Interpretable Machine Learning for Child Stunting and Wasting Risk: A Cross-National Evidence Base for Programme Monitoring and Targeting

Craig Carlos Ouma*[1]

[1]Independent Researcher

August 10, 2026

Abstract

Programme teams that monitor child nutrition outcomes typically rely on periodic household surveys that arrive months after data collection and are reported at a level too coarse to guide week-to-week targeting decisions. This paper asks a narrower, programme-relevant question: given the country-level structural and survey-design information already available to a monitoring team at the time a survey is planned or released, can a model flag which country-survey observations are likely to fall in the World Health Organization's high or very-high public-health-significance category for child stunting, and which factors drive that classification. Using 880 country-survey observations covering 150 countries between 1983 and 2019, compiled from national MICS, DHS, and SMART surveys and joined with World Bank Open Data covariates, an XGBoost classifier separates high-burden (stunting prevalence 20 % or above) from lower-burden observations with an AUROC of 0.948 and an AUPRC of 0.975. A SHAP (SHapley Additive exPlanations) analysis of the model identifies a composite WASH (water, sanitation, and hygiene) access indicator, GDP per capita, and under-5 population size as the three dominant drivers, with income classification and a female-literacy proxy for maternal education contributing a secondary but consistent signal. A parallel model for high-burden wasting (prevalence 15 % or above) achieves an AUROC of 0.774, reflecting wasting's more volatile, seasonally-driven character relative to the more structural determinants of stunting. The paper applies the USAID Data Quality Assessment framework (Validity, Integrity, Precision, Reliability, Timeliness) to the panel used here, and translates each SHAP-identified driver into a candidate output or outcome indicator that a programme monitoring system could track. Because no individual or household-level dataset cleared a pre-specified data quality checklist without requiring gated access to microdata, the analysis is scoped honestly as a country-level driver and burden-classification study rather than an individual-child determinants analysis, and the paper states this scope limitation directly rather than simulating individual records to avoid it.

1 Introduction

A programme team responsible for monitoring child nutrition outcomes across multiple countries or regions faces a structural information problem before it faces an analytical one. Nationally

*Correspondence: craigcarlos95@gmail.com

representative anthropometric data comes from household surveys such as the Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), or SMART surveys, each of which takes months to field, clean, and publish. By the time a stunting prevalence estimate is released, the underlying conditions it describes may already have shifted. A programme officer deciding where to direct scarce monitoring visits, or which countries warrant closer review before the next reporting cycle, often has to make that call using whatever structural information is available at the moment the survey lands: the country's income classification, its recent GDP trajectory, WASH access statistics that update on a different cycle than the anthropometric survey itself, and the demographic scale of the population at risk.

This paper builds a model that uses exactly that kind of structural information, available at or before the point a stunting survey is released, to classify whether the resulting observation is likely to fall in the World Health Organization's high or very-high burden category, and it uses SHAP to make the basis for that classification transparent. The goal is not to replace the survey itself: nothing here substitutes for direct anthropometric measurement. The goal is to give a monitoring system a triage layer that flags which country-survey observations warrant prioritized review, and a defensible, auditable account of why, before the full survey report is in hand.

The paper is deliberately scoped at the country-survey level rather than the individual child. Section 3 explains why: no individual or household-level dataset that met a pre-specified data quality checklist was available without going through a gated, multi-day microdata request process. Rather than simulate individual records to manufacture a richer-looking analysis, the paper uses the real, ungated panel that was actually available, and states the resulting scope limitation directly. Section 2 places this analysis within a standard theory-of-change framework so its role in a monitoring system is explicit rather than implied.

2 Background: Theory of Change

Monitoring and evaluation practice distinguishes inputs, activities, outputs, outcomes, and impact along a programme's causal chain:

$$\text{Inputs} \rightarrow \text{Activities} \rightarrow \text{Outputs} \rightarrow \text{Outcomes} \rightarrow \text{Impact}$$

Figure 1 places this analysis on that chain. A nutrition programme's **outputs** include the households and children actually reached and measured, and the indicator datasets that measurement produces. Its **outcomes** include a measurable reduction in stunting or wasting prevalence within the target population. The model in this paper sits between the two: it is a targeting and screening layer that uses output-adjacent data (who was surveyed, under what structural conditions) to anticipate outcome-level classification (is this a high-burden observation) before the outcome itself is fully processed and reported. In results-framework language, this model strengthens output-to-outcome measurement rather than measuring impact directly.

