *145*, Article 107291.

Chen, T., Radke, J. D., Lang, W., & Li, X. (2020). Environment resilience and public health: Assessing healthcare's vulnerability to climate change in the san Francisco Bay Area. *Growth and Change*, *51*(2), http://dx.doi.org/10.1111/grow.12379.

Coleman, N., Esmalian, A., & Mostafavi, A. (2020). Equitable resilience in infrastructure systems: empirical assessment of disparities in hardship experiences of vulnerable populations during service disruptions. *Natural Hazards Review*, *21*(4), Article 04020034.

Comfort, L. K., Boin, A., & Demchak, C. C. (2010). *Designing resilience: Preparing for extreme events*. University of Pittsburgh Pre.

Conrad, P., Creitaru, I., & United Nations Development Programme (2024). Getting ahead of the crisis curve: UNDP's approach to multi-hazard early warning and preparedness. UNDP Practice Note, September 2024, https://www.undp.org/sites/g/files/zskgke326/files/2024-09/undp-getting-ahead-of-the-crisis-curve.pdf.

Corscadden, L., Levesque, J., Lewis, V., Strumpf, E., Breton, M., & Russell, G. (2018). Factors associated with multiple barriers to access to primary care: an international analysis. *International Journal for Equity in Health*, *17*(1), 28.

Cutter, S. L., & Emrich, C. T. (2017). *Social vulnerability index (SoVI®): Methodology and limitations* (pp. 1–6). Columbia, SC: Hazards and Vulnerability Research Institute.

Department of Drinking Water and Sanitation (2022). National report: Functionality assessment of household tap connection - 2022. *Technical report*, New Delhi, India: Department of Drinking Water and Sanitation, Ministry of Jal Shakti, Government of India, Prepared by HTA-Kantar Public for the National Jal Jeevan Mission, URL https://jaljeevanmission.gov.in/sites/default/files/2022-10/national_report_of_functionality_assessment_2022.pdf.

Duenas-Osorio, L., & Vemuru, S. M. (2009). Cascading failures in complex infrastructure systems. *Structural Safety*, *31*(2), 157–167.

ESRI India (2023). India police station locations 202. Dataset, ESRI India, URL https://www.arcgis.com/home/item.html?id=35371bda12c74ae39b19e5c45c91cc19. (Accessed 5 November 2024).

ESRI India (2024). India: Road network bharatmala. Dataset, ESRI India, URL https://www.arcgis.com/home/item.html?id=e33dfe4d58ec4a0c9c331a6d92894ba5. (Accessed 12 November 2024).

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, *110*, Article 104743.

Fang, Y. p., Zhao, C., Rasul, G., & Wahid, S. M. (2016). Rural household vulnerability and strategies for improvement: An empirical analysis based on time series. *Habitat International*, *53*, 254–264.

Feofilovs, M., Romagnoli, F., Gotangco, C. K., Josol, J. C., Jardeleza, J. M. P., Litam, J. E., Campos, J. I., & Abenojar, K. (2020). Assessing resilience against floods with a system dynamics approach: a comparative study of two models. *International Journal of Disaster Resilience in the Built Environment*, *11*(5), 615–629.

Firdaus, S., & Uddin, M. A. (2015). A survey on clustering algorithms and complexity analysis. *International Journal of Computer Science Issues (IJCSI)*, *12*(2), 62.

Fox, S., Agyemang, F., Hawker, L., & Neal, J. (2024). Integrating social vulnerability into high-resolution global flood risk mapping. *Nature Communications*, *15*(1), 3155.

George, A., & Sharma, P. (2022). Socioeconomic and infrastructural vulnerability of Indian population: a district level study. *GeoJournal*, *88*(2), 1841–1871. http://dx.doi.org/10.1007/s10708-022-10712-6.

Guidotti, R., Gardoni, P., & Rosenheim, N. (2019). Integration of physical infrastructure and social systems in communities2019 reliability and resilience analysis. *Reliability Engineering & System Safety*, *185*, 476–492. http://dx.doi.org/10.1016/j.ress.2019.01.008.

Guo, C., Zhou, W., Jing, C., & Zhaxi, D. (2024). Mapping and measuring urban-rural inequalities in accessibility to social infrastructures. *Geography and Sustainability*, *5*(1), 41–51.

Hallegatte, S., Rentschler, J., & Rozenberg, J. (2019). *Lifelines: The resilient infrastructure opportunity*. Washington, DC: World Bank, http://dx.doi.org/10.1596/978-1-4648-1330-6, URL https://openknowledge.worldbank.org/handle/10986/31805.

