

Socio-Behavioral and Spatial Determinants of HIV/AIDS Incidence in Ghana: An Ecological Cross-Sectional Study with Explainable Machine Learning

Valentine Golden Ghanem^{1,2,3} 

¹Department of Public Health and Wellbeing, School of Health and Social Work, University of Suffolk, Ipswich, UK

²School of Architecture, Computing and Engineering, University of East London, London, UK

³Cocoa Clinic, Ghana Cocoa Board Medical Department, Accra, Ghana

Email: valentineghanem@gmail.com

How to cite this paper: Ghanem, V.G. (2025) Socio-Behavioral and Spatial Determinants of HIV/AIDS Incidence in Ghana: An Ecological Cross-Sectional Study with Explainable Machine Learning. *Advances in Infectious Diseases*, 15, 669-694. <https://doi.org/10.4236/aid.2025.154050>

Received: August 29, 2025

Accepted: October 8, 2025

Published: October 11, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Although there has been a drop in cases nationally, Ghana still suffers from severe subnational gaps in HIV/AIDS incidence because of educational inequities, stigma, HIV awareness, and access to antiretroviral therapy (ART). The inequities that lie beneath the surface are frequently concealed by national averages, which reduce the effectiveness of public health interventions. High-incidence and regionally vulnerable areas were identified through spatial clustering and choropleth mapping using a unified, regionally defined dataset from 2000 to 2022 that incorporated socio-behavioral and health infrastructure factors. The Random Forest and XGBoost models, explained through SHAP values, PDPs, and counterfactuals, showed the highest HIV incidence in the Greater Accra, Ashanti, and Central regions, strongly associated with urbanization, stigma, and limited ART access. The analyses further indicated that incidence declined with greater educational access and ART coverage, supporting the need for policies that curb stigma, address regional inequities, and improve access.

Keywords

HIV/AIDS, Ghana, Spatial Epidemiology, Sociobehavioral Determinants, Explainable Machine Learning, Public Health Disparities, ART Coverage, Stigma, Education, SHAP

1. Background

The incidence of HIV/AIDS in Ghana exhibits heterogeneous impacts across re-

gions and populations rather than a uniform distribution. Although the national HIV prevalence is less than 2% [1], regional disparities in the number of cases and access to healthcare services remain. While the national Human Immunodeficiency Virus (HIV) prevalence dropped from 3.1% in 2004 to 2.4% in 2016 [1], some districts, such as Lower Manya Krobo Municipality (LMKM), still have high incidence rates, as shown by the 5.64% reported in 2018 [2]. Despite being highly urbanized, the Greater Accra and Ashanti regions still have an unusually high HIV prevalence [3], implying that long-term public health interventions did not yield uniform regional epidemic outcomes.

Many researchers in sub-Saharan Africa (SSA) have found that HIV vulnerability is influenced by complex socio-behavioral factors such as education, awareness, stigma, and gender roles. Higher educational attainment among Ghanaian women correlates with increased HIV testing and treatment uptake, whereas young rural men remain persistently underserved in diagnosis rates, underscoring demographic disparities in healthcare access [4]. The stigma and legal barriers faced by men who have sex with men (MSM) in Ghana prevent them from accessing services, thus exacerbating regional disparities in HIV outcomes [5].

These discrepancies emphasize the need for spatially targeted interventions to address granular inequities. A recent analysis of Demographic and Health Survey (DHS) data using spatial interpolation found that subregional testing in the western region was as low as 5% in some places and over 30% in others [6]. Regional variations in disease prevalence are often obscured by national averages, which may result in inappropriate policies and misuse of resources. Spatial epidemiology helps to uncover and target areas where HIV risk is hidden.

Ghana's HIV surveillance system still relies on administrative reports that lack detailed geographic information and exclude behavioral factors. Traditional regression models used in HIV policy forecasting often assume spatial independence and are poorly suited for capturing the multidimensional, nonlinear interactions that drive regional epidemics. Therefore, many high-risk populations remain underserved, whereas programmatic responses continue to be overly centralized.

New developments in spatial analysis and interpretable Machine Learning (ML) can help solve these issues. SHapley Additive exPlanations (SHAP) and Partial Dependence Plots (PDPs) can be used to study the role of factors such as education, ART coverage, and stigma in HIV incidence and transmission. When paired with clustering and choropleth techniques, the ML results can make interventions more equitable by tailoring them to different areas.

Few studies in Ghana have attempted to combine these approaches to examine how both geography and behavior influence HIV incidence. Moreover, few studies have examined how policy changes in HIV awareness or education may disproportionately affect regions. This study aimed to address this issue by directing its analysis to the spatial and behavioral factors that impact the incidence of HIV/AIDS. Using ecological modeling, this study merges spatial analysis with ex-

plainable ML to guide precise public health strategies for combating HIV/AIDS incidence in Ghana.

2. Methods

2.1. Study Design and Rationale

This study used a retrospective ecological cross-sectional approach to analyze how socioeconomic and spatial factors affect the incidence of HIV/AIDS in Ghana between 2000 and 2022. The framework focuses on explaining HIV incidence through spatial, structural, and behavioral factors, using machine learning models to support, but not replace, this explanation.

2.2. Study Area and Units of Analysis

The study was conducted across ten legacy administrative regions to maintain consistency in the data before and after the 2018 redistricting. Monthly data were gathered from 2000 to 2022, providing a panel dataset of 9,792 observations (12 months \times 23 years \times 10 regions).

2.3. Data Sources and Variable Construction

Data were compiled into a region-level panel for Ghana's ten legacy regions (2000-2022): monthly HIV incidence, ART coverage, and testing counts from the Ghana Health Service (GHS) and Ghana AIDS Commission (GAC); socio-behavioral indicators (education, HIV knowledge/awareness, stigma, condom use, literacy, youth unemployment) from the Ghana Statistical Service (GSS), Demographic and Health Surveys (DHS), and UNAIDS; and geospatial boundaries from geoboundaries.org. The outcome is monthly new HIV/AIDS cases per 100,000 population.

Definitions and scales

1) Education access index = percentage of regional population aged ≥ 15 with at least secondary education (0 - 100%);

2) HIV awareness index = composite percentage of correct responses to core DHS/UNAIDS HIV knowledge items (0 - 100%), standardized (z-score) for modeling;

3) Regional stigma index = aggregate of DHS stigma/attitude items rescaled to 0 - 1 (0 = minimal reported stigma; 1 = maximal stigmatizing responses).

Survey-derived indicators (education, awareness, stigma) were available intermittently and were aligned to the monthly panel via linear interpolation or held constant across inter-survey periods when interpolation was not appropriate; administrative indicators were used at their native monthly frequency. Prior to modeling, survey indices were smoothed with 3-month rolling averages and standardized; continuous predictors were winsorized at the 1st and 99th percentiles and transformed to z-scores for clustering and ML. Missing values were examined and imputed (primary approach: KNN, $K = 5$) with MAR diagnostics and sensitivity checks using MICE reported in Section 3.6 and the Supplement. The index con-

struction scripts, variable dictionary (observed min/max and regional means), and analysis code are archived in the Zenodo repository (DOI: 10.5281/zenodo.15292209) for reproducibility.

2.4. Spatial Data Harmonization

Regional geographic boundaries and surveillance data were harmonized to a common monthly time step and to the ten legacy administrative regions used throughout the study. Spatial weights were constructed using Queen contiguity on the regional polygon layer, and both Global Moran's I and Local Moran's I (LISA) were computed to assess spatial autocorrelation in HIV incidence and candidate predictors. To reduce bias from unobserved, region-specific factors, we included region fixed effects and a spatially lagged incidence term (the weighted mean incidence of neighboring regions) among the predictors used for modeling. After fitting models, residuals were re-tested with Global Moran's I and LISA to confirm that no substantive spatial clustering remained; these spatial diagnostics and the mapping procedures are reported in the Results and Supplementary Materials. All spatial operations were performed using standard spatial-analysis libraries (PySAL/GeoPandas), and the code repository contains the exact scripts used for harmonization and testing.

2.5. Spatial and Behavioral Analysis

HIV incidence and ART coverage maps were developed using Folium and Altair to identify places where social stigma and illiteracy overlap. Using K-means clustering with three clusters, this study identified different patterns across the country for educational access, ART coverage, stigma, and awareness. The elbow method was used to determine the optimal number of clusters, while hierarchical clustering with Ward's linkage was used to ensure that the results were robust. Global Moran's I was used to identify the overall clustering pattern, while Local Moran's I (LISA) was used to identify hotspots and coldspots of HIV incidence. Model residuals were mapped to identify regions that performed better or worse than expected, with values exceeding ± 1.5 indicating potential structural inconsistencies. The relationship between HIV infection and key variables (HIV awareness, ART coverage, educational access, stigma index, and condom use) was explored using boxplots and scatterplots.

The optimal number of K-means clusters was determined using multiple quantitative criteria: the elbow plot of total within-cluster sum of squares (SSE), average silhouette scores, and the gap statistic. All three metrics indicated an inflection at $K = 3$ (elbow in SSE, peak/near-peak average silhouette, and gap statistic favoring 3 clusters). Hierarchical clustering with Ward linkage produced a concordant three-group solution.