It is worth being precise about which variables in the model are output indicators and which are outcome-adjacent. Survey sample size and the year a survey was conducted are output-side variables: they describe how the monitoring activity itself was carried out. GDP per capita, WASH access, income classification, and the female-literacy proxy for maternal education are structural covariates that sit upstream of both outputs and outcomes; they describe the conditions a programme is operating in rather than anything the programme itself produced. The stunting and wasting burden classifications are the outcome indicators the model is trying to anticipate. Keeping this distinction explicit matters for programme use: a model that scores well by leaning on output-side variables (for instance, a spurious correlation between which countries get surveyed

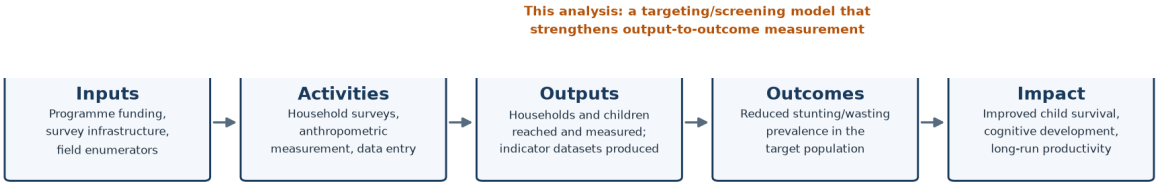

Figure 1: Theory of change. This analysis is a targeting and screening layer positioned between programme outputs (households and children reached and measured) and outcomes (reduced stunting and wasting prevalence), strengthening output-to-outcome measurement rather than measuring downstream impact directly.

more often and which are lower-burden) would be diagnosing something about survey coverage, not about nutrition programming. Section 6 returns to this concern directly when interpreting the SHAP results.

3 Data

3.1 Dataset selection

A determinants analysis of child stunting is best done with individual or household-level records: age, sex, birth order, breastfeeding status, maternal education, antenatal care attendance, household wealth, and WASH access, joined to an anthropometric outcome for the same child. A prespecified checklist required any such dataset to contain individual or household-level records, at least one anthropometric outcome, at least five of the standard covariates used in the published determinants literature, at least 1,000 records, and to be downloadable without a gated multi-day approval process, with clearly documented provenance.

Live searches against this checklist, run across the standard DHS-derived covariate patterns for Ethiopia, Pakistan, Bangladesh, India, Indonesia, and Kenya, surfaced several candidate datasets, none of which cleared the checklist. Pure-anthropometry datasets contained age, sex, height, weight, and a derived nutrition-status label, but none of the maternal or household determinants the analysis needed. A cluster-level dataset built for a geospatial poverty-prediction study contained rich satellite-derived covariates but no maternal or household variables and a different unit of analysis (the survey enumeration area, not the child or household). The individual and household-level microdata that would satisfy the checklist in full exists inside the DHS Program's own repository, but only behind a registration request with a review lag, which the checklist explicitly excludes. The full account of what was tried and why each candidate failed is recorded in `DECISIONS.md` in the project repository.

Rather than fabricate individual child records to sidestep this outcome, the analysis uses the dataset the brief itself names as the appropriate fallback: a real, ungated country-survey panel, reframed honestly as a country-level driver and burden-classification study rather than an individual-child determinants analysis.

3.2 Sources and coverage

The outcome data is `malnutrition-across-the-globe`, a public Kaggle compilation of 924 country-survey observations spanning 152 countries and the years 1983 to 2019, drawn from national MICS, DHS, SMART, and other government nutrition surveys and cross-referenced against UNICEF/WHO/World Bank Joint Malnutrition Estimates. Each row is one survey round for one country and reports stunting, wasting, severe wasting, overweight, and underweight prevalence, alongside income classification, Least Developed Country (LDC) status, landlocked or small-island developing state status, survey sample size, and under-5 population. Every field in this panel is a pre-aggregated, country-survey-level statistic; no individual child or household record, and no personal identifier of any kind, appears anywhere in the data used for this analysis.

This panel was joined to four World Bank Open Data indicators by country and nearest available survey year, matched within a five-year window: GDP per capita (current US$), basic drinking water access (% of population), basic sanitation access (% of population), and adult female literacy rate (% of women aged 15 and above), used here as a country-level proxy for maternal education, since individual maternal education is unavailable at this level of aggregation.

