

Explainable Digital Twin Modelling of Semi-Urban Community for Multidimensional Poverty Governance Intelligence

Taiwo Olapeju Olaleye

agsobaolaleyetaiwo@gmail.com

Federal university of Agriculture Abeokuta <https://orcid.org/0000-0001-6222-7575>

Oluwasefunmi Arogundade

Federal University of Agriculture Abeokuta <https://orcid.org/0000-0001-9338-491X>

Taiwo David Ajayi

Federal University of Agriculture Abeokuta <https://orcid.org/0009-0004-7920-9700>

Oluwasegun Dada

Federal University of Agriculture Abeokuta <https://orcid.org/0000-0002-0304-9032>

Research Article

Keywords: Digital Twin, Multidimensional Poverty Index, Sustainable Development Goals, Explainable AI, Vulnerability Mapping, Spatial Autocorrelation, Machine Learning

Posted Date: September 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11020829/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Explainable Digital Twin Modelling of Semi-Urban Community for Multidimensional Poverty Governance Intelligence

Taiwo Olaleye¹, Oluwasefunmi Arogundade¹, Taiwo Ajayi¹, Oluwasegun Dada²

¹*Software Engineering and Information Systems Department, Federal University of Agriculture Abeokuta.*

²*Information and Communication Technology and Information Technology Department, Federal University of Agriculture Abeokuta.*

olaleyeto; arogundadeot; ajayitd; dadaov{@funaab.edu.ng}

Corresponding Author: T. O. Olaleye
olaleyeto@funaab.edu.ng

Abstract

Conventional poverty assessment and governance frameworks are continually challenged by the dynamic nature of multidimensional inequality within African peri-urban communities. Existing static and descriptive approaches are fast becoming insufficient, incapable of availing adaptive management. This study presents an explainable Digital Twin (DT) framework that integrates supervised, un-supervised and explainable machine learning methodologies, counterfactual simulation, cost-aware policy optimization, and synthetic population modelling to support multidimensional poverty governance of a semi-urban community. This was motivated by the growing concerns over Multidimensional Poverty Index rate in Nigeria, and functionalities offered by the DT paradigm. The study developed a computational digital replica of Isolu vulnerability dynamics by applying the Alkire-Foster methodology to enumerate household deprivation states. This availed K-Prototypes the modelling of households into distinct vulnerability stratification and spatial analytical techniques to investigate poverty hotspots and derive inequality surface maps. The DT simulation modelled counterfactual household states under multiple policy interventions, returning food support as the most efficient standalone intervention across the vulnerability clusters. Further simulations unveiled largest absolute reductions in multidimensional poverty through multi-policy strategy, integrating electricity, internet, and water access interventions as superior system-level poverty transformation factors. Optimization of intervention packages across graduated populations revealed that uniform safety-net initiatives generated varied outcomes, demanding cluster-specific optimization strategies. Empirical evidences also showed that though poverty was socially structured in the case study, it is however spatially dispersed. Findings validated the effectiveness of the conceptual framework as a computational policy laboratory capable of identifying context-sensitive and economically efficient intervention pathways for intelligent poverty governance.

Keywords: Digital Twin, Multidimensional Poverty Index, Sustainable Development Goals, Explainable AI, Vulnerability Mapping, Spatial Autocorrelation, Machine Learning

1. Introduction

Developing African communities are experiencing deepening multidimensional inequality, particularly peri-urban settlements which are usually characterized by fragile infrastructure (Egba et al., 2026), informal economies (Carriho et al., 2025), and uneven access to essential services (Atigaku et al., 2026). Income-based deprivation metrics have been over-flogged by conventional studies as the traditional causative factor of poverty rate but emerging evidences suggests that vulnerability within underserved communities is increasingly multidimensional (Belete, 2024) and behaviourally dynamic (Angai, 2026). Therefore, complex interactions between socioeconomic conditions, infrastructure inequality, spatial distribution, and household resilience metrics cannot be efficiently modelled by static, non-adaptive poverty assessment tools, which could undermine the efficiency of intended policy interventions (Vidya et al., 2025). In Nigeria, sub-urban settlements adjoining university communities represent a particularly important but underexplored category of vulnerable ecosystems. Such communities specifically experience rapid and dynamic demographic expansion, which are driven by student migration (Anukam et al., 2026), informal housing development pattern (Sharma, 2026), unstable livelihood structures (Majumder, 2026), and unabating pressure on public utilities (Ezzati et al., 2025). Isolu, a peri-urban suburb to the Federal University of Agriculture Abeokuta (FUNAAB), characterize these observations. As a residential and economic extension of the university environment, there exist an obvious substantial inequality pattern in access to electricity, internet connectivity, food security, healthcare, educational stability, and environmental resilience. The close proximity of an academic institution of FUNAAB status (7th best University of Agriculture in the world by the Times Higher Education (THE) World University Rankings of 2024) usually hides such vulnerability pockets because of the assumption of relative socioeconomic stability. However, a layered vulnerability structure is apparent due to the coexistence of student populations, informal workers, low-income and mud-built households, illiterate villagers, and transient residents, thereby necessitating an adaptive analytical framework for interrogating its vulnerability dynamics.

Components of computing sciences have been widely applied to poverty governance but Artificial Intelligence (AI) remained the most broadly useful segment, availing an intelligent approach to optimizing expert knowledge for managing vulnerability dynamics (Saha *et al.* 2025). However, the use cases of AI; Machine Learning, Natural Language Processing, Expert System, Fuzzy Logic, etc., focuses on theoretical capabilities or isolated lab experiments, which may not fully guarantee an integrated system for an adaptive methodology. Applied Intelligence bridges this gap by providing a nexus between technology and execution to deliver measurable outcomes (Iqbal *et al.*, 2020). It is defined as the practical use of AI, advanced data analytics, and automation to solve real-world operational problems (Dantas *et al.*, 2024), with Digital Twin (DT) as a prime use case example. The operational brain of DT adopts Applied Intelligence framework to produce an intellectual expert system for solving domain problems. Therefore, simulation of an object's digital copy through its feature representations and AI algorithms creates a complete Applied Intelligence loop. A DT methodology is an adaptive computational replica of a real-world system, capable of

integrating data (real-time or observed) with predictive simulation and decision intelligence (Tao et al., 2022). This provides opportunities for modelling complex human-related systems beyond the expertise of traditional descriptive analysis (Hassani and MacFeely, 2025), including simulation of future possibilities. Although DT have gained substantial adoption for smart cities (Deng, et al., 2021), manufacturing (Onaji et al., 2022), and industrial systems (Xu et al., 2023) etc., its application to multidimensional poverty management is rare, especially for university community context. Most importantly, existing poverty mapping studies lack integrated use cases for counterfactual policy simulation and intervention optimization, and explainable decision support. This study therefore provides insights on how digital twin modelling of semi-urban African university communities could support intelligent poverty governance initiatives that improves quality of livelihood, thereby enabling cost-efficient and result-oriented social safety-net intervention ideas. A DT-driven geospatial poverty analytics framework was experimented for policy refinement and optimization in underserved African community. The framework operationalized the quantification of Multidimensional Poverty index (MPI), geospatial hotspot analysis, eXplainable Artificial Intelligence (XAI), machine learning, and an optimization-oriented counterfactual simulation, within a unified computational architecture. Household-level primary survey data was deployed to test the framework, to eventually model a ranking intervention strategy according to cost-efficiency and MPI reduction potential. The main aim of the study was therefore to simulate household-level interventions to manage poverty for a university suburb, by a way of intelligent poverty governance.

Specific objectives are to:

- i. Implement a Alkire-Foster-based multidimensional poverty profiling of surveyed households in Isolu university community.
- ii. Determine the vulnerability stratification of the households using evaluated K-prototypes by unsupervised learning
- iii. Investigate the possibility of poverty hotspots in the community using Moran's I and Getis-ord Gi* through a spatially digital representation of vulnerability clusters.
- iv. Implement, test and optimize counterfactual intervention simulation to quantify the minimum and profitable feature-adjustments needed to transition households from high vulnerability state to lower deprivation states.
- v. Develop a Gaussian Copula-based synthetic population generation framework to evaluate the scalability and generalizability of baseline outcomes across expanded household populations

The study makes the following original contributions to knowledge:

- i. An ensemble and explainable Digital Twin framework was developed for multifaceted poverty governance by integrating Digital Twin technology, unsupervised, supervised and explainable machine learning, counterfactual reasoning, and policy optimization into a unified decision-support architecture.

- ii. A counterfactual intervention modelling framework was proposed that enables virtual experimentation on household socioeconomic states, allowing policymakers to evaluate alternative intervention scenarios before real-world implementation while quantifying their effects on multidimensional poverty transitions.
- iii. A cost-aware policy optimization model was introduced to identify intervention packages that simultaneously maximize multidimensional poverty reduction and resource efficiency, thereby providing an evidence-based mechanism for optimal allocation of limited public resources.
- iv. A cluster-adaptive poverty governance strategy was established by demonstrating that intervention effectiveness varies significantly across vulnerability strata, providing empirical evidence that uniform poverty alleviation programmes are suboptimal compared with vulnerability-specific intervention policies.
- v. A scalable synthetic population validation framework based on the Gaussian Copula Synthesizer was developed to expand the household population from 720 to 5,000 and 10,000 households while preserving the multivariate dependency structure, thereby demonstrating the robustness and generalizability of the proposed Digital Twin framework beyond the original survey sample.
- vi. The study advances the application of Digital Twin technology to social policy, extending its traditional use in engineering and industrial systems to multidimensional poverty governance and demonstrating its potential as an intelligent policy simulation and optimization platform for achieving the Sustainable Development Goals, particularly SDGs 1, 2, 6, 7, 8, 9, and 10.

The rest of the study is structured in the following way: existing methods in literature are discussed in chapter 2 while the methodological approach of this study is presented in chapter 3. Experimental results are presented and discussed in chapter 4 and the study is concluded in chapter 5.

2. Literature review

The origin of Digital Twin technology has been traced to a Product Lifecycle Management at the University of Michigan, in 2002 by Michael Grieves (Singh *et al.*, 2021). The proposal was composed of the real space, virtual space and the linking mechanism for free-flow of data between the two, thereby described as ‘Mirrored Spaces Model’ (Grieves, 2020). A similar concept called ‘Mirror Worlds’ was conceptualized by David Gelernter in 1991, in which software models mimic reality through physical world information (Gelernter, 1993). By 2006, Grieves’ conceptual model was improved into ‘Information Mirroring Model’ (Grieves, 2020), showing improvements on concepts and prospects. The phrase ‘Digital Twin’ was then identified in NASA’s draft version of the technological roadmap in 2010 (Singh *et al.*, 2021), where it was defined as “an integrated multi-physics, multi-scale, and probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin”. Subsequently, authors have devised means of DT description including digital model (Rosen *et al.*, 2015), counterpart (Negri *et al.*, 2017), doppelganger (Wired Brand Lab, 2020), clone (Banerjee *et al.*, 2017), software analogue (Volodin and Tolokonskii, 2019), information constructs (Kritzing *et al.*, 2018), or sometimes simulation (Shafto *et al.*, 2010). A common idea in most

definitions is the bidirectional sharing of data between the physical and digital counterpart of an entity, other than just being described as virtual representation of a physical object. Identified data sources also include both quantitative and qualitative, historical, environmental, but most importantly, a real-time data (Singh *et al.*, 2021). Digital Twin can then be described as a functioning system, representing the digital representation of an actual (or intended) physical entity (or environment), which collects data, processes it, and makes decisions based on the current, continuous and future patterns analyzed from the data.

The DT architecture is built to continuously bridge a physical entity with its virtual counterpart, through layers that process real-world data into actionable virtual insights. The layers broadly include:

- i. Physical Layer: The real-world entity, including a hardware, machinery, or environment, etc. that is being twinned or intended for virtual representation.
- ii. Digital Model Layer: The core of the twin, modelling the geometry, physics, logic, orientation, behavioral definitions, etc. that mirror the real entity.
- iii. Analytics and AI Layer: The engine running simulations, machine learning, and *what-if* scenarios to predict failures, detect anomalies, or optimize operations.
- iv. Visualization & Application Layer: The user interface where operators interact with the twin via 3D graphics, XR environments, or operational dashboards.

Digital Twin has been largely described in literature within the manufacturing context but its functional specifications are indeed viable for Sustainable Development Goal (SDG) use case. This is made possible by the dramatic technological revolution across all aspect of human life (Cordes et al., 2024), typically by deploying captured moments for simulating future possibilities. The popularity of DT and SGD are concurrent, as phenomenon with global popularity and interest. A Google Trend analysis of 4th June, 2026 to check their search-trend admiration within 24 hours is presented in Fig1. Withing the day, SDG secured a 100% peak popularity at 6:36am GMT-1 when DT was 41% as popular at its peak, whereas, DT peaked 56% at 12:28am GMT-1 when SDG net 41%. Both paradigms represent popular phenomenon advancing livelihoods across diverse demography. In order to aggregate the applicability of DT for SDG actualization, Hassani *et al.* (2022) conducted a review on the enablement of DT for achieving United Nations SDGs. Authors identified significant improvements in sustainable production, maintenance, logistics and circular economy with the aid of DT methodology. Smart construction and building, energy and resource management, urban planning and smart water infrastructure are other SDG constructs empowered by the DT paradigm. Expansion of DT application to other fields like agriculture and livestock farming, etc., were researched where concepts such as smart agriculture (Ayazi *et al.*, 2019), smart farming (Navarro *et al.*, 2020) and smart fishery (Ebrahimi *et al.*, 2021) emerged.

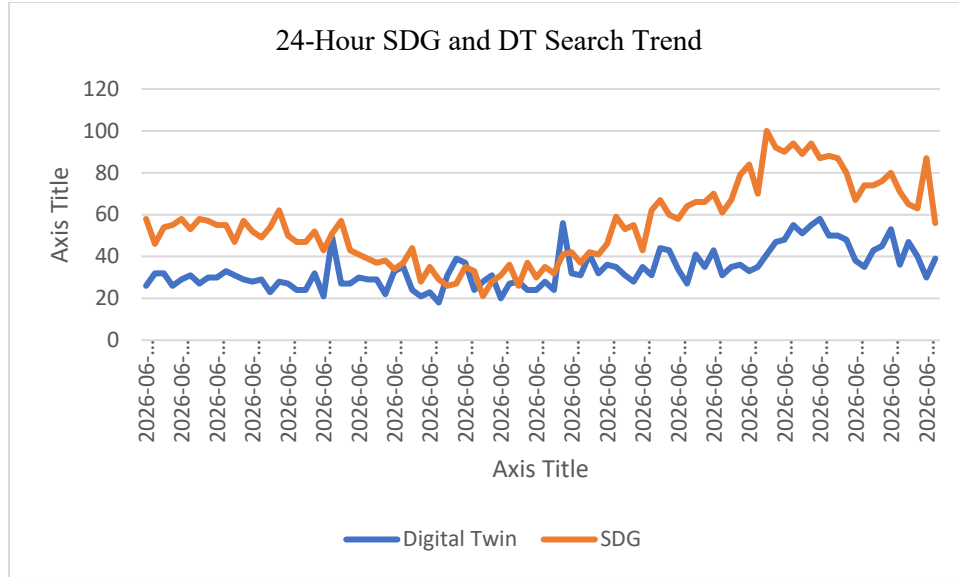


Fig.1. Popularity trend of Digital Twin and SDG concepts

Poverty in Africa has been widely described as multidimensional (Cordes et al., 2024), spatially characterized (David et al., 2018)], and usually linked to quality of living (Khan and Sloboda, 2023), food security (Mathenge et al., 2023), healthcare (Khan and Sloboda, 2023), infrastructure (Suhardono et al., 2025), employment (Liu et al., 2023), digital access (Ruiu and Ragnedda, 2024), etc. Despite global mitigation commitments, especially under the SDGs, particularly SDG 1 (No Poverty) (Çelik et al., 2025), many African communities continue to grapple with unabating vulnerability driven by several validated factors including rapid urbanization (Ahimah-Agyakwah et al., 2022), lingering economic inequality (Thorbecke, 2023), climate stress (Leichenko and Silva, 2024), weak institutional capacity (Appiah-Otoo et al., 2022), and limited social protection systems (Nowak-Garmer, 2023). The interconnected nature of deprivation captured in SDG constructs implies that poverty cannot be addressed through isolated interventions (Motadi, 2025), as deficiencies in one sector often reinforce vulnerabilities in others. Consequently, recent studies increasingly advocate for data-driven (Onsay and Rabajante, 2025), geospatial (Arya et al., 2025), and Artificial Intelligence-enabled (Sheik and Kabanda, 2026) methodologies, with the capability of modelling complex poverty interactions and dynamics. The SDGs view a host university community like Isolu as a grassroots priority for sustainable transformation (Rankakane et al., 2025) with the pledge to “leave no one behind”, making suburban and rural neighbourhoods prime targets for inclusive development. The interaction, relationship and expected impact of SDG constructs on Isolu can be understood through targeted focus goals like SD1 (No Poverty) and SDG 2 (Zero Hunger) (Parotto and Pablos-Méndez, 2023), seeking to bridge the gap between academic innovation and community welfare. Sustainable Cities and Communities (SDG 11) likewise envision localized planning to ensure adjacent communities to universities enjoy safe housing, reliable infrastructure (including clean water and sanitation), and sustainable waste management (Phdungsilp, 2022), for a student-tenanted, suburban hub of that nature. Quality Health and Well-being (SDG 3), Quality Education (SDG 4) and Partnerships for the Goals (SDG 17) are all signaling Isolu possibilities, advocating access to primary healthcare and sanitation, lifelong learning and skills acquisition, town and gown collaborations (Karunaratne et al., 2024). An adaptive and intelligent-based system is therefore imperative for mapping wellbeing of residents

who usually exhibit heterogeneous vulnerability dynamics owing to the transitional nature of student-population.