Han, X., Su, J., & Thia, J. P. (2019). Impact of infrastructure investment on developed and developing economies. Working Paper, Asian Infrastructure Investment Bank (AIIB), URL https://www.aiib.org/en/news-events/media-center/working-papers/pdf/2019_Infrastructure-Impact_AIIB-Working-Paper-Revision.pdf. (Accessed 10 August 2025).

Hartwig, T., & Nguyen, T. T. (2023). Local infrastructure, rural households' resilience capacity and poverty: evidence from panel data for southeast Asia. *Journal of Economics and Development*, *25*(1), 2–21.

Hassan, E. M., & Mahmoud, H. (2021). Healthcare and education networks interaction as an indicator of social services stability following natural disasters. *Scientific Reports*, *11*(1), 1664.

Henshel, D. S., & Ashby, J. L. (2023). Rethinking critical infrastructure in the United States from a community-based perspective. *PLoS Climate*, *2*(4), Article e0000178. http://dx.doi.org/10.1371/journal.pclm.0000178.

Jaafari, A., Mafi-Gholami, D., & Choubin, B. (2025). Spatiotemporal dynamics of social vulnerability to natural hazards: Trends and projections from 2002 to 2030 in northwestern Iran. *Sustainable Cities and Society*, *120*, Article 106172.

Koliou, M., van de Lindt, J. W., McAllister, T. P., Ellingwood, B. R., Dillard, M., & Cutler, H. (2018). State of the research in community resilience: progress and challenges. *Sustainable and Resilient Infrastructure*, *5*(3), 131–151. http://dx.doi.org/10.1080/23789689.2017.1418547.

Kumar, A. (2015). Rural households' access to basic amenities in India: Deprivation and socio-economic exclusions. *Social Change*, *45*(4), 561–586.

Mahapatra, B., Chaudhuri, T., & Saggurti, N. (2022). Climate change vulnerability, and health of women and children: Evidence from India using district level data. *International Journal of Gynecology & Obstetrics*, *160*(2), 437–446. http://dx.doi.org/10.1002/ijgo.14515.

Mathur, S. (2024). *Development charges: Funding urban infrastructure in India and the global south*. Cambridge University Press.

McGlade, J., Bankoff, G., Abrahams, J., Cooper-Knock, S., Cotecchia, F., Desanker, P., Erian, W., Gencer, E., Gibson, L., Girgin, S., et al. (2019). *Global assessment report on disaster risk reduction 2019*. UN Office for Disaster Risk Reduction.

MoE (2023). Unified district information system for education plus (UDISE+). Dataset, Government of India, URL https://udiseplus.gov.in. (Accessed 7 January 2025).

MoHFW (2018). National hospital directory with geo code and additional parameters. Dataset, Government of India, URL https://www.data.gov.in/resource/national-hospital-directory-geo-code-and-additional-parameters-updated-till-last-month. (Accessed 12 November 2024).

MoHFW (2019). All India health centres directory. Dataset, Government of India, URL https://www.data.gov.in/resource/all-india-health-centres-directory-7th-october-2016. (Accessed 12 November 2024).

MoHFW (2022). *National family health survey (NFHS - 5), 2019–21: India report vol II: Technical report*, New Delhi, India: Ministry of Health and Family Welfare, Government of India, URL https://pdf.usaid.gov/pdf_docs/PA00ZBQT.pdf. (Accessed 7 November 2024).

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17–23.

Mundle, S., Chowdhury, S., & Sikdar, S. (2016). Governance performance of Indian states: Changes between 2001-02 and 2011—12. *Economic and Political Weekly*, 55–64.

National Infrastructure Advisory Council (2010). A framework for establishing critical infrastructure resilience goals: Final report and recommendations. U.S. Department of Homeland Security, October 19, 2010, https://www.dhs.gov/xlibrary/assets/niac/niac-a-framework-for-establishing-critical-infrastructure-resilience-goals-2010-10-19.pdf.

Nicoletti, L., Sirenko, M., & Verma, T. (2023). Disadvantaged communities have lower access to urban infrastructure. *Environment and Planning B: Urban Analytics and City Science*, *50*(3), 831–849.

Nowell, B., Bodkin, C. P., & Bayoumi, D. (2017). Redundancy as a strategy in disaster response systems: A pathway to resilience or a recipe for disaster? *Journal of Contingencies and Crisis Management*, *25*(3), 123–135.

OpenStreetMap (2024). OpenStreetMap data for India. Dataset, https://download.geofabrik.de/asia/india.html. (Accessed 12 November 2024).

Pandey, B., Brelsford, C., & Seto, K. C. (2022). Infrastructure inequality is a characteristic of urbanization. *Proceedings of the National Academy of Sciences*, *119*(15), Article e2119890119.