3.3 Cleaning and honest exclusions

Thirty-seven of the original 924 rows had no stunting estimate and were dropped, since stunting is the primary outcome. Of the remaining 887 rows, 7 had no GDP-per-capita match within the five-year join window and were dropped, since GDP per capita has no defensible proxy. Basic water access, basic sanitation access, and female literacy were missing for 149, 149, and 169 rows respectively, almost entirely because the World Bank's underlying JMP and UIS series do not extend before roughly 2000 for many countries; these were imputed with the median value for that country's income classification rather than dropped, to avoid systematically removing older, typically higher-burden survey rounds.

The `Survey Sample (N)` field required a further correction. After stripping thousands-separator commas, several values exceeded 5 million, implausible for a single national nutrition survey and consistent with digit-grouping or transcription artifacts in the underlying source field rather than genuine sample sizes. Values above 200,000 (77 rows) were treated as invalid, and, combined with 59 rows genuinely missing, 136 values in total were imputed with the panel median of 4,421. This mirrors the kind of sentinel-value cleanup that real secondary data regularly requires, and is exactly the sort of issue the Precision and Integrity dimensions in Section 7 are built to catch.

The final analysis panel contains 880 country-survey observations across 150 countries, 1983 to 2019. Figure 2 summarises the class balance of the primary outcome and the missingness that the cleaning steps above addressed. Because the underlying surveys are disproportionately fielded in lower- and middle-income countries where stunting is a public health priority, 589 of the 880 observations (67 %) fall in the high or very-high burden category; this is a genuine feature of where nutrition surveys get conducted, not a representative sample of global stunting prevalence, and is treated explicitly as a Validity limitation in Section 7.

4 Methods

4.1 Feature engineering

Following the determinants literature, three covariates were combined or transformed rather than used raw. A WASH access composite averages basic water and basic sanitation access, since the two move together and both capture exposure to the same faecal-oral disease pathway implicated

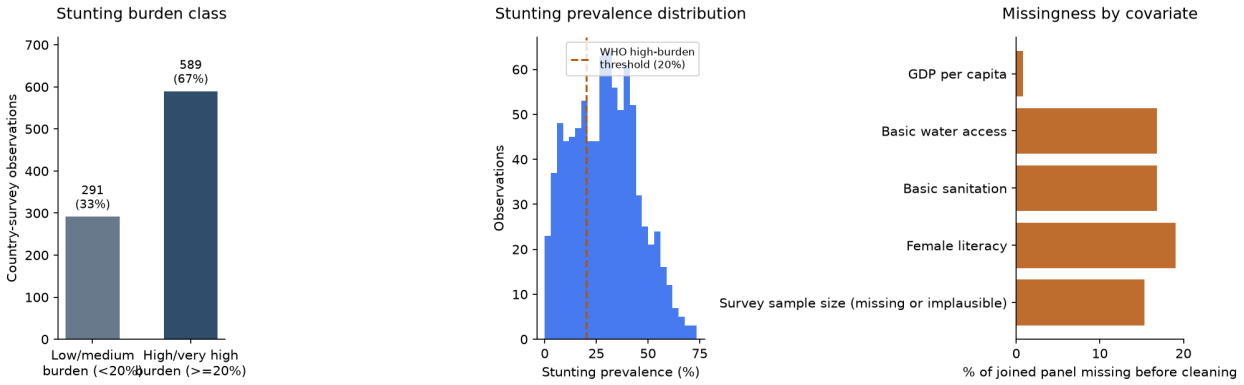

Figure 2: Class balance for the primary outcome (left), the underlying stunting prevalence distribution against the WHO high-burden threshold (centre), and missingness by covariate before cleaning (right).

in stunting. Income classification, already ordinally coded by the World Bank from low to high income, serves the same analytical role a household wealth quintile plays in an individual-level study. Female literacy was binned into quartiles to create a maternal-education proxy consistent with how maternal education is typically encoded as an ordinal variable in the DHS-based literature, rather than used as a raw continuous percentage. GDP per capita and under-5 population were log-transformed to address their strong right skew. The final feature set contains nine variables: the WASH access composite, log GDP per capita, log under-5 population, survey year, income classification, the maternal-education proxy, log survey sample size, LDC status, and landlocked or small-island status.

4.2 Target definition

Rather than a continuous prevalence regression, the model classifies a binary outcome: whether a country-survey observation falls in the WHO's high or very-high public-health-significance category for stunting, defined as prevalence at or above 20 %, following the thresholds set out in de Onis et al. [4] and used in the UNICEF/WHO/World Bank Joint Malnutrition Estimates. This threshold was chosen over a continuous regression because a monitoring system's actual decision is rarely "predict the exact prevalence"; it is closer to "does this observation cross the line that would trigger a closer look or a resource reallocation." A parallel target for high-burden wasting uses the WHO's 15 % threshold for that indicator.

