Title of the Paper

The Baseline of Fragility and Capability: Hedging Socio-Environmental Impacts of Char Dham Expressway in Uttarakhand, India.

Author Details

Name: Saransh Kumar

E mail: saransh@nias.res.in

Affiliation: Manipal Academy of Higher Education (MAHE);

National Institute of Advances Studies, Bengaluru.

The Baseline of Fragility and Capability: Hedging Socio-Environmental Impacts of Char Dham Expressway in Uttarakhand, India.

Abstract

Conventional socio-environmental impact assessments (SEIA) generally rely on linear evaluations that fail to accommodate the complex, multi-scalar realities of infrastructure expansion. In ecologically fragile regions, development phases frequently stall or overlap, creating a temporal parallelity, which disrupts standard temporal baselines required for accurate longitudinal tracking. To navigate this uncertainty, this study examines the Char Dham Mahamarg project in the Uttarakhand Himalayas. It proposes a methodological convergence by integrating Geographic Information Systems (GIS) with a TabNet machine learning (ML) architecture. Using this socially informed and ML automated architecture, the study mapped cross-thematic interactions among 49 socio-economic indicators across 463 local households.

Rather than forcing a static baseline onto a highly dynamic system, this framework accounts for the chaotic sequencing of actual construction. The model classified household capabilities with 80% accuracy, revealing that reserve land assets, educational access, and economic mobility act as non-linear drivers for registering community resilience. When spatialized against environmental time-series data, these capability scores reveal latent impact trajectories, visible across the physical landscape. Moving away from procedural stock-taking, this approach provides a diagnostic tool for policy makers to evaluate shifting vulnerabilities and support anticipatory planning in data-scarce developmental contexts.

Keywords: Development, Machine Learning, Spatial Analysis, Socio-Environmental Impact Assessment, Temporal Parallelity.

1. Introduction

Contemporary development challenges, inherent to modern infrastructure projects, are driving a necessary convergence between sustainability paradigms and Impact Assessment (IA) studies. As an aspired principal, sustainability is not realized as an end goal but an adaptive philosophy critical for capacity addition. It entails that true monitoring of progress cannot stay static. In cases of large-scale infrastructures, it requires dynamic indicators, indicative of interactive state of systems. As a result, integrated indicators act not only as a composite measure but necessary methodological choice in understanding coupled human and physical systems.

Although linear in design, infrastructures such as transport networks introduce complex, non-linear disturbances across the landscapes they traverse. They occasionally modify fragile environments through habitat fragmentation, induced structural instabilities, and shifting human-nature dynamics. In such cases, interactions among multiple domains of development require assessment designs that are communicative, not exclusive. Recent focus on integrated IA models such as Socio-Environmental Impact Assessments (SEIAs) recognise thematic indicators as better represenation of compound realities. Traditionally organised as binary domains namely, the Environmental Impact Assessments (EIA) and Social Impact Assessment (SIA), they merge their respective evaluative baskets to move beyond outcome-oriented polaroids (Morrison-Saunders et al., 2021). These are increasingly tasked with probing the dynamic nature of development challenges, in order to engineer consistent and resilient resolutions (Slootweg and Jones, 2020).

However, the effective functioning of these studies is often circumscribed by rigid regional regulations. A reasoned scientific requirement is deemed necessary only to align them with regulatory frameworks. Consequently, in contexts characterized by limited desire to expand institutional capacity for independent formulation, development proponents often prioritize 'templated', 'stencil', or 'isomorphic' framework designs (DiMaggio & Powell, 1983). Although studies show their functional merits such as professionalism, standardization and simplification, especially in regions where assessment practice is nascent, there is growing concern for manipulating these studies for gaining credence for good practice beyond just legal consents. This effectively reduces the process to a procedural obstacle rather than a substantive diagnostic tool (Paliwal, 2006).

1.1. Constriction of Indicators in regulatory spaces

Regulators and the development agencies responsible for interventions often share organizational interdependencies. Project finance, stake holding in development product, mutual ideocracies generally advocates component replication. Studies () indicate these practices developing into an assessment crisis, Consequently, specific development projects in geographically varied (and variegated), ecologically fragile, and culturally (in)significant[1] expanses use context independent indicators which reduces their reliability. Particularly in India, development imperatives outpace assessment mechanisms and the methodical procedures dependent on such measures cumulate in

[1] Cultural Insignificance in context of development projects in accessibility scarce geographies is understood not as a blatant disregard, but as a dilution of identity, values, meanings as lived experiences of the 'project affected people'. It increases in spaces where heterodox praxis of development drives the social influence. In addition, the awareness creation and sharing are centripetal or inward.

forms of environmental resistance from civil society and stakeholders. This further politicizes policy led impact management in later stages.

Continued reliance on traditional frameworks in India results in a reductionist understanding. The chosen indicator template privilege certain themes of development while reducing the weightage of others. This is witnessed in favoring environmental or economic outcomes while largely externalizing the social, geographical, and temporal dimensions of change (Penadés-Plà et al., 2020; Woldesenbet et al., 2024). Even when these frameworks incorporate multi-dimensional indicators, they fall short in capturing existing non-linearities and spatial fragmentation of impacts across the life cycle of infrastructure interventions (Zhang et al., 2021). Not engaging a holistic scenario in such cases invoke policy disjunction (Nita et al., 2024), abatement parallels for the product, misrepresentation of loss (Castelblanco et al., 2023), community alienation (Mambiravana & Umejesi, 2023), and skewed resource accumulation (Blue et al., 2021; Wang et al., 2023).

1.2. Existing inclusiveness in IA based comprehensive models.

This, however, does not discount frameworks that advocate comprehensiveness. Cumulative, Collective, and Strategic assessments have been conceived around inclusive principles and practices with a notion of epistemic closure. These efforts towards holistic assessments have been driven by systems-based approaches (Alan-Ehrlich, 2022) and chaotic embeddedness (Preiser et al., 2018) that extend beyond subsystems, often (re)producing dependencies unique to parent vectors. They have also witnessed targeted spatiality, such as eco-sensitive zones or developments intersecting with spaces of indigeneity and exoticness, where assessments focus more on changes rather than a proactive sublimation.

Across the integrative designs, a loss of interactive meaning is observed where the impact assessment paradigm suffers from aggregation and attribution problems (Esteves & Vanclay, 2012; Larsen et al., 2018). Especially in India, impact information is generalized overlooking context-specific realities. In addition, project reports often assigns impact drivers without clarity or accountability (Chang et al., 2023). Beyond factoring a single cause, prevailing frameworks rarely account for cumulative and time-lagged indicators. Regulatory procedures, such as geotechnical surveys and land acquisition measurements, are regularly conflated with comprehensive assessments (Afroosheh & Askari, 2024). This leads to critical dimensions like social values, local acceptance, and contextual interpretations of risk and benefit remaining peripheral to the evaluation process (Vitali et al., 2023).

1.3. Parallel Development(s) and paralysis

Another form of challenge is methodological where the dominant nature of development of these projects is simultaneous. Within dynamic developmental contexts, the frequency of an overlap, stall, or sequence jump is large. Processes like screening, baselining, and resource acquisition are observed simultaneously in different project phases, leading to dilution of phase boundaries. Development outcomes in such cases are hard to ascertain. This mechanism embeds its components in a state of perplexity or *parallelity***,** where temporal space loses meaning in determining the phase or state of the development. Time, often the chief component against which outcomes are measured, remains no longer defined, leading to fractures in longitudinal approaches. This continued paralysis contributes to segmented policy design and grievances.

Spatiality, as a manifold space of all existing interactions of themes of development, gives a unique character to the affective region studied. Similarly, the temporality modifies the parameters and assumptions predisposed for any assessment model.

Universal suppositions for culture and context produce irregularities in implementing the frameworks.

As a proposed solution, this paper introduces a novel methodological framework which addresses the alignment of temporal paralysis of development evaluation space with environmental data which remains longitudinal. It recognizes this complexity as a bi-directional temporal wormhole able to transfer knowledge from the predevelopment phase to post development and vice versa.

1.4. Research Objectives

This study addresses the longitudinal limitation of socio-economic data by introducing machine learning (ML) methods as a temporal bridge. It utilizes interpretability of Tabular Networks or TabNet (Arik & Pfister, 2019) to map out a significant group of indicators. Parallely, it reasons longitudinal environmental change using spatial analyses of land change indices procured using remote sensing data to generate a regional perspective of change in different development phases. Finally, it formulates an epoch-based synthesis matrix that evaluates the socio-environmental capability of households in project-affected regions.

These objectives strengthen the impact assessment process as a temporally dynamic and socially embedded process. It fills the gap between machine and social learning by bringing a reasoned interface in design. The remainder of this paper is structured to bring out a detailed methodology, establishing a theoretical context of existing parallelity and socio-economic vulnerability, followed by designing an integrative apparatus, comprising the TabNet architecture and spatial projections. It further analyses the quantitative and spatial results of the dimensional interactions and suggests broader policy implications for monitoring linear developments.