The study of Cordes *et al.* investigated the potential role of DT for poverty alleviation in Tetegu, a sub-sahara Ghana community. An aerial and community surveys of the built environment including the socioeconomic profiling of businesses and households were acquired to inform a virtual representation of the study area. Business intervention of an e-commerce business enterprise was introduced to simulate impact on the community, and a limited view obtained from a small population sample revealed that online transactions increased at the intervention, indicating the willingness of businesses and customers to engage e-commerce given incentives. Similarly, Shehadeh *et al.* (2024) explored the application of DT to advance the SDGs within the water-energy-food-environment in refugee community. The study modelled the Irbid camp in 3D Revit in combination with real-time data and community insights, operationalized with machine learning algorithms. Examination of four hundred and fifty qualitative surveys revealed an 80% knowledge level regarding conservation of water practices but 35% satisfaction rate was achieved with the current management of resources. A 25% increase in water scarcity was estimated by predictive analytics with an 18% surge in energy demand projected for the next decade. Resource allocation was simulated by the DT engine alongside environmental impacts, and urban development planning was achieved in line with SDGs 6 (Clean Water and Sanitation). Saha *et al.* (2025) modelled household level data from Tanzania for predicting multidimensional poverty using predictive analytics. Longitudinal data was employed by authors from the 2014/15 and 2020/21 public data from the Tanzania National Bureau of Statistics, fitted on predictive algorithms including SVM, Random Forest, Logistic regression, Decision tree, Gradient Boosting, etc. Performance evaluation returned efficient performances across metrics. In another vein, multidimensional poverty modeling of Benin was the aim of the work of Arranhado *et al.* (2025) combining fractional regression and random forest modelling. White-box modelling of the framework was achieved by the Shapley Addictive eXplanations and the ALE plots, thereby returning education as the major impact on poverty with traces of gender differences and natural conditions, household structure, etc. Rodermund *et al.*, (2024) presented a multi-layered and agent-based DT for modelling sustainable future of rural areas in Germany. Authors conceptualized a framework designed to simulate scenarios and intervention for fostering pro-environmental behaviours. Also, a risk-informed DT simulation of buildings was implemented by Alibrandi (2022) for a resilient urban community. Authors introduced a novel DT methodology that is risk-informed. In the approach, predictions are expectedly data-driven with algorithmic operationalization. Multiple uncertainty dynamics challenged the simulation lifecycle. The study prides in its statistical and machine learning-based risk analysis.

Observations from existing studies reveal that suburban settlements adjoining university communities are seldom investigated for vulnerability dynamics and socioeconomic deprivation severity. Also, existing poverty mapping studies lack integrated use cases of counterfactual policy simulation and intervention optimization, with explainable AI decision support.

3. Research methodology

This study conceptualized a research methodology based on the functionalities of DT framework, domesticated in the conceptual framework presented in **Fig.2**. The framework operationalizes

household-level interventions to mitigate poverty, and subsequently quantify the minimum feature-adjustments needed to shift individual households from high vulnerability state to low deprivation status. Initially, vulnerability profiling of households and poverty mapping was defined by predictive analytics for the case study. This implies a small-area estimation methodology, using statistical technique to evaluate welfare indicators for disaggregated geographical units such as the peri-urban Isolu community. The framework includes the multidimensional poverty indexing of households based on their socioeconomic differentials, and an unsupervised geospatial modelling of the feature space to achieve a spatially explicit digital representation of household vulnerability states. The framework further quantifies and implements the minimum intervention thresholds that are required to shift households out of their current vulnerability states, as intervention package, through counterfactual simulation. Subsequently, optimization phase identified prominent features driving deterioration or improvements, then searched for better intervention combinations in order to determine the most cost-efficient policy package. Feedbacks from the experimental simulations were tested for generalizability through feature engineering of surveyed data into graduated synthetic populations of 5000 and 10000 respectively. Explainability and transparency were achieved through Shapley Addictive Explanations (SHAP) and supervised Decision Tree analytics.

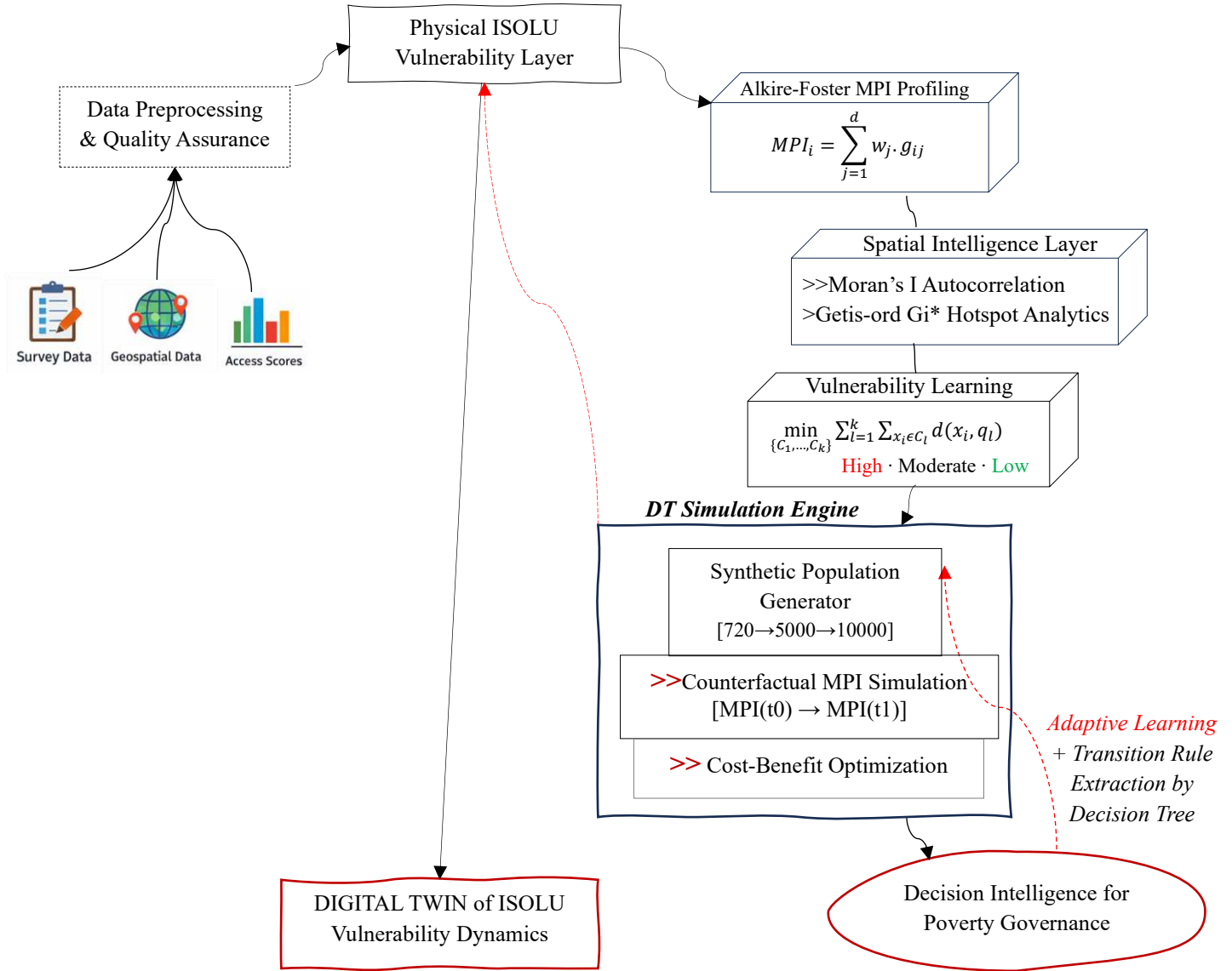


Fig.2. Framework of the DT-based research methodology

3.1 Physical Community Layer

The Isolu community characterizes the physical layer, a university environ housing a high population of FUNAAB students, being the earliest settlement with close proximity to the university (Google map in **Fig.3**). As at 2017, the population of Isolu was put at 1,500 [21], which has actually experienced exponential growth, due to the continued expansion of the university curricular ecosystem. Thirty-two (32) socioeconomic and geospatial features were captured in a survey conducted between April and December 2025 using Google form survey instrument, validated by a sociologist. The survey captures the MPI indicators of health, education, living standard metrics, geospatial data of respondents, their perceived deprivation access scores, etc. In

all, seven hundred and twenty (720) participants, representing different household datapoints, participated. The feature space is described in Table 1, including Timestamp (S/N 1) and consent of participation expressed by the respondents (S/N 34).

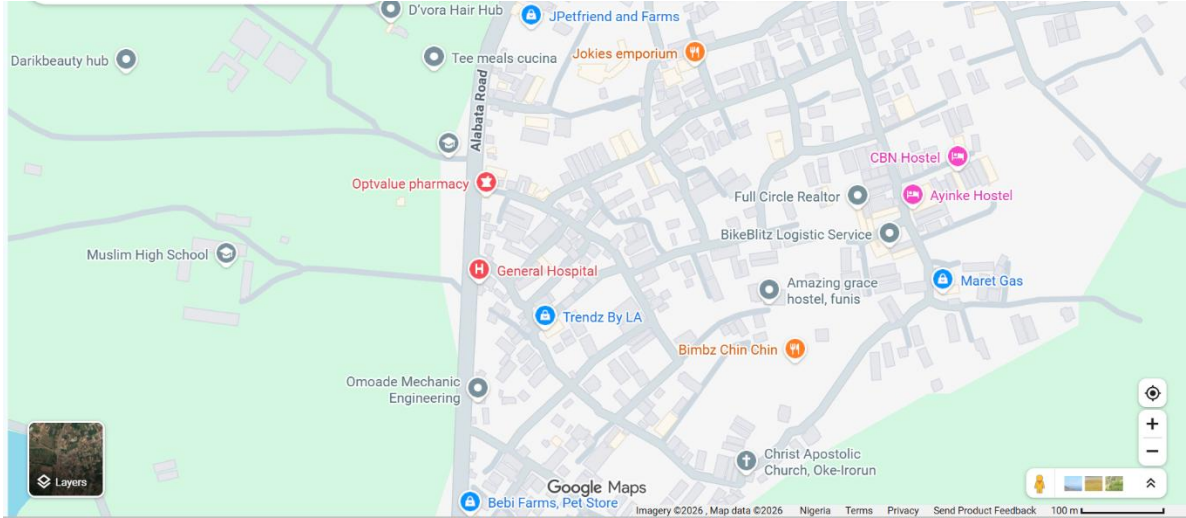


Fig.3. The geospatial map of Isolu community as captured on Google map

Table 1. Overview of surveyed feature space

S/N	Feature description	Variable name	Data type
1	Timestamp		Continuous
2	Age	Age	Numeric
3	Gender	Gender	Categorical (Y/N)
4	Size of households,	House_size	Numeric
5	Dependability of households on care-giver	Care_giver_inde	Categorical (primary caregiver/dependent)
6	Source of income	Income_source	Categorical
7	Actual income	Income	Numeric (in naira)
8	Employment status	Employment_st	Categorical (Y/N)
9	Access to financial services	Acc_fin_ser	
10	Number of meals per day	Meals_day	Numeric
11	History of meal skipping	Skip_meal	Categorical (Y/N)
12	Availability of food aids	Food_aid	
13	Access to farmland	Acc_farm	
14	Access to healthcare	Acc_healthcare	
15	Management of chronic infection	Chro_health_con	

16	Children access to immunization	Immunization	
17	Availability of payment options for health services	Diff_payment_for_health_serv	Categorical (rarely/sometime/never)
18	Education level	Edu_level	Categorical
19	Years of education	Yrs_Edu	Numeric
20	Proximity to children school	Child_hou_enr_sch	Categorical (Y/N)
21	Participation of women in decision making	Women_dec_making	
22	Traces of discrimination of social exclusion	Discrim_socialexclusion	
23	Cases of domestic violence	Domestic_vio	
24	Water access score	Water_acc_score	
25	Health access score	Health_acc_sco	Numeric (1-10)
26	Internet access score	Internet_acc_sco	
27	Electricity access score	Elect_acc_sco	
28	Food security score	Food_security_sco	
29	Climate resilience options	Climate_rel_iss	Categorical (Y/N)
30	Fire charcoal cooking	Fire_charcoal_cooking	
31	Access to sanitation	Acce_sanitation	
32	Latitude	Latitude	Decimal
33	Longitude	Longitude	
34	Participant consent		Categorical (Y/N)

A rigorous data preprocessing was done on the survey. Various mismatches were observed with the Age variable where participants responded with either numeric or categorical data types. The same noise was prominent with the Income feature with responses including notations like comma (,), full stop (.), etc. to indicate thousands. Missing values were observed across the numerical and categorical features which were addressed by computing the mean and mode values respectively. The phase addressed the various anomalies and inconsistent inputs, standardized categorical variables, convert messy numeric fields, etc. This is an essential logic for survey-type feature space of this study. The correlation coefficient and Variance Inflation Factor (VIF) are both implemented to quantify the presence of multicollinearity in the data layer, a germane quality assurance practice in data science-based modelling. Multicollinearity usually introduce redundancy in numerical variables thereby undermining the ability of modelling socioeconomic dynamics of the feature space by the unsupervised K-Prototypes, by distorting the centroid estimation and clustering assignment of the model. Let the Isolu numerical feature space be represented as $X = \{x_1, x_2, \dots, x_p\}$, if two variables such as the electricity access score and the water access score exhibit strong linear dependence, that is:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ik} - x'_i)(x_{jk} - x'_j)}{\sqrt{\sum_{k=1}^n (x_{ik} - x'_i)^2 \sum_{k=1}^n (x_{jk} - x'_j)^2}} \quad (1)$$

where $r_{ij} \rightarrow 1$, they would be contributing identical information to the Euclidean distance component of the K-Prototypes such that:

$$d(x_i, c_j) = \sum_{m=1}^q (x_{im} - c_{jm})^2 + \gamma \sum_{m=q+1}^p \delta(x_{im}, c_{jm}) \quad (2)$$

Therefore, over-weighting of the correlated variables would occur, causing an erroneous cluster compactness and unstable segmentation boundaries, during distance minimization. therefore, Pearson correlation is often used for multicollinearity detection by evaluating pairwise linear dependency between two variables but a higher order dependency inflation is better suited for the purpose of this study as it rigorously quantifies the severity level of correlation using the VIF:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (3)$$

where R_i^2 is the coefficient of determination obtained when feature x_i is regressed against all remaining predictors. Therefore, large VIF values (eg. >5 >10) signals severe multicollinearity between two variables, implying they contribute redundant variance which may bias centroid formation, reduce cluster interpretability, and weaken the reliability of the vulnerability segmentation during K-Prototypes analytics.

The pairwise correlation (ρ_{ij}) existing between two variables x_i and x_j is also defined by:

$$\rho_{ij} = \frac{Cov(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}} \quad (4)$$

where $\rho_{ij} \in [-1, 1]$.