4.3 Models

XGBoost is the primary classifier, trained with `scale_pos_weight` set to the training set's class imbalance ratio. LightGBM is trained as a secondary comparison model, consistent with the modelling approach used in the prior credit-scoring paper in this portfolio. Both models were evaluated on a stratified 20 % holdout (176 observations). A third XGBoost model, trained identically but against the wasting high-burden target, provides a secondary comparison across indicators rather than across model families.

4.4 SHAP analysis

SHAP values were computed using the TreeExplainer [1], exact for tree-based models, on the primary XGBoost stunting classifier's holdout predictions. The analysis reports mean absolute SHAP as a global importance ranking, a beeswarm plot showing the direction of each feature's effect, and dependence plots for the three highest-ranked features.

5 Results

5.1 Model performance

Table 1 reports holdout performance for both models on the primary stunting target and for the secondary wasting model. XGBoost and LightGBM perform almost identically on the stunting task, with XGBoost marginally ahead on AUROC and AUPRC and LightGBM achieving slightly higher sensitivity at the default 0.5 probability threshold. Both separate high-burden from lower-burden observations cleanly: an AUROC above 0.94 at this level of aggregation reflects the fact that a country's structural conditions (income, WASH infrastructure, population scale) are strong, near-deterministic correlates of whether its stunting prevalence crosses 20 %, in a way that individual-level determinants studies, which model much noisier child-to-child variation, typically do not achieve. This performance gap between ecological and individual-level prediction is itself a finding worth stating plainly: a model this accurate at the country level says relatively little about which specific children within a high-burden country are most at risk, and Section 9 returns to why that distinction matters for how this model should and should not be used.

Table 1: Holdout performance. The stunting task classifies observations at or above the WHO's 20 % high-burden threshold (176 holdout observations, 67 % high-burden); the wasting task uses the 15 % threshold on a separately stratified holdout. Sensitivity and specificity are reported at a 0.5 probability threshold.

Model (target)	AUROC	AUPRC	Sensitivity	Specificity
XGBoost (stunting)	**0.948**	**0.975**	0.873	0.879
LightGBM (stunting)	0.947	0.974	**0.907**	0.879
XGBoost (wasting, secondary)	0.774	0.550	—	—

At the default threshold, the stunting model correctly identifies 87 % of genuinely high-burden observations while correctly clearing 88 % of lower-burden ones. A monitoring team could move this threshold lower to catch a larger share of true high-burden cases at the cost of more false alarms, or raise it if review capacity is scarce and false alarms are costly; the AUPRC of 0.975 indicates the model retains high precision across most of that recall range, so this is a genuine operating choice rather than a forced trade-off. Figure 3 shows the full ROC and precision-recall curves for both stunting models.

The wasting model's lower AUROC (0.774) against the same feature set is itself informative rather than a modelling failure. Wasting is driven substantially by acute, short-term shocks (seasonal food insecurity, disease outbreaks, localized conflict) that the mostly slow-moving structural covariates used here cannot capture, whereas stunting reflects chronic, cumulative deprivation that tracks closely with the same structural conditions year over year. The gap between the two models is consistent with that distinction and is discussed further in Section 6.