2. Methods

2.1. Study Area

This paper takes the Char Dham Expressway Project for engaging parallel interventions that challenge the traditional time dimension of infrastructure projects. Conceived as a major connectivity project, the project aims to upgrade an 890 km single-lane pilgrimage route into a two-lane, all-weather highway (Figure 1). Situated entirely within the Himalayan state of Uttarakhand, India, the project assumes a dendritic spread, branching into four high-altitude valleys leading to the sacred Hindu *dhams*: Yamunotri, Gangotri, Kedarnath, and Badrinath. The documented objectives of the project span spiritual tourism, strategic defence, and regional development; however, its scope expands beyond mere connectivity, drawing multiple, often conflicting, themes of development onto the horizon. Geographically, the alignment traverses an altitudinal range from 800 to over 3500 meters. It transits diverse ecological zones (mixed forest vegetation to high altitude tundra) and seismically sensitivity configurations (boundary and main continental thrusts). This terrain is characterized by unconsolidated rock, steep gradients, and active seismicity, conditions highly prone to landslides (Nguyen et al., 2024).

Figure_1: Study Area depicting the extent of Char Dham Expressway (*Mahamarg*) project with its development agencies. It also shows three development phases as Cluster 1 (Pre-Development), Cluster 2 (Construction Phase) and Cluster 3 (Post-Development Cluster).

Three development clusters representing were identified on highway sections in Uttarkashi and Rudraprayag districts, in the north section of the project. For identifying the phase of development, these clusters were defined according to the nature of disruption *planned, in process, or that had happened* across certain stretches of the highway during the time of gathering the data. Indicators representing themes of development and household characteristics were captured from settlements around the highway sections NH 34 and NH 107.

2.2. Pluralistic Parallel Modelling Approach

The study uses a mixed-method approach (Figure_2) to tackle multi-data gaps and alignment of redundancies. The approach fits between Convergent Parallels and Sequential explanatory designs to navigate the complex nature of the data as elaborated by previous studies (Wenger-Trayner et al., 2017). It cojoins quantitative (measurement) data structures captured at the household level with qualitative insights inferred from the community level through focused group discussions to improve the interpretability of the regional socio-economic and environmental interactions.

Figure_2: Methodological Workflow Diagram.

2.2.1. Field Site Selections

For selecting the field site, the paper utilises a multistage purposive sampling process for site selection. Drawing on geohydrological and social consistency, it sequentially refines the referent region by criteria like association with geographic features, demographic and social composition, and economic dependence. The study utilized primary social profile construction from 463 validated households using the survey method. The data was collected using the Open Data Kit (ODK) digital platform, which supports the collection and geotagging of various data types. The survey assumed the village as a primary unit of impact block, since the policies are pushed through villages as the deepest leaf. Consequently, the population sample with an acceptable rejection rate of 10 per cent was designed to fit these settlements. In addition to maintaining statistical coherence, the sample sites were selected to mimic social homogeneity (in terms of community compositions, values, and language shared). The sites also shared geographic conformity in terms of association with geographic features such as valleys, streams (rivers, springs), distance from the nearest town (district headquarters), existing economic activities, and aspirations. Eight villages were selected for the survey, composing the final sample size of 150 samples in each cluster.

2.3. Data Sources and Preprocessing

Primary socio-economic data was gathered through a detailed questionnaire covering seven themes, namely, demography, social status, economic status, social asset ownership, health, education, and migration aspirations. These represent the existing nature and the capacity of the household to sustain medium to long-term changes and endure geohydrological and socio-economic shocks. Drawing on the Modified Capacities and Vulnerabilities Index (CVA) framework given by (Anderson & Woodrow 1989) and the Alkire-Foster multidimensional methodology (2011; 2025), this study processed 49 discrete variables (see Supplementary Appendix A) into contextualized weighted values. These were aggregated to produce standardized capability scores ranging from 0 to 1, where 0 denotes a total absence of capability and 1 represents the theoretical (existing) maximum. These variables encapsulated themes such as Economic capital, Health-Education accessibility and resilience; Land Ownership and functionality, Community's social and environmental values and Welfare Services. As for the environmental datasets, multispectral imagery at 10m resolution observed in project lifetime (from 2016-2024) were analyzed at four benchmarks. These were sourced using the Sentinel 2 Platform of the European Space Agency (ESA) and processed using Google Earth API Engine. A detailed comparative account of both groups of data is listed below: [Table 1].

2.4. Data Analysis

Given that development datasets are complex and often packed of multiple themes, they carry high dimensionality which conventionally limits any holistic account of the analytical process. However, with the advent of machine learning and artificial intelligence frameworks, computational capacities have become decentralized and accessible. Utilizing this capacity, this paper compared a combinatorial space, assisted by tabular learning architectures such as Google® TabNet and a modified version the TabPFN (Tabular Pre-Fitted Network) which are specifically designed to investigate the information contained in gridded structures. Developed and tested for various applications (Nguyen & Byeon, 2024; Aalianvari & Jahanmiri, 2025 ; Hollmann, 2025), these models are attunable and scalable with capabilities to integrate multi-scalar and non-linear dimensions of socio-economic information.

For testing the overall capability of the household, the indicators were factored based on the existence of specific capacities based on thematic discipline of learning. In this way, the process was supervised by ML rather than a pure mathematical disposition. Once converted, the composite capability score (CI_i) for a sample (i) is calculated using a linear addition given by;

$$CI_i = \sum_{j=1}^{n} x_{ij}$$

Where:

- n = 49 (your total number of variables).
- x_{ij} = *The converted score of variable j for sample i.*

Once generated, the total capability scores were grouped as a five-scale index based on the natural Jenks distribution of the data. Below are the corresponding Y labels:

1 – Extreme Precarity

2 – Vulnerable

3 – Fairly Capable

4 – Moderately Capable

5 – Highly Resilient

2.4.1. Quantifying the Qualitative Variables

This approach built upon Transformer-based architectures that processed data through an encoder-decoder framework, fundamentally differing from the sequential processing of Recurrent Neural Networks (RNNs). It utilizes a sequential multi-step architecture where sparse attention is used to select which features to reason from each decision step (Arik & Pfister, 2021). Although being a deep learning framework, the model shows integrity even on smaller data sets, however, such limitation is compensated by making it a supervised architecture explained in previous section. By integrating Gated Linear Units (GLU) and a self-attentive transformer, the model balances high-degree of representation-based learning and enhances the credibility of quantifying subjective variables. This strengthens the argument for classification formats which are built on a similar level of household capabilities. The model identified which features, such as land tenure or access to welfare are influential in determining the capability crust for a household within the flux of the project's lifecycle. Below is a representative diagram for the model (Figure 3):

Figure_3: TabNet Architecture for Classifying Tasks.

These impact trajectories (contribution of indicator and theme groups) are further mapped as spatialized capability scores and reasoned with GIS-based indices specifically for the Land Use Classificationm, visualising impact jumps or overlap across the geographical hyperspace of the project.

2.5. Data Validation: Weight Distribution compared to Rankings

Beyond assimilating multiple methods, this methodology utilized Weight Distributions (built in the model as feature importance) over simple rankings. While the ranking implies a linear hierarchy that often fails to capture 'one to many' connectivity in chaotic embeddedness. Weight distribution, conversely, allows for a non-linear understanding of how a single variable (e.g., loss of a kitchen garden) can disproportionately collapse a household's capability score when combined with temporal uncertainty. Further, as far as the model interpreted information is concerned, the thematic relationships can further be validated using community and development stakeholders' perceptions.

3. Research Findings

3.1. Basic character of household's socio-economic status

Among the 453 households, the natural distribution of the sample arises to follow homogeneity. Following a social value-based logistic thresholding, indicator level differentiation occurs across thematic profiles. Certain indicators are indicative of two or more variables. For instance, the indicator for overcrowding (TS5) is ascertained as relatively capable as a function of available space for individuals in the house. Similar reasoning for each indicator is listed as supplementary at the end. Given below is a descriptive statistic of the data analyzed: [Table 2]

As seen from the figures below (Figure 4. a, b, c), the data is highly supported by higher categories in the capability score distribution. To retain the fundamental quotient behind the distribution, the skewness was carried forward as guided by the contextual realities. From the figure, almost 70 per cent of households hold medium to high socio-economic capability as compared to regional growth.

Figure_4**:** Capability Class Distribution across the clusters: a) Histogram Distribution of the Scores; b) Thresholding Classes based on the existing scores and c) proportional distribution among the 462 households surveyed (in per cent).

3.2. Model Performance and Thematic Contributions

Within indicator influence across the development clusters, Income [Ec2] (Figure 5) showed the most influence in the model which makes it important determinant for classifying capabilities. As for the differences between development clusters, it does not show such significance. This can be attributed to wealth accumulation as a long-term process despite likely differences in the manner of wealth is distributed across the clusters. It acts as the primary driver for capability differentiation globally in the dataset, enabling other acquisitions like social asset holding or access to better education. Similar development spillovers may be seen in terms of ability to afford critical medical care, which increases as more private facilities increase in the post-development cluster. In terms of ownership of social assets, having an appliance such as a washing machine, television, or mobility assets like a car or motorcycle prominently differentiates households in terms of inherent capacity. It also resembles mobility means as the chief object defining the household's capacity.