2.6. Data Preprocessing

Temporal filtering was applied to retained data from 2000 to 2022. Missing values

were imputed using the K-nearest neighbor method ($K = 5$) based on correlated variables such as urbanization and facility density. Indicators for condom use and HIV awareness were smoothed using 3-month rolling averages. The analytic dataset is monthly ($n = 9792$; 12 months \times 23 years \times 10 regions). Many administrative indicators (incidence, ART reporting, testing counts) were available monthly from GHS/GAC. Survey-based indicators (education, awareness, stigma) were available intermittently (DHS waves); these were interpolated to monthly frequency using linear interpolation and held constant between survey years when interpolation was not appropriate. To reduce reverse causality and capture short-term implementation effects, predictor variables used in the ML models were aligned to the outcome using a 3-month lag (predictor at month t used to predict incidence at month $t + 3$) and smoothed with 3-month rolling averages as described above. Sensitivity checks were performed with 0- and 6-month lags; model performance and feature ranks were robust to these alternatives. To mitigate the effect of outliers, the winsorizing technique was applied at the 1st and 99th percentiles. All variables were normalized using Z-scores for the modeling and clustering steps.

Missing values were imputed using KNN ($K = 5$) on variables with correlated predictors (e.g., urbanization, facility density). To evaluate the plausibility of Missing-At-Random (MAR), (a) missingness patterns were examined by variable and year, (b) logistic regressions were run predicting missingness indicators using observed covariates (urbanization, region, year), and (c) Little's MCAR test to differentiate MCAR from MAR/MNAR was conducted. These diagnostics suggested that missingness was not MCAR and was plausibly MAR—*i.e.*, missingness was associated with observed region-level characteristics rather than unobserved outcomes. As a sensitivity analysis, we re-ran the principal models using multiple imputation by chained equations (MICE) and with simple median imputation; model ranks (SHAP importances) and counterfactual experiment directions were robust across imputation methods.

2.7. Interpretable Machine Learning Analysis

Instead of a single 80/20 temporal split, a rolling-origin (time series) cross-validation for hyperparameter tuning and model selection was used. Specifically, an expanding-window TimeSeriesSplit with five folds was applied: the initial training window covered 2000-2011, and each successive fold expanded the training set and validated on the next contiguous period. Hyperparameter tuning was performed within each fold using GridSearchCV. To preserve a full recent-period test, final holdout evaluation was performed on 2020-2022. Results from rolling-origin CV (mean \pm SD) and the final holdout metrics (R^2 , RMSE, MAE, MAPE) are reported.

Tree-based models (Random Forest, XGBoost) were trained and interpreted using SHAP values and partial dependence plots to evaluate variable importance and functional relationships. To respect temporal structure and avoid information

leakage, predictors were aligned to the outcome using a 3-month lag and smoothed with 3-month rolling averages; sensitivity checks using 0- and 6-month lags are reported. We assessed the robustness of feature importances to preprocessing choices by re-running models under alternative imputation strategies (KNN, MICE) and by examining model residuals for spatial autocorrelation (see Section 3.4). Counterfactual simulations and regional-level effect summaries reported in the Results are derived from the final time-aware models and SHAP-based explanations.

2.8. Streamlit Dashboard Deployment

A dashboard app was developed using Python with HyperText Markup Language (HTML) and Cascading Style Sheet (CSS) elements. The app was deployed via Streamlit to enable the interactive exploration of HIV incidence maps, SHAP scores, cluster patterns, and counterfactual forecasts. The app is version-controlled through GitHub and archived on Zenodo under DOI: 10.5281/zenodo.15292209. The app can be publicly accessed at

<https://dashboardapppy-3ryuuaxnlmrsrrkqoxpcfw.streamlit.app>.

3. Results

3.1. Geographic Distribution of HIV Incidence

Choropleth mapping revealed pronounced spatial disparities in HIV incidence across the ten administrative regions of Ghana. Regions such as Greater Accra, Ashanti, and Central consistently recorded the highest incidence rates, surpassing 220 new HIV cases per 100,000 people annually (**Figure 1**). In contrast, the Upper East, Northern, and Volta regions had lower incidence levels, averaging below 140 per 100,000, although recent data indicate increasing trends in some of these areas.

These geographic variations persisted over the period 2000-2022, suggesting that structural and behavioral inequities have remained entrenched. **Figure 1** illustrates this trend through a regional choropleth map of the average incidence rates over the study period, thereby capturing the temporal stability of these disparities.

3.2. Regional Clustering by Sociobehavioral Profiles

To better understand the heterogeneity in the incidence patterns, unsupervised machine learning (K-means clustering, $K = 3$) was employed to classify regions into three socioepidemiological clusters on the basis of the standardized values of education access, ART coverage, HIV awareness, and stigma index. The clustering algorithm grouped Greater Accra, Ashanti, and Central into Cluster A, characterized by a high HIV burden alongside high urbanization, ART coverage, access to education, and relatively elevated stigma index scores. Cluster B, composed of the Brong-Ahafo, Western, and Eastern regions, presented intermediate incidence rates and a mixture of socio-behavioral profiles, and was hence classified as transitional. Cluster C, which included the northern, upper-east, and upper-west

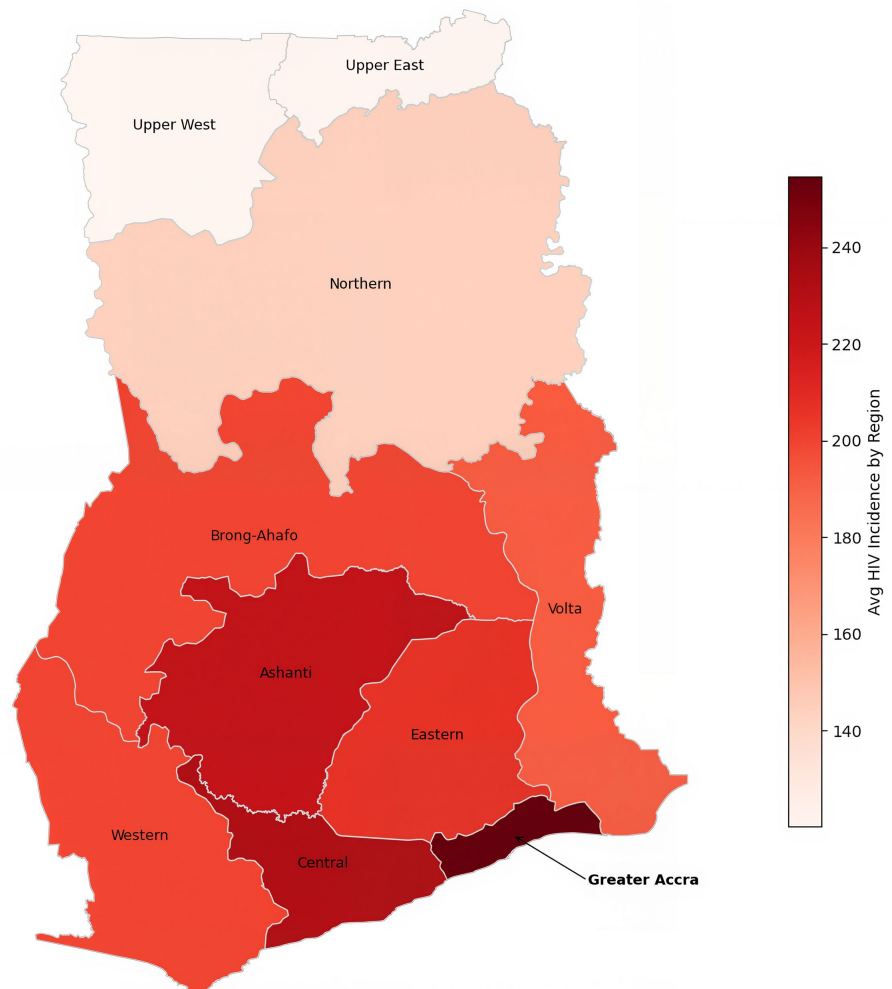


Figure 1. Choropleth map of average HIV incidence.

regions, was characterized by a lower incidence, but also lower levels of ART access, education, and HIV awareness, combined with higher stigma indices.

These observations are supported by **Table 1**, which details the regional-level HIV and TB averages, and **Table 2**, which summarizes malaria incidence, education access, and urbanization levels across the regions. Regional typologies are shown in **Figure 2**.

3.3. Sociobehavioral Correlations with HIV Incidence

Bivariate correlation analysis demonstrated strong and statistically significant associations between HIV incidence and several key socio-behavioral predictors. Access to education was positively correlated with HIV incidence ($r = 0.71$), indicating a higher incidence in regions with greater coverage of formal education. This seemingly paradoxical correlation may be explained by differential case detection; better-educated regions are likely to have higher testing penetration, case ascertainment, and surveillance infrastructure, which can inflate the reported incidence despite potentially lower actual transmission. While counterintuitive at

Table 1. Spatial analysis—HIV & TB summary.

Region	HIV incidence (Mean)	Std_Dev	Min	Max	TB incidence (Mean)
Ashanti	224.1	33.71	154.35	288.78	159.89
Brong-Ahafo	199.97	29.76	140.46	260.1	147.87
Central	231.78	35.44	163.72	298.04	152
Eastern	207.73	31.01	148.43	267.18	144.06
Greater Accra	254.44	35.92	184.7	300.98	167.26
Northern	144	21.41	97.13	188.14	136.07
Upper East	120.46	17.44	94.77	155.66	128.23
Upper West	120.27	17.53	94.77	154.8	128.29
Volta	191.92	28.69	132.32	251.86	148.01
Western	200.04	29.74	138.11	259.86	151.99

Table 2. Regional-level summary statistics for malaria incidence, education access, and urbanization (2000-2022).

