A Socio-Spatial Approach to Visualize Phases of Development in One Timeline

Saransh Kumar

National Institute of Advanced Studies, Bengaluru (India)

Saransh@nias.res.in

Abstract

Discretized Impact Assessments requires at least two points of data to establish a change. The monitoring process establishes this change over baseline spread across time and space which is often complex for the impact pathways in socio-environmental interactions.

This paper aims to simplify and understand these impact pathways using consumption vectors in order to build a comprehensive foundation for evidence collection. We use machine deep learning based TabNet architecture to analyze and GIS based methods to spatialize the impacts of linear infrastructure projects over one timeline. The resultant impact scenarios deconstruct context specific issues subject to different starting point in the development phases. Using Char Dham Mahamarg Project (Uttarakhand, India) as a case study, this paper thus, explores the scope of impact assessment planning for data sourcing and ways to operationalize it in the field.

The results suggests that the approach is successful in extrapolating nonlinearity of the impact pathways of linear infrastructures across development intervention. Incorporating these insights in developing scenarios show an in depth understanding of the impacts which can be drawn to the household levels.

Key Words: Consumption Pathways, Development, Field Visualization, Linear Infrastructures, TabNet.

Introduction

Impact Assessments were evolutionary conceived to bring out the intricated associations that exist within bio-physical and social systems. The procedure of conducting these studies is highly determined by the regional regulations which prioritizes template formats of assessments as a way to fasten the process of getting clearances from regulators. This leads to dynamic and multi-scalar impacts of linear infrastructure projects in ecologically fragile and culturally significant regions being inadequately understood, particularly in countries like India where development imperatives often outpace assessment mechanisms. This has been seen in many cases which resulted in environmental resistance from civil society and further politicization of the intervention initially sought.

Traditional frameworks for impact assessment (IA) tend to adopt a reductionist view, privileging either environmental or economic outcomes while largely externalizing the social and temporal dimensions of change (Penadés-Plà et al., 2020; Woldesenbet et al., 2024). These assessments typically rely on static before-and-after comparisons, failing to capture the nonlinearity and spatial fragmentation of impacts across the life cycle of infrastructure interventions (Zhang et al., 2021).

This paper examines the Char Dham Expressway Project in Uttarakhand, a Himalayan state in northern India, which aims to upgrade a 900-km single-lane pilgrimage route into a two-lane all-weather highway connecting four sacred Hindu shrines. The project's justification spans spiritual tourism, strategic defense, and regional development. It traverses altitudes from 800 to over 3500 meters, cutting through diverse ecological zones and settlements, from river valley towns to near-abandoned ridge-top villages. This varied terrain is marked by unconsolidated rock and debris, steep gradients, and active seismicity—conditions highly prone to landslides and slope failures (Nguyen et al., 2024). The study is focused specifically on Uttarkashi and Rudraprayag districts, both bordering China and deeply embedded in geopolitical concerns. The region is economically dependent on seasonal tourism, and socio-environmentally entangled in challenges of uneven resource distribution, cultural marginalization, and increasing ecological vulnerabilities (Mambiravana & Umejesi, 2023). Fieldwork conducted across three development clusters—spanning acquisition and alignment, construction activity, and early maintenance phases—captured over 400 samples of household and settlement-level data. Within this empirical context, the research critiques how the Indian IA system suffers from both the aggregation and attribution problem: either overgeneralizing complex, context-specific realities or assigning impact causes without clarity or accountability (Chang et al., 2023). Moreover, prevailing frameworks rarely account for cumulative and time-lagged effects. Regulatory procedures, such as geotechnical surveys and land acquisition measurements, are often mistaken as comprehensive assessments (Afroosheh & Askari, 2024). Consequently, social values, local acceptance, and contextual interpretations of risk and benefit remain peripheral to the evaluation process (Vitali et al., 2023).

The rationale for this research lies in addressing the methodological limitations of traditional impact assessments when applied to dynamic infrastructure developments. Existing models fail to capture micro-level changes and nonlinear pathways, often leading to under-representation of ground realities in decision-making processes. To address these limitations, this paper proposes an integrated modeling approach using TabNet, a deep learning architecture that performs well on tabular data and offers interpretability in complex classification tasks (Arik & Pfister, 2019).