The model gave a test accuracy of 0.796 per cent (Table 3) with a random split of 0.2 with Entmax gradient activation. The Entmax function which generalizes the softmax and sparsemax activations, allow for sparse output probabilities (some exact zeros), demonstrated by **(**Niculae, Martins & Smith, 2019**)** to suit interpretable models like TabNet. (Peters M E., et al. (2019). This helps in high dimensionality of the dataset, assisted with batch normalization to compensate scaling bias.

Figure_5: Feature Importance in the Model.

Although the model shows certain limitations in terms of overfitting character (Figure 6), the rationale behind retaining its use may be pruned with context-specific learning (which was incorporated into assumptions of indicator capabilities). Another method which may improve model performance is reclassifying the capability of weights. However, since the motive of the method was to retain the information structure over model efficiency, it is conceived as is. In this way, the method offers flexibility to decision-makers in defining the scope of machine learning-based substance to support policy decisions.

Figure_6: Tab Net model learning and interpretation across development themes.

4. Discussion

The approach shows promise in explaining socio-economic information with varied data structures and high dimensionality (feature classes or indicators). However, the model robustness is likely to improve as more data points are added. As a demonstration of exercise for use by development agencies, the approach is likely to nudge the agencies to choose data-driven decision-making. For instance, one of the prominent issues arising in sustaining the gains from expressway developments is the case of (in)access across the alignment. By factoring in the requirements based on highly agrarian/industrial settlement regions, the decisions may be reasoned to develop the orthogonal paths also. Multi method approaches show merit in data complementing processes with major advantage lying in supplementing the information. In certain data deficient sub clusters within the data distribution, it shows latent factors enhancing household vulnerability. For instance, holding agricultural land on high degree slopes (> 30 degree) render farming unsustainable. This extends to its existence even as a survival asset. Further, certain households with high fertile and gentle slope agricultural landholding report lower income in regions due to persistent threat of wildlife interactions.

Although the method provides information, engaging mixed methods have also resulted in decision conflict scenarios. One is where the results arising from a method conflict with others. Often in complex impact states, findings go against each other. In certain cases, significant feature classes differ from model to model. On the contrary, certain cases also result in conflating evidence. Inflation in terms of indicator resonances may enhance bias towards certain sets of feature classes while ignoring the rest. To rationalize a comprehensive interpretation, evidence thus requires an interactive design. The results from models were integrated with other data structures like qualitative interviews, focused group discussions, and observations from the field. Also, since contextuality warrants an integration of environmental parameters, thematic interpretation becomes key in merging evidence. The results are interpreted in the following manner

4.1. How does the approach improve thematic understanding?

Land as a thematic variable shows maximum contribution in ascertaining a household's socio-economic capability. Contemporary studies often cite it as one of the major aspects in impact assessments for its acquisition for development purposes such as roads. However, it is to be noted that despite the regional economy away from the agrarian setup amidst land scarcity, it still carries enough social value to influence household positionality. In addition, it shows a transition where the land acts as a reserve capital asset rather than a revenue asset, driving its significance for household prosperity. It also nudges decision-makers to comprehend a conspicuous increase in land value as the development expands towards certain land patches.

Development Clusters also show a significant contribution to education as a medium of awareness and to reserve a livelihood of opportunity. Since the regions studied rarely had institutional capacities, especially the latter, the ability to afford quality education (and training) forms a defining factor influencing household statuses. Land, as a conversion asset, allows many households, a flexibility to choose quality education. In the absence of revenue, it is better suited for brokerage. Likewise, the economic status, as a functional triad of availability of livelihood opportunities, multiplicity of jobs, and a higher income bracket, influences the economic capabilities and, to an extent, the social and political status.

Health accessibility, as a theme, contributes to least among the six themes engaged. It aligns with the paucity of medical infrastructure with robust diagnosis and emergency treatment facilities. Across the clusters, households depend on hospitals in the city, no matter the severity of the medical case. Even for the delivery of infants, the preferred choice is guided by the available facilities. Although highway development significantly reduces the time and distance to avail in the medical facilities, it is considerably associated with the existence of mobility within the household (Figure 7).

As for the inter-theme interaction, Infrastructure Synergy can be seen in terms of a strong positive covariance (red) between housing components like Roof and Floor, and between Water and Toilet access, indicating that improvements in one often coincide with the other. Further, Asset Density shows a noticeable cluster of positive covariance among high-end appliances (AC, Washing Machine, Iron), suggesting these are concentrated in a specific household**.**

Figure_7: Intensity of Association between the indicators.

4.2. Spatializing Capability Scores vis a vis project alignment.

In addition to the statistical performance of the model framework, the distribution (variation) of capabilities across the geographical space matters in defining the scope and intensity of impact. Char Dham expressway across the three development phases shows actual and potential impact as defined by the existing capabilities. Spatializing these capabilities also indicates the transitioning socio-economic status of impact-affected people. In the absence of a common temporal basis, approaches such as space for time substitution show merit in simultaneously visualizing phases as segments or life stages of the project. It enables an organic thought process built on stimuli and feedback systems. In cases of linear project interventions, the alignments act coherently in generating impact pathways. As per the capability score measured across 49 feature classes, the distribution across the clusters shows moderate to high existing socio-economic capability as shown in the maps [Figure 8] below.

Figure_8: Spatialized Capability Scores across three clusters studied.

This may be attributed to the anticipation created for the development which is desired in the region in general. However, evidently, the impact of tourism growth on household capacities is clearly visible in the upper group of the pre-development phase [Map A] while distributed scores represent the lower group where the alignment is proposed. Similarly, high-capacity households are abundant in phase two and three with an attribution differential. While phase two is marked by the parallel contour road and thus not attributable to the project, phase three shows significant dependency and aligns with the theoretical understanding of development interventions on household capacities as studied in the contemporary literature. In this manner, spatialized impacts not only challenge established knowledge but also complement it.

4.3. Environmental Interactions and Information Transfers

In addition to the household-centered impacts in terms of socio-economic dimensions of development, rupture in the temporal base negatively affects the quality of socio-environmental impacts. Spatial frameworks, as one of the mechanisms to overcome this deficit, assist in the baseline information as differentiated regional reality. Drawing from the major development cycle of the expressway development, three timelines were conceived to show a time series change at a periodicity of four years. Evidence in terms of the progressive use of change points towards multiple meanings of change classes. Visualizing these changes (Figure 9) shows an imminent increase in built-up areas across the clusters. However, it is more prominent in Pre and Post development clusters. One of the reasons that may be attributed to this can be the proximal distance to larger service settlements and relief. Both of these factors invite potential for exponential and disproportionate growth in their service region. However, if the growth factor is expropriated, the growth dampens in favor of the service center.

The predevelopment phase shows an increase in (red) settlement zones in the upper village with existing alignment, increasing at the expense of agricultural land. As a variable change, forest fires and landslides modify vegetative zones in favor of barren/degraded shrub land. Contrary to the Construction phase where daily discomfort is a micro-level change, the change in built-up area is scattered, nonlinear, and indeterminate. Since the relief differential in this phase is relatively high, there is no significant change in the agricultural class. The post-development phase shows an increase in built-up regions as hospitality and residential real estate progressively. Since the valley is wide, three-level conversions are evident from agricultural land (yellow) to barren (orange) scape to settlement reds. An increase in construction demand facilitates land degradation and furthers their change in use.

Figure_9: Time series change of Land Cover and Land Use across clusters as five classes.

4.4. Policy Relevance and Limitations

The study finds inter-dimensional interactions between themes of development which register application of impact information beyond project extent. Since the expressway extends across the state of Uttarakhand, the information arising out of it intersects with welfare goods distribution, inspecting localized progress in the valleys and shifts in migration tendency, geographically (from hill to valley) or administratively (outside constituency, district or state).

However, pluralistic designs should be cautiously interpreted in cases of conflicting findings. Machine learning methods on smaller datasets often exhibit limitations and overfitting tendencies. Policy decisions in such cases may involve more validation of sources such as stakeholder engagement and in-situ studies. In addition, a certain degree of optimization may be performed to reduce the overfitting character ranging from apparatus side (for instance, SMOTE) or from social domain knowledge which strengthens the context. Validation of information using context represented indicators thus can take decision making outside the binaries of heuristics and technical exclusions.

5. Conclusion

This study advances a novel methodological framework tailored for assessing the complex socio-economic impacts of linear infrastructure projects. It utilizes the Char Dham Mahamarg project in the Uttarakhand Himalayas as an empirical case. It engages pluralistic method approach to navigate pervasive challenges of temporal discontinuity that exists presently in Indian road infrastructure impact assessments with absence of baseline reference information, common in such large-scale implementations. In contexts of severe information scarcity, integrating spatial projections with advanced machine learning techniques offers a n interpretable mechanism for data distribution and validation which reduces subjective bias especially in environmental decision-making. Precisely, the deployment of the TabNet architecture serves as a powerful integrative framework which illuminates the complex, cross-thematic interactions among 49 distinct development indicators. Through EntMax activation, the model captures the intricate resonance between social and economic dimensions. This quantitative coupling signals fundamental shifts within a broader capabilities and vulnerabilities framework, providing a high-resolution understanding of how household vulnerability statuses transition over time. Crucially, this methodology demonstrates significant utility in extracting dynamic impact transitions from non-longitudinal datasets, complementing traditional case study approaches and enriching the spatial explanation of developmental impacts on the ground.