Region	Malaria incidence (Mean)	Education access index (Mean)	Urbanization level (Mean)
Ashanti	199.64	56.84	62.94
Brong-Ahafo	216.1	53.38	50.76
Central	207.88	54.21	48.15
Eastern	199.9	58.66	56.87
Greater Accra	129.12	65.6	78.71
Northern	304.14	43.76	39.34
Upper East	359.51	48.11	34.98
Upper West	367.21	45.52	36.72
Volta	216.06	52.49	45.47
Western	224.07	55.16	52.52

first glance, further analysis suggests that this relationship may be driven by higher levels of testing, urbanization, and surveillance intensity in better-educated regions.

Similarly, both the urbanization level and the HIV awareness index were positively correlated with incidence ($r = 0.65$ and $r = 0.59$, respectively), reinforcing the notion that areas with more infrastructure and outreach services are more likely to detect and report cases. Condom use rates and the regional stigma index showed complex patterns. While condom use was moderately associated with reduced incidence, stigma appeared to correlate with underreporting and reduced service uptake, especially in lower-burden regions where disclosure remains a challenge.

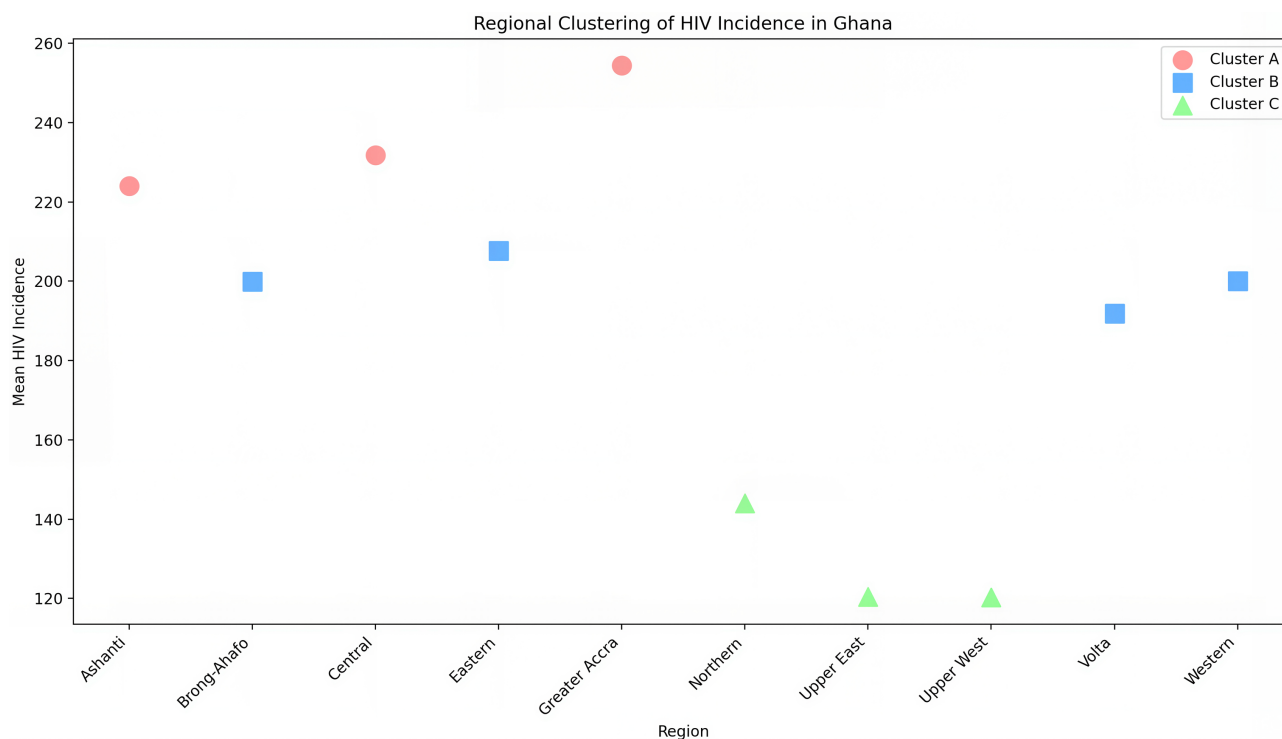


Figure 2. K-means cluster map of Ghana's ten regions.

These patterns became particularly salient when urban and rural geographies were compared. Urbanized regions, such as Greater Accra and Ashanti, despite having high education access and ART coverage, also recorded the highest incidence rates. This paradox may be explained by intensified surveillance, elevated partner concurrency, transactional sex, and increased network density. In contrast, rural settings, such as the Lower Manya Krobo Municipality (LMKM) in the Eastern Region, demonstrated localized hyperendemicity. Agormanya, a sentinel site within the LMKM, reported an HIV prevalence of 19.2%, among the highest in Ghana, despite the region's relatively modest health infrastructure (Ocran, 2022; Dias, 2021).

These rural-urban contrasts align with the SHAP and PDP outputs from the predictive models, which ranked urbanization, educational access, and stigma among the most influential variables (Figures 3-5). Evidence suggests that regional differences in the HIV burden are not binary but instead are shaped by layered contextual factors, including behavioral norms, population mobility, stigma gradients, and access visibility.

Figure 3 presents scatterplots illustrating the relationships between HIV incidence and selected predictors, and Table 3 summarizes the strength and direction of these associations using Pearson correlation coefficients. Table 4 provides the background on the quality and transformation of these variables.

3.4. SHAP-Based Interpretation of Structural Predictors

To further explore the drivers of HIV incidence, random forest and XGBoost

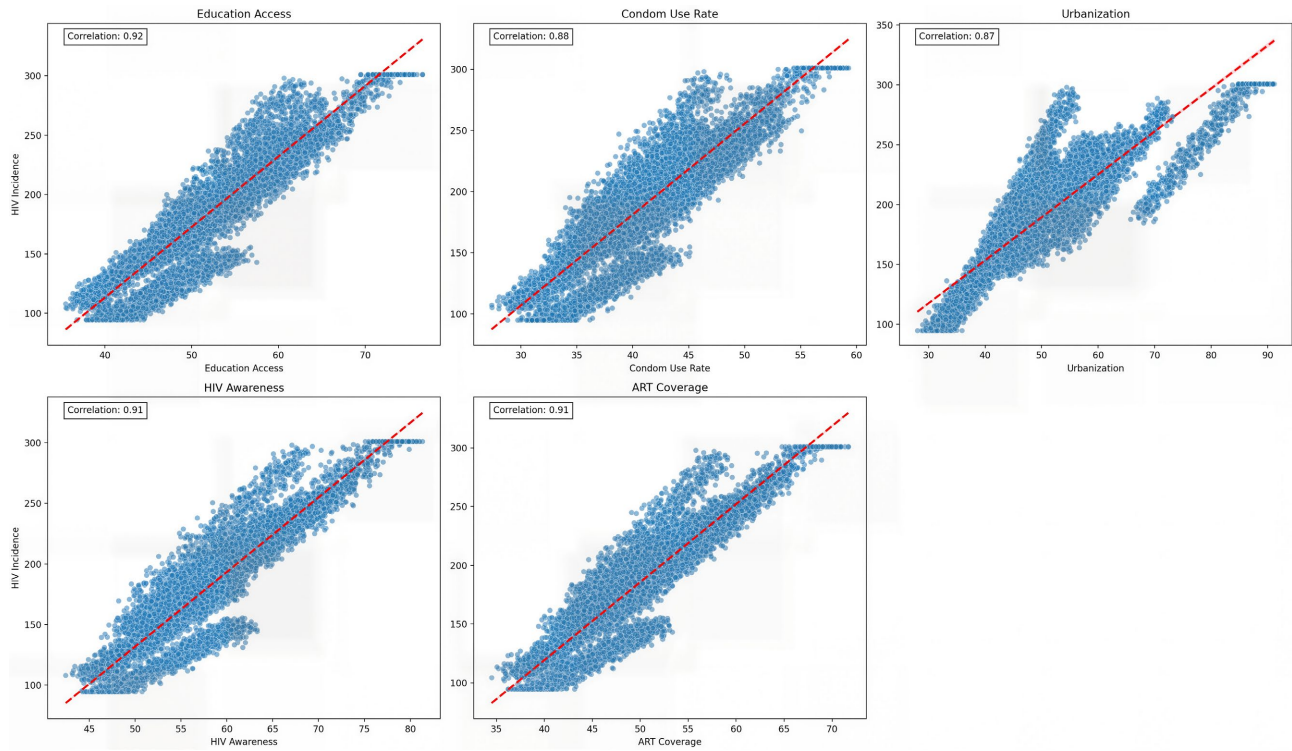


Figure 3. Scatterplots of HIV incidence vs. sociobehavioral variables.

Table 3. Key feature importance and SHAP summary for HIV_incidence.