By leveraging consumption vectors—data representations that track changes in household-level behaviors, livelihoods, and access to resources—the study aims to uncover latent patterns in how impacts evolve across space and time. Spatial visualization through GIS techniques complements the model by contextualizing impact trajectories within geographies of terrain, habitation, and infrastructure (Oshri et al., 2018).

The primary contribution of this research lies in reconceptualizing IA as a temporally dynamic and socially embedded process, especially within the infrastructural politics of fragile borderlands. The intersection of development, defense, and divinity in the Char Dham project exemplifies how national agendas impose a standardized logic onto highly localized contexts, often at the cost of ecological stability and cultural continuity. This work seeks to identify which variables are most responsive to developmental interventions and how these can be systematically incorporated into scenario-based assessments for more grounded and anticipatory planning.

Methodology

The Study utilizes the data from 456 households spread evenly across three development clusters which shows distinct characteristic of the phase of development at the time of data collection. From a detailed questionnaire across nine themes, five objective themes on demography, social status, economic status, assets ownerships, status and aspiration to migrate formed basic data. We used 49 variables and preprocessed them based on contextualized weighted values to arrive at ranged capabilities scores (0 being no capability while 1 being most capability). The data variables carried themes from Economic, Health and Education, Land, Social Assets, Social Status and Welfare Services.

To analyze this dataset, we used Google® Tab Net, a nascent hybrid model as a way to integrate dimensions of development. The paper builds up from core concepts of transformers – architectures that process data in one go, fundamentally differing from how an RNN would work. And it applies to data structures with tabular formats where socio-economic data comes in. Two studies [Arik, S. Ö., & Pfister, T. (2021), Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2023)] brought out the use and applications of architectural mechanism of Gated Linear Units for the output.tasks of classification and regression problems. It integrated decision tree framework with neural networks to balance out interpretation with scaling. Here the model is tested for the relative capabilities held within each household sampled.

From input data, important features are extracted and numerous decisions are made until no more value can be generated. It shows which features that are most influential in the model. We utilize the model for regression task where factor relationships of the variables across the clusters have been established. Below is the Model Workflow.