While the current framework defines the scope of socio-economic assessments and identifies critical dimensional interactions, it allows for future iterations that systematically integrate qualitative data structures such as in-depth interviews and focus group discussions to capture lived experiences.

Given the highly sensitive and ecologically fragile context of the Himalayan landscape, subsequent studies in this research may incorporate disaster risk metrics and climatic parameters to make it more holistic.

Ultimately, this research offers a vital diagnostic approach for policymakers to identify and navigate existing societal stratifications, acknowledging that the benefits and burdens of infrastructural development are distributed unevenly. It allows for such decisions to be made even when phrasal boundaries of development projects are blurred. It also forms a temporal bridge for learnings to be transferred across the phases. Further, for mega-projects with extended gestation periods, this framework provides a dynamic foundation for ongoing monitoring and evaluation. Beyond serving as a static reference, it enables the formulation of targeted, micro-level policy interventions capable of adapting to the evolving socio-ecological realities of affected communities.

6. References

1. Afroosheh, S., & Askari, M. (2024). Fusion of deep learning and GIS for advanced remote sensing image analysis. *arXiv preprint arXiv:2401.12345*.
2. Alan Ehrlich (2022) Collective impacts: using systems thinking in project-level assessment, Impact Assessment and Project Appraisal, 40:2, 129-145, DOI:10.1080/14615517.2021.1996901

3. Aalianvari, A., Jahanmiri, S. (2025). A Comparative Study of TabNet and Classical Machine Learning Models for Landslide Prediction, 17 October 2025, Preprint, Research Square, https://doi.org/10.21203/rs.3.rs-7694229/v1
4. Alkire, S., Mishra, R., Selden, L., & Suppa, N. (2025). *The Global Multidimensional Poverty Index (MPI) 2025: Disaggregation results and methodological note*. Oxford Poverty and Human Development Initiative (OPHI).
5. Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, *95*(7–8), 476–487. https://doi.org/10.1016/j.jpubeco.2010.11.006
6. Anderson, M. B., & Woodrow, P. J. (1989). *Rising from the Ashes: Development Strategies in Times of Disaster*. Routledge
7. Arik, S. O., & Pfister, T. (2019). TabNet: Attentive interpretable tabular learning. *arXiv preprint arXiv:1908.07442*.
8. Blue et al., 2021. Beyond distribution and participation: a scoping review to advance a comprehensive environmental justice framework for impact assessment. Environ. Impact Assess. Rev., 90 (2021), Article 106607.
9. Castelblanco, G., et al. (2023). "Environmental Impact Assessment Effectiveness in Public–Private Partnerships: Study on the Colombian Toll Road Program." Journal of Management in Engineering.
10. Chang, Z., Wang, Y., Liu, Q., & Zhou, M. (2023). Application of TabNet for landslide susceptibility assessment using hybrid data sources. *Remote Sensing of Environment, 295*, 113007. https://doi.org/10.1016/j.rse.2023.113007
11. DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organisational fields. *American Sociological Review*.

12. Esteves, A. M., Franks, D., & Vanclay, F. (2012). "Social impact assessment: the state of the art." Impact Assessment and Project Appraisal.

13. Hollmann, N., et al. (2025). "Accurate predictions on small data with a tabular foundation model." *Nature*. DOI: 10.1038/s41586-024-08328-6

14. Larsen, S. V., et al. (2018). "The role of EIA in addressing cumulative impacts: A case study of practice." Environmental Impact Assessment Review.

15. Mambiravana, T., & Umejesi, I. (2023). Infrastructure development and community perceptions of environmental risk: A case study of the Wild Coast, South Africa. *Jàmbá: Journal of Disaster Risk Studies, 15*(1), a1377. https://doi.org/10.4102/jamba.v15i1.1377

16. Morrison-Saunders, A., et al. (2021). The role of Environmental Impact Assessment in moving towards the Sustainable Development Goals. Environmental Impact Assessment Review.

17. Nguyen HV and Byeon H (2024) A hybrid self-supervised model predicting life satisfaction in South Korea. Front. Public Health 12:1445864. doi: 10.3389/fpubh.2024.1445864

18. Nguyen, T. H., Le, T. T., & Pham, V. H. (2024). Integrated flood and landslide hazard assessment using machine learning and geospatial data. *Environmental Science and Pollution Research, 31*(4), 42467–42483. https://doi.org/10.1007/s11356-024-30584-9

19. Nita, A., et al. (2024). Climate change impact assessment: the state of the art. *Impact Assessment and Project Appraisal*

20. Oshri, B., Hu, A., Adelson, P., Chen, X., Dupas, P., Weinstein, J., Lobell, D., & Ermon, S. (2018). Infrastructure quality assessment in Africa using satellite imagery and deep learning. *arXiv preprint arXiv:1805.11374*.

21. Paliwal, R. (2006). EIA practice in India and its evaluation using SWOT analysis. *Environmental Impact Assessment Review*, 26(5), 492-510.

22. Penadés-Plà, V., Martínez-Muñoz, D., García-Segura, T., Navarro, I. J., & Yepes, V. (2020). Environmental and social impact assessment of optimised post-tensioned concrete road bridges. *Sustainability, 12*(10), 4265. https://doi.org/10.3390/su12104265

23. Peters, M. E., Niculae, V., Martins, A. F. T., & Smith, N. A. (2019). Sparse Sequence Modelling with Entmax. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP). https://doi.org/10.18653/v1/D19-1006

24. Preiser, R., et al. (2018). "Social-ecological systems as complex adaptive systems: organising principles for advancing research methods and approaches." Ecology and Society.

25. Slootweg, R., & Jones, M. (2020). Resilience thinking in environmental assessment. Impact Assessment and Project Appraisal.

26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2023). Attention is All You Need. Proceedings of the Neural Information Processing Systems (NIPS 2017). arXiv preprint arXiv:1706.03762v7, 2023.

27. Vitali, E. R., Barbanera, M., Morini, E., Ranzi, E., & Cecchini, A. (2023). Mapping urban green transitions using multi-source data fusion and deep learning. *Sensors, 23*(8), 3805. https://doi.org/10.3390/s23083805

28. Wang et al., 2023. Environmental justice, infrastructure provisioning, and environmental impact assessment: evidence from the California Environmental Quality Act. Environ. Sci. Pol., 146 (2023), pp. 66-75.

29. Wenger-Trayner, B., Wenger-Trayner, E., Cameron, J., Eryigit-Madzwamuse, S., & Hart, A. (2017). Boundaries and boundary objects: An evaluation framework for mixed methods research. *Journal of Mixed Methods Research*, 13(3), 321–338. https://doi.org/10.1177/1558689817732225

30. Woldesenbet, T. A., Gebremedhin, T. G., & Kassa, S. M. (2024). Sustainable infrastructure and social inclusion for energy access in Sub-Saharan Africa. *Journal of Cleaner Production, 470*, 143230. https://doi.org/10.1016/j.jclepro.2024.143230

31. Zhang, L., Li, X., Zheng, D., Zhang, K., & Ge, Y. (2021). Rainfall forecast model based on the TabNet model and its comparison with other machine learning algorithms. *Water, 13*(9), 1272. https://doi.org/10.3390/w13091272

7. Tables

Table 1: Types and Forms of Data Used in the Analysis –

	Data	Data Type	Resolution	Source	Platform
1.	Socio-Economic (Thematic) Status	Geo-Coordinated (Attribute – Point)	Granular (Accuracy < 3 m)	Primary Field Survey	ODK Architecture
2.	Environmental Data	Spatial Time Series (Pixel)	Spectral - 10 m Temporal – 4 Yrs (2016-2024)	ESA	Sentinel 2