Features	Data type	Missing values	Missing %	Unique values	Outliers detected	Post winsorisation
region	object	0	0	10	64	0
date	object	0	0	612	9	0
tb_outlier	bool	0	0	1	0	0
malaria_outlier	bool	0	0	2	379	0
urban_rural_sum_pct	float64	0	0	1	0	0
age_sum_pct	float64	0	0	3	57	0
migration_rate	float64	0	0	9261	N/A	0
regional_stigma_index	float64	0	0	6579	0	0
urbanization_level	float64	0	0	9792	112	0
health_facility_density	float64	0	0	9761	N/A	0
testing_coverage_pct	float64	0	0	9792	N/A	0
access_to_art_pct	float64	0	0	9792	0	0
hiv_awareness_index	float64	0	0	9792	0	0
youth_unemployment_rate	float64	0	0	9792	0	0
female_literacy_rate	float64	0	0	9792	52	0
condom_use_rate	float64	0	0	9792	69	0

Continued

education_access_index	float64	0	0	9792	45	0
region_fixed	object	0	0	10	0	0
population_rural_pct	float64	0	0	9531	52	0
population_urban_pct	float64	0	0	9531	N/A	0
population_15_64_pct	float64	0	0	9792	334	0
population_65_plus_pct	float64	0	0	9781	N/A	0
population_0_14_pct	float64	0	0	9751	0	0
population_total	float64	0	0	9631	44	0
hiv_incidence	float64	0	0	9598	53	0
tb_incidence	float64	0	0	9598	0	0
malaria_incidence	float64	0	0	9598	59	0
month	int64	0	0	12	N/A	0
year	int64	0	0	51	92	0
hiv_outlier	bool	0	0	1	2	0

Feature name	Description
education_access_index	% of the population with secondary education or higher
condom_use_rate	Percentage consistently using condoms
female_literacy_rate	Literacy among women aged 15+
youth_unemployment_rate	Youth (15 - 24) unemployment rate
hiv_awareness_index	Composite score measuring HIV knowledge and awareness
access_to_art_pct	% of HIV-positive individuals receiving ART
testing_coverage_pct	% of population tested for HIV
health_facility_density	Number of health facilities per 10,000 people
regional_stigma_index	0 - 1 index quantifying HIV-related stigma in regions
urbanization_level	% of the population living in urban areas
migration_rate	Net migration rate per 1000 people

Table 4. Correlation coefficients between HIV incidence and key predictors.

Predictor	HIV incidence	Malaria incidence	TB incidence	Education access	Condom use rate	Female literacy	Youth unemployment	HIV awareness	ART coverage	HIV testing coverage	Facility density	Urbanization level
HIV incidence	1	-0.41	0.85	0.92	0.88	0.91	-0.31	0.91	0.91	0.91	0.2	0.87
Malaria incidence	-0.41	1	0.08	-0.31	-0.25	-0.27	0.86	-0.18	-0.25	-0.21	-0.1	-0.53

Continued

TB incidence	0.85	0.08	1	0.83	0.82	0.84	0.12	0.89	0.86	0.87	0.17	0.68
Education access	0.92	-0.31	0.83	1	0.95	0.97	-0.29	0.95	0.96	0.96	0.21	0.89
Condom use rate	0.88	-0.25	0.82	0.95	1	0.95	-0.26	0.95	0.95	0.95	0.21	0.89
Female literacy	0.91	-0.27	0.84	0.97	0.95	1	-0.25	0.93	0.94	0.95	0.2	0.85
Youth unemployment	-0.31	0.86	0.12	-0.29	-0.26	-0.25	1	-0.19	-0.26	-0.2	-0.09	-0.48
HIV awareness	0.91	-0.18	0.89	0.95	0.95	0.93	-0.19	1	0.97	0.96	0.21	0.88
ART coverage	0.91	-0.25	0.86	0.96	0.95	0.94	-0.26	0.97	1	0.96	0.21	0.9
HIV testing coverage	0.91	-0.21	0.87	0.96	0.95	0.95	-0.2	0.96	0.96	1	0.21	0.87
Facility density	0.2	-0.1	0.17	0.21	0.21	0.2	-0.09	0.21	0.21	0.21	1	0.22
Urbanization level	0.87	-0.53	0.68	0.89	0.89	0.85	-0.48	0.88	0.9	0.87	0.22	1

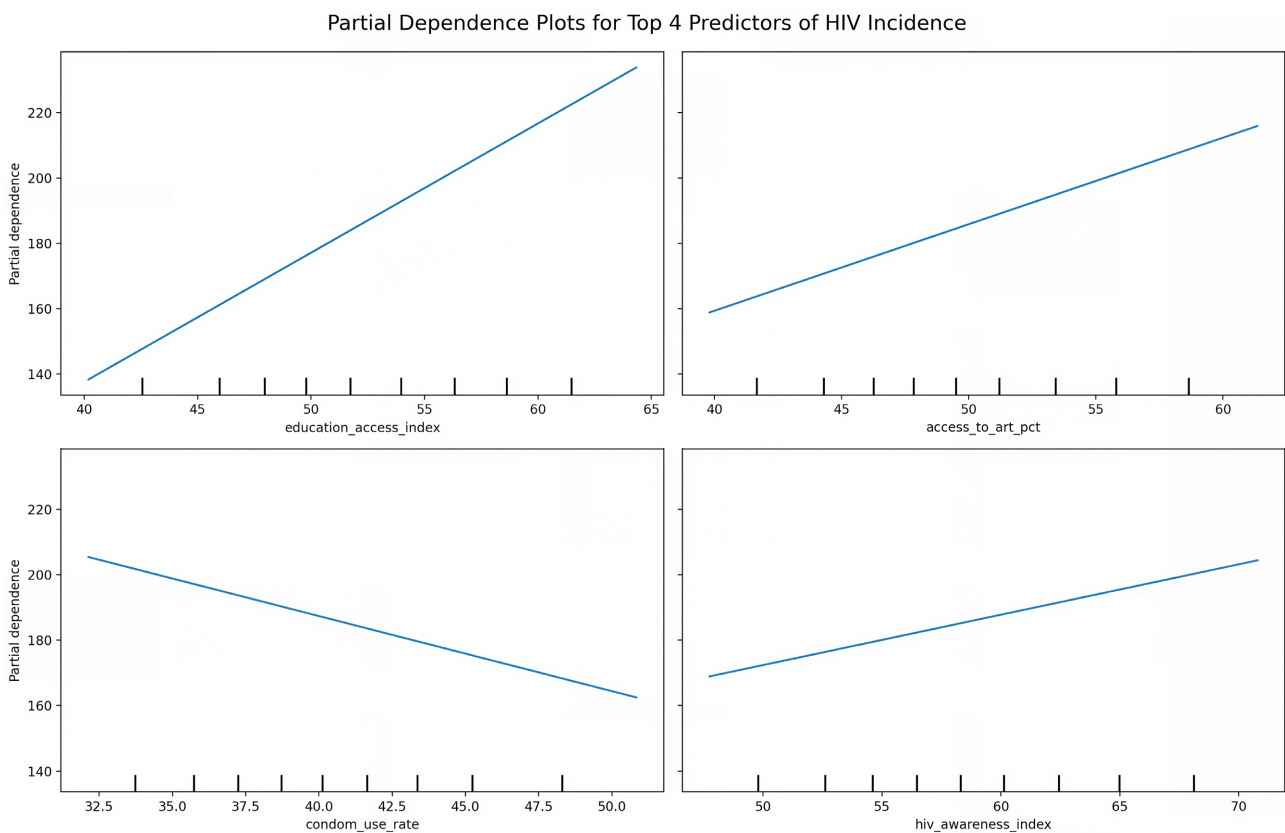


Figure 4. Partial dependency plots (Top 4 Figures).

models were trained on the harmonized dataset and interpreted by Shapley Additive Explanations (SHAP) values. The SHAP summary plot ranks the features based on their mean absolute contribution to the model predictions. ART coverage, educational access, HIV awareness, urbanization, and TB co-infection were the most significant predictors.

Several threshold effects were observed. Notably, regions with an education index above 0.6 experienced a steep increase in the observed incidence, possibly re-

flecting increased detection due to improved health literacy and testing uptake. Conversely, ART coverage exceeding 45% was associated with a substantial decline in incidence, highlighting the preventive impact of treatment-as-prevention approaches. HIV awareness levels above 70% exhibited diminishing returns, suggesting that beyond a certain threshold, increased awareness alone did not yield further reductions in transmission.