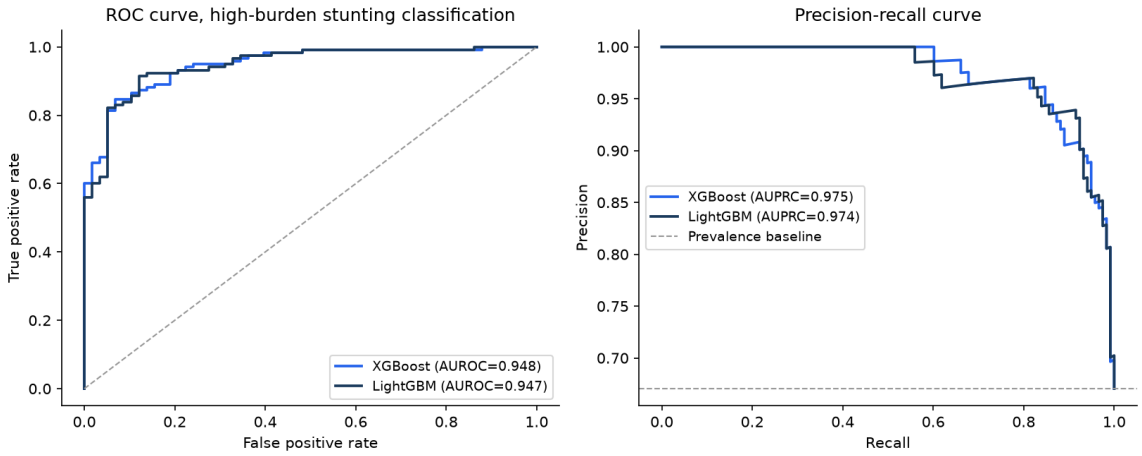

Figure 3: ROC (left) and precision-recall (right) curves for the primary stunting classification task, holdout set. Both models trace nearly identical curves.

5.2 SHAP feature importance

Figure 4 and Figure 5 show the global SHAP ranking for the primary stunting model. The WASH access composite dominates, with a mean absolute SHAP value of 2.122, more than double the second-ranked feature. Log GDP per capita follows at 0.983, then log under-5 population at 0.659, survey year at 0.533, income classification at 0.480, and the maternal-education proxy at 0.382. Survey sample size, landlocked or small-island status, and LDC status contribute comparatively little once the higher-ranked features are accounted for.

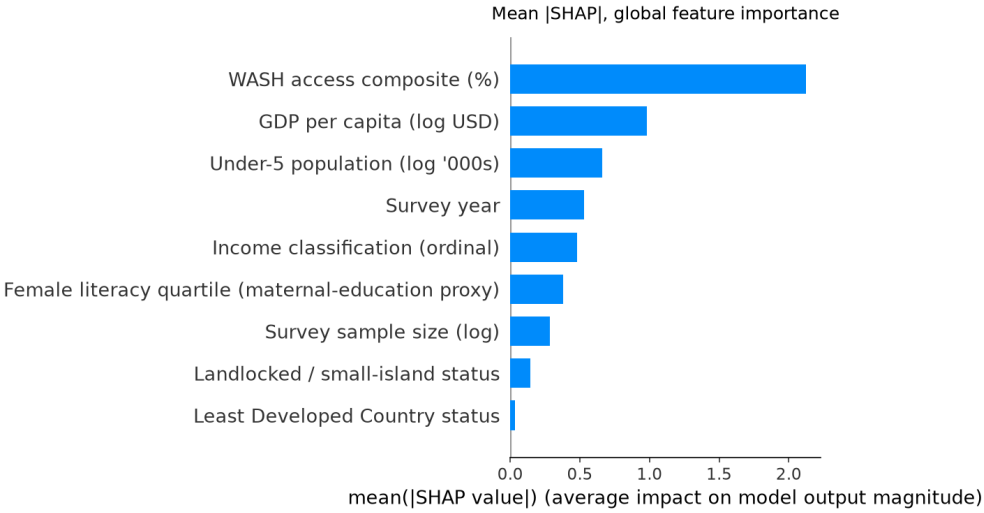

Figure 4: Global feature importance measured by mean absolute SHAP value. The WASH access composite dominates, consistent with WASH access being both a direct exposure pathway and a strong proxy for broader public-health infrastructure.

The beeswarm plot tells a coherent, programme-legible story. Low WASH access (blue, left side of the WASH row) pushes predictions firmly toward high burden, while high WASH access pushes firmly the other way, with a visible threshold-like clustering rather than a smooth gradient.