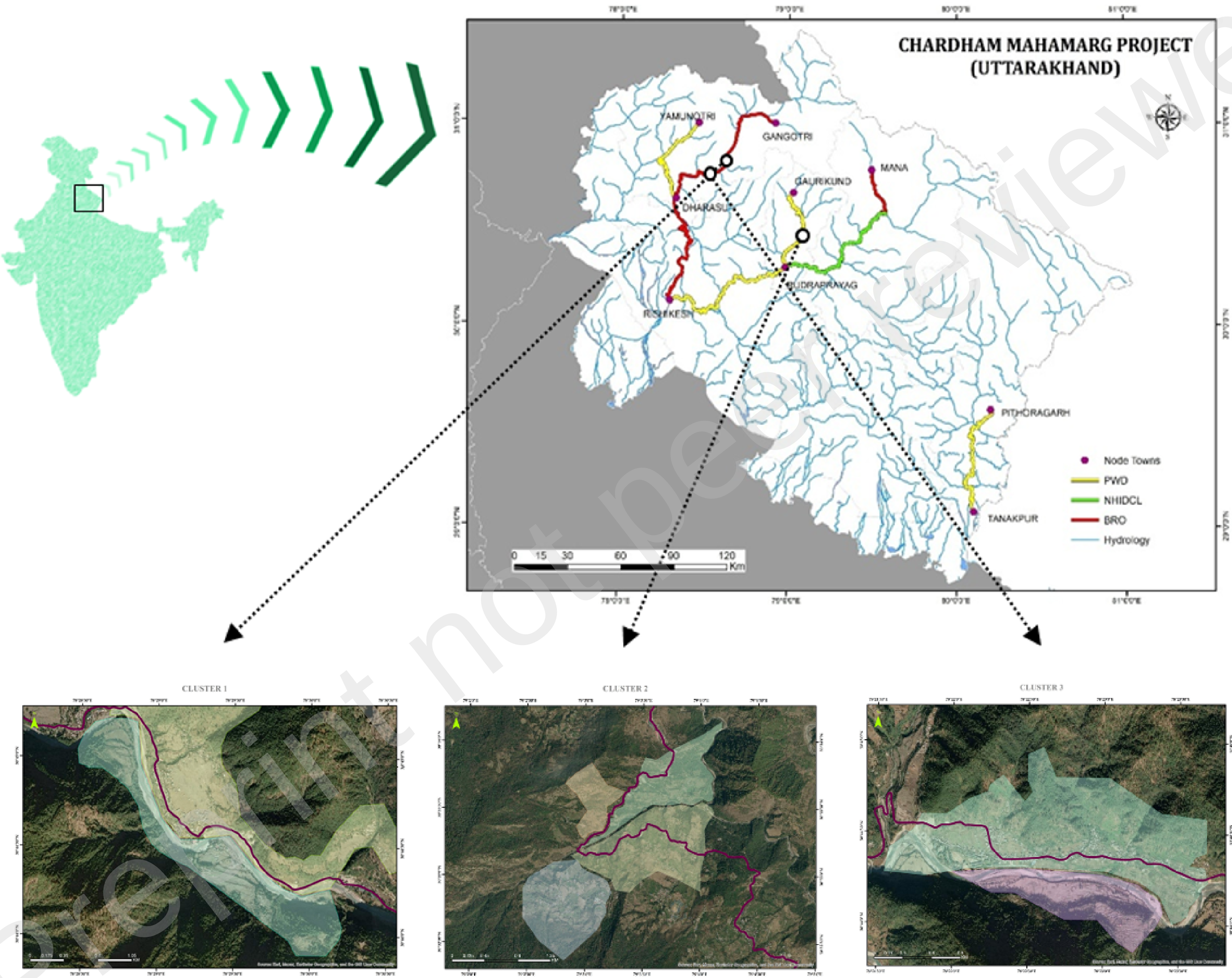
Table 2: Basic socio-economic data distribution statistics

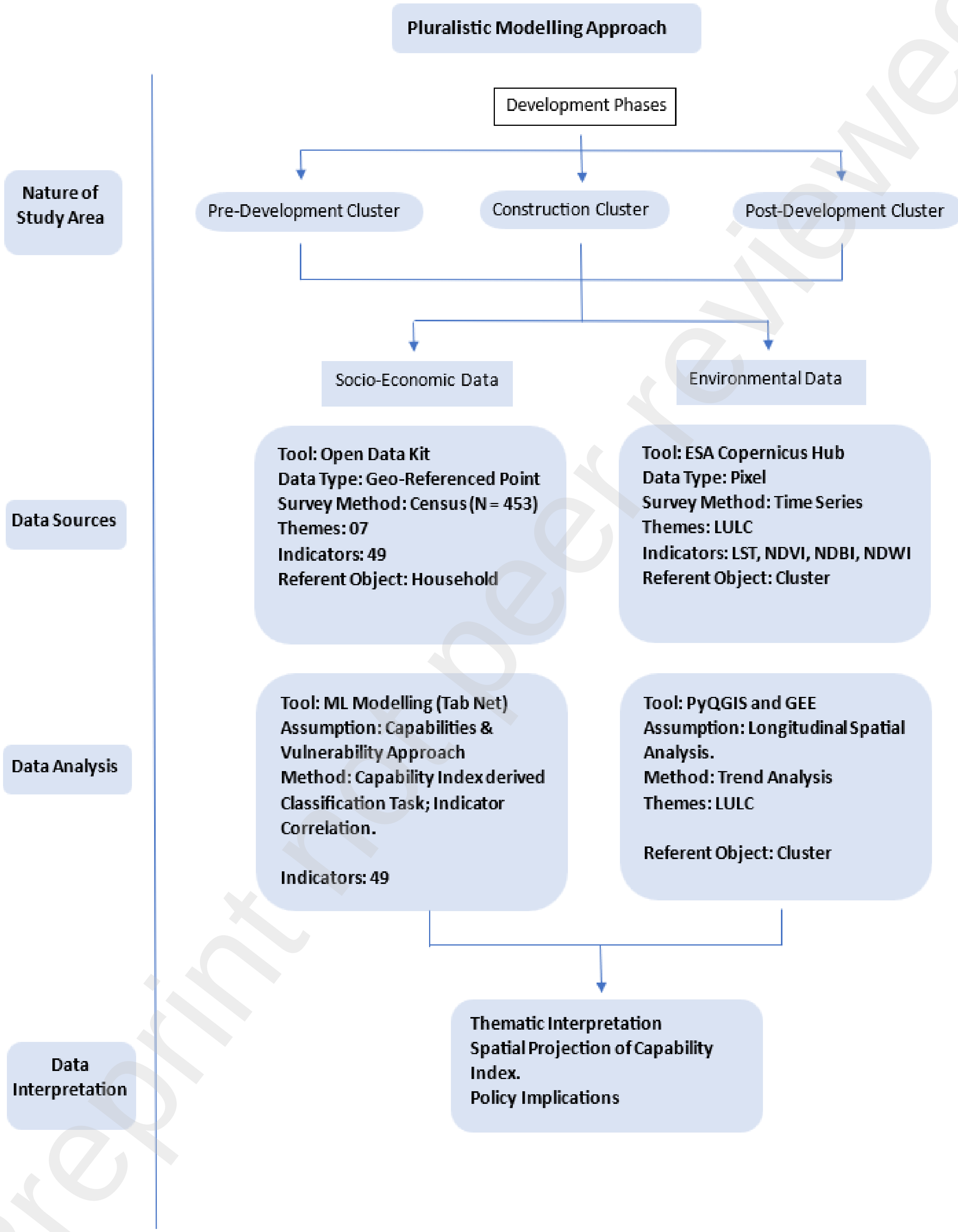
	N	Range	Min	Max	Mean (Std. Error)	Std. Deviation (Variance)
Y (Capability Scores)	462	26.00	13.00	39.00	30.85 (0.185)	3.97831 (15.83)

Table 3: Model Accuracy for Impact transition (Indicators) across the Clusters.

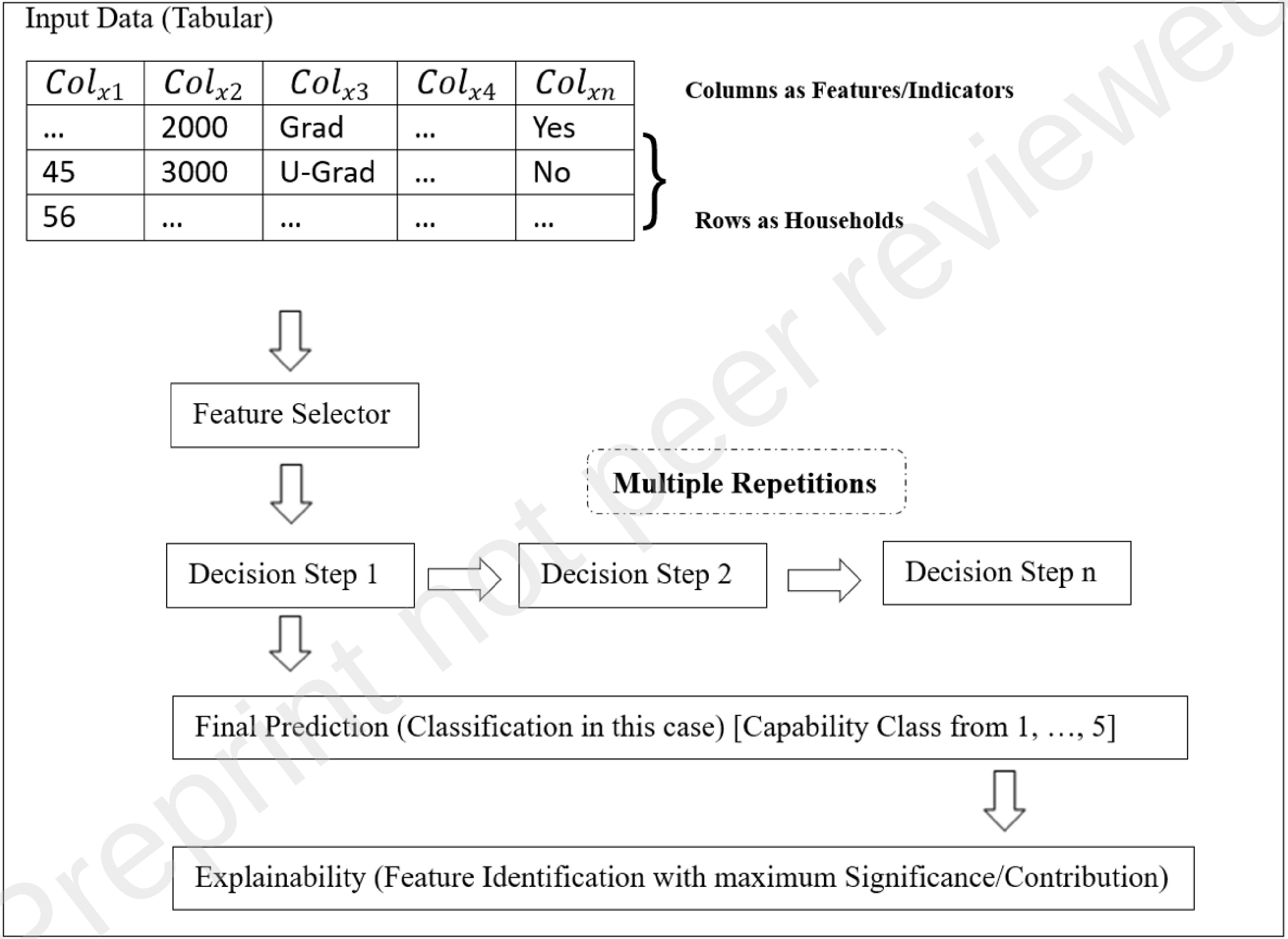
Capability Class	Precision	Recall	F1 Score	Support
1	0	0	0	1
2	0	0	0	2
3	0.70	0.64	0.67	22
4	0.87	0.87	0.87	53
5	0.79	0.94	0.86	16
TabNet Accuracy			0.80	94
Weighted Average.	0.79	0.80	0.79	94