Figure 4 displays the global SHAP summary plot, whereas **Figure 5** offers detailed SHAP dependence plots illustrating how changes in education and ART coverage influenced the predicted incidence across regions. **Table 5** summarizes

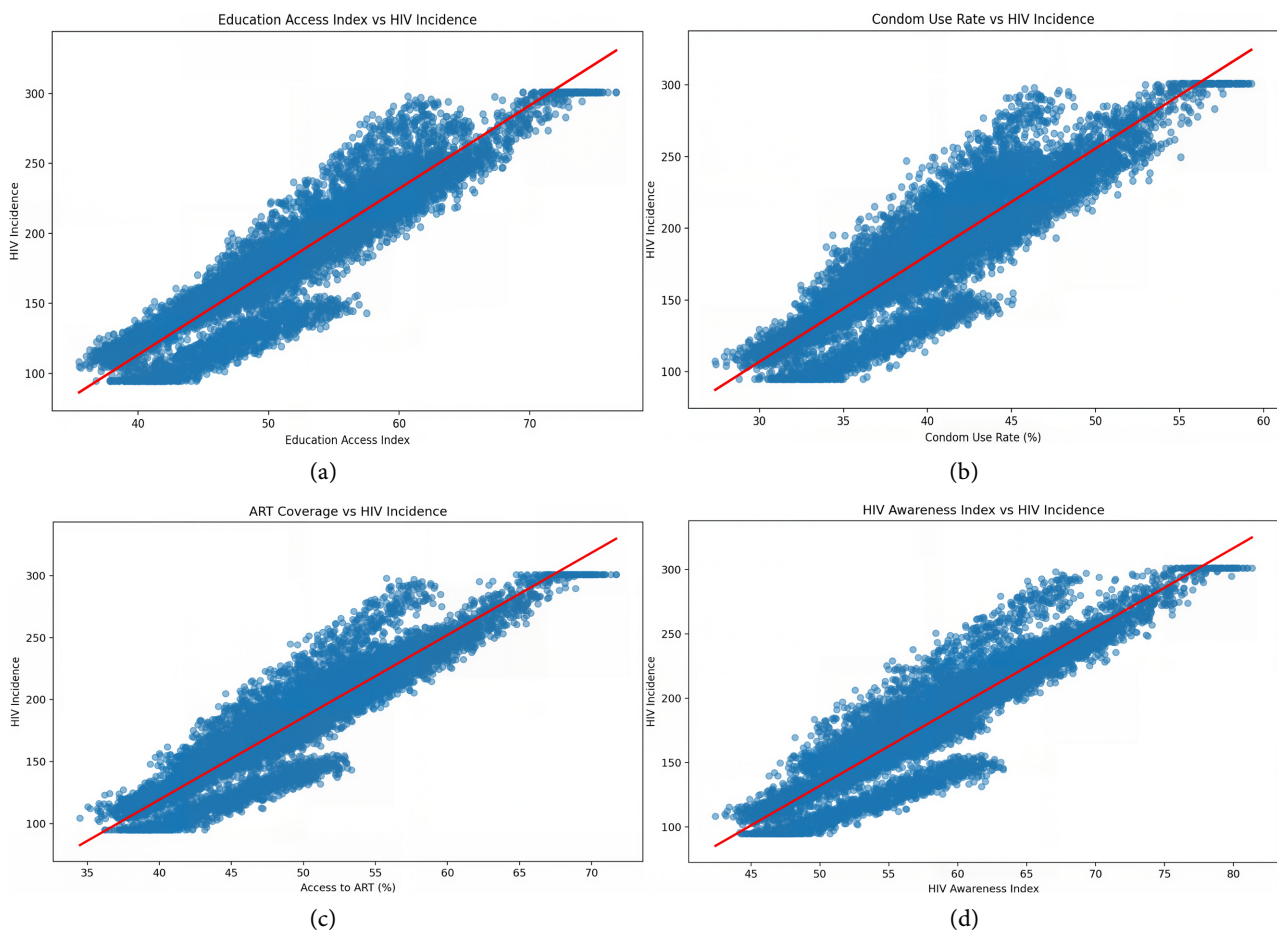


Figure 5. (a) SHAP Dependence for Education Access Index; (b) SHAP Dependence for Condom Use Rate; (c) SHAP Dependence for ART Coverage; (d) SHAP Dependence for HIV Awareness.

Table 5. Key feature importance and SHAP summary for HIV incidence.

Feature	SHAP score	Key interaction highlighted
SHAP for condom use rate	0.127	Positive effect increases with literacy.
SHAP for education access index	0.241	Nonlinear jump effect around score ~55
SHAP for HIV awareness index	0.198	Steep increase above awareness ~65
SHAP for TB incidence	0.113	Moderate rise, especially with higher urbanization

the SHAP-based rankings and explanations.

3.5. Counterfactual Simulation: Impact of Modifying Key Predictors

A counterfactual analysis was performed using the trained XGBoost model to simulate the potential effects of policy interventions. A hypothetical 10% increase in either HIV awareness or educational access was applied uniformly across all regions. The results revealed that high-burden regions such as Greater Accra, Ashanti, and Western benefited the most, with projected reductions in incidence ranging from 13% to 16.3%.

The transitional and lower-access regions demonstrated modest projected declines between 7% and 10%. Notably, the simulations revealed that regions with lower baseline levels of education or awareness experienced the greatest relative gains, indicating greater elasticity of response. These findings support the idea that targeted structural improvements, particularly in educational access and public health awareness, can yield meaningful reductions in HIV burden when directed toward vulnerable regions.

These counterfactual results highlight the policy sensitivity of behavioral determinants in Ghana's HIV response. Importantly, they illustrate that the effectiveness of interventions such as education and awareness is not uniform but context-dependent, amplified in transitional zones with mid-level infrastructure and underutilized testing. This confirms a "diminishing returns" pattern, where high-incidence urban centers with saturated services gain less from blanket messaging, whereas transitional or underresourced regions respond more elastically to marginal investment. Thus, counterfactual simulation does not merely estimate statistical change; it functions as a strategic tool for public health targeting, guiding scalable, cost-efficient investments where they are likely to yield the greatest marginal benefit.

These findings are grounded in the predictor distributions described in **Table 2** and SHAP simulation logic outlined in **Table 5**. **Figure 6** shows the projected changes in incidence across regions.

3.6. Residual Mapping and Latent Factors

Residual analysis comparing the predicted and observed incidence revealed spatial patterns, suggesting that latent factors were not captured in the model. The central and eastern regions exhibited systematic underprediction, possibly indicating the presence of protective community norms, localized prevention programs, or reporting gaps that were not accounted for in the dataset. This discrepancy may reflect unmeasured behavioral resilience or under-documented public health interventions.

Conversely, the Greater Accra and Western regions showed consistent overprediction, which could signal higher-than-expected transmission dynamics or limitations of the model in capturing highly mobile populations, transactional sexual behaviors, or complex social and sexual network dynamics. These insights are vis-

ualized in **Figure 7**, which maps the regional distribution of the average residuals. Thus, residual analysis serves as a diagnostic tool, guiding future efforts to refine both surveillance and predictive modeling.

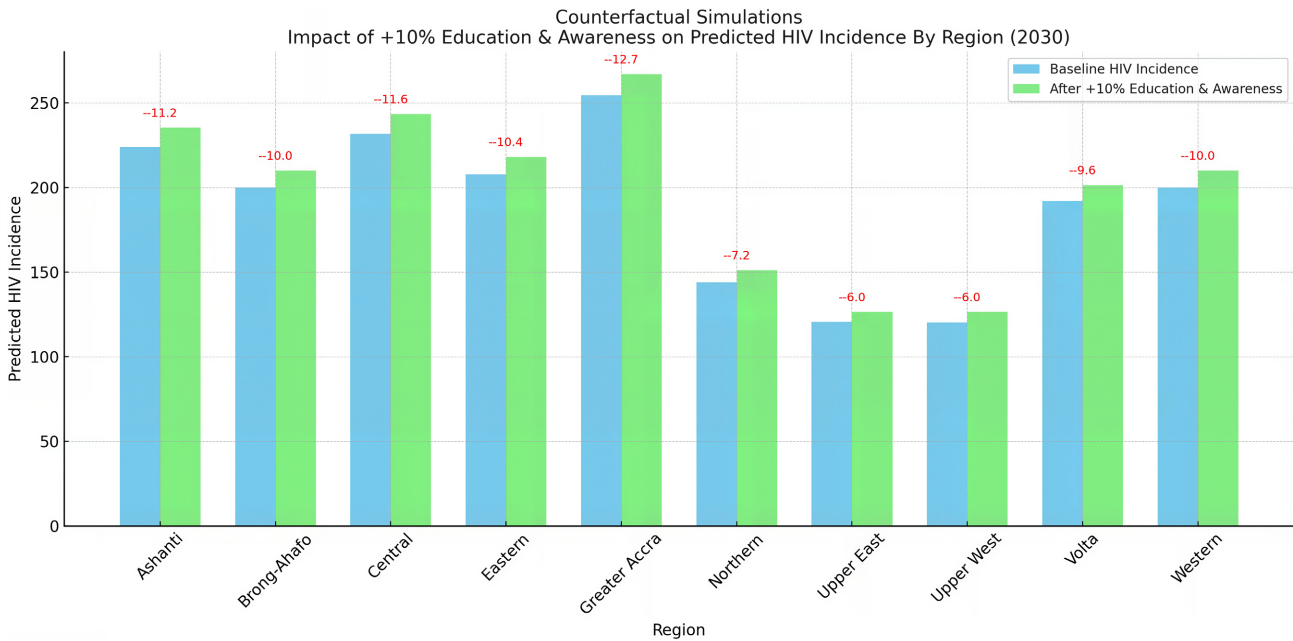


Figure 6. Counterfactual simulations.