Pandey, B., Brelsford, C., & Seto, K. C. (2025). Rising infrastructure inequalities accompany urbanization and economic development. *Nature Communications*, *16*(1), 1193.

Puka, L. (2011). Kendall's tau. In *International encyclopedia of statistical science* (pp. 713–715). Springer.

Rendon, C., Osman, K. K., & Faust, K. M. (2021). Path towards community resilience: Examining stakeholders' coordination at the intersection of the built, natural, and social systems. *Sustainable Cities and Society*, *68*, Article 102774.

Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine*, *21*(6), 11–25.

Rives, J. M., & Heaney, M. T. (1995). Infrastructure and local economic development. *Journal of Regional Analysis and Policy*, *25*(1), 58–73.

Saxena, M., Chotia, V., & Rao, N. M. (2018). Estimating the efficiency of public infrastructure investment: A state-wise analysis. *Global Business Review*, *19*(4), 1037–1049.

Smit, S., Musango, J. K., Kovacic, Z., & Brent, A. C. (2017). Conceptualising slum in an urban african context. *Cities*, *62*, 107–119. http://dx.doi.org/10.1016/j.cities.2016.12.018.

Song, W., Li, Y., Cheng, J., Chen, R., Wu, J., & Jia, N. (2024). Enhancing social vulnerability assessment with energy resilience: A comprehensive study of the netherlands. *Sustainable Cities and Society*, *103*, Article 105251.

Suharto, S., Lukiyanto, K., Ali, H., Atutik, W. S., Aswirawan, M. Y. M. S. K., et al. (2025). Infrastructure development inequality: When big projects sacrifice local acess. *Oikonomia: Journal of Management Economics and Accounting*, *2*(2), 72–83.

Taghizadeh-Hesary, F., Sarker, T., Yoshino, N., Mortha, A., & Vo, X. V. (2021). Quality infrastructure and natural disaster resiliency: A panel analysis of Asia and the Pacific. *Economic Analysis and Policy*, *69*, 394–406.

Tripathy, N., Srikanth, M., & Aravalath, L. M. (2016). Infrastructure investment and economic growth: Evidence from India. *Journal of International Business and Economy*, *17*(1), 91–111.

Tu, Y., Chen, B., Liao, C., Wu, S., An, J., Lin, C., Gong, P., Chen, B., Wei, H., & Xu, B. (2025). Inequality in infrastructure access and its association with health disparities. *Nature Human Behaviour*, 1–14.

Valinejad, J., Mili, L., & Van Der Wal, C. N. (2022). Multi-agent based stochastic dynamical model to measure community resilience. *Journal of Social Computing*, *3*(3), 262–286.

Verschuur, J., & Balakrishnan, S. (2024). Brief communication: Evidence of the impacts of climate extremes on power system outages in India. *EGUsphere*, *2024*, 1–14.

Wagale, M., Singh, A. P., & Sarkar, A. K. (2020). Impact of rural road construction on the local livelihood diversification: evidence from Pradhan Mantri Gram Sadak Yojana in Jhunjhunu district, India. *GeoJournal*, *85*(4), 961–978.

Wu, R. M., Zhang, Z., Yan, W., Fan, J., Gou, J., Liu, B., Gide, E., Soar, J., Shen, B., Fazal-E-Hasan, S., Liu, Z., Zhang, P., Wang, P., Cui, X., Peng, Z., & Wang, Y. (2022). A comparative analysis of the principal component analysis and entropy weight methods to establish the indexing measurement. *PLoS One*, *17*, Article e0262261. http://dx.doi.org/10.1371/JOURNAL.PONE.0262261.

Yadav, V., & Kalambe, R. N. (2022). *Corridor development in India: impact on land acquisition*. Routledge India.

Yang, Y., Ng, S. T., Xu, F. J., & Skitmore, M. (2018). Towards sustainable and resilient high density cities through better integration of infrastructure networks. *Sustainable Cities and Society*, *42*, 407–422.

Zhang, Q., Li, C., Wen, D., Kang, J., Chen, T., Zhang, B., Hu, Y., Yin, J., & Slater, L. (2025). Global south shows higher urban flood exposures than the global north under current and future scenarios. *Communications Earth & Environment*, *6*(1), 594.

Zhou, Y., Li, X., Chen, W., Meng, L., Wu, Q., Gong, P., & Seto, K. C. (2022). Satellite mapping of urban built-up heights reveals extreme infrastructure gaps and inequalities in the global south. *Proceedings of the National Academy of Sciences*, *119*(46), Article e2214813119.