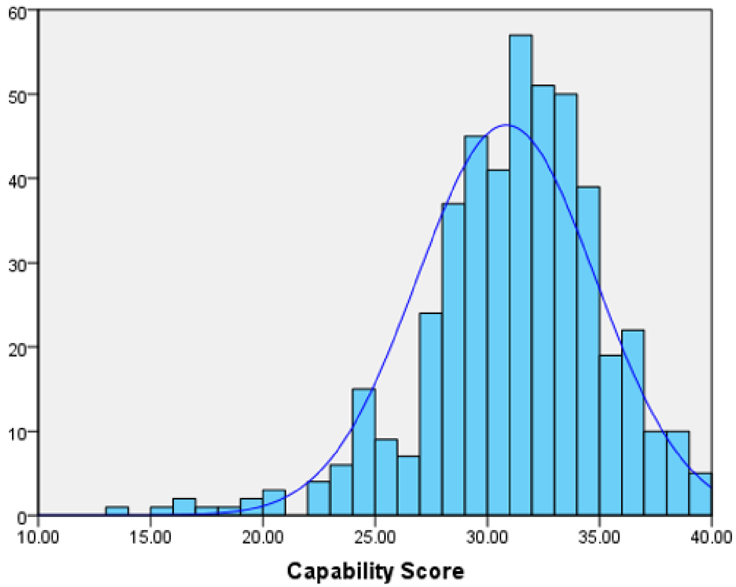
The model accuracy acknowledges the limitation in interpretability for certain classes (1 and 2) due to limited survey samples.

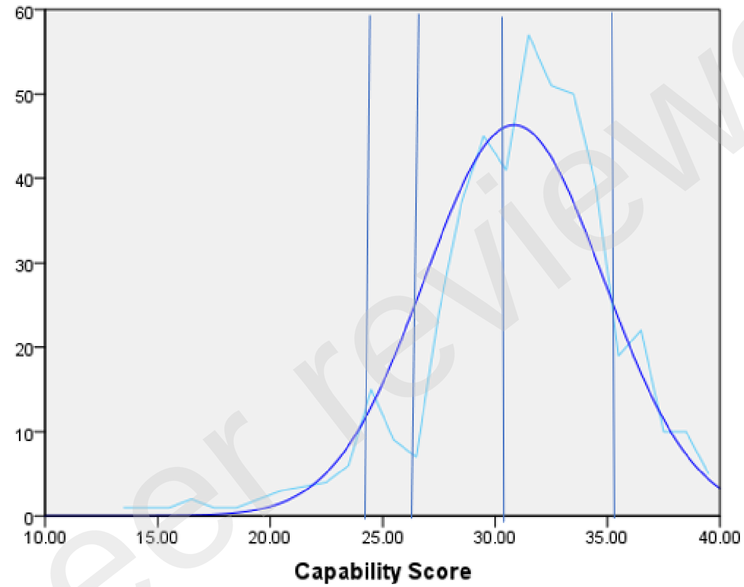

CHARDHAM MAHAMARG PROJECT
(UTTARAKHAND)
YAMUNOTRI
GANGOTRI
MANA
GAURIKUND
DHARASU
RUDRAPRAYAG
RISHIKESH
PITHORAGARH
TANAKPUR
Node Towns
PWD
NHIDCL
BRO
Hydrology
0 15 30 60 90 120
Km
CLUSTER 1
CLUSTER 2
CLUSTER 3

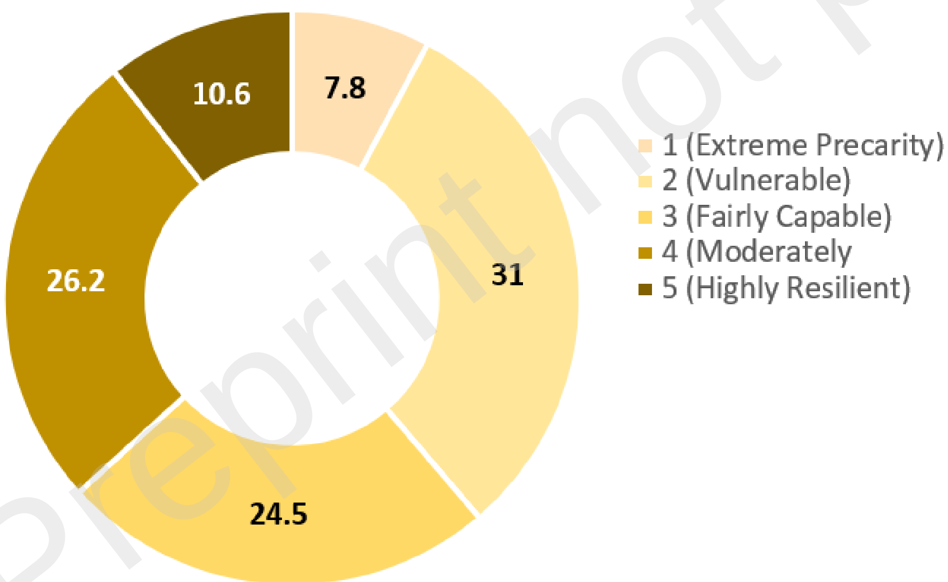


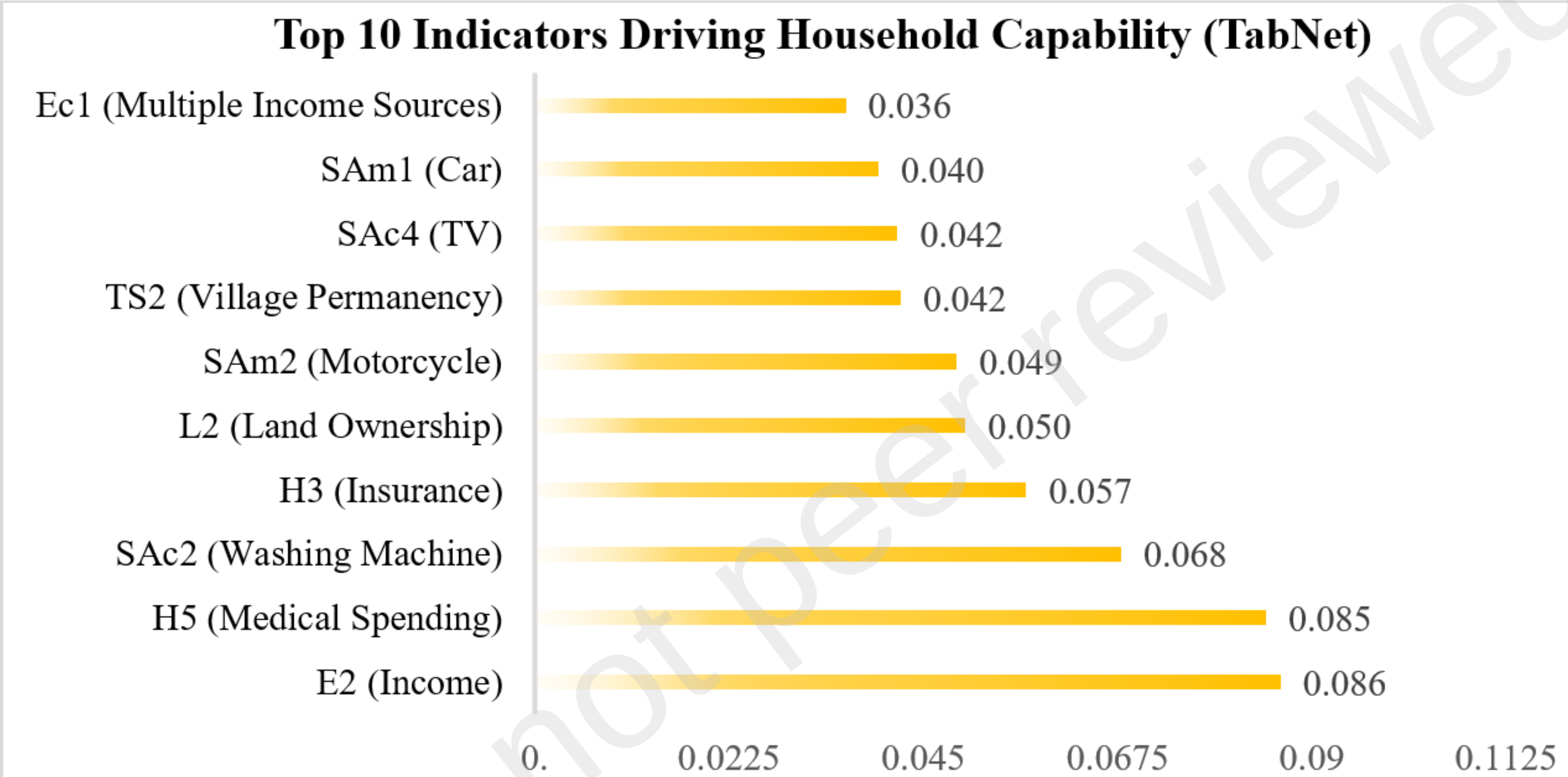
Pluralistic Modelling Approach
Development Phases
Nature of Study Area
Pre-Development Cluster
Construction Cluster
Post-Development Cluster
Socio-Economic Data
Environmental Data
Data Sources
Tool: Open Data Kit
Data Type: Geo-Referenced Point
Survey Method: Census (N = 453)
Themes: 07
Indicators: 49
Referent Object: Household
Tool: ESA Copernicus Hub
Data Type: Pixel
Survey Method: Time Series
Themes: LULC
Indicators: LST, NDVI, NDBI, NDWI
Referent Object: Cluster
Data Analysis
Tool: ML Modelling (Tab Net)
Assumption: Capabilities & Vulnerability Approach
Method: Capability Index derived Classification Task; Indicator Correlation.
Indicators: 49
Tool: PyQGIS and GEE
Assumption: Longitudinal Spatial Analysis.
Method: Trend Analysis
Themes: LULC
Referent Object: Cluster
Data Interpretation
Thematic Interpretation
Spatial Projection of Capability Index.
Policy Implications