3.1.1 Multidimensional poverty profiling

The MPI is computed on each of the household datapoints, based on the Alkire-Foster method. General MPI model is formulated as:

$$MPI_i = \sum_{j=1}^d w_j \cdot g_{ij} \quad (5)$$

where MPI_i is the poverty score for an individual/household i ; d is the number of indicators; w_j is the weight of an indicator j ; and $g_{ij} \in \{0, 1\}$ is the deprivation status of the individual/household.

For this study, an explicit mapping is done:

$$MPI_i = \frac{1}{9}(H_i + N_i + E_i + W_i + El_i + I_i + S_i + Inc_i + Emp_i) \quad (6)$$

where H_i (health access deprivation), N_i (Nutrition deprivation), E_i (education deprivation), W_i (Water access), El_i (Electricity), I_i (Internet), S_i (Sanitation), Inc_i (income), and the Emp_i (Employment). This is a weighted deprivation index that indicates a continuous vulnerability score.

3.1.2 Spatial Intelligence and Vulnerability Learning

The spatial autocorrelation of households is achieved alongside the K-Prototypes unsupervised modelling for determining the spatial vulnerability states of households. The K-Prototypes extends the K-Means model with the ability to analyze both categorical and numeric data types. The objective function is to partition the dataset into k clusters such that the total dissimilarity between each datapoint and its assigned cluster prototype (Low, Moderate and High vulnerability states) is minimized. Let k be the number of clusters and $\mathcal{C}_l$ be the set of data points in cluster l . The prototype of cluster l is the q_l , which is the mean for numerical and mode for categorical. Therefore, the optimization objective is achieved by:

$$\min_{\{C_1, \dots, C_k\}} \sum_{l=1}^k \sum_{x_i \in C_l} d(x_i, q_l) \quad (7)$$

The Euclidean distance is implemented for the numeric (num) features and the Hamming distance for the categorical (cat) data types with the distance function:

$$D(x_i, C_k) = \sum_{m \in \text{num}} (x_{im} - \mu_{km})^2 + \gamma \sum_{m \in \text{cat}} 1(x_{im} \neq \mu_{km}) \quad (8)$$

where γ assures weight balancing (numeric vs categorical) and D is the spatial input.

The structural risk analysis is therefore achieved by grouping the individual households into the Low, Moderate and High-risk clusters. Therefore, each household i is represented its state:

$$State_i = (MPI_i, Cluster_i)$$

forming a DT representation of vulnerability states. Clustering validity is evaluated using the silhouette score and Davies–Bouldin index to estimate separation and compactness of vulnerability clusters. Additionally, cluster-wise MPI distributions and spatial centroids were used to assess socio-spatial coherence.

3.2 Digital Model Layer

The digital representation of the physical Isolu community is achieved in this layer. Based on the earlier clustering, a geo-aware clustering is done to achieve the spatial component of the earlier phase:

$$D_{spatial} = (lat_i - lat_k)^2 + (lon_i - lon_k)^2 \quad (9)$$

using the Latitude (lat) and the Longitude (lon) component of the features.

The Moran's I and the Getis-ord G_i^* will check presence of spatial autocorrelation and hotspot concentration in the geospatial analysis of the household deprivation dynamics. This will achieve the objective that seeks to investigate the possibility of poverty hotspots, and avail information as to whether poor residents in Isolu are concentrated or scattered. The MPI is treated as a normalized deprivation score between $0 \leq MPI_i \leq 1$, while the Moran's I (spatial autocorrelation) would test if similar MPI values are geographically clustered using the MPI, Latitude and Longitude features:

$$I = \frac{N}{W} \cdot \frac{\sum_i \sum_j w_{ij} (MPI_i - MP'I) (MPI_j - MP'I)}{\sum_i (MPI_i - MP'I)^2} \quad (10)$$

where N is the total households, MPI_i is the poverty score of households I , $MP'I$ is the mean MPI, w_{ij} is the spatial weight based on distance between coordinates (Lat_i, Lon_j) and (Lat_j, Lon_j) and $W = \sum_i \sum_j w_{ij}$.

Possibilities:

- $I > 0$ indicating clustering
- $I \approx 0$ indicating random spatial pattern
- $I < 0$ indicating dispersion

Local poverty hotspots would be identified by the getis-ord G_i^* by:

$$G_i^* = \frac{\sum_j w_{ij} MPI_j - MP'I \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - (\sum_j w_{ij})^2}{n-1}}} \quad (11)$$

where MPI_j is the neighbouring household poverty values, w_{ij} is the spatial proximity weight and S is the standard deviation of MPI.

Possibilities:

- High positive G_i^* indicates hotspot
- Large negative G_i^* indicates a coldspot and
- Near zero G_i^* means a neutral and no hotspot zone

3.3 Analytics and Digital Twin Simulation Layer

Given the geospatial and vulnerability profiling of households, the DT simulation takes effect. The vulnerability score (MPI) and the clustering classification are key inputs into the simulation layer, together with relevant feature attributes. Two functionalities are critical to this phase:

- i. the Counterfactual MPI Simulation and
- ii. the Cost-Benefit Optimization stages.

However, a synthetic population generation is built into the DT engine to test outcome generalizability. This avails a cross-population and scalability analysis needed to know if conclusions from the survey would extend beyond the original sample size of Isolu case study.

3.3.1 Synthetic population generation

A synthetic population expansion is employed to validate the outcome of the experimentations on a graduated household population, scaling from 720 to 5000 and then 10000 datapoints. The Gaussian Copula Synthesizer from the Synthetic Data Vault (SDV) is employed involving marginal distribution estimation of each feature and dependency modelling using a Gaussian copula. For marginal transformation, each feature X_j is transformed into a uniform variable using its empirical cumulative distribution function (CDF):

$$U_j = F_j(X_j) \quad (12)$$

where X_j is the original feature, $F_j(\cdot)$ is the empirical CDF and $U_j \in [0,1]$. The uniform variables are then mapped into Gaussian space:

$$Z_j = \phi^{-1}(U_j) \quad (13)$$

where ϕ^{-1} is the inverse standard normal CDF and Z_j becomes the normally distributed. The synthesizer estimates covariance structure as:

$$\Sigma = Cov(Z) \quad (14)$$

and this captures relationships among the variables:

$$\Sigma = \begin{matrix} & \sigma_{11} & \sigma_{12} & \dots \\ \sigma_{21} & & \sigma_{22} & \dots \\ \vdots & & \vdots & \ddots \end{matrix}$$

thereby preserving dependencies. The synthetic latent samples are thereafter drawn from $Z_{synthetic} \sim N(0, \Sigma)$ and an inverse transformation is implemented to map generated Gaussian samples back:

$$U_j = \phi(Z_j) \quad (15)$$

and then $X_j^{synthetic} = F_j^{-1}(U_j)$, thus preserving feature distribution, socioeconomic structure, spatial variability and the inter-feature correlations. This would return a final synthetic graduated dataset:

$$D_{synthetic} = \{X_1^{synthetic}, X_2^{synthetic}, \dots, X_n^{synthetic}\} \quad (16)$$

that can act as a statistically consistent digital population expansion of the baseline 720 surveyed households.

It is important to verify nonlinear relationships and joint distributions captured in the baseline from the synthetic populations. A Mutual Information (I) between two dataset variables X and Y is therefore defined as:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) \quad (17)$$

where $p(x, y)$ is the joint probability distribution and the $p(x)$ and $p(y)$ are the marginal probability distributions. Specifically, the objective is to prove that:

$$I_{720}(X_i, X_j) = I_{5000}(X_i, X_j) \quad (18)$$

and

$$I_{720}(X_i, X_j) = I_{10000}(X_i, X_j) \quad (19)$$

for all interacting feature pairs:

$$X_i, X_j \in \{Income, Yrs_{Edu}, Int_{Acc_{Sco}}, etc.\} \quad (20)$$

Formulations examples include:

$$I(Income; MPI)$$

$$I(Yrs_{Edu}; MPI)$$

$$I(Inte_{Acc_{Sco}}; MPI)$$

$$I(Latitude, Longitude; MPI)$$

Therefore, the goal of the validation is:

$$\Delta I = |I_{original} - I_{synthetic}| \rightarrow 0 \quad (21)$$

suggesting that dependency structure, multidimensional poverty relationships and spatial interactions remained statistically stable after synthetic population scaling.

3.3.2 Counterfactual intervention phase

This phase of the DT engine generates *what-if* scenarios for each household through the modification of deprivation-related attributes within the surveyed feature space, thereby re-computing a post-intervention MPI (MPI_{t1}). The objective is to quantify the basis for household transition across vulnerability states, within the improved simulation of their socioeconomic conditions. The phase takes on a household attribute for modification, signaling an intervention area, with an intended effect. The interventions thereby produce a counterfactual outcome:

$$MPI_{change} = MPI_{t1} - MPI_{t0} \quad (22)$$

and a resulting vulnerability transition depicted as:

$$Vulnerability_Class \rightarrow Class_{t1}$$

where MPI_{t0} is the initial MPI and the classes indicate the initial vulnerability class (High, Low, or Moderate) and the newly derived class after the transition aided by the change in MPI.

3.3.3 Policy Optimization Phase

Following the counterfactual simulation above, a cost-benefit policy optimization engine analytically evaluates the combinations of intervention packages under budget constraints to identify the most cost-effective poverty mitigation intelligence. Subsequently, an informed poverty-mitigation package is assumed across critical features in order to optimize earlier achievements recorded during the counterfactual simulation. Table 2 compares the two phases for clarity.

Table 2. Overview of the DT simulation phases

Phase	Objective	Output
Counterfactual policy simulation	To determine the behaviour of households to hypothetical improvements on their socioeconomic livelihood	(MPI _{t1}), (MPI _{change}), vulnerability transitions
Policy optimization	To determine exact social safety-net intervention package that delivered the largest MPI reduction per unit cost	Optimal policy package, cost of intervention, efficiency of the intervention, degree of MPI reduction

For this purpose, each individual household i is represented as a state vector:

$$X_i = [x_1, x_2, \dots, x_p] \quad (23)$$

where X_i is the socioeconomic, health and access features, while MPI is a derived state function as:

$$MPI_i = f(X_i) \quad (24)$$

The system then constructs a DT environment for the intervention modelling. The modelling returns the new state replica of households that allows simulation of policy or feature changes, thereby observing resulting shifts in vulnerability state. For such individual households with high vulnerability, we *find* ΔX_i such that $MPI_i^{new} < \tau$, where τ is the threshold for low deprivation and ΔX_i is the minimum intervention changes.

For the counterfactual MPI transition modelling on the other hand, intervention scenarios are applied across the critical features to recompute the their MPI, evaluate their transition and optimize the problem.

We define:

$$MPI_i^{new} = f(X_i + \Delta X_i) \quad (25)$$

and the goal constraint:

$$MPI_i^{new} < \tau$$

The marginal impact of each feature is estimated as:

$$S_j = \frac{\partial MPI}{\partial x_j} \quad (26)$$

such that a higher S_j indicate a stronger influence on poverty reduction, thereby S_j is used to prioritize the interventions. The threshold $MPI^{new} < \tau$ would be derived from the K-Prototypes determined vulnerability classification boundary, which represent the target condition for exiting poverty vulnerability.

Simulation procedure:

For each household with high vulnerability:

- i. Identify baseline state X_i^{high}
- ii. Apply intervention scenarios (eg. increment income)
- iii. Recompute MPI: $MPI_i^{simulated}$
- iv. Evaluate Transition: $\Delta MPI_i = MPI_i^{before} - MPI_i^{after}$

The core DT engine is then achieved by the optimization engine as we formulate the policy design as:

$$\min_{\Delta X_i} MPI_i^{new} = f(X_i + \Delta X_i) \quad (27)$$

subject to cost $(\Delta X_i) \leq B$

where B is the intervention budget constraint and $cost$ is the resource requirement for intervention. For the cluster transition condition, a household would exit high vulnerability if

$$MPI_i^{new} < \mu_{cluster_low}$$

i.e, practically if $MPI_i^{new} < \tau$ as earlier formulated.

3.2.4 Explainability Layer

The Shapley Addictive exPlanations (SHAP) model is employed for a white-box analysis of the impact of data features and the direction of those impacts on the determination of the vulnerability clusters. This is essential to further offer perspectives for the counterfactual simulation. SHAP would reveal the data components that moved a household into any of the vulnerability classes, based on Decision Tree modelling. For each feature x_i in the dataset, the Shapley value $\phi_{(x_i)}$ for a household j is expressed by Vimbi *et. al.* (2024) as:

$$\phi_{(x_i)} = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \quad (28)$$

where:

F is the set of all households; S is a subset of features excluding x_i ; $f(S)$ is the model output when only features in S are considered and $(S \cup \{x_i\})$ is the model output when feature x_i is added. The impact and direction of influence of a particular feature x_i is thereby quantified.

To clarify household shifts from their baseline vulnerability state to the post-counterfactual intervention state, the Decision Tree classifier recursively partitioned the feature space using

threshold-based rules that maximized class purity. At each node, the optimal split is selected by minimizing impurity measured through the Gini Index:

$$G(t) = 1 - \sum_{k=1}^K P_k^2 \quad (29)$$

where $G(t)$ denotes impurity of node t , K is the number of vulnerability classes (i.e Low, Moderate and High), and P_k is the proportion of households belonging to class k at node t . The best split s produces the maximum reduction in impurity such as:

$$\Delta G = G(t) - \left(\frac{N_L}{N} G(t_L) + \frac{N_R}{N} G(t_R) \right) \quad (30)$$

where N is the number of households in the parent node, N_L and N_R are the households in left and right child nodes, and the resulting child nodes are the t_L and t_R . Within the study, the predictor vector for a household i was defined as:

$$X_i = \{Income, Yrs_Edu, Int_Acc_Sco, Elect_Acc_Sco, Water_Acc_Sco, etc.\} \quad (31)$$

with the target variable $Y_i = Class_{t1}$, which represents the vulnerability state after counterfactual intervention. The decision tree then finally learns a mapping

$$f: X_i \rightarrow Class_{t1},$$

resulting into an interpretable decision rule of the form:

$$If \text{Income} < \tau_1 \wedge Food_{sec_sco} < \tau_2 \wedge MPI_{t0} > \tau_3 \Rightarrow Class_{t1} = High$$

where τ_1, τ_2, τ_3 represents the optimal thresholds learned from the survey data.

Feature importance is computed to rank, by quantification, the prominence of variables. This is determined by the cumulative impurity reduction expertise contributed by each across the splits, as:

$$FI_j = \sum_{m \in S_j} \frac{N_m}{N_{root}} \Delta G_m \quad (32)$$

where the importance of feature j is depicted as FI_j , the set of tree nodes split by the feature is set as S_j , the ΔG_m is the impurity reduction achieved at node m , and the total number of households by the N_{root} . This will avail detection of dominant drivers of vulnerability transitions (in addition to the instance-based relative approach of SHAP), enumerate the comparative contribution of prominent variables in the feature space, and importantly reveal threshold conditions governing movement between vulnerability states. These are in addition to the provision of transparent policy-relevant rules explaining household transitions under the Digital Twin counterfactual intervention framework.

4. Experimental results and discussion

All analytical components of the study were implemented using Python tool within a modular DT analytics framework, on the Google Colab environment. The initial counterfactual simulation was implemented as a rule-based engine that generated alternative household states through the

modification of deprivation-related attributes within the surveyed feature space, and recomputing the resulting MPI. For each household i , the baseline state vector X_i was replicated and subjected to predefined intervention rules (ΔX_i) through conditional logic. The updated state $X_i + \Delta X_i$ was then propagated through the MPI function to obtain MPI_i^{new} , enabling the evaluation of household transitions across vulnerability classes. This process created a counterfactual representation of how each household would perform under improved socioeconomic and infrastructure conditions without altering the observed baseline data.

4.1 Feature space overview and quality analysis

A total of 420 (58%) males and 300 (42%) females participated in the survey and turned out that half (360; 50%) are primary caregivers. This avails a balanced representation across dependent/independent scenario. Their sources of income court across informal employment (64%), formal employment (23%), farming (11%), and informal (2%) sectors of the economy. A total of 424 (59%) had skipped meals severally while only 150 (21%) had benefited from food aid. A total of 282 (39%) nurses/nursed chronic health condition with 404 (56%) having healthcare access. Respondents possess varying educational qualifications including Bachelors (63%), Secondary/High school certificate (23%), Primary/Basic school certificate (7%), Masters degree (4%) and Doctorate (3%). From the summary statistics presented in Table 3, mean age of respondents was 25 years (mostly youths, though with a deviation of 5.85), while the youngest and oldest respondents were 17 and 45 years old. However, 75% of respondents were less than or 26 years. An average household had 5 people with the maximum at 8, while the interquartile range revealed that 75% of households have up to 6 family members. It is also observed that households took an average of 2 meals per day.

