Ritika Singh
Indian Institute of Technology Delhi
Statement of Research Interest

Research Overview My research focuses on learning interpretable latent structure from noisy, heterogeneous, and spatially and temporally evolving data, with the goal of understanding and addressing unequal health and developmental outcomes. Across my doctoral work, I have developed probabilistic, spatial-temporal, and machine-learning models that infer unobserved processes-such as access, engagement, risk, and developmental vulnerability-from large-scale observational data. A defining feature of my work is the use of latent variable modeling as a tool for substantive inference, not merely dimensionality reduction or prediction. Rather than treating outcomes as isolated labels, I model them as the observable consequences of underlying latent processes shaped by social, structural, and contextual factors. This approach allows predictive models to remain interpretable, uncertainty-aware, and aligned with real-world decision-making.

Doctoral Research: Decoupling Drivers of Unequal Outcomes

My doctoral research addressed the challenge of identifying zero-dose and under-immunized children in low- and middle-income countries, where data are often sparse, biased, and irregularly observed. Instead of treating under-vaccination as a single failure event, I focused on decoupling its latent drivers using a combination of generative modeling, spatial-temporal inference, and interpretable machine learning. This work comprises four interconnected projects.

Latent Variable Modeling of Behavioral Drivers: A central contribution of my work is the development of latent variable models to infer unobserved behavioral and structural drivers when direct measurement is unavailable. Using large-scale survey data across multiple national contexts, I developed generative Bayesian frameworks that model observed outcomes as the interaction of latent processes. In work on vaccine uptake, I used Item Response Theory-based latent traits to infer individual-level healthcare access from routine indicators, and subsequently inferred a latent demand trait governing acceptance conditional on access. Embedding these components within a hierarchical Bayesian model enabled decomposition of under-vaccination into access-constrained versus demand-constrained populations. Importantly, model-derived demand estimates showed strong agreement with independent, real-world attitudinal survey data, strengthening their substantive interpretability.

Large-Scale Spatial and Spatiotemporal Modeling: To quantify disparities at fine geographic scales, I developed large-scale spatial and spatiotemporal models using spatial generalized linear models (spGLMs) and SPDE-based formulations implemented via Integrated Nested Laplace Approximation (INLA). Working with national health surveys containing geographic identifiers, I modeled spatial dependence, regional heterogeneity, and measurement uncertainty directly, producing hyper-local estimates with principled uncertainty quantification. A key methodological focus was the *treatment of maternal recall bias* in immunization data. Rather than excluding recall-based observations, I modeled them probabilistically, demonstrating that recall data meaningfully improves coverage estimates when incorporated with appropriate uncertainty bounds. This work reinforced my broader modeling philosophy: treating space and time as latent processes, rather than nuisance variation, enables more faithful inference from imperfect data.

Interpretable Machine Learning for Sequential Risk Processes: Building on these probabilistic foundations, I developed interpretable machine-learning frameworks to study risk across sequential processes, such as progression through the immunisation cascade. Rather than optimizing prediction alone, I designed analyses to explicitly separate model performance from substantive interpretation. This work combined model-based comparisons between linear and nonlinear methods, cascade-level analyses across outcomes, and region-specific modeling to assess geographic heterogeneity in inferred mechanisms. By examining feature contributions across models and contexts, I identified non-linear and stage-specific relationships that are obscured in aggregate analyses. Conceptually, this reflects my broader interest in interpretable representation learning for sequential decision problems, where understanding how and why risk evolves over time is as important as forecasting outcomes.

Latent Risk Modeling for Early Childhood Development: I extended this framework to early childhood development (ECDI), where outcomes are multidimensional, age-dependent, and weakly observed. Using household- and child-level data, I developed predictive models to identify children at risk of being developmentally off-track, with careful attention to outcome construction, domain aggregation, and uncertainty. In addition to modeling, I actively participated in the development and contextual adaptation of ECDI survey instruments for Indian village settings. This process involved translating abstract developmental constructs into contextually meaningful, behaviorally grounded measures and assessing measurement validity in situ. This experience strengthened my ability to reason about construct alignment, systematic noise, and the relationship between measurement and modeling, reinforcing my treatment of developmental status as a latent trajectory rather than a fixed label.

Moving forward, my goal is to extend this experience and learning to richer, more temporally resolved data, while preserving an emphasis on interpretability and inference. I am particularly interested in developing probabilistic and deep latent dynamical models that infer meaningful latent states from irregular, multimodal data streams. I view upcoming positions not as a departure from my doctoral research, but as a natural extension of a coherent modeling framework-one that treats observed data as noisy projections of underlying processes and prioritizes understanding alongside prediction.