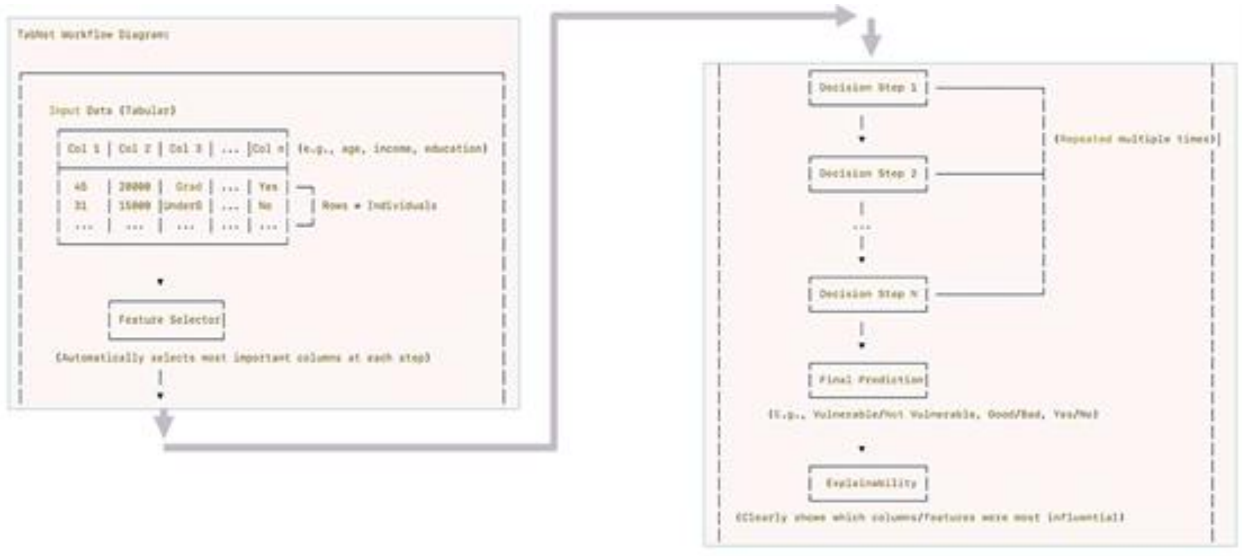
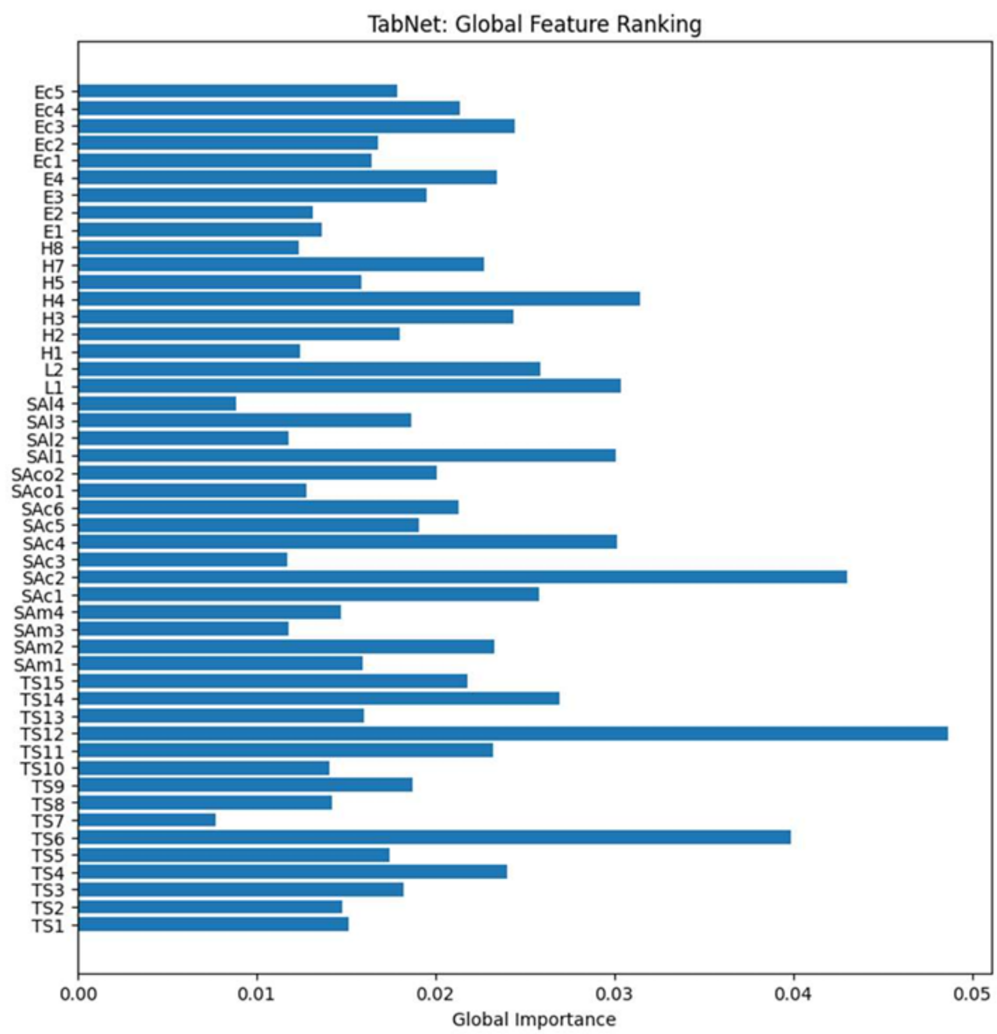


Fig. 1: TabNet Model Reasoning Chart Flow.

Results and Discussion

The data variables carried themes from Economic, Health and Education, Land, Social Assets, Social Status and Welfare Services. The variables like Food Security [Sac1] [Fig.2] showed the most influence in the model, meaning, in between cluster changes, there are likely differences in the manner people rely on rations for their nutritional upkeep. Likewise, development spillovers may be seen in terms of ownership of material assets, the housing structures, medium of cooking, a change in ownership of traditional goods like livestock (cow and buffaloes). On the other hand, having a car or bicycle in the house doesn't contribute significantly, or the availability and accessibility to electricity or the language. Social status and welfare services as an aggregate thematic block show maximum effect across the clusters among the households surveyed followed by the ownership of material assets. However, this discounts the importance of the theme areas. For instance, on the accounts of health access, all the samples irrespective of cluster has low access since the nearest hospital with good tertiary care in 5 hours away in the state headquarter. Likewise, the region has low aspirational employability, low transitioning of women to jobs even after high education.



<u>Fig. 2</u>: Feature Importance in the Model.