The pairwise correlation coefficient analysis of the feature space is captured on the heatmap of **Fig.4**. Observation shows that water_access_score and food_access_score exhibit positive correlation (0.61), just as the internet_access_score and the food_security-score with correlation coefficient of 0.69. Other features with notable coefficients include the water_access_score vs internet_access-score (0.57). The observations indicated that some of the predictive features exhibit pairwise moderate correlation, which is not a severe case. Output of the rather robust VIF analysis is presented in Table 4 and as can be observed, the highest VIF of 2.484138 (internet_access_score), 2.366478 (food_access_score) and 1.959112 (water_access_score) indicates a mild multicollinearity of predictors and are typically considered acceptable in clustering-related feature analysis. The computed Minimum Redundancy Maximum Relevance (mRMR) returned prominent features out of the surveyed feature spaces:

['Internet_Acc_Sco', 'Acc_Farm', 'Meals_day', 'Water_Acc_Score', 'Elect_Acc_Score', 'Food_Security_Sco', 'Yrs_Education', 'Edu_Level', 'Skip_Meal', 'Employment_St', 'Fire_Charcoal_Cooking', 'Discrim_SocialExclusion', 'Income_Source', 'Climate_Rel_Iss', 'Domestic_Vio']

as the most efficient for the DT simulation. The trio of the internet access score, water access score and the food security score were highly relevant to proposed vulnerability prediction as well as exhibiting low redundancy within the feature space, validating their mild multicollinearity status.

Table 3. Summary statistics of the surveyed Isolu household numeric feature space

Features	Mean	Std.	Min	25%	50%	75%	Max
Age	25	5.85	17	20	24	26	45
House Size	5	1.43	2	4	5	6	8
Meals day	2	0.49	2	2	3	3	4
Yrs Education	13	3.35	0	12	12	15	20
Water Acc Score	6	1.83	2	6	6	7	10
Elect Acc Score	7	4.28	1	4	6	8	18
Health Acc Sco	1280.74	6886.7	0	5	7	1280.74	50000
Internet Acc Sco	6.32	2.65	1	4	7	9	10
Food Security Sco	6.18	2.37	1	4	6	8	10

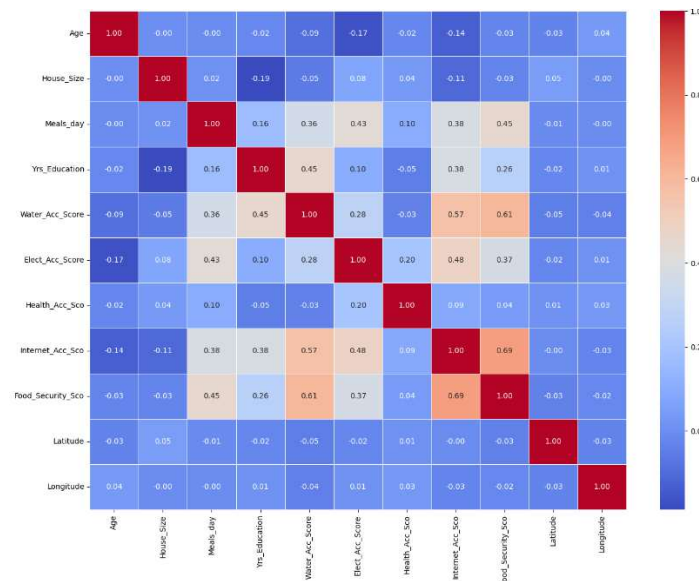


Fig.4. Pairwise correlation analysis of the surveyed feature space

Table 4. Distribution of variance inflation factor of the predictive attributes

	Features	VIF
1	Age	1.058857
2	House Size	1.065572
3	Meals day	1.420952
4	Yrs Education	1.370242
5	Water Acc Score	1.959112
6	Elect Acc Score	1.548554
7	Health Acc Sco	1.058201
8	Internet Acc Sco	2.484138
9	Food Security Sco	2.366478
10	Latitude	1.009790
11	Longitude	1.006801

4.2 Household vulnerability learning and profiling

K-Prototypes modeled the hidden multidimensional similarity structures within the feature space to reveal latent deprivation and socioeconomic similarity zones within the surveyed households, as operationalized by the MPI. The analytics returned three clusters with reliable performance metrics. The model returned 0.309, 1.0491, and 316.484 respectively on the metrics silhouette, Davies-Bouldin and Calinski-Harabasz. These indicated that there are no strict structures within the clusters (Silhouette farther from 1), even though there was a better between-cluster and within-cluster dispersion (higher Calinski-harabasz). The lower Davies-Bouldin is likewise positive, indicating lower average similarity between each cluster with respect to their closest neighbour. The model clustered 286 (40%) households with Low Vulnerability (LV), 247 (34%) with High Vulnerability (HV) and 187 (26%) with Moderate Vulnerability (MV), implying a LV majority. Table 5 indicates that male respondents dominated households with medium vulnerability while the low and high vulnerability classifications showed largely even distribution across both genders. Majority of households in the high vulnerability class does not get food aid, neither do they have access to health facilities, just as they are hardly educated, while households with LV exhibited lowest mean MPI (0.170330), strong infrastructure access across all domains, highest education level (mean=14.85 years) and highest internet access (mean=7.59). On the other hand, MV households exhibited intermediate mean MPI levels (0.183346), partial deprivation patterns which are structurally closer to the low MPI vulnerability, and a transitional socioeconomic condition including partial deprivation patterns and transitional socioeconomic condition. HV households exhibited highest MPI mean (0.583574), strong deprivation concentration, severe food insecurity (mean=4.11), low electricity access (3.34 mean value) and low internet access (mean=3.84). The cluster-wise mean feature profile of the deprivation classes is presented in the heatmap of **Fig.5**. Evidently, MPI alone does not fully explain vulnerability structure since latent multidimensional clustering indicates concealed inequality substructures within similar MPI bands, which speaks to the lower silhouette score earlier presented.

Table 5. Summary statistics of MPI distribution across the vulnerability stratification

Cluster	Gender (%)		Employment status (%)		Food Aid (%)		Access to Health (%)	
	Male	Female	Y	N	Y	N	Y	N
LV	52.80	47.20	60.49	39.51	2.80	97.20	88.81	11.19
MV	72.73	27.27	86.63	13.37	35.29	64.71	19.25	80.75
HV	53.85	46.15	16.60	83.40	30.77	69.23	45.75	54.25

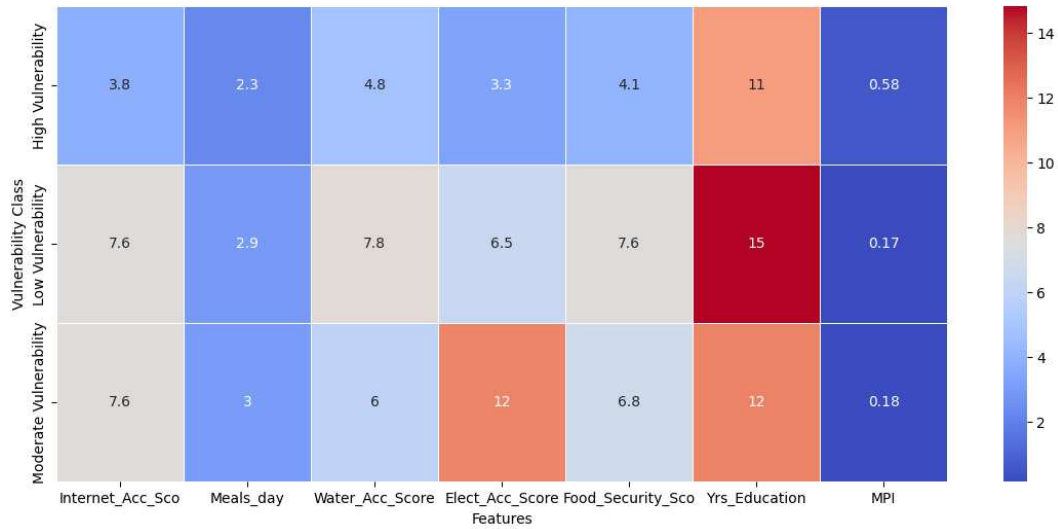


Fig.5. Cluster-wise heatmap for mean feature profiling of household vulnerability dynamics

The Moran and Getis-Ord G_i^* modelling of the spatial vulnerability dynamics of the feature space was quite revealing, as visualized in the hotspot detection visualization in **Fig.6**. Moran's returned I : 0.0079 and p-value: 0.258, which is close to zero and statistically insignificant, implying that MPI is spatially random in the Isolu study area. Therefore, there was no evidence of spatial clustering of poverty or even vulnerability. Hotspot result was also returned as “Neutral” as the Getis-Ord G_i^* test confirmed no statistically significant hotspots or coldspots. Therefore, vulnerability is not geographically concentrated in clusters at the current resolution. We infer that even though individuals differ in their MPI, poverty is socially structured but spatially dispersed, not spatially clustered. Therefore, there are no poverty zones in the investigated households since those with high vulnerability are not all located together in one specific area. This signals a micro-level inequality domination in Isolu, over geographic inequality. Consequently, interventions or safety-net initiatives should be household-targeted, not area-focused. Consequently, the digital geospatial vulnerability twin of investigated Isolu households is presented in **Fig.7**.

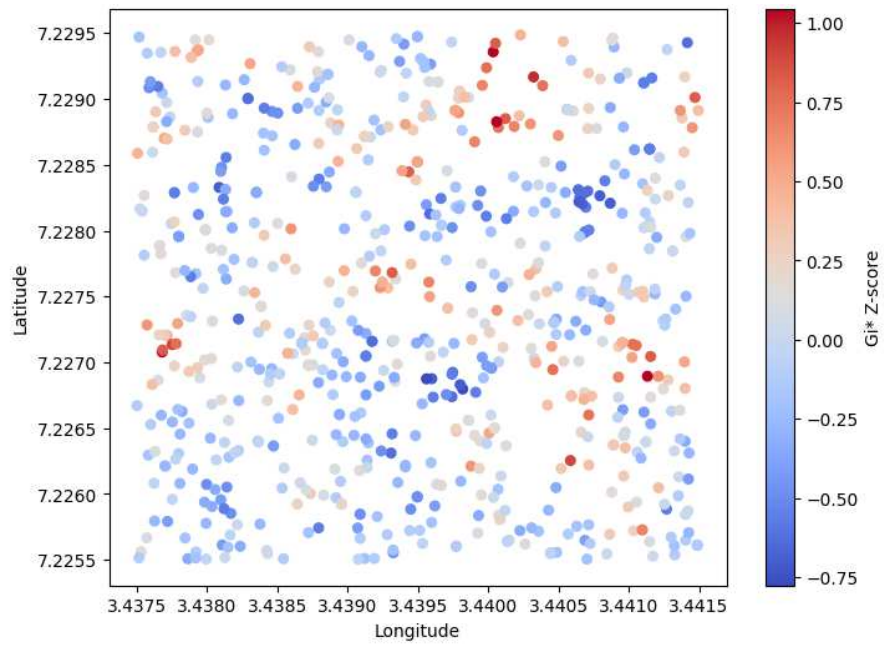


Fig.6. Spatial hotspot modelling of investigated Isolu household showing spatially dispersed vulnerability instances

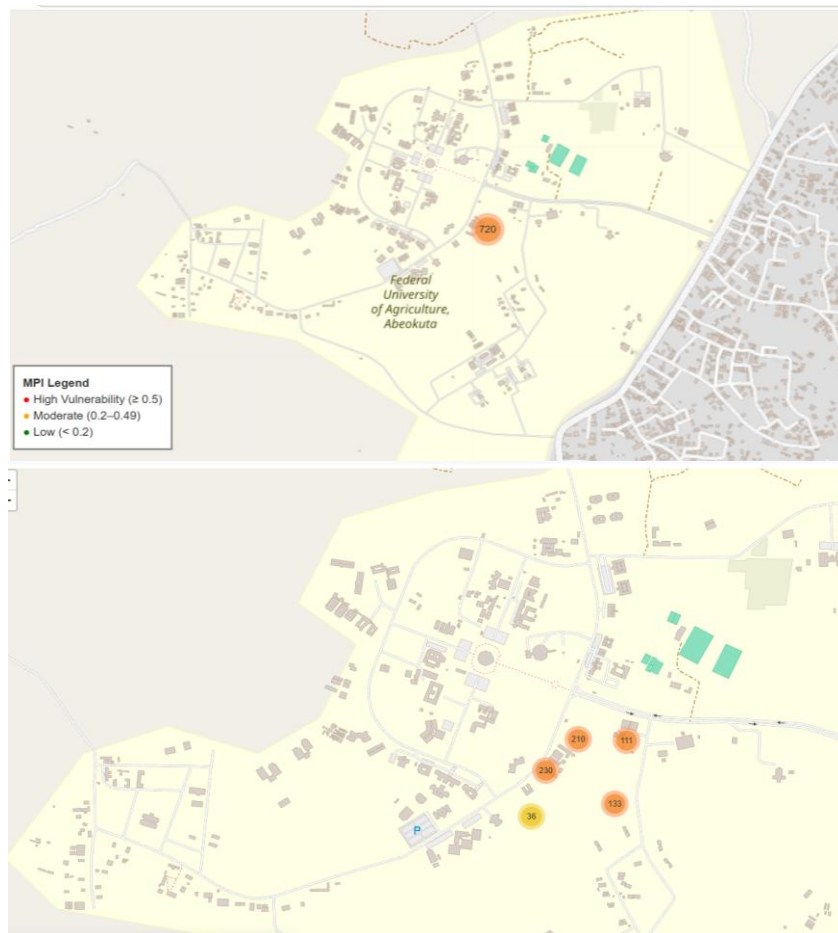




Fig.7. Computational geospatial vulnerability twin of investigated Isolu households

4.3 Counterfactual simulation for poverty governance

The vulnerability transition matrix (**Fig.8**), the MPI shift plot (**Fig.9**), summary of observed intervention dynamics (Table 6) and the transition probability matrix (**Fig.10**) summarize the behaviour of households to the counterfactual intervention simulation. The objective of this phase was to evaluate whether simulated changes in critical deprivation indicators could shift households from higher vulnerability states toward lower deprivation regimes, as determined by the recomputed poverty condition MPI_i^{new} . These interventions were applied individually and in combinations, and the simulation values were informed by:

- i. Observed survey scale ranges, since for example, all the access scores already existed on bounded ordinal scales
- ii. Realistic policy direction as in the case of ‘years of education’, since one intervention cycle is unlikely to raise education by 10 years instantly
- iii. Simulation stability by ensuring that intervention packages produce measurable but non-exaggerated MPI shifts.

Though aggregate MPI improved, an uneven benefit distribution was experienced, indicating policy inefficiency at instance level despite macro-level gains. The mobility dynamics showed a 75-household High → High transition, indicating strong persistence of deprivation; possibility of a structural poverty trap. A partial escape is observed with a 68-household High → Moderate transition, showing a limited mobility, while a 19-household Low → Moderate setback is revealing of a slight downward mobility, suggesting a vulnerability spillover risk. Only seven (7) households escaped from High to Low, which is a rare full escape. The Moderate → Low (124 households) mobility is the strongest improvement pathway in the DT system, though majority remained in their vulnerability states, (Low=327, Moderate=106, and High=71), despite the interventions. In all, High-vulnerability households exhibited strongest structural inertia, confirming the presence of a poverty trap scenario, but with a high stability in the Low → Low scenario (94.5%). Therefore, the policy package is highly effective at aggregate level (as can be observed from the MPI

distribution shift plot), but introduces a small adverse effect group, indicating uneven policy sensitivity.