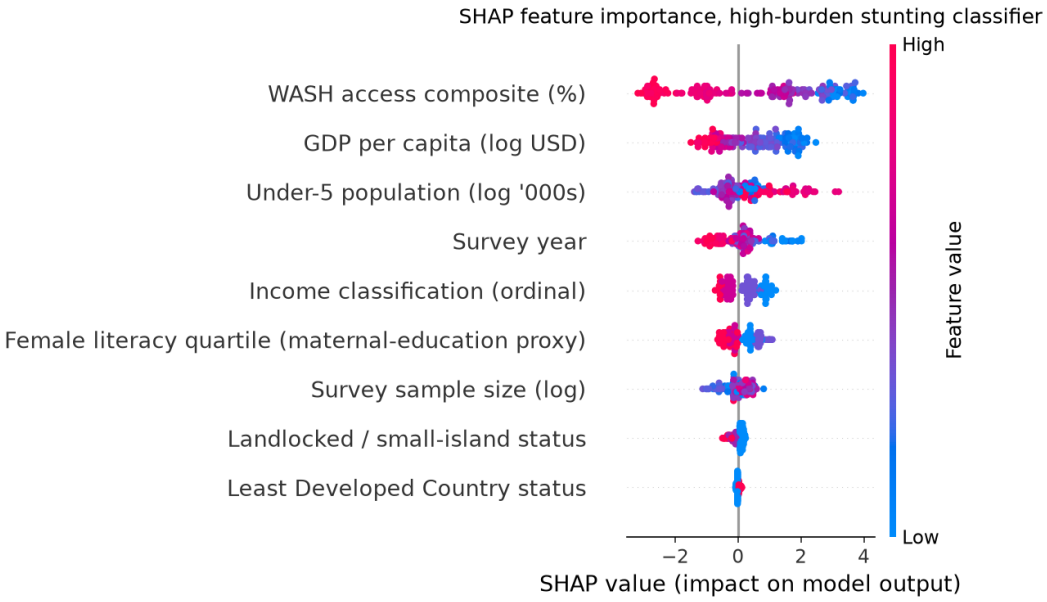

Figure 5: SHAP beeswarm plot. Each point is one holdout country-survey observation. Colour indicates feature value (pink = high, blue = low); horizontal position shows the direction and magnitude of that observation's push toward or away from a high-burden classification.

GDP per capita and under-5 population show the same directional pattern: poorer, more populous countries are pushed toward high-burden classification, consistent with both established nutrition literature [5] and with a simpler structural reality, that WASH infrastructure investment tracks national income closely across this period.

5.3 SHAP dependence plots

Figure 6 examines the three top-ranked features in detail. WASH access shows a genuinely threshold-like relationship: SHAP values stay strongly positive (pushing toward high-burden) until access crosses approximately 60 %, then fall sharply and turn negative above roughly 80 %. This is a sharper transition than a simple linear relationship would predict, and it lines up with WASH programming practice, where access below roughly two-thirds of a population is generally treated as a public-health priority threshold rather than a marginal concern. GDP per capita shows a smoother but still clearly monotonic decline in SHAP value as income rises, flattening out above roughly USD 3,000 per capita, log-scaled. Under-5 population shows a less clean pattern: very large under-5 populations (strongly correlated with a small number of populous low- and middle-income countries) carry a consistently positive SHAP value, but the relationship is noisier at moderate population sizes, where the model is evidently leaning on the other features more heavily.

6 Interpreting the Drivers as Programme Indicators

The value of this analysis to a monitoring team is not the AUROC figure; it is the translation from a statistical driver into something a programme can actually track and act on. Each of the three dominant SHAP drivers maps onto an indicator category a results framework would already recognise.

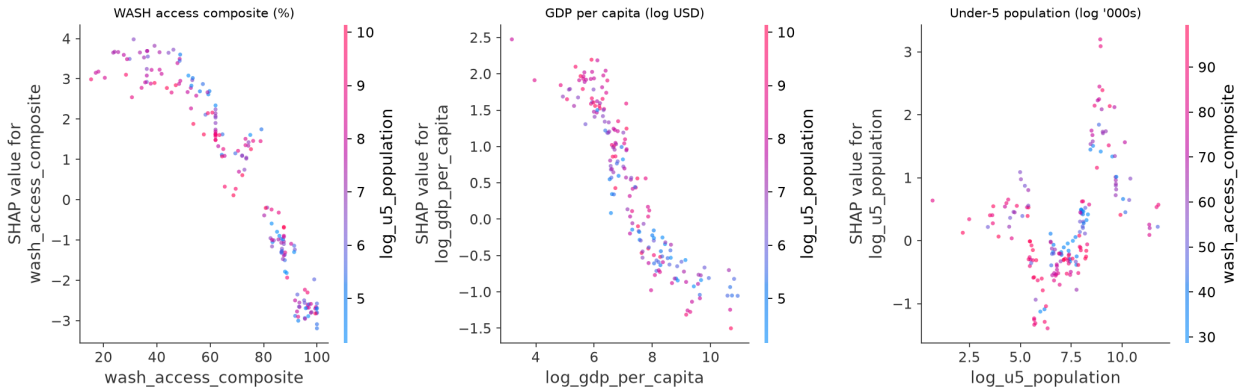

Figure 6: SHAP dependence plots for the three highest-ranked features. WASH access (left) shows a threshold-like transition around 60-80 % access; GDP per capita (centre) shows a smooth monotonic decline that flattens above roughly USD 3,000; under-5 population (right) shows a consistently positive effect at the largest population sizes with more noise at moderate sizes.