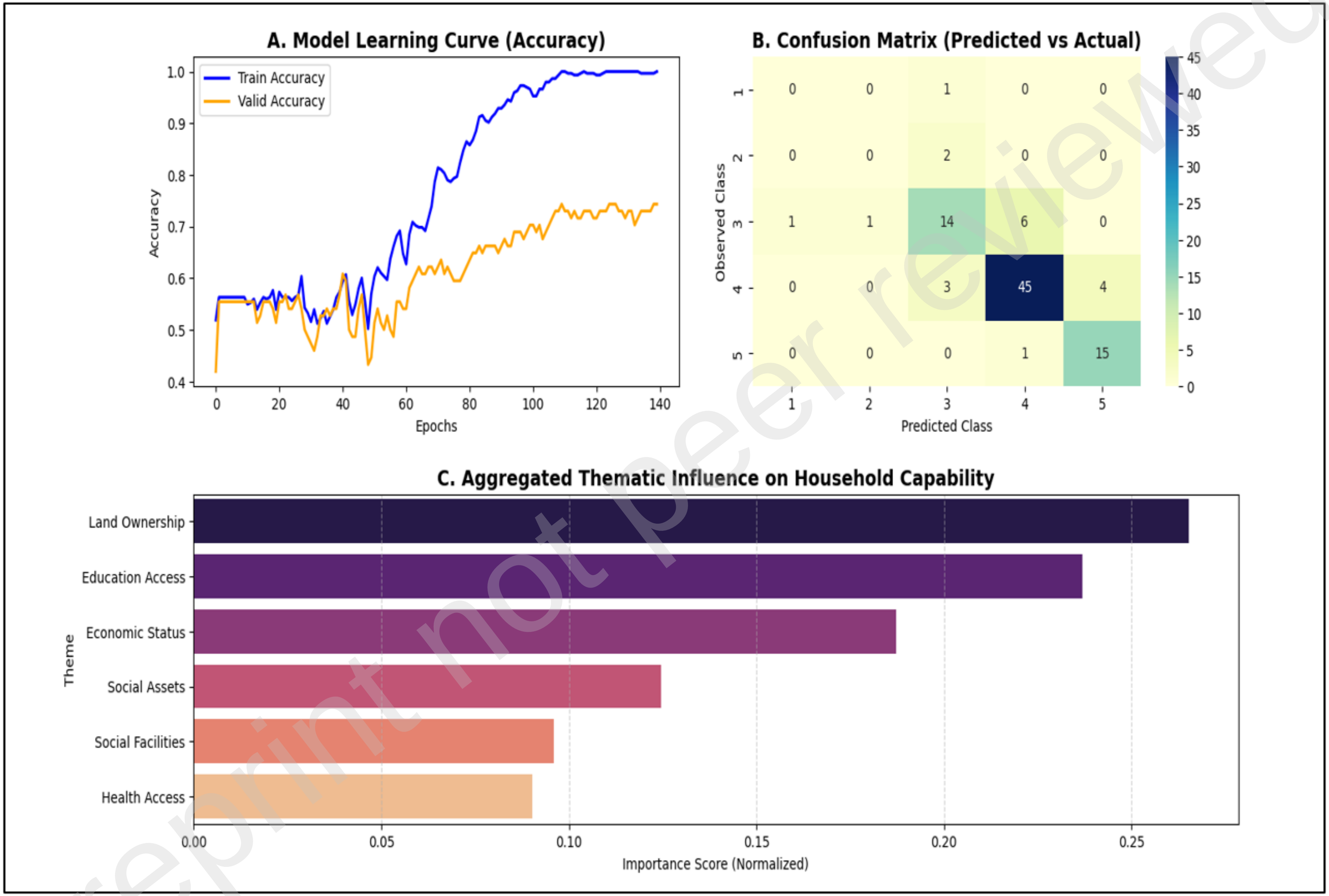

Input Data (Tabular)
Col_{x1} Col_{x2} Col_{x3} Col_{x4} Col_{xn}
... 2000 Grad ... Yes
45 3000 U-Grad ... No
56 ... ...
Columns as Features/Indicators
Rows as Households
Feature Selector
Multiple Repetitions
Decision Step 1
Decision Step 2
Decision Step n
Final Prediction (Classification in this case) [Capability Class from 1, …, 5]
Explainability (Feature Identification with maximum Significance/Contribution)