By implication, intervention effects were strongly concentrated among High and Moderate vulnerability households, indicating that poverty reduction returns increase with the severity of baseline deprivation. Observations from the transition probability matrix indicated that among households initially classified as High vulnerability however, 52.4% remained trapped in the High state, while 42.7% transitioned to Moderate vulnerability and only 4.9% achieved a direct transition to the Low-vulnerability state. This suggests that although the intervention generated measurable poverty reduction effects, a large proportion of highly deprived households remained above the threshold τ , indicating the persistence of multidimensional deprivation. In contrast, households initially classified as Moderate vulnerability exhibited the most favorable response, with 53.7% transitioning into the Low-vulnerability state and only 0.4% deteriorating into the High-vulnerability regime. The Low-vulnerability group demonstrated strong state stability, with 94.5% remaining Low after intervention and only 5.5% transitioning to Moderate vulnerability. As indicated in the matrix, counterfactual intervention was more effective in facilitating downward poverty transitions among moderately vulnerable households than among severely deprived households, highlighting the greater difficulty of moving households with entrenched deprivation below the critical poverty threshold. This underscores the need for more intensive and targeted interventions for households residing in the High-vulnerability regime.

The intervention effectiveness assessment (**Fig.11**) reveals a clear gradient in poverty reduction across the vulnerability classes. Households within the High-vulnerability cluster achieved the largest average MPI reduction, exceeding 0.20, indicating that the interventions generated substantial improvements among the most deprived households. The Moderate cluster recorded a mean MPI reduction of approximately 0.10, suggesting modest but meaningful improvements in multidimensional welfare. In contrast, the Low-vulnerability cluster exhibited a negligible MPI reduction, resulting in an almost invisible bar in the visualization. This outcome is expected because households in the Low-vulnerability regime already possessed relatively low deprivation levels at baseline, leaving limited scope for further MPI improvement. The results therefore suggest that the intervention framework was appropriate for households experiencing higher deprivation, where the potential marginal gains from policy actions were lowest. The poverty driver profiling analysis (**Fig.12**) reveals substantial socioeconomic disparities across the vulnerability classes. Among all examined indicators, income emerged as the most visually dominant differentiating factor, with mean income levels of approximately ₦15,839 for the High-vulnerability cluster, ₦14,928 for the Low-vulnerability cluster, and ₦8,030 for the Moderate-vulnerability cluster. The large numerical scale of income relative to the remaining indicators compressed the visualization of education, meals per day, water access, electricity access, and internet access, rendering their bars comparatively indistinguishable on the common axis. The results indicated that vulnerability patterns cannot be attributed to income alone. Households in the High-vulnerability cluster simultaneously exhibited lower levels of educational attainment, food consumption, electricity access, water access, and internet connectivity, reflecting multidimensional deprivation. Conversely, the Low-vulnerability cluster demonstrated consistently superior access across these welfare dimensions, supporting its relatively favorable MPI profile. The Moderate-vulnerability cluster occupied an intermediate position, suggesting partial deprivation across multiple indicators. The findings therefore reinforce the multidimensional nature of poverty within the study area, demonstrating that vulnerability is

driven by the combined effects of economic, educational, and infrastructural deficits rather than by income deprivation alone. This observation provides empirical justification for integrated intervention strategies targeting multiple poverty dimensions simultaneously rather than focusing exclusively on income enhancement.

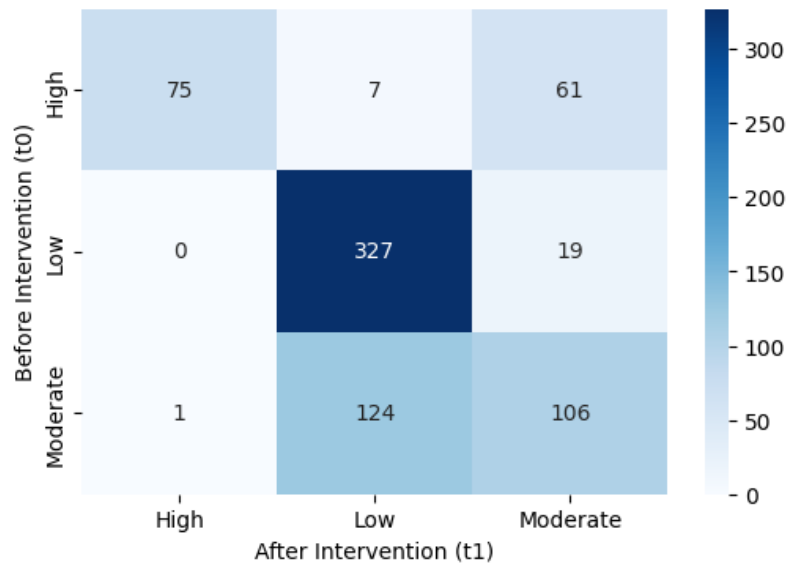


Fig.8. Household vulnerability transition matrix of the baseline population (n=720)

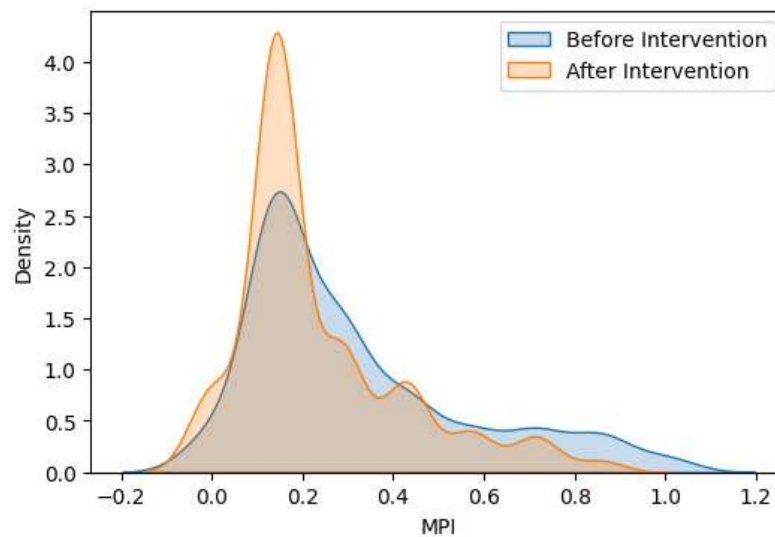


Fig.9. MPI distribution shift plot before and after intervention simulation

Table 6. Summary of intervention dynamics across vulnerability strata.

Vulnerability Class	Count	Mean MPI	Std. Dev.	Mean MPI at (t1)	Mean MPI Reduction
High	143	0.743	0.142	0.520	0.223
Moderate	231	0.334	0.068	0.226	0.108
Low	346	0.126	0.046	0.126	0.0004

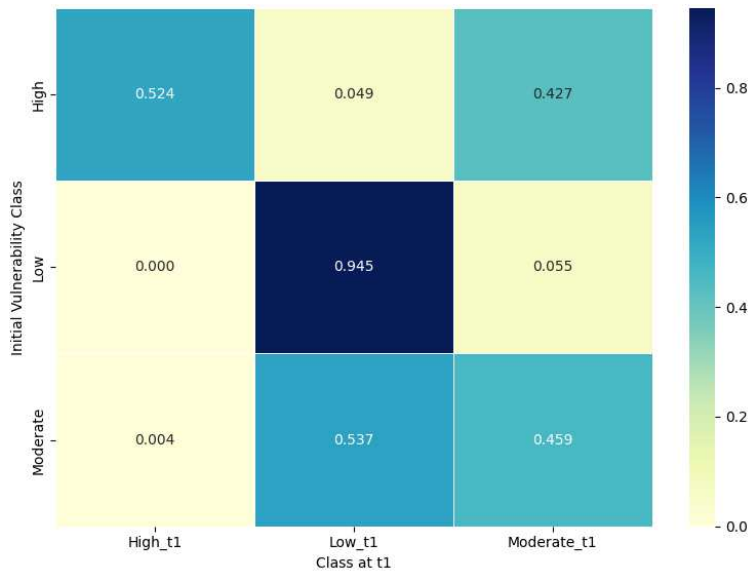


Fig.10. Transition probability matrix of the counterfactual simulation

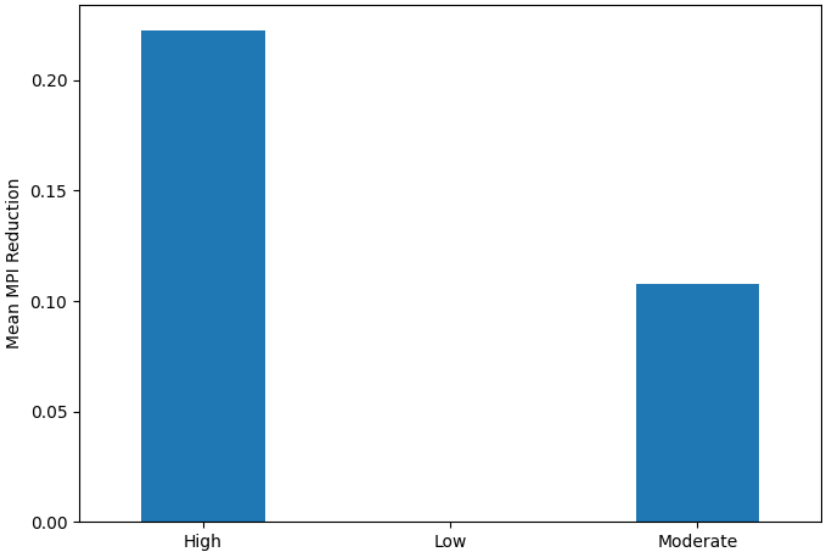


Fig.11. MPI reduction rate across vulnerability stratification

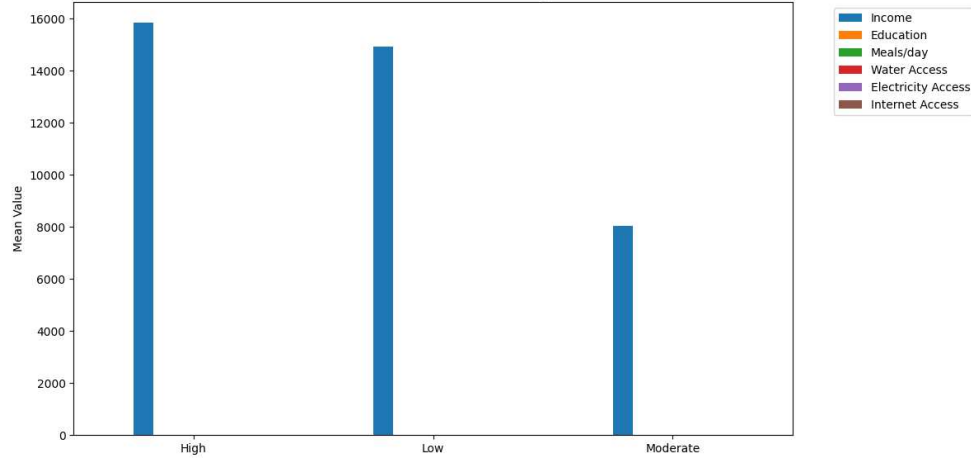


Fig.12. Exhibition of prominent vulnerability drivers

4.4 Counterfactual policy optimization layer

Following the counterfactual simulation, an optimization was undertaken to identify the most cost-effective policy combinations capable of maximizing poverty reduction under a constrained intervention budget. Unlike the initial simulation, which primarily assessed whether households could transition to a less vulnerable state after hypothetical interventions, the optimization framework simultaneously minimized intervention cost while maximizing the reduction in MPI, thereby ranking all feasible policy packages according to their economic efficiency. The policy actions on Table 7 constitute the intervention library (policy packages) from which the optimization algorithm selected and combined the most cost-effective interventions subject to the budget constraint. Each policy action directly speak to a target variable in the feature space, relating to a simulated effect and cost simulation weight (which were assigned to create a relative cost hierarchy for optimization). The policy optimization engine was operationalized by creating a policy library where intervention packages are captured. An MPI function is triggered during sequential policy and combinatorial package applications. The optimization engine thereafter activates the budget limit for each of the package across the vulnerability stratification. The limits were administered per package and the engine subsequently determines the top policy package, the best policy per cluster and the policy intelligence summary.

Table 8 presents the ranked policy packages evaluated for the High-vulnerability cluster. The optimization reveals a clear trade-off between intervention effectiveness and implementation cost. Although combining *food support*, *electricity expansion*, and *water upgrade* achieved the largest MPI reduction (0.3377), this required a substantially higher average cost (3200 units), resulting in a lower efficiency score (0.000106). In contrast, the standalone *food_support* intervention achieved an MPI reduction of 0.1928 at a considerably lower cost of only 900 units, producing the highest efficiency score (0.000214) among policy packages. Similar trends were observed for other intervention combinations, where additional policy components further reduced MPI but yielded diminishing marginal returns relative to their increased implementation costs. These findings demonstrate that the optimization framework successfully identifies intervention portfolios that balance effectiveness with economic feasibility rather than simply maximizing MPI reduction.

Table 9 summarizes the optimal policy selected for each vulnerability cluster. The results indicate that the standalone food_support intervention consistently emerged as the globally optimal policy across the vulnerability groups. Specifically, the High-vulnerability cluster experienced a reduction in mean MPI from 0.7433 to 0.5505 ($\Delta\text{MPI} = 0.1928$), while the Moderate cluster recorded a reduction from 0.3537 to 0.2511 ($\Delta\text{MPI} = 0.1027$). In contrast, the Low-vulnerability cluster exhibited only a marginal improvement ($\Delta\text{MPI} = 0.0083$), suggesting that households already experiencing relatively low deprivation derive comparatively limited benefits from additional welfare interventions. The consistently higher efficiency values obtained for the High and Moderate clusters further indicate that directing limited resources toward severely deprived households yields substantially greater poverty reduction per unit cost. The optimal policy on three sample instances from the entire population is presented on Table 10. For the High vulnerability cluster, the optimal intervention was a single food_boost policy, yielding the highest efficiency (0.000211) and a meaningful MPI reduction (≈ 0.121), indicating that basic consumption-level interventions provided the most immediate welfare gains in severely deprived household number 87. In contrast, the Moderate cluster responded best to a combined electricity and food intervention, achieving substantial MPI reduction (≈ 0.112), suggesting that household number 60 benefited most from the ensemble of infrastructure and basic needs, reflecting transitional poverty dynamics. However, household number 29 exhibited Low vulnerability with a negative policy efficiency and a negative MPI reduction under the tested composite policy, implying over-intervention or misaligned resource allocation, where additional inputs do not translate into welfare gains and may even introduce inefficiencies. Consequently, while the initial counterfactual analysis provided diagnostic insight into household transition behaviour, while the optimization framework translated those insights into actionable policy recommendations by identifying the intervention strategy that delivers the highest return on investment. The two phases collectively demonstrated that DT-driven policy simulation should not terminate at transition prediction alone but should be complemented with optimization to support evidence-based and economically sustainable intervention planning.