The WASH access composite is, in results-framework terms, closest to an **output indicator**: the percentage of a target population with access to basic water and sanitation is something a WASH programme produces directly through infrastructure investment, and it is already routinely tracked by national statistics offices and WASH-sector programmes alike. Given how sharply this analysis finds WASH access separating high- from lower-burden observations, a monitoring system with a stunting-reduction objective should treat WASH access coverage, tracked at a sub-national resolution where possible, as a leading indicator worth reviewing on a shorter cycle than the anthropometric survey itself, rather than waiting for the full nutrition survey to confirm what the WASH data may already be signalling.

GDP per capita and under-5 population, by contrast, function as **context indicators** in a theory of change: they describe the environment a programme operates in, not something a programme's activities change directly. Their role in a monitoring system is not targeting within a single country, but flagging, across a portfolio of countries, which contexts are structurally predisposed toward higher risk and therefore warrant more intensive monitoring investment relative to lower-risk contexts. Income classification and the female-literacy maternal-education proxy play a similar, secondary role: income classification changes slowly and rarely provides new targeting information within a single reporting cycle, while female literacy, as a national aggregate, is a coarse stand-in for maternal education that a household-level survey would measure far more precisely. Both are worth retaining as contextual features in a screening model, but neither is something a programme can influence on the timescale that matters for the **outcome indicator** this analysis is ultimately trying to anticipate, the stunting or wasting prevalence itself.

This distinction matters most where the model could otherwise mislead. A programme team looking at this analysis and concluding that raising GDP per capita is the intervention would be reading a structural correlate as an actionable lever, when the actual actionable lever the analysis points toward is WASH access, an output a programme can move directly and that this model finds carries the largest share of the predictive signal.

7 Data Quality Assessment

The USAID Data Quality Assessment framework evaluates data along five dimensions: Validity, Integrity, Precision, Reliability, and Timeliness (VIPTR) [7]. Applying it to the panel used here surfaces exactly the kind of concerns a programme monitoring specialist would be expected to flag before trusting a model built on it.

Validity. The panel is not a representative sample of all 195 countries; it is a compilation of countries where a MICS, DHS, or SMART survey happened to be fielded, which skews toward lower- and middle-income countries where child nutrition is already a monitoring priority. The 67 % high-burden rate in this dataset should not be read as a global stunting rate; it reflects where surveys get conducted, not where stunting occurs. Female literacy, used here as a maternal-education proxy, is also a validity compromise: it measures a national aggregate that stands in for an individual-level construct the country-level data cannot capture directly.

Integrity. The `Survey Sample (N)` field's implausible multi-million-child values, described in Section 3, are a direct integrity issue in the underlying source data, most likely introduced during data entry or format conversion rather than reflecting the survey design itself. Any monitoring system ingesting this kind of compiled, multi-source panel needs an automated range check on fields like this one before they reach an analyst, rather than relying on a one-off manual review as was done here.

Precision. Anthropometric measurement in field conditions carries known measurement error, particularly for height-for-age in very young children, where small measurement inaccuracies shift a child across the stunting threshold. This paper cannot assess that source of imprecision directly, since it works with already-aggregated prevalence estimates rather than raw anthropometric measurements, but any programme relying on individual-level anthropometric data downstream of this kind of country-level screening model should budget for enumerator training and measurement standardisation as a precision safeguard.

Reliability. The underlying survey methodologies (MICS, DHS, SMART, and national surveys) differ in sampling design, questionnaire structure, and implementing agency, and the compiled panel used here does not fully harmonise across them. A stunting estimate from a SMART survey and one from a full DHS round are not perfectly comparable measurement instruments, even when both report the same underlying indicator.

Timeliness. This is the most consequential dimension for how this model should be used. Survey-based nutrition data is fundamentally a lagging indicator: the newest observation in this panel is from 2019, and even a freshly published survey describes conditions that existed months before release. The value proposition of the model in this paper is precisely that it uses more timely, structurally available covariates (GDP trajectories, WASH access statistics, demographic data) to anticipate what the lagging survey indicator will eventually show, which is a Timeliness mitigation, not a Timeliness solution: the model's own covariates update on cycles measured in months to years, not days or weeks.