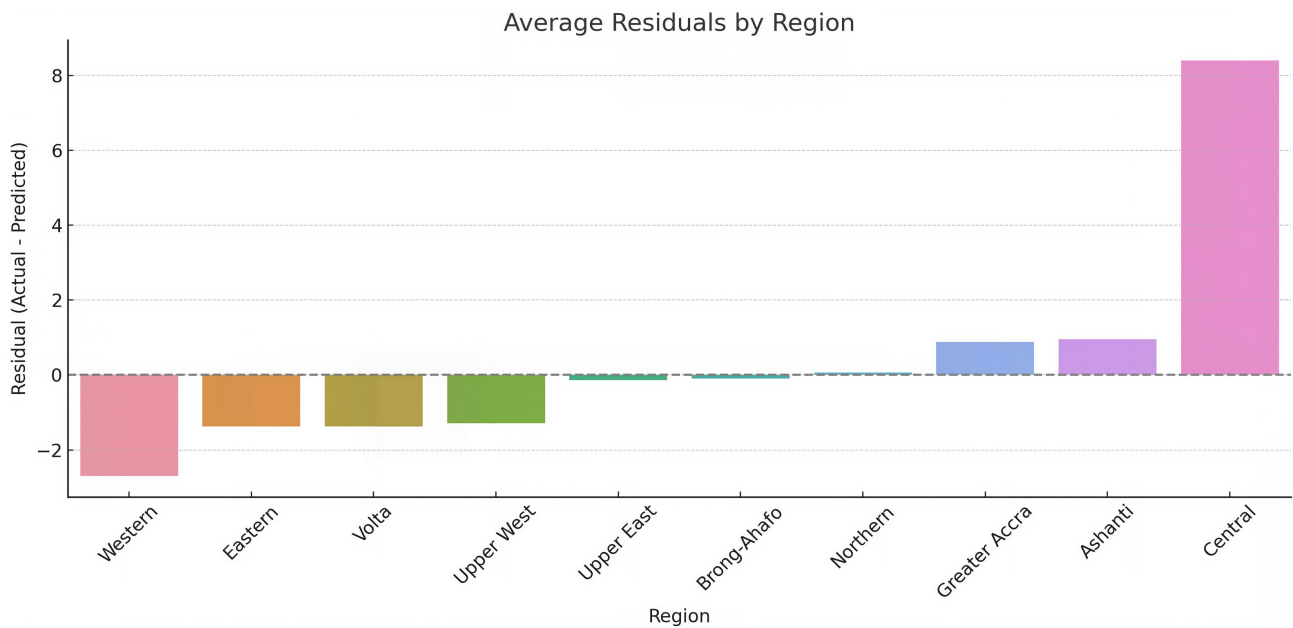


Figure 7. Average residuals by region.

4. Discussion

4.1. Principal Findings and Interpretations

The study applied ecological machine learning and spatial clustering to investi-

gate the socio-behavioral and structural determinants of HIV incidence across ten administrative regions of Ghana. According to the findings, HIV incidence is highest in Greater Accra, Ashanti, and some areas in the Central and Eastern regions, which have more schools, larger towns, and greater access to ART. In addition, these results mirror a broader phenomenon in sub-Saharan Africa (SSA), where highly urbanized areas tend to have a higher rate of HIV among youth and young women, despite having better healthcare [7] [8]. These urban “hotspots” often align with zones of heightened economic activity and risky sexual behavior [8].

Paradoxically, these urban areas still had high incidence rates. SHAP and PDP visualizations in this study suggest that while ART coverage of over 45% is negatively correlated with HIV incidence rates (**Figure 4** & **Figure 5**), this trend was not significant in urban areas. ART saturation may reduce marginal gains unless accompanied by behavioral shifts [9]. This is in agreement with studies showing that biomedical interventions alone are insufficient without complementary behavioral or structural shifts, especially among high-risk female populations, such as bar workers [10].

Figure 2 (Cluster A: high burden, high access; Cluster B: transitional; Cluster C: low burden, low access) reflects the ideas proposed by previous studies [6] [11], demonstrating that HIV outcomes are shaped not only by infrastructure but also by layered behavioral, economic, and spatial inequalities.

4.2. Urban-Rural Inequities and Behavioral Dynamics

Although urban areas such as Accra demonstrate higher ART coverage and testing rates, they paradoxically sustain an elevated incidence. This urban-rural paradox mirrors broader regional findings where urban youth, particularly girls aged 15 - 24, face higher HIV risks due to mobility, economic survival strategies, and social vulnerabilities [7] [8]. SHAP outputs (**Figures 5(a)-(d)**) revealed that this mismatch arises from high-risk behaviors, population mobility, and structural gaps such as inconsistent education programming. In contrast, rural zones such as Lower Manya Krobo (LMKM) report hyper-local epidemics—Agormanya’s 19.2% prevalence [2]—despite the regional averages being lower. Such local disparities reflect “micro-epidemic” dynamics observed across Eastern and Southern Africa, where high-prevalence pockets are not necessarily aligned with broader regional averages [8]. **Table 1**, **Table 2**, and **Table 4** provide supporting evidence of these structural differences.

This spatial vulnerability reflects the underlying behavioral and structural vulnerabilities. Male reluctance toward HIV testing and under-targeted adolescents have been observed as drivers of under-diagnosis in rural zones [6]. Similarly, Gu *et al.* (2021) found that an autonomy-supportive healthcare climate was a significant positive predictor of linkage to HIV care for men who have sex with men (MSM) in Ghana, while perceived community stigma (felt normative stigma) was a negative predictor (pp. 5-6). Their research highlights the critical role of the healthcare

environment and social stigma in influencing engagement with HIV treatment services among this key population [12]. These findings are consistent with our model residuals (Figure 7), which were under-predicted in the Eastern and Central Regions, potentially because of stigma-suppressed disclosure [13] or local resilience factors.

Educational access and HIV awareness (SHAP dependence plots, Figure 5(a) & Figure 5(d)) only reduced the incidence above the 0.6 threshold. This resonates with findings that general awareness rarely leads to preventive action unless paired with structural empowerment, particularly among populations such as female bar workers, who often report high HIV awareness but lack negotiating power for condom use [10]. Moreover, some scholars argue that general awareness (~98%) often does not translate into comprehensive knowledge [1]. This supports a threshold-based policy: boosting access in underserved areas, where gains would be exponential.

4.3. Simulation of Structural Levers and Policy Impacts

Counterfactual modeling demonstrated that a +10% increase in HIV awareness or educational access can yield incidence reductions of up to 16.3% in high-burden regions (Figure 6). This observation aligns with real-world results from initiatives such as the Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) program, which showed up to a 40% reduction in HIV incidence in targeted adolescent groups when structural and behavioral components were integrated [7]. The observed responsiveness of the model to education-based interventions underscores the need for structural tailoring. For instance, Dambach *et al.* (2020) highlighted that, in bar settings, even moderate wage improvements or access to sexual health counseling could shift behavior away from transactional sex. These are scalable levers that can complement national awareness campaigns [10].

The SHAP simulations (Figure 5 & Figure 6) also demonstrated that this effect (education boost reduces HIV in hotspots) is most pronounced in transitional regions, where the baseline indicators are moderate. These model-derived insights correspond with the feature importance results in Table 5 and data validation metrics in Table 3, and support the case for a paradigm shift toward non-biomedical drivers in HIV prevention.

These simulations reinforced the need for differentiated programming. Regions such as Ashanti and Western responded sharply to educational improvements. These findings echo those of a 2021 study that linked female education and literacy programs to durable HIV prevention outcomes [14].

Such evidence also supports the findings of previous research from 2021, which advocated for region-targeted interventions in SSA, highlighting that one-size-fits-all strategies are inefficient and ethically problematic in heterogeneous epidemics [11].

These findings suggest that public health investments in educational infrastruc-

ture and awareness campaigns could have the greatest impact when strategically allocated to transitional or underserved regions. In effect, simulations serve as policy scenario-testing tools, enabling decision-makers to visualize tangible returns on specific structural reforms.

4.4. Comparison with Prior Literature

The findings from this investigation are broadly aligned with those of prior studies, but introduce new granularity. While past studies [1] [6] have highlighted spatial disparities in HIV testing, this study quantifies behavioral thresholds and identifies saturation effects. Notably, the nonlinear impact of education—initial increases in incidence due to testing bias, followed by later declines—has not been previously modeled in Ghana.

A policy on abstinence-only education in LMKM contradicts the Ministry of Health's comprehensive messaging, creating mixed narratives that impede behavioral change [2]. These contradictory curricula are rarely captured in national datasets, but they significantly shape local epidemic trajectories.

Unlike traditional regression-based models [15], the SHAP-driven ML analysis (Figures 5(a)-(d)) in this study uncovered nonlinear and threshold effects across the structural predictors. While it does not model spatial lag explicitly, the feature behavior patterns provide richer insight than models that assume linear, independent effects.

4.5. Study Limitations

This study has several limitations that should be considered when interpreting the findings. First, the use of an ecological design based on region-level aggregates introduces the risk of an ecological fallacy. As a result, the associations observed between regional predictors and HIV incidence cannot be assumed to apply to individuals within these regions [16]. While this is valid, recent spatial studies such as Bulstra *et al.* (2020) have used similar regional aggregation techniques and were still able to identify robust patterns of micro-epidemics, suggesting that the method retains analytical value even if individual-level inferences are limited [8].

Although K-Nearest Neighbors (KNN) imputation was employed to address missing data (Table 3), the assumption that data were missing at random may not hold for certain variables, particularly sensitive behavioral indicators such as condom use. Previous studies [17] [18] have underscored the potential for social desirability bias in such self-reported behaviors, which could distort the accuracy of the imputed values. Furthermore, this limitation is well justified, especially in light of studies such as Dambach *et al.* (2020), who documented social desirability bias in self-reporting among female bar workers, resulting in significant under-reporting of risky sexual behaviors [10].