Table 7. Description of the intervention policy library

Policy Action	Target Feature	Simulated Effect (ΔX)	Cost Weight	Cost	Policy Interpretation
Income subsidy	Income	+ ₦20,000	5	2500	Household financial aids
Employment program	Employment_St	Unemployed $\rightarrow$ Employed	4	1800	Job creation Empowerments
Internet access	Internet_Acc_Sco	+2 units	3	1000	Enhanced digital connectivity
Electricity expansion	Elect_Acc_Score	+2 units	3	1200	Improved electricity infrastructure
Water upgrade	Water_Acc_Score	+2 units	3	1100	Portable water interventions
Food support	Meals_day / Food_Security_Sco	+1 meal level / improved food security at +2	4	900	Nutritional and welfare assistance
Education support	Yrs_Education	+1 year	4	900	Education support aids

Table 8. Ranked policy packages evaluated during counterfactual optimization for the high vulnerability cluster

Rank	Cluster	Policy Package	Avg. Cost	MPI Before	MPI After	MPI Reduction	Efficiency
1	High	(food_support)	900	0.743257	0.550450	0.192807	0.000214
2	High	(food_support, internet_access)	1900	0.743257	0.483516	0.259740	0.000137
3	High	(food_support, water_upgrade)	2000	0.743257	0.479520	0.263736	0.000132
4	High	(food_support, electricity_expansion)	2100	0.743257	0.476523	0.266733	0.000127
5	High	(food_support, internet_access, water_upgrade)	3000	0.743257	0.412587	0.330669	0.000110
6	High	(food_support, electricity_expansion, internet_access)	3100	0.743257	0.409590	0.333666	0.000108
7	High	(food_support, electricity_expansion, water_upgrade)	3200	0.743257	0.405594	0.337662	0.000106
8	High	(employment_program, food_support)	2700	0.743257	0.550450	0.192807	0.000071
9	High	(employment_program, food_support, internet_access)	3700	0.743257	0.483516	0.259740	0.000070
10	High	(employment_program, food_support, water_upgrade)	3800	0.743257	0.479520	0.263736	0.000069
11	High	(employment_program, food_support, electricity_expansion)	3900	0.743257	0.476523	0.266733	0.000068
12	High	(internet_access)	1000	0.743257	0.676324	0.066933	0.000067
13	High	(internet_access, water_upgrade)	2100	0.743257	0.605395	0.137862	0.000066
14	High	(water_upgrade)	1100	0.743257	0.672328	0.070929	0.000065
15	High	(electricity_expansion, internet_access, water_upgrade)	3300	0.743257	0.531469	0.211788	0.000064

Table 9. Distribution of optimal policy per cluster

Cluster	Best Policy Package	Avg Cost	MPI Before	MPI After	MPI Reduction	Efficiency
High	('food_support')	900	0.7433	0.5505	0.1928	0.000214

Moderate	('food_support')	900	0.3537	0.2511	0.1027	0.000114
Low	('food_support')	900	0.1346	0.1263	0.0083	0.000009

Table 10. Optimal policy optimization outcome across three household instances

Household	Cluster	Policy package	Cost	MPI_reduction	Efficiency
87	High	(food_boost,)	572	0.12088	0.00021
29	Low	(income_boost, electricity_boost, food_boost)	4152	-0.0085	-0.00006
60	Moderate	(electricity_boost, food_boost)	1617	0.11245	0.00007

4.4.1 Minimum-cost intervention

The optimization demonstrated food_support constituted as the minimum-cost intervention capable of producing the highest cost-effectiveness across all vulnerability clusters. Despite an average implementation cost of only 900 units, it consistently generated the highest efficiency scores, indicating that nutritional support offers the greatest poverty reduction per unit investment and represents the most economically scalable short-term intervention.

4.4.2 Best intervention combination

Although the standalone food_support policy achieved the highest efficiency, combining food_support, electricity_expansion, and water_upgrade produced the largest absolute MPI reduction for the High-vulnerability cluster ($\Delta\text{MPI} = 0.3377$). This finding highlights the trade-off between maximizing intervention impact and minimizing implementation cost, with multi-sector interventions providing greater poverty reduction at the expense of reduced cost efficiency.

4.4.3 Cluster-level intervention strategy

The optimization further revealed that intervention effectiveness varies across vulnerability clusters. High-vulnerability households derived the greatest benefit from immediate food-based interventions, Moderate-vulnerability households exhibited meaningful improvements under the same policy, whereas Low-vulnerability households experienced only marginal MPI reductions, indicating diminishing returns. These findings support a targeted, cluster-specific resource allocation strategy in which intensive interventions are prioritized for severely deprived households while minimizing unnecessary expenditure on relatively resilient populations.

4.5 Policy optimization on expanded household population

The counterfactual simulation was operationalized on a graduated household population of up to 10000 synthetic instances. The intention was to check if findings on the $n=720$ baseline are generalizable on a larger population. The Gaussian Copula Synthesizer was fitted on a joint probability distribution directly on the 720 observed points to generate in-distribution synthetic versions of $n=5000$ and $n=10000$ households. We however validated whether the synthetic populations preserved the statistical dependency structure of the baseline using Mutual

Information (MI)-based dependency preservation validation. A MI preservation scores of 0.8795 for the 720→5000 synthetic expansion and 0.8811 for the 720→10000 expansion were recorded, indicating structural consistency of high degree between the baseline and expanded populations. This suggests that during scaling, there was preservation of nonlinear dependencies and joint probabilistic relationships among feature variables. Therefore, no significant distributional distortion or collapse in dependency was introduced by increasing the synthetic population size, as observed from the minimal deviation between the two preservations. Latent intrinsic multidimensional vulnerability patterns in the investigated Isolu household dataset therefore remained consistent. A mutual information preservation analysis was further done, and the following results validated the feature scaling.

4.5.1 Preservation of socioeconomic dependency pattern

MI matrixes revealed the preservation of dominant dependency structure observed in the baseline population. In the baseline, the strongest relationships were prominent among core infrastructure and welfare variables, as observed from **Fig.13(a)**. For example, the MI interaction between *Elect_Acc_Score* and *Water_Acc_Score* was high at 0.6397, while *MPI* and *Elect_Acc_Score* reached 0.5009. Similarly, *MPI* sustained strong associations with *Internet_Acc_Sco* (0.4204), *Water_Acc_Score* (0.3856), and *Food_Security_Sco* (0.3192). Observations showed that the 5000 and 10000 household populations retained these interaction patterns notwithstanding reductions in absolute MI magnitude. The 5000-household matrix (**Fig.13(b)**) retained strongest dependency between *Elect_Acc_Score* and *Water_Acc_Score* (at 0.0482), while *MPI* conserved associations with *Yrs_Education* (0.0932) and *Meals_day* (0.0776). The 10000-household dataset (**Fig.13(c)**) similarly maintained strong MPI interactions with *Meals_day* (0.0952), *Yrs_Education* (0.0943), and *Elect_Acc_Score* (0.0775). Relative dependency hierarchy is therefore preserved by the synthetic generation processes across feature variables, even under large-scale population expansion.

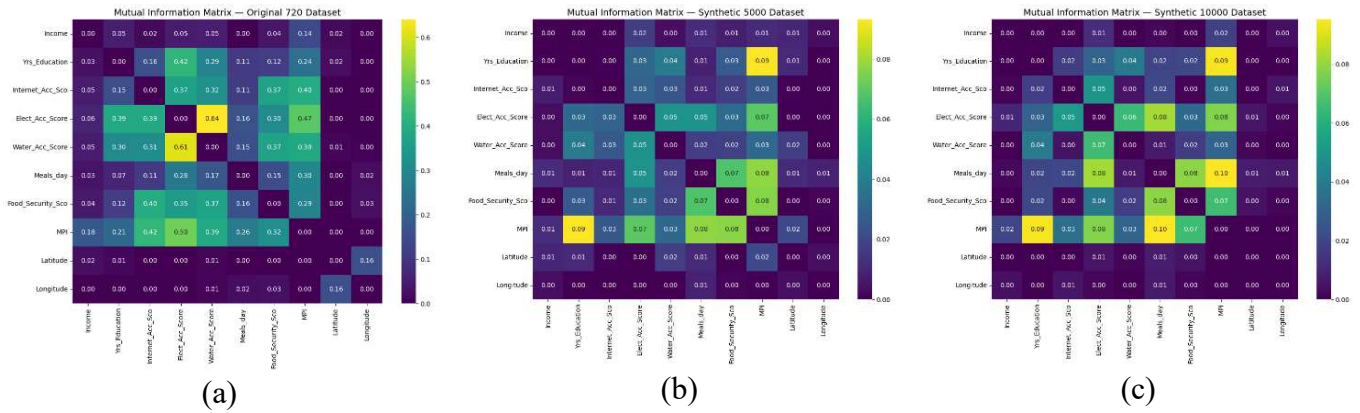


Fig.13. Matrixes showing mutual information and investigation of stability across populations

The MI difference matrixes further confirmed the statistical alignment between the synthetic populations and the baseline survey distribution. Observed pairwise differences remained relatively small, signifying slight structural distortion during synthesis. For instance, the difference

between the baseline and 5000-household populations for $Income \leftrightarrow MPI$ was 0.1320 (**Fig.14(a)**), while the corresponding difference for the 10000-household reduced slightly to 0.1249 (**Fig.14(b)**). Similarly, the $MPI \leftrightarrow Elect_Acc_Score$ difference declined from 0.4312 (5000) to 0.4235 (10000), indicating notable consistency across scaling levels. Similar characteristics was observed for critical welfare relationships, e.g. baseline $Food_Security_Sco \leftrightarrow MPI$ interaction was 0.3192, while the corresponding preservation deviations were 0.2422 (5000) and 0.2538 (10000). Notwithstanding occurrence of attenuation, direction of dependency and relative influence remained stable. This is expected in probabilistic synthetic generation because smoothing effects typically reduce sharp local dependencies while preserving global structural behaviour. Notably, the geospatial variables constantly exhibited very low MI interactions with majority of the deprivation indicators across all datasets. For instance, $Latitude \leftrightarrow MPI$ remained approximately 0.0208 (5000) and 0.0000 (10000), empirically reinforcing the Moran's I earlier findings that poverty vulnerability in Isolu is socially differentiated but not strongly spatially clustered.

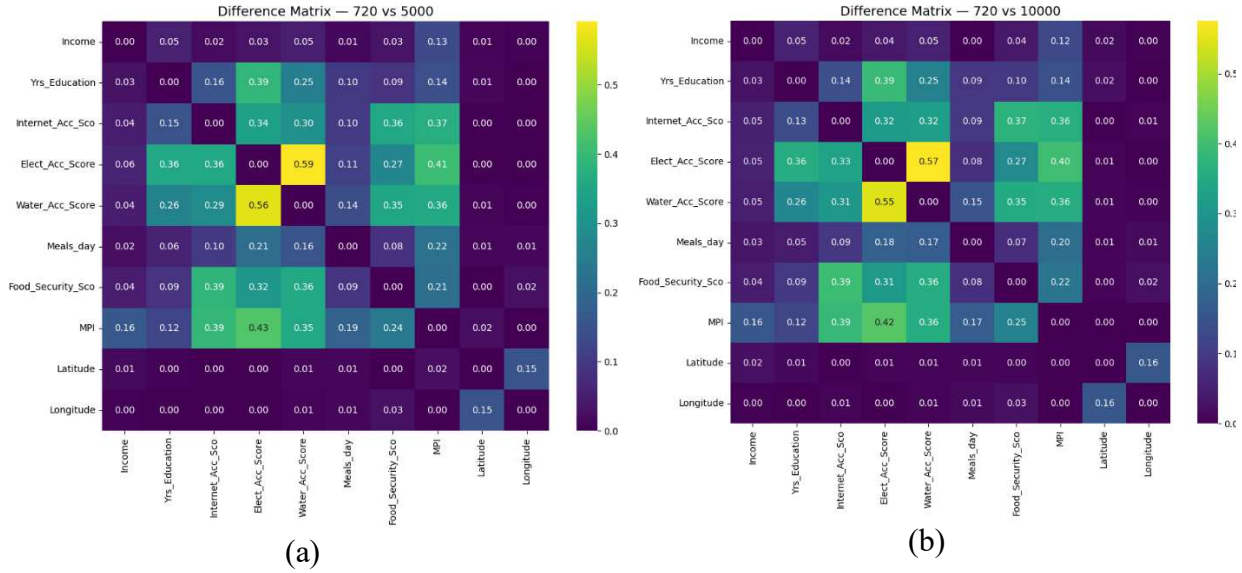


Fig.14. Pairwise evaluation of statistical alignment across study populations

4.5.5 Contextualization of the generalization for DT reliability

The established preservation of inter-variable dependency patterns has major implications for the credibility of the DT framework. Since policy optimization, scenario simulation, and intervention modelling rely on realistic socioeconomic interactions, maintaining the original MI structure ensures that the synthetic populations behave similarly to the observed community. The consistent preservation of relationships involving MPI , $Food_Security_Sco$, $Elect_Acc_Score$, and $Internet_Acc_Sco$ confirms that the expanded datasets remain behaviorally and statistically representative of the real Isolu population. Furthermore, the close similarity between the 5000-household and 10000-household preservation results demonstrates scalability robustness.

Therefore, increasing the synthetic population size did not fundamentally alter the dependency topology of the poverty system. This validates the suitability of the synthetic datasets for large-scale counterfactual simulation, intervention testing, and policy optimization within the proposed digital twin framework for investigated Isolu community.

4.5.6 Counterfactual simulation of initial interventions on expanded population

The initial counterfactual simulation were subsequently implemented on the graduated populations. Results are presented in Table 11 (n=5000) and Table 12 (n=10000) respectively. From observations, same optimal policy selection took place across the graduated populations, with food_support consistently retaining the most dominant welfare-maximizing intervention across the HV, MV, and LV clusters. This indicates policy stability under distributional scaling, signaling the informed intervention-response as not sample-specific. MPI consistently declined across population clusters, and efficiency values remained of comparable magnitude and positive. This confirmed directional consistency of welfare gains. However, key dissimilarities emerged in effect sizes and cluster separability. In the baseline data, clusters exhibited strong heterogeneity in MPI levels (e.g., High: 0.743 vs Low: 0.134), whereas in higher populations this gap compressed, with values converging upward across clusters (e.g., 0.50-0.61 range). This indicated that the Gaussian Copula synthesis induced distributional smoothing, reducing sharp vulnerability stratification observed in the real sample. Additionally, MPI reductions in higher data were more uniform across clusters (≈ 0.137 – 0.172), compared to the more dispersed baseline outcomes (0.008 – 0.193), suggesting weakening of extreme responsiveness. The implication is that the policy conclusions are directionally robust but magnitude-sensitive. The persistence of food_support as the optimal intervention supported strong generalizability of policy ranking. However, the reduced heterogeneity and compressed effect sizes in higher populations imply that policy impact may be more heterogeneous and extreme than synthetic projections suggest, particularly for HV groups. Therefore, while the DT framework demonstrated credible population-level stability, the findings are best interpreted as conservative lower-bound estimates of intervention effects rather than precise population replicates.

Following the initial counterfactual simulation across the populations, Table 13 presents a population-level distribution of household vulnerability class transitions. The transition analysis demonstrated a high degree of behavioural preservation across the original and synthetic populations, indicating that the scaling procedure successfully retained the underlying vulnerability dynamics. The most dominant transition across all populations was Low→Low, accounting for 45.42%, 36.46%, and 36.24% of households in the 720, 5,000, and 10,000 populations respectively, suggesting strong resilience among initially low-vulnerability households. Importantly, the beneficial Moderate→Low transition remained highly stable, increasing slightly from 17.22% to 18.27%, while the High→Low recovery pathway expanded from 0.97% in the original population to 8.68% in both synthetic populations, indicating that upward mobility patterns were preserved at scale. Although some deterioration pathways emerged in the larger populations, such as Low→High (4.16%) and Moderate→High (3.73%), the substantial decline in persistent High→High transitions from 10.42% to 2.74% suggested a net reduction in entrenched deprivation. The findings provided evidence that the synthetic populations preserved both positive and negative transition mechanisms while maintaining the overall structure of vulnerability evolution, thereby supporting the robustness and scalability of the Digital Twin framework.

Table 11. Optimal policy performance on the 5000-population

Cluster	Optimal Policy Package	Avg. Cost	MPI Before	MPI After	MPI Reduction	Efficiency
High	(food_support)	900	0.692174	0.504958	0.187215	0.000208
Low	(food_support)	900	0.502900	0.365843	0.137057	0.000152
Moderate	(food_support)	900	0.600740	0.428849	0.171891	0.000191

Table 12. Optimal policy performance on the 10000-population

Cluster	Optimal Policy Package	Avg. Cost	MPI Before	MPI After	MPI Reduction	Efficiency
High	(food_support)	900	0.6938	0.5088	0.1850	0.000206
Low	(food_support)	900	0.5017	0.3638	0.1378	0.000153
Moderate	(food_support)	900	0.6083	0.4371	0.1712	0.000190

Table 13. Population-level distribution of household vulnerability class transitions

Class Transition	n= 720	n= 5000	n= 10000	Rate (720) %	Rate (5000) %	Rate (10000) %	Preservation observation with implication
Low→Low	327	1823	3624	45.42	36.46	36.24	High retention observed with stable resilience
Low→Moderate	19	365	705	2.64	7.30	7.05	Moderate increase in vulnerability with emerging vulnerability
Low→High	0	201	416	0.00	4.02	4.16	New deterioration per increase in population size, signaling poverty escalation
Moderate→Low	124	888	1827	17.22	17.76	18.27	Highly preserved status with successful uplift
Moderate→Moderate	106	479	929	14.72	9.58	9.29	Moderate reduction with partial persistence
Moderate→High	1	178	373	0.14	3.56	3.73	Increased deterioration with growth in vulnerability
High→Low	7	434	868	0.97	8.68	8.68	Improved recovery with consequent poverty reduction
High→Moderate	61	519	984	8.47	10.38	9.84	Preserved transition with partial recovery
High→High	75	113	274	10.42	2.26	2.74	Strong reduction and reduced deprivation

4.6 Shapley Decision Tree validation and explainability

SHAP visualization on **Fig.15** (High (class 0); Low (class 1; Moderate (class 2)) identified electricity, internet, and water access features as strongest factors that influenced pattern recognition by the Decision Tree modelling. High vulnerability was enabled by lower values in

the access scores, reflecting clear threshold effects where minimal service access sharply increases deprivation risk. However, meals per day and food security scores show moderate influence, with reduced consumption patterns and weaker food security shifting predictions toward higher poverty states, but with weaker nonlinearity. Years of education exhibited a directional effect where lower schooling levels increased vulnerability likelihood, reinforcing human-capital constraints. Income showed comparatively weak marginal influence, suggesting poverty classification was more infrastructure- and access-driven than purely monetary. Skip_Meal_Yes contributed negligibly, implying limited discriminatory power in the survey. In terms of SHAP directionality, positive SHAP contributors (electricity access, internet access, income, food security) generally indicated movement away from High vulnerability, whereas negative contributions (water access, meals per day, and education) pushed predictions toward higher deprivation. The pattern confirms strong non-linear threshold behaviour, where falling below basic service access levels (especially electricity and internet) triggered sharp increases in predicted poverty status, while improvements above these thresholds yielded disproportionately large reductions in vulnerability risk.