8 Implications for Programme Monitoring Systems

Four concrete steps follow from this analysis for a monitoring system built on the tools commonly used in this space: SurveyCTO for field data collection, Salesforce or a similar system for indicator tracking, and Tableau or Power BI for dashboard reporting.

First, treat WASH access coverage as a leading indicator with its own dashboard panel, updated on whatever cycle the underlying WASH-sector data allows, rather than folding it into an annual

nutrition survey review cycle. Given how strongly this analysis finds WASH access separating high- from lower-burden observations, a sustained decline in WASH access coverage in a given country or region is a legitimate early-warning signal worth flagging before the next full nutrition survey confirms it.

Second, build the burden-classification model itself into a screening layer inside the monitoring dashboard, scored against whatever structural covariates are already being tracked for other purposes, so that a country or region can be flagged for closer review as soon as its GDP, WASH, or demographic profile shifts, rather than only after the next survey round lands.

Third, apply an automated range check to survey design fields such as sample size before they enter any downstream indicator system; the transcription artifact found in the `Survey Sample (N)` field here is exactly the kind of integrity issue that a lightweight validation rule would catch immediately rather than requiring a manual data-cleaning pass.

Fourth, pair every dashboard-level output indicator (WASH access, survey coverage, sample size) with an explicit note on which outcome indicator it is meant to anticipate, and over what expected lag. The VIPTR review in Section 7 found that Timeliness, not model accuracy, is the binding constraint on how useful this kind of model can be in practice, and a dashboard that is explicit about expected lag helps a programme officer calibrate how much weight to place on a screening flag before the underlying survey confirms it.

9 Limitations

The central limitation of this paper is the one it states openly rather than engineers around: it analyses country-survey observations, not individual children, because no individual or household-level dataset cleared a pre-specified quality checklist without gated access. A model that separates high- from lower-burden countries with an AUROC above 0.94 says very little about which children within a high-burden country are most at risk; individual-level targeting still requires individual-level data, and nothing in this paper substitutes for that. A second, related limitation is temporal: the panel spans 1983 to 2019 with uneven coverage across that range, and a model trained on this full window may weight structural relationships that have shifted since the most recent observations. Third, the World Bank covariates were joined by nearest available year within a five-year window rather than the exact survey year in every case, which introduces a modest, documented timing mismatch on top of the underlying survey timing lag discussed in Section 7. Finally, the strong performance of the stunting model relative to the wasting model is itself a limitation worth naming directly: it suggests this modelling approach, built on slow-moving structural covariates, is better suited to chronic, cumulative outcomes like stunting than to acute, shock-driven outcomes like wasting, and a monitoring system that needs to anticipate acute nutritional crises would need a different set of faster-moving covariates than the ones used here.

Reproducibility Statement

The full data pipeline, from the live dataset search log through feature engineering, model training, and SHAP analysis, is available as a single executable notebook, along with the trained model and the exact figures used in this paper, at `https://github.com/craigouma/child-nutrition-shap`.

References

[1] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[3] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[4] M. de Onis, E. Borghi, M. Arimond, P. Webb, T. Croft, K. Saha, M. De-Regil, D. Thuita, C. Heidkamp, A. Krasevec, S. Hayashi, and J. Flores-Ayala, "Prevalence thresholds for wasting, overweight and stunting in children under 5 years," *Public Health Nutrition*, vol. 22, no. 1, pp. 175–179, 2019.

[5] R. E. Black, C. G. Victora, S. P. Walker, Z. A. Bhutta, P. Christian, M. de Onis, M. Ezzati, S. Grantham-McGregor, J. Katz, R. Martorell, and R. Uauy, "Maternal and child undernutrition and overweight in low-income and middle-income countries," *The Lancet*, vol. 382, no. 9890, pp. 427–451, 2013.

[6] B. J. Akombi, K. E. Agho, J. J. Hall, D. Wali, A. Renzaho, and A. Merom, "Stunting, wasting and underweight in sub-Saharan Africa: A systematic review," *International Journal of Environmental Research and Public Health*, vol. 14, no. 8, p. 863, 2017.

[7] United States Agency for International Development, *Data Quality Assessment: Tool for USAID-Supported Sources*, ADS Chapter 201 Additional Help, USAID, 2018.

[8] World Bank, *World Development Indicators*, World Bank Open Data, 2024, `https://data.worldbank.org`.

[9] W. K. Kellogg Foundation, *Logic Model Development Guide*, W.K. Kellogg Foundation, 2004.