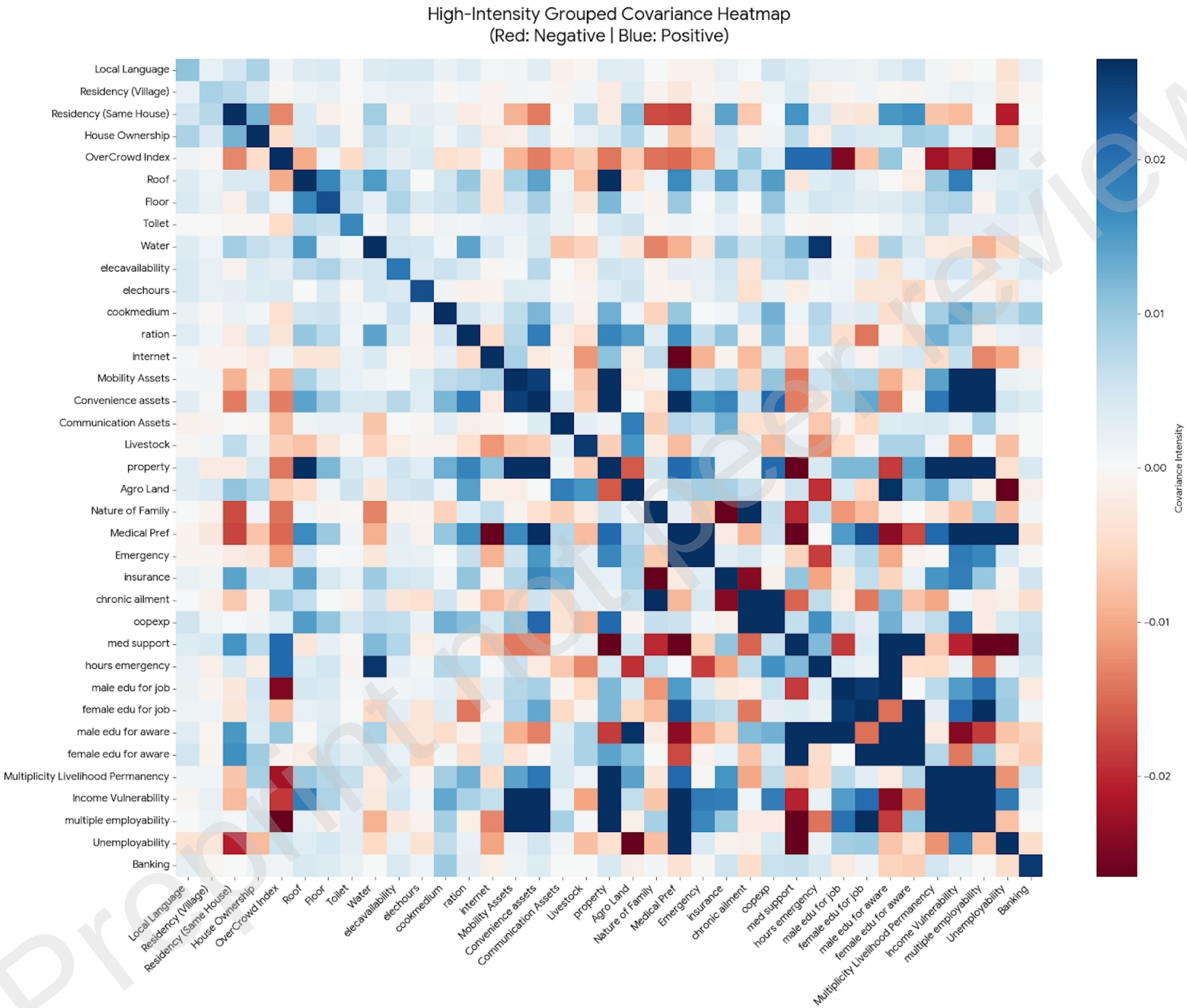

60
50
40
30
20
10
0
10.00
15.00
20.00
25.00
30.00
35.00
40.00
Capability Score


60
50
40
30
20
10
0
10.00
15.00
20.00
25.00
30.00
35.00
40.00
Capability Score


10.6
7.8
31
26.2
24.5
1 (Extreme Precarity)
2 (Vulnerable)
3 (Fairly Capable)
4 (Moderately
5 (Highly Resilient)


Top 10 Indicators Driving Household Capability (TabNet)
Ec1 (Multiple Income Sources)
SAm1 (Car)
SAc4 (TV)
TS2 (Village Permanency)
SAm2 (Motorcycle)
L2 (Land Ownership)
H3 (Insurance)
SAc2 (Washing Machine)
H5 (Medical Spending)
E2 (Income)
0.036
0.040
0.042
0.042
0.049
0.050
0.057
0.068
0.085
0.086
0.
0.0225
0.045
0.0675
0.09
0.1125

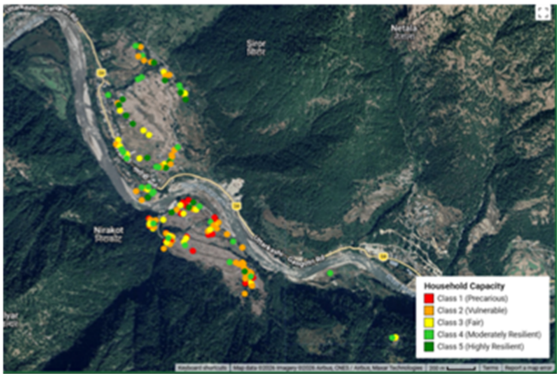

A. Model Learning Curve (Accuracy)
Train Accuracy
Valid Accuracy
Accuracy
Epochs
B. Confusion Matrix (Predicted vs Actual)
Observed Class
Predicted Class
C. Aggregated Thematic Influence on Household Capability
Land Ownership
Education Access
Economic Status
Social Assets
Social Facilities
Health Access
Theme
Importance Score (Normalized)



High-Intensity Grouped Covariance Heatmap
(Red: Negative | Blue: Positive)
Local Language
Residency (Village)
Residency (Same House)
House Ownership
OverCrowd Index
Roof
Floor
Toilet
Water
elecavailability
elechours
cookmedium
ration
internet
Mobility Assets
Convenience assets
Communication Assets
Livestock
property
Agro Land
Nature of Family
Medical Pref
Emergency
insurance
chronic ailment
copexp
med support
hours emergency
male edu for job
female edu for job
male edu for aware
female edu for aware
Multiplicity Livelihood Permanency
Income Vulnerability
multiple employability
Unemployability
Banking
Covariance Intensity
0.02
0.01
0.00
−0.01
−0.02

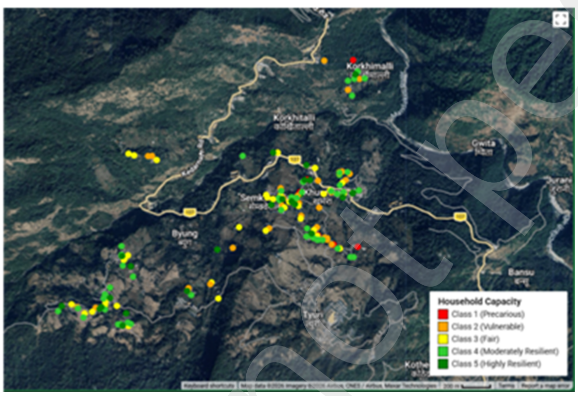
Cluster 1: Pre Development Phase

Composed of two villages. Upper circle denotes Village A with existing single lane road not yet doubled. Lower circle denotes Village B with new alignment planned.



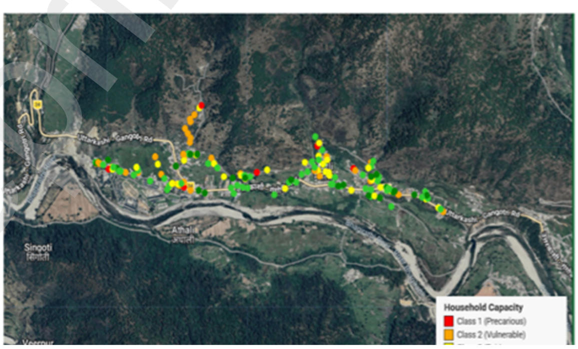
Cluster 2: Construction Phase

Composed of three village settlements highly scattered in terms of relief range and proximity to the main alignment. Settlement away from road in the south-east corner of the map is serviced with an existing alignment and shows no major attribution with the lower (main road).



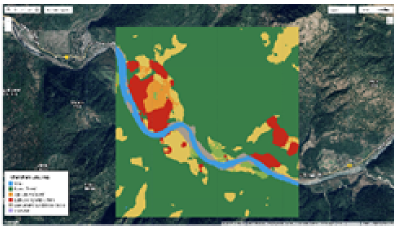
Cluster 3: Post-Development Phase

Composed of a single village, outgrown with influx of migrant population. As per current administrative records, show an estimated, five times increase from the last census survey in 2011. The village shows an increase in high-capacity households near the highway road.

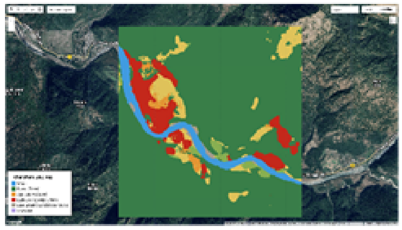


2016	2020	2024

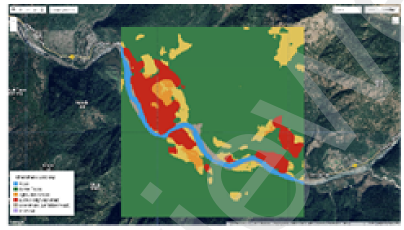
1. Pre-Development Cluster



1a

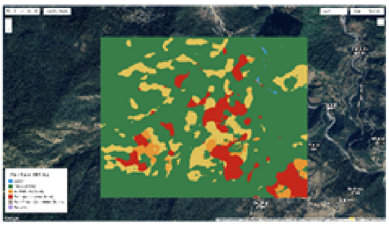


1b

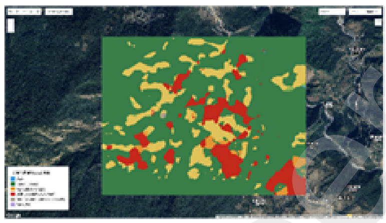


1c

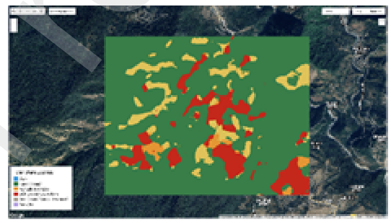
2. Construction Cluster



2a

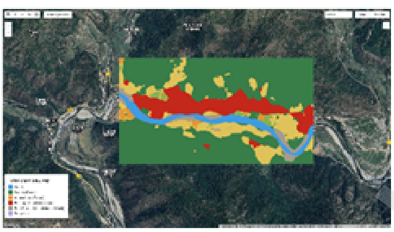


2b

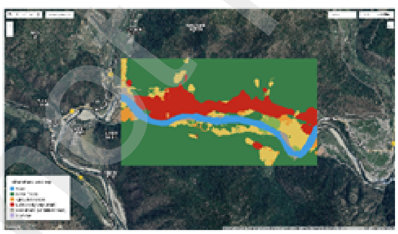


2c

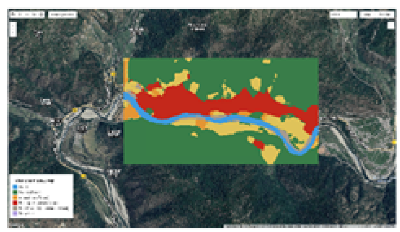
3. Post-Development Cluster



3a



3b



3c