The model gave a test accuracy of 0.83 per cent (Fig. 3) with a random split of 0.2 with Entmax gradient activation. The Entmax function generalizes the softmax and sparsemax activations, allowing for sparse output probabilities (some exact zeros), making it suitable for interpretable models like TabNet (Peters M E., et al. (2019).

The Entmax transformation with parameter $\alpha \in [1, 2]$ is defined as:

$$\text{Entmax}_\alpha(\mathbf{z}) = \arg\max_{\mathbf{p} \in \Delta^K} \left(\langle \mathbf{p}, \mathbf{z} \rangle - H_\alpha(\mathbf{p}) \right)$$

Where:

- $\mathbf{z} \in \mathbb{R}^K$ is the input vector
- $\Delta^K = \{\mathbf{p} \in \mathbb{R}^K : p_i \geq 0, \sum_{i=1}^K p_i = 1\}$ is the probability simplex
- $H_\alpha(\mathbf{p})$ is the **Tsallis α-entropy**:

$$H_\alpha(\mathbf{p}) = \frac{1}{\alpha(\alpha - 1)} \left(1 - \sum_{i=1}^{K} p_i^\alpha \right)$$

Special cases:

- $\alpha = 1$ → **Softmax**
- $\alpha = 2$ → **Sparsemax**
- $\alpha = 1.5$ → **Entmax 1.5** (commonly used in TabNet)

	Precision	**Recall**	**F1 Score**	**Support**
Cluster 1	0.77	0.75	0.76	31
Cluster 2	0.85	0.76	0.80	28
Cluster 3	0.83	0.77	0.78	30
Accuracy	0.83			89
Macro Av.	0.81	0.76	0.79	89
Weighted Av.	0.81	0.76	0.79	89

Overall Test Accuracy: 0.82697

Fig. 3: Model Accuracy for Impact transition (Indicators) Across the Clusters

It shows promise in explaining socio-economic data sets with varied data structures. However, the model robustness is likely to improve as more data points are added. As this was a demonstration test for use for development agencies, it is likely to nudge the agencies to choose data driven decision making. However, there is a need to integrate other data structures like, qualitative interviews, focused group discussions. Also, the contextuality warrants an integration of disaster data and climatic parameters which is likely to follow after this.

The variables like Food Security showed the most influence in the model, meaning, in between cluster changes, there are likely differences in the manner people rely on rations for their nutritional upkeep. Likewise, development spillovers may be seen in terms of ownership of material assets, the housing structures, medium of cooking, a change in ownership of traditional goods like livestock (cow and buffaloes). On the other hand, having a car or bicycle in the house doesn't contribute significantly, or the availability and accessibility to electricity or the language. Social status and welfare services as an aggregate thematic block show maximum effect across the clusters among the households surveyed followed by the ownership of material assets. However, this discounts the importance of the theme areas. For instance, on the accounts of health access, all the samples irrespective of cluster has low access since the nearest hospital with good tertiary care in 5 hours away in the state headquarter. Likewise, the region has low aspirational employability, low transitioning of women to jobs even after high education.

There is a need to integrate other data structures like, qualitative interviews, focused group discussions. Also, the contextuality warrants an integration of disaster data and climatic parameters which is likely to follow after this. As a contribution, this study does serve purpose for policy makers in identification of existing differences in the society. As we say, impact affects differently both in terms of benefits and burdens. Over a period of time, projects with longer gestation period can utilize the model in monitoring and evaluation and of course micro level policies could be sought not only to serve as a reference documents, but building over it. Further this study helps into identifying the cross thematic variables and their couplings.

Conclusion

This study shows dimensional interactions in between indicators of development at a household level. By using TabNet as a model for integration, the framework explains cross thematic interaction among the 49 indicators. Using EntMax activation, the model explains the resonance between the social and economic themes which establish quantitative coupling signaling an overall change in the status of vulnerabilities of the households surveyed. Further, this approach proves handy in explaining impact transition in non-longitudinal data sets. In addition to the case study methodology this compliments the spatial explanations of impacts of development in the field.

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. Arik, S. O., & Pfister, T. (2019). TabNet: Attentive interpretable tabular learning. *arXiv preprint arXiv:1908.07442*.

3. 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

4. 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

5. 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

6. 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*.

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

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

9. 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.

10. 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

11. 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

12. 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