A Decision Tree classifier extracted explicit decision rules governing transitions between baseline vulnerability classes (*Vulnerability_Class*) and post-intervention classes (*Class_{it}*). The analysis revealed the dominant socioeconomic thresholds and feature interactions associated with upward and downward mobility across household vulnerability states. A comparative regime-shift across population scales is evaluated and presented on Table 14. Across the three population scales, the decision structures revealed a consistent but progressively refined regime shift in multidimensional vulnerability, governed primarily by MPI_{t1} as the principal thresholding variable, with secondary modulation from food security, electrification access, and education. At lower sample size ($n = 5,000$), vulnerability transitions were highly sensitive to narrow MPI bands ($\approx 0.21\text{--}0.50$), where small increments in MPI_{t1} rapidly shifted households between Class 1 and Class 2, indicating a relatively sharp poverty boundary dominated by immediate deprivation indicators. As the sample expanded ($n = 7000$), the system exhibited greater structural layering, with additional conditioning on MPI_{change} and MPI_{t0} , suggesting that temporal poverty dynamics began to govern transitions, where households moved between states not only based on static deprivation but also on trajectory persistence or improvement. At $n=10000$, the regime became substantially more nonlinear and interaction-driven: thresholds such as $MPI_{t1} \approx 0.35\text{--}0.58$ acted as critical branching points, beyond which food security and electrification jointly determined whether households remained in transitional (Class 1) states or descended into persistent deprivation (Class 2). Notably, higher education and improved meals/day function as stabilizing absorptive factors, while low electrification and weak food security amplified downward mobility into vulnerability. This decision paths are as visualized on **Fig.16** ($n=720$), **Fig.17** ($n=5000$) and **Fig.18** ($n=10000$) respectively.

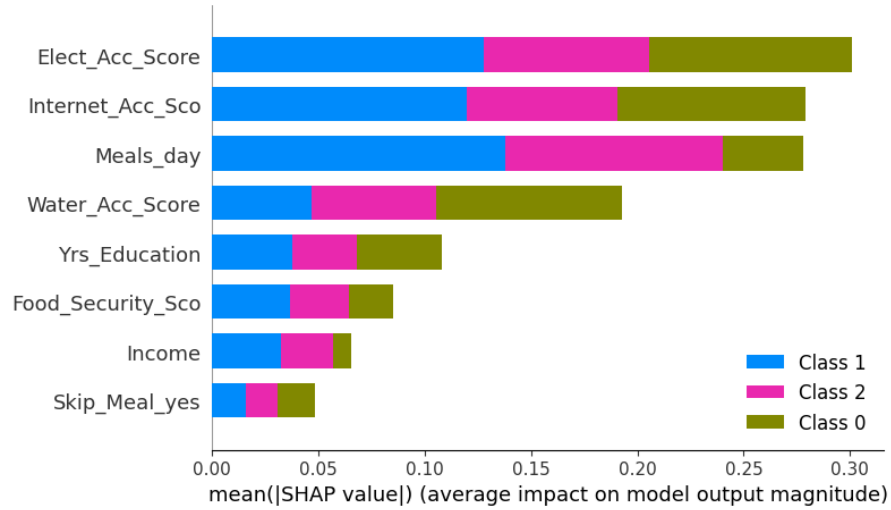


Fig.15. SHAP plot showing the direction of feature impact on household vulnerability stratification

Table 14. Comparative regime-shift analysis across population scales

Size	Hierarchy of Influence	Key Decision Threshold	Observed Vulnerability Transition Pattern	Primary Drivers of Class Shift	Regime Interpretation
n = 5000	$MPI_{t1} \rightarrow$ Food_Security_Sco $\rightarrow$ Elect_Acc_Score $\rightarrow$ Yrs_Education $\rightarrow$ MPI_change	MPI_{t1} : 0.21–0.50 (critical band)	Sharp binary-like transitions between Class 1 and Class 2 with minimal intermediate stability	food security constraint	Hard-threshold poverty regime: classification is dominated by static deprivation intensity with limited interaction effects; vulnerability shifts are abrupt and MPI-driven.
n $\approx$ 7000	$MPI_{t1} \rightarrow$ Food_Security_Sco $\rightarrow$ MPI_change $\rightarrow$ $MPI_{t0} \rightarrow$ Yrs_Education	MPI_{t1} : $\sim$ 0.05–0.53; MPI_{change} : -0.07 to +0.05 critical pivot zone	Emergence of transitional states; households move between classes based on poverty trajectory stability	Temporal dynamics; education as secondary stabilizer	Dynamic transition regime: vulnerability becomes path-dependent, where improvement or deterioration trajectories influence class mobility beyond static MPI levels.
n = 10000	$MPI_{t1} \rightarrow$ Food_Security_Sco	MPI_{t1} : 0.35–0.58	Strong multi-dimensional	Interaction effects:	Coupled-deprivation regime: vulnerability

	→ Elect_Acc_Score → Yrs_Education → Meals_day → MPI _{change}	bifurcation zone; Elect_Acc_S core: ~2.6– 9.2 dispersion threshold	clustering; smoother but more nonlinear transitions between classes	MPI × food security × electrificati on × education	emerges from interacting socioeconomic constraints, with nonlinear reinforcement across multiple welfare dimensions.
--	---	---	---	--	---

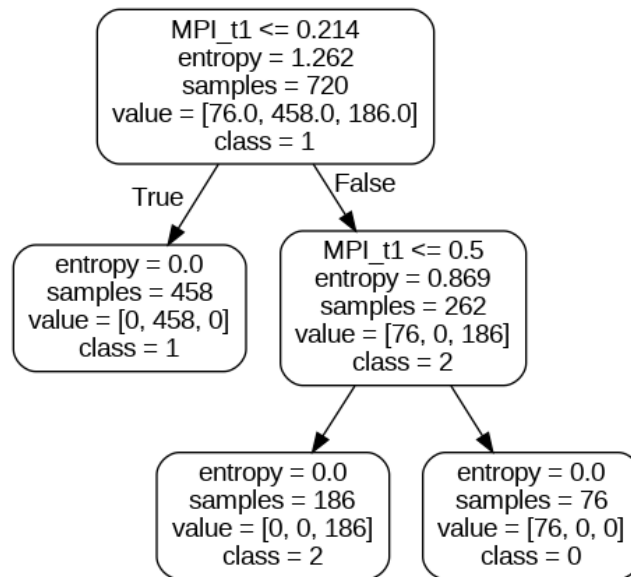


Fig.16. Decision path on the baseline population

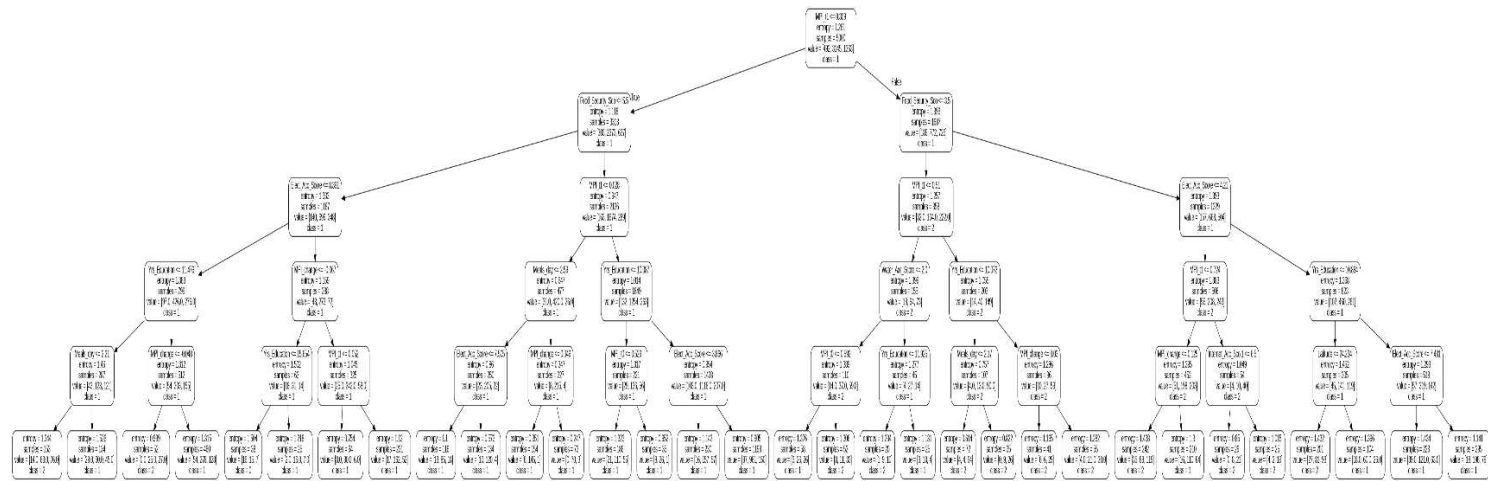


Fig.17. Decision path on the n=5000 population

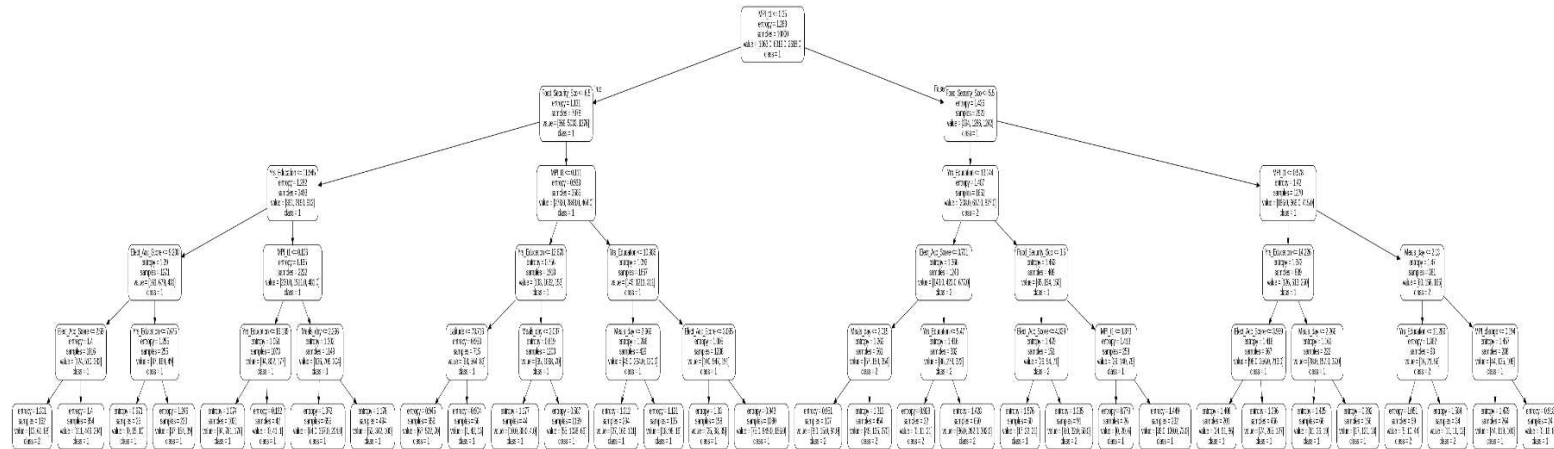


Fig.18. Decision path on the n=10000 population

4.7 Implications on SDG attainment and poverty governance

The ensemble of findings from the Digital Twin simulations and Decision Tree modelling provided several important implications for SDG implementation, particularly SDG 1 (No Poverty), SDG 2 (Zero Hunger), SDG 6 (Clean Water and Sanitation), SDG 7 (Affordable and Clean Energy), SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation and Infrastructure), and SDG 10 (Reduced Inequalities) within peri-urban communities in Nigeria and the broader West African region.

I. Multidimensional poverty is infrastructure-driven

The study demonstrated that multidimensional poverty within peri-urban communities is predominantly infrastructure and service-access driven rather than income-driven, as evidenced by the comparatively stronger influence of electricity access, internet connectivity, water access, food security, and meal frequency over household income in both the SHAP and Decision Tree analyses. This observation contrasts with many conventional poverty intervention programmes in Nigeria that remain heavily oriented toward cash-transfer mechanisms. Instead, the present findings suggest that accelerating progress towards SDGs 6, 7 and 9 may simultaneously generate significant improvements in SDG 1 by addressing multiple deprivation dimensions concurrently.

II. Integrated intervention initiatives are better social safety-net strategy

Although the food-support intervention consistently achieved the highest cost-efficiency, the largest absolute reductions in multidimensional poverty were obtained through integrated intervention packages combining food support, electricity expansion, internet access and water improvement. This demonstrates that poverty reduction exhibited strong systems behaviour, where simultaneous improvements across multiple deprivation dimensions generate substantially larger welfare gains than isolated sectoral interventions. Poverty governance should therefore increasingly adopt integrated multi-sector development planning instead of fragmented intervention programmes administered independently by different government agencies.

III. Effective poverty mitigation strategy is heterogenous across vulnerability stratification

The pronounced heterogeneity observed across vulnerability clusters indicates that uniform poverty alleviation policies are unlikely to achieve optimal resource utilization. High-vulnerability households exhibited the greatest intervention elasticity and therefore generated the highest MPI reductions per unit investment, whereas Low-vulnerability households displayed intervention saturation, with several intervention packages producing negligible or even negative efficiency values. These findings provide empirical support for cluster-adaptive and precision-targeted social protection frameworks, where intervention intensity is dynamically matched to household vulnerability rather than distributed uniformly across beneficiaries.

IV. Severity of deprivation should determine resource allocation

The optimization framework further demonstrated that resource allocation efficiency improves substantially when interventions are prioritized according to deprivation severity. Directing limited public resources toward severely deprived households consistently produced greater poverty reduction per intervention cost than distributing equivalent resources across relatively resilient households. This has direct implications for Nigeria's social investment programmes, suggesting

that future poverty governance should shift from beneficiary-count maximization toward maximizing welfare gains per unit expenditure.

V. Non-existence of poverty zone

Observations showed that vulnerability may not necessarily be geographically concentrated. Even though individuals differ in their MPI, poverty is socially structured but spatially dispersed, not clustered. Therefore, there are seldom poverty zones semi-urban communities since those with high vulnerability are not all located within a specific area. This signals a micro-level inequality domination in similar communities, over geographic inequality. Therefore, intervention should be instance-based rather than a one-fits-all policy intervention.

VI. Efficiency of Digital Twin for evidence-driven policy recommendations

The findings also reinforce the growing literature advocating multidimensional rather than monetary poverty mitigation strategies, while extending existing evidence by demonstrating how Digital Twin-enabled counterfactual optimization can operationalize such assessments into actionable policy recommendations. Recent multidimensional poverty studies have largely focused on identifying deprivation profiles or predicting poverty status; however, they rarely provide mechanisms for evaluating thousands of alternative intervention scenarios under explicit budget constraints. The present study advances this literature by integrating Digital Twin simulation, counterfactual reasoning, explainable artificial intelligence, and policy optimization within a single decision-support framework capable of identifying both the most cost-efficient interventions and the largest-impact policy combinations before real-world implementation.