Furthermore, the spatial aggregation process, in which Ghana's 16 newly demarcated regions were merged into the ten legacy administrative zones, may have obscured intraregional disparities. This could result in the underestimation of lo-

calized epidemics, especially in emerging high-risk zones not historically identified as hotspots [6]. The study also lacked the ability to control for latent confounders, such as transactional sex, economic migration, and displacement due to infrastructure development, such as dam construction. These factors, although not captured in the current dataset, are known to influence HIV transmission patterns, particularly in high-burden districts, such as Agormanya [2]. Moreover, this limitation is critical, as other studies have highlighted how mobility and economic factors drive risk, particularly among informal labor sectors such as market traders and bar workers [10] [12] [19]. In regions with infrastructure-induced displacement, HIV burden may be significantly underestimated if these transient populations are excluded.

Furthermore, the omission of key populations, such as sex workers and mobile traders, may lead to structural under-diagnosis in high-burden zones. Several studies have reported that these groups often operate outside routine surveillance systems and require targeted outreach strategies [12] [20]. Their exclusion weakens the generalizability of the current findings and limits their predictive sensitivity for areas experiencing rapid urban growth or seasonal migration.

4.6. Policy and Programmatic Implications

These findings highlight the urgent need for a more granular, equity-focused approach to HIV prevention and control in Ghana. Moving beyond national averages, the regional stratification observed in this study calls for precise public health strategies that are responsive to the unique sociostructural profiles of each region.

First, there is a clear need to expand community-based HIV testing services, particularly for men and adolescents who continue to exhibit lower testing rates. Gu *et al.* (2021) showed that adolescent testing uptake can be significantly improved through school-based and peer-driven interventions, particularly when stigma and peer influence are addressed simultaneously [12]. Mobile testing units, home-based testing kits, and self-testing options may help close this diagnostic gap between under-served northern and rural zones [6]. These interventions should be prioritized in areas where the testing prevalence is below the threshold values identified in the SHAP modeling framework (**Figure 5(d)**).

Second, policy realignment is necessary in regions in which abstinence-only education continues to dominate. In Lower Manya Krobo Municipality, for example, policy contradictions between the Ghana Education Service and the Ministry of Health have led to mixed messaging between educators and students [2]. This mirrors the structural misalignment reported by Sambah *et al.* (2020), where conflicting institutional narratives on youth sexuality undermined HIV prevention among female students [21]. Integrating comprehensive, evidence-based sex education, which includes condom use, partner reduction, and biomedical prevention, is essential for curbing localized epidemics. Moreover, cross-border trading corridors—identified as risk zones for mobile women traders—should incorpo-

rate livelihood support with HIV prevention messaging and services [11]. However, Cane *et al.* (2021) argued that livelihood support is insufficient unless paired with efforts to address gendered power imbalances that sustain women's vulnerability to coercion and HIV risk [7].

Resource allocation decisions should also be informed by SHAP-guided clustering (Figure 2) and partial dependence plots (Figure 4), which help to identify "transitional" regions with moderate coverage but high responsiveness to interventions. This data-driven targeting aligns with the calls for fine-scale geospatial mapping seen in Bulstra *et al.* (2020), who emphasized that localized epidemic zones often escape national policy radar and require precision prevention approaches [8].

Finally, implementing SHAP-informed or AI-assisted programming assumes a digital infrastructure and analytic capacity, which may be lacking in rural health systems. As Owusu *et al.* (2020) highlighted, the gap in digital literacy and uneven access to data platforms could hinder the practical integration of these tools, reinforcing existing inequities if not addressed through training and decentralization efforts [22].

5. Conclusions

This study analyzed the influence of spatial and socio-behavioral factors on HIV/AIDS incidence in Ghana using a combination of ecological modeling, spatial clustering, and explainable machine learning techniques. It was shown that both biomedical factors, such as ART availability, and socio-structural characteristics, such as education, awareness, stigma, and urban density, contribute to regional variations in HIV incidence.

The results revealed three distinctive region-level groups (Figure 2) characterized by varying levels of HIV risk and available resources. These analyses collectively reveal that HIV risk in Ghana exhibits distinct spatial and contextual variations. Areas with high HIV prevalence, such as Greater Accra, enjoy better access to resources but are also exposed to greater behavioral risk. The lower incidence in rural areas could be attributed to either genuine safety or lack of reporting (Figure 1).

The simulations revealed that targeted social policy interventions could substantially impact HIV incidence. Increasing education or awareness by 10% in high-incidence areas could reduce new infections by as much as 16% (Figure 6). These results highlight the importance of structural interventions for tackling HIV epidemics. Investing in education and awareness, particularly in areas experiencing a transition from low to moderate risk, has been shown to have a greater impact on reducing the incidence than increasing biomedical interventions. These findings support the reorientation of policies to address the root causes of HIV infections.

This study presents a methodology that enables researchers and policymakers to analyze spatial disparities, tailor interventions to specific regions, and predict the

public health benefits of evidence-based social policies. This advocates the adoption of regionally specific, evidence-based, and socially responsive interventions.

6. Policy Recommendations

The results highlight the importance of regionally specific HIV policies in Ghana. Improving education and awareness by small margins could lead to a 16% decrease in HIV incidence in areas undergoing economic and demographic transformations (Figure 6).

Eastern and Brong-Ahafo. These regions are particularly amenable to behavioral interventions, indicating that efforts such as HIV education, peer support, and school-based initiatives should receive greater attention. Communication should be culturally sensitive and should address misunderstandings and social barriers.

Areas such as Greater Accra and Ashanti face a conundrum, as they have high ART coverage but persistently high HIV incidence. The incidence of HIV remains elevated despite high ART uptake. SHAP dependence plots (Figures 5(a)-(d)) suggest that this may be due to persistent behavioral risks and saturation effects. Addressing behavioral factors, stigma, and network-based transmission is necessary to improve HIV control in these areas. Outreach to mobile youth, MSM, and sex workers should be a crucial component of prevention efforts. Approaches could include expanding Pre-Exposure Prophylaxis (PrEP) availability, conducting mobile testing, and running digital campaigns that leverage behavioral science to encourage participation and compliance.

Tools such as SHAP and geospatial clustering (Figure 4, Figure 5 and Figure 2) should be formally integrated into the decision-making processes of regional health directorates. These dashboards help inform planning and enable the timely identification of increasing risk. Integrating transparent ML outputs into district-level decisions improves transparency and responsiveness.

HIV services should be closely linked to maternal health, TB, and gender-based violence programs, particularly in under-resourced cluster C areas (Figure 2). A coordinated effort involving public, private, and donor resources is required to address geographical disparities and ensure lasting reforms. Applying preemptive, data-driven, and tailored approaches in each region is crucial for achieving equitable HIV reduction.

Declarations

Ethical approval or consent was not required as this study used only publicly available data. Publicly available aggregated data were analyzed. Personally identifiable information and samples collected from individuals were excluded.

Availability of Data and Materials

All materials used in this study were made openly accessible at the Zenodo repository under a CC-BY 4.0 license. The Zenodo archive contains detailed documen-

tation, metadata, and replicability-verification files. There were no limitations to the use of the data provided.

All necessary files and information required for reproducing the analysis are included in the Zenodo repository.

- Cleaned regional-level dataset (ghana_infectious_disease_model_dataset_cleaned.csv)
- Geospatial boundary files (GHA_10regions_merged_final.geojson)
- Model code and forecasting script
- Documentation and SHA-256 verification

Data were gathered from publicly available national reports and statistical summaries provided by organizations such as the GHS, GAC, MoH, GSS, UNAIDS, and the World Bank. The data were carefully cleaned, organized, and prepared for forecasting. This study did not use individual-level human data. The analysis employed regional-level data for forecasting.

Authors' Contributions

VG contributed to all aspects of the study, including the conceptualization, methodology, data collection and analysis, software development, validation, visualization, and drafting and editing of the manuscript.

Acknowledgments

The authors gratefully acknowledge the contributions of the Ghana Health Service, Ghana AIDS Commission, Ghana Statistical Service, and the Humanitarian Data Exchange (HDX) platform for sharing valuable epidemiological and demographic datasets that enabled the advancement of this research. I acknowledge the efforts of individuals who share open geospatial data through the geoboundaries' platform. This research is dedicated to the memory of my sister, Imelda Farr, who motivated me to pursue this academic goal.