Within the context of Nigeria and other West African economies characterized by persistent infrastructure deficits, rapid peri-urban expansion, and constrained public budgets, Digital Twin framework provides an evidence-driven governance mechanism capable of supporting proactive poverty management. Rather than implementing generalized welfare programmes, policymakers can simulate alternative intervention portfolios, quantify expected multidimensional poverty reductions, evaluate cost-effectiveness, and prioritize investments that maximize SDG attainment under available fiscal constraints. This represents a shift from reactive poverty alleviation toward adaptive data-driven poverty governance, which aligns closely with emerging digital government and smart sustainable development agendas.

5. Conclusion

This study presented a Digital Twin (DT)-driven framework for modelling, simulating, and optimizing multidimensional poverty dynamics within peri-urban communities using household-level socioeconomic, infrastructure, and geospatial indicators. The aim was to investigate the extent to which Digital Twin methodology could efficiently help with poverty governance intelligence. By integrating K-Prototypes un-supervised clustering, counterfactual simulation, supervised and explainable artificial intelligence (SHAP and Decision Tree), synthetic population scaling, and policy optimization, the DT framework moved beyond conventional poverty assessment to provide a predictive decision-support system capable of evaluating alternative intervention strategies before implementation. The synthetic population expansion from 720 to 5000 and 10000 households preserved the underlying multivariate dependency structure, demonstrating the robustness and scalability of the proposed framework for larger population

analyses. The counterfactual simulation revealed substantial heterogeneity in household transition behaviour, with High-vulnerability households exhibiting the greatest structural persistence but also the highest responsiveness to well-targeted interventions. Although the initial intervention scenarios produced measurable improvements, they also highlighted that uniform intervention strategies may unintentionally worsen outcomes for a considerable proportion of households, thereby necessitating a more intelligent optimization process. The subsequent optimization framework demonstrated that intervention effectiveness is strongly cluster-dependent, with standalone food support producing the highest cost-efficiency, while integrated intervention packages combining food support, electricity expansion, water improvement, and internet access generated the greatest absolute reductions in multidimensional poverty. Furthermore, the explainable learning models established that deprivation within the study area is driven primarily by deficiencies in infrastructure and essential service access rather than income alone, emphasizing the multidimensional nature of poverty. Findings demonstrated that Digital Twin technology provides a practical mechanism for transforming multidimensional poverty measurement into actionable policy intelligence by enabling virtual experimentation, counterfactual reasoning, and resource optimization under realistic budget constraints. The proposed framework therefore contributes a scalable and explainable decision-support approach for evidence-based poverty governance and offers a viable pathway for accelerating progress toward the Sustainable Development Goals in resource-constrained environments.

6. Statements and Declarations

This section presents the various declarations we have concerning this manuscript.

6.1 Funding

Authors declare that there is no funding from any individual or organization (public or private) for this research work.

6.2 Ethical approval

This particular study does not require nor involve human or animal subjects. Therefore, ethical approval was not required

6.3 Informed consent

Respondents expressed consent towards the use of their responses during survey, for research purpose only.

6.4 Data Availability Statement

The data employed for this study are available upon request

6.5 Conflict of Interest

Authors declare that there are no competing interests and also that there are no financial interests that are directly or indirectly related to this work. Authors did not receive any sponsorship whatsoever for the implementation of the experiments.

6.6 Author Contributions

T. O. reviewed existing literature, conceptualized the research design and method, implemented the conceptual framework by python and wrote the manuscript.

O. T. supervised the formulation of the framework, research design and method, and edited the manuscript.

T. D co-reviewed literature, co-implemented the framework, and co-prepared the manuscript.

O. V co-reviewed literature, co-implemented the framework, and co-prepared the manuscript.

References

- Egba, V. N., Ukeje, I. O., Ayanwale, M. A., Robinson, D. F., Ejem, E. N., Nwaigwe, H. C., & Osuebi, K. T. (2026). Digital governance and citizen trust in Nigeria's public sector: urban and semi-urban residents' perceptions of socio-technical dynamics of inclusivity and institutional transformation. *SN Social Sciences*, 6(3), 101.
- Carrilho, J., Balas, M., Dgedge, G., & Trindade, J. (2025). Fast growing informal peri-urbanization in Africa: The role of local practices in assessing sustainability and planning. *Journal Of Urban Management*.
- Atigaku, A. K., Tchakala, I., Agbovi, K. K., Kouyole, C. N., & Tsipoaka, K. (2026). Community governance and access to drinking water in semi-urban Togo: Dynamics and impacts in Kovié and Noépé. *Asian Journal of Water, Environment and Pollution*, 23(2).
- Belete, D. H. (2024). Sustaining escapes from multidimensional poverty: employing the zero poverty tripods. Panel data evidence from rural Ethiopia. *GeoJournal*, 89(5), 227.
- Angai, C. K. (2026). Understanding the intersection of behavioural science and power: a case study of the application of behavioural change approaches to anti-corruption projects in Nigeria (Doctoral dissertation, University of Sussex).
- Vidya, R., Deepan, P., & Arul, N. (2025). An Adaptive Framework for Empowering Marginalized Communities Through Human-Centric Artificial Intelligence (HCAI) for Decision-Making in Poverty Reduction Using Regression Techniques. In *Sustainable Economy Models in the Age of Industry 5.0* (pp. 199-215). Singapore: Springer Nature Singapore.
- Anukam, C. C., Udah, T., Udah, H., & Francis, A. (2026). Experiences of Vulnerability and Resilience Among Migrant Youth: A Scoping Review. Available at SSRN 6213390.
- Sharma, S. N. (2026). Precarious Urbanism: Housing Vulnerability and Spatial Inequality in Informal Settlements of the Global South. In *Housing Vulnerability and Disaster Risk in the Global South* (pp. 135-166). IGI Global Scientific Publishing.

Majumder, S. (2026). Not Just Climate—Caste, Class, and Capacity Matter: Exploring the Dynamics of Livelihood Vulnerability in an Indigenous South Asian Community. *The Professional Geographer*, 1-27.

Ezzati, F., Xiao, Q., Dong, Z.S., Jiao, J., Vargas, A., Yeh, V., Ptak, T. and Pan, K., 2025. Power outage-risk integrated social vulnerability analysis highlights disparities in small residential communities. *Communications Earth & Environment*, 6(1), p.294.

Tao, F., Xiao, B., Qi, Q., Cheng, J., & Ji, P. (2022). Digital twin modeling. *Journal of Manufacturing Systems*, 64, 372-389.

Hassani, H., & MacFeely, S. (2025). Digital twin and official statistics: A paradigm shift?. *Statistical Journal of the IAOS*, 41(2), 347-359.

Deng, T., Zhang, K., & Shen, Z. J. M. (2021). A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of management science and engineering*, 6(2), 125-134.

Onaji, I., Tiwari, D., Soulatiantork, P., Song, B., & Tiwari, A. (2022). Digital twin in manufacturing: conceptual framework and case studies. *International journal of computer integrated manufacturing*, 35(8), 831-858.

Xu, H., Wu, J., Pan, Q., Guan, X., & Guizani, M. (2023). A survey on digital twin for industrial internet of things: Applications, technologies and tools. *IEEE Communications Surveys & Tutorials*, 25(4), 2569-2598.

Cordes, D., Sefah, P., & Marinova, D. (2024). Developing digital twins of urban low-income communities in Sub-Saharan Africa: a case study in Ghana, West Africa. *Discover Analytics*, 2(1), 18.

Onsay, E. A., & Rabajante, J. F. (2025). From data to decision: Alleviating poverty and promoting development through measuring the unmeasurable economic numbers. *Societal Impacts*, 6, 100138.

Arya, P. K., Sur, K., Dhote, S., Siral, H., Kundu, T., Mehta, B. S., & Srivastava, R. (2025). Integrating Multi-Source Satellite Imagery and Socio-Economic Household Data for Wealth-Based Poverty Assessment of India: A GIS and Machine Learning Based Approach: Arya et al. *Social Indicators Research*, 179(2), 653-676.

Sheik, I., & Kabanda, G. (2026). The Role of Artificial Intelligence in Poverty Governance: A Systematic Literature Review of Innovations and Implementation Challenges. *Administrative Sciences*, 16(6), 269.

Parotto, E., & Pablos-Méndez, A. (2023). From MDGs to SDGs. In *Global health essentials* (pp. 463-468). Cham: Springer International Publishing.

Rodermund, S. C., Janzso, A., Bae, Y. E., Kravets, A., Schewerda, A., Berndt, J. O., & Timm, I. J. (2024). Driving Towards a Sustainable Future: A Multi-Layered Agent-Based Digital Twin Approach for Rural Areas. In *ICAART* (1) (pp. 386-395).

Alibrandi, U. (2022). Risk-informed digital twin of buildings and infrastructures for sustainable and resilient urban communities. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 8(3), 04022032.

Cordes, D., Sefah, P., & Marinova, D. (2024). Developing digital twins of urban low-income communities in Sub-Saharan Africa: a case study in Ghana, West Africa. *Discover Analytics*, 2(1), 18.

David, A., Guilbert, N., Hamaguchi, N., Higashi, Y., Hino, H., Leibbrandt, M., & Shifa, M. (2018). Spatial poverty and inequality in South Africa: A municipality level analysis.

- Khan, S. U., & Sloboda, B. W. (2023). Spatial analysis of multidimensional poverty in Pakistan: Do income and poverty score of neighboring regions matter?. *GeoJournal*, 88(3), 2823-2849.
- Mathenge, M., Sonneveld, B. G., & Broerse, J. E. (2023). Mapping the spatial dimension of food insecurity using GIS-based indicators: A case of Western Kenya. *Food Security*, 15(1), 243-260.
- Suhardono, S., Fitria, L., Prayogo, W., Lee, C. H., & Suryawan, I. W. K. (2025). Enhancing community engagement with digital twins: Technological adoption in marine debris management. *Journal of Urban Management*.
- Liu, M., Ge, Y., Hu, S., & Hao, H. (2023). The spatial effects of regional poverty: Spatial dependence, spatial heterogeneity and scale effects. *ISPRS International Journal of Geo-Information*, 12(12), 501.
- Ruiu, M. L., & Ragnedda, M. (2024). From poverty to digital poverty. *Digital-Environmental Poverty: Digital and environmental inequalities in the post-covid era*, 15-39.
- Çelik, S., Öztürk, Ö. F., Akkucuk, U., & Şaşmaz, M. Ü. (2025). Global sustainability performance and regional disparities: A machine learning approach based on the 2025 SDG Index. *Sustainability*, 17(16), 7411.
- Ahimah-Agyakwah, S., Nketiah-Amponsah, E., & Agyire-Tettey, F. (2022). Urbanization and poverty in Sub-Saharan Africa: evidence from dynamic panel data analysis of selected urbanizing countries. *Cogent Economics & Finance*, 10(1), 2109282.
- Thorbecke, E. (2023). The interrelationships among growth, inequality and poverty in Africa. *Journal of African Economies*, 32(Supplement_2), ii81-ii86.
- Leichenko, R., & Silva, J. A. (2024). Climate change and poverty. In *Routledge Handbook of Climate Change and Society* (pp. 374-387). Routledge.
- Appiah-Otoo, I., Chen, X., Song, N., & Dumor, K. (2022). Financial development, institutional improvement, poverty reduction: The multiple challenges in West Africa. *Journal of Policy Modeling*, 44(6), 1296-1312.
- Nowak-Garmer, R. (2023). Social protection in Africa. In *Handbook on social protection and social development in the global south* (pp. 112-129). Edward Elgar Publishing.
- Motadi, M. S. (2025). A strategy for poverty eradication towards sustainable development. *OIDA International Journal of Sustainable Development*, 18(09), 141-154.
- Rankakane, N. P., Mathaulula, M. A., & Francis, J. (2025). Enabling Harmonious Relationships between Grassroots Communities and Rural-based Universities in South Africa. *African Journal of Development Studies (formerly AFFRIKA Journal of Politics, Economics and Society)*, 15(4), 173-194.
- Phdungsilp, A. (2022). Waste management and its contribution to the sustainable development goals at Dhurakij Pundit University, Thailand. *Journal of Sustainability Perspectives*, 2(1), 65-72.
- Karunaratne, P. V. M., Jayalath, U. I., & Bhagya, K. P. N. (2024). Heritage of living entity and living skills: Advancing sustainability goals 1 and 4 in handloom textiles. In *Textiles of Sri Lanka: Sustainable innovations and practices* (pp. 49-96). Singapore: Springer Nature Singapore.
- Shehadeh, A., Alshboul, O., & Arar, M. (2024). Enhancing urban sustainability and resilience: employing digital twin technologies for integrated WEF nexus management to achieve SDGs. *Sustainability*, 16(17), 7398.

- Saha, S., Ruganzu, L., & Kar, S. (2025). Prediction of multidimensional poverty status with machine learning classification at household level: Empirical evidence from Tanzania. *IEEE Access*, 13, 23461-23471.
- Arranhado, E., Barbosa, L., & Bastos, J. A. (2025). Multidimensional poverty in Benin. *Socio-Economic Planning Sciences*, 99, 102181.
- Vimbi, V., Shaffi, N., and Mahmud, M. (2024). Interpreting artificial intelligence models: A systematic review on the application of LIME and SHAP in Alzheimer's disease detection. *Brain Informatics*, 11(1), 10.
- Singh, M., Fuenmayor, E., Hinchy, E. P., Qiao, Y., Murray, N., & Devine, D. (2021). Digital twin: Origin to future. *Applied System Innovation*, 4(2), 36.
- Grieves, M. Origins of the Digital Twin Concept. 2016. Available online: https://www.researchgate.net/publication/307509727_Origins_of_the_Digital_Twin_Concept. (accessed on 25 September 2020).
- Gelernter, D. *Mirror Worlds: Or: The Day Software Puts the Universe in a Shoebox. How it Will Happen and What it Will Mean*; Oxford University Press: Oxford, UK, 1993.
- Rosen, R.; Von Wichert, G.; Lo, G.; Bettenhausen, K.D. About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine* **2015**, 48, 567–572.
- Negri, E.; Fumagalli, L.; Macchi, M. A review of the roles of digital twin in cps-based production systems. *Procedia Manuf.* 2017, 11, 939–948.
- Wired Brand Lab; Digital Twin: Bridging the Physical-Digital Divide. 2017. Available online: <https://www.ibm.com/blogs/internet-of-things/iot-digital-twin-enablers/> (accessed on 11 October 2020).
- Banerjee, A.; Dalal, R.; Mittal, S.; Joshi, K.P. Generating digital twin models using knowledge graphs for industrial production lines. *UMBC Inf. Syst. Dep.* 2017.
- Volodin, V.; Tolokonskii, A. Concept of instrumentation of digital twins of nuclear power plants units as observers for digital NPP I&C system. *J. Phys. Conf. Ser.* 2019, 1391, 012083
- Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* 2018, 51, 1016–1022.
- Shafto, M.; Conroy, M.; Doyle, R.; Glaessgen, E.; Kemp, C.; LeMoigne, J.; Wang, L. Draft modeling, simulation, information technology & processing roadmap. *Technol. Area* 2010, 11.
- Ayaz, M.; Ammad-Uddin, M.; Sharif, Z.; Mansour, A.; Aggoune, E.H.M. Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE Access* 2019, 7, 129551–129583.
- Navarro, E.; Costa, N.; Pereira, A. A systematic review of IoT solutions for smart farming. *Sensors* 2020, 20, 4231.
- Ebrahimi, S.H.; Ossewaarde, M.; Need, A. Smart fishery: A systematic review and research agenda for sustainable fisheries in the age of AI. *Sustainability* 2021, 13, 6037.
- Taiwo, O. T., Sam-Wobo, S. O., Idowu, O. A., Talabi, A. O., & Taiwo, A. M. (2017). Comparative assessment of intestinal helminths prevalence in Water, Sanitation and Hygiene (WASH) intervention and non-intervention communities in Abeokuta, Nigeria. *Asian Pacific Journal of Tropical Biomedicine*, 7(6), 524-532.

Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2020). Big data analytics: Computational intelligence techniques and application areas. *Technological Forecasting and Social Change*, 153, 119253.

Dantas, P. V., Sabino da Silva Jr, W., Cordeiro, L. C., & Carvalho, C. B. (2024). A comprehensive review of model compression techniques in machine learning: PV Dantas et al. *Applied Intelligence*, 54(22), 11804-11844.