Authors' Information

VG is a Biomedical Scientist working at the Cocoa Clinic, a Medical Department of the Ghana Cocoa Board in Accra, Ghana. He earned an MSc in Data Science from the University of East London and is currently pursuing an MSc in Public Health through distance learning at the University of Suffolk. He is interested in developing models for infectious diseases, making predictions using epidemiological data, and using machine-learning techniques to inform public health strategies.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Fenny, A.P., Crentsil, A.O. and Asuman, D. (2017) Determinants and Distribution of Comprehensive HIV/AIDS Knowledge in Ghana. *Global Journal of Health Science*,

- 9, 32-46. <https://doi.org/10.5539/gjhs.v9n12p32>
- [2] Ocran, B., Talboys, S. and Shoaf, K. (2022) Conflicting HIV/AIDS Sex Education Policies and Mixed Messaging among Educators and Students in the Lower Manya Krobo Municipality, Ghana. *International Journal of Environmental Research and Public Health*, **19**, Article No. 15487. <https://doi.org/10.3390/ijerph192315487>
- [3] Ba, D.M., Ssentongo, P. and Sznajder, K.K. (2019) Prevalence, Behavioral and Socio-economic Factors Associated with Human Immunodeficiency Virus in Ghana: A Population-Based Cross-Sectional Study. *Journal of Global Health Reports*, **3**, e2019092. <https://doi.org/10.29392/joghr.3.e2019092>
- [4] Dey, N.E.Y., Owusu Ansah, K., Norman, Q.A., Manukure, J.M., Brew, A.B.K., Dey, E.A., *et al.* (2022) HIV Testing among Sexually Active Ghanaians: An Examination of the Rural-Urban Correlates. *AIDS and Behavior*, **26**, 4063-4081. <https://doi.org/10.1007/s10461-022-03731-4>
- [5] Fiorentino, M., Coulibaly, B., Couderc, C., Keita, B.D., Anoma, C., Dah, E., *et al.* (2023) Men Who Have Sex with both Men and Women in West Africa: Factors Associated with a High Behavioral Risk of Acquiring HIV from Male Partners and Transmission to Women (Cohmsm ANRS 12324—Expertise France). *Archives of Sexual Behavior*, **53**, 757-769. <https://doi.org/10.1007/s10508-023-02715-2>
- [6] Nutor, J.J., Duah, H.O., Duodu, P.A., Agbadi, P., Alhassan, R.K. and Darkwah, E. (2021) Geographical Variations and Factors Associated with Recent HIV Testing Prevalence in Ghana: Spatial Mapping and Complex Survey Analyses of the 2014 Demographic and Health Surveys. *BMJ Open*, **11**, e045458. <https://doi.org/10.1136/bmjopen-2020-045458>
- [7] Cane, R.M., Melesse, D.Y., Kayeyi, N., Manu, A., Wado, Y.D., Barros, A., *et al.* (2021) HIV Trends and Disparities by Gender and Urban-Rural Residence among Adolescents in Sub-Saharan Africa. *Reproductive Health*, **18**, Article No. 120. <https://doi.org/10.1186/s12978-021-01118-7>
- [8] Bulstra, C.A., Hontelez, J.A.C., Giardina, F., Steen, R., Nagelkerke, N.J.D., Bärnighausen, T., *et al.* (2020) Mapping and Characterising Areas with High Levels of HIV Transmission in Sub-Saharan Africa: A Geospatial Analysis of National Survey Data. *PLOS Medicine*, **17**, e1003042. <https://doi.org/10.1371/journal.pmed.1003042>
- [9] Jahagirdar, D., Walters, M.K., Novotney, A., Brewer, E.D., Frank, T.D., Carter, A., *et al.* (2021) Global, Regional, and National Sex-Specific Burden and Control of the HIV Epidemic, 1990-2019, for 204 Countries and Territories: The Global Burden of Diseases Study 2019. *The Lancet HIV*, **8**, e633-e651. [https://doi.org/10.1016/s2352-3018\(21\)00152-1](https://doi.org/10.1016/s2352-3018(21)00152-1)
- [10] Dambach, P., Mahenge, B., Mashasi, I., Muya, A., Barnhart, D.A., Bärnighausen, T.W., *et al.* (2020) Socio-Demographic Characteristics and Risk Factors for HIV Transmission in Female Bar Workers in Sub-Saharan Africa: A Systematic Literature Review. *BMC Public Health*, **20**, Article No. 697. <https://doi.org/10.1186/s12889-020-08838-8>
- [11] Dias, B.R.L., Rodrigues, T.B., Botelho, E.P., de Oliveira, M.F.V., Feijão, A.R. and Polaro, S.H.I. (2021) Integrative Review on the Incidence of HIV Infection and Its Socio-Spatial Determinants. *Revista Brasileira de Enfermagem*, **74**, e20200905. <https://doi.org/10.1590/0034-7167-2020-0905>
- [12] Gu, L.Y., Zhang, N., Mayer, K.H., McMahon, J.M., Nam, S., Conserve, D.F., *et al.* (2021) Autonomy-Supportive Healthcare Climate and HIV-Related Stigma Predict Linkage to HIV Care in Men Who Have Sex with Men in Ghana, West Africa. *Journal of the International Association of Providers of AIDS Care*, **20**, 1-8.

- <https://doi.org/10.1177/2325958220978113>
- [13] Adam, A., Fusheini, A., Ayanore, M.A., Amuna, N., Agbozo, F., Kugbey, N., *et al.* (2021) HIV Stigma and Status Disclosure in Three Municipalities in Ghana. *Annals of Global Health*, **87**, 49. <https://doi.org/10.5334/aogh.3120>
- [14] Mannoh, I., Amundsen, D., Turpin, G., Lyons, C.E., Viswasam, N., Hahn, E., *et al.* (2021) A Systematic Review of HIV Testing Implementation Strategies in Sub-Saharan African Countries. *AIDS and Behavior*, **26**, 1660-1671. <https://doi.org/10.1007/s10461-021-03518-z>
- [15] Adetokunboh, O.O. and Are, E.B. (2024) Spatial Distribution and Determinants of HIV High Burden in the Southern African Sub-Region. *PLOS ONE*, **19**, e0301850. <https://doi.org/10.1371/journal.pone.0301850>
- [16] Ibrahim, N.K. and Asghari, A. (2016) Ecological Fallacy and Individual Risk. *Frontiers in Public Health*, **4**, Article No. 46.
- [17] Sevugu, J.T. (2021) Socio-Demographic, Behavioural and Biomedical Factors Contributing to HIV Spread among Adults in Ghana—A Case-Control Study. *Texila International Journal of Academic Research*, **8**, 64-79. <https://doi.org/10.21522/tijar.2014.08.03.art006>
- [18] Koomson, F. and Teye-Kwadjo, E. (2020) How Much Do We Really Know about Sociosexuality in Ghana? *Sexuality & Culture*, **25**, 167-188. <https://doi.org/10.1007/s12119-020-09764-y>
- [19] Giguère, K., Eaton, J.W., Marsh, K., Johnson, L.F., Johnson, C.C., Ehui, E., *et al.* (2020) Trends in Knowledge of HIV Status and Efficiency of HIV Testing Services in Sub-Saharan Africa, 2000-20: A Modeling Study Using Survey and HIV Testing Programme Data. *The Lancet HIV*, **8**, e284-e293. <https://doi.org/10.1101/2020.10.20.20216283>
- [20] Wand, H., Morris, N. and Reddy, T. (2021) Temporal and Spatial Monitoring of HIV Prevalence and Incidence Rates Using Geospatial Models: Results from South African Women. *Spatial and Spatio-Temporal Epidemiology*, **37**, Article ID: 100413. <https://doi.org/10.1016/j.sste.2021.100413>
- [21] Sambah, F., Baatiema, L., Appiah, F., Ameyaw, E.K., Budu, E., Ahinkorah, B.O., *et al.* (2020) Educational Attainment and HIV Testing and Counselling Service Utilisation during Antenatal Care in Ghana: Analysis of Demographic and Health Surveys. *PLOS ONE*, **15**, e0227576. <https://doi.org/10.1371/journal.pone.0227576>
- [22] Owusu, A.Y. (2020) A Gendered Analysis of Living with HIV/AIDS in the Eastern Region of Ghana. *BMC Public Health*, **20**, Article No. 1114. <https://doi.org/10.1186/s12889-020-08702-9>
-

List of Abbreviations

AIC	Akaike Information Criterion
AIDS	Acquired Immunodeficiency Syndrome
AI	Artificial Intelligence
API	Application Programming Interface
ART	Antiretroviral Therapy
CI	Confidence Interval
CSS	Cascading Style Sheets
CSV	Comma-Separated Values
CV	Cross-Validation
DAG	Directed Acyclic Graph
DHS	Demographic and Health Survey
DREAM	Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe
GAC	Ghana AIDS Commission
GHS	Ghana Health Service
GIS	Geographic Information System
GridSearchCV	Grid Search with Cross-Validation
GSS	Ghana Statistical Service
HDX	Humanitarian Data Exchange
HIV	Human Immunodeficiency Virus
HTML	HyperText Markup Language
KNN	K-Nearest Neighbors
LISA	Local Indicators of Spatial Association
LMKM	Lower Manya Krobo Municipality
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAR	Missing At Random
ML	Machine Learning
MoH	Ministry of Health

MSM	Men who have Sex with Men
OR	Odds Ratio
ORCID	Open Researcher and Contributor ID
PCA	Principal Component Analysis
PDP	Partial Dependence Plot
PrEP:	Pre-Exposure Prophylaxis
PySAL	Python Spatial Analysis Library
QGIS	Quantum Geographic Information System
R ²	Coefficient of Determination
RMSE	Root Mean Square Error
SDG	Sustainable Development Goal
SHAP	SHapley Additive explanations
SHAP-ML	SHAP-based Machine Learning
SSA	Sub-Saharan Africa
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SVR	Support Vector Regressor
SVM	Support Vector Machine
TB	Tuberculosis
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization
Zenodo	Open-access data repository platform