Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Lifesaving Maternal Commodity Stock outs in Ethiopian Primary Healthcare Facilities Using Supervised Machine Learning Classifiers: Evidence from EDHS 2024/2025

Yenehun Demelash Gelaw

yenehun6101@gmail.com

Teda Health Science College

Ashebir Mamay Gebiru

Teda Health Science College

Article

Keywords: EDHS 2024/2025, Maternal Health Commodities, Stock out Prediction, Machine Learning, XGBoost, Decision Curve Analysis, SHAP, Ethiopia

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11221251/v1

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

Additional Declarations: No competing interests reported.

Predicting Lifesaving Maternal Commodity Stock outs in Ethiopian Primary Healthcare Facilities Using Supervised Machine Learning Classifiers: Evidence from EDHS 2024/2025

Authors and affiliations

Yenehun Demelash Gelaw[1*], Ashebir Mamay Gebiru[2]

[1]Department of Pharmacy, Teda Health Science College, Gondar, Ethiopia

[2]Department of biostatistics and epidemiology, Teda Health Science College, Gondar, Ethiopia

Corresponding author*

Yenehun Demelash Gelaw

Email: yenehun6101@gmail.com

Abstract

Background; Obstetric hemorrhage and hypertensive disorders drive Ethiopia's high maternal mortality ratio (267/100000 live births), requiring continuous availability of lifesaving maternal commodities (LMCs). Conventional statistics offer retrospective insights rather than facility level risk forecasting. This study developed and validated machine learning classifiers to predict LMC stock outs in Ethiopian primary healthcare facilities using EDHS 2024/2025 data.

Methods; A secondary analysis evaluated 1245 primary facilities. The binary outcome measured physical absence, expiration or recent stock outs of any 5 WHO tracer commodities. Nineteen features spanning logistics, human resources and infrastructure were analyzed. Five algorithms Logistic Regression, SVC, MLP-ANN, Random Forest and XGBoost were implemented, handling class imbalance via SMOTE and evaluating performance via ROC-AUC, PR-AUC, Decision Curve Analysis (DCA), probability calibration and SHAP interpretability.

Results; Overall, 45.8% of facilities experienced stock outs. XGBoost outperformed other models on the test set, achieving a ROC-AUC of 0.894, Accuracy of 88.8%, Sensitivity of 89.5% and Specificity of 88.1%. It exhibited high reliability (Brier Score of 0.084) and superior net clinical benefit via DCA across thresholds from 10% to 85%. SHAP analysis identified non-functional cold chain equipment, lack of IPLS reporting, absence of APTS auditing, irregular bin card reconciliation and lack of professional pharmacy personnel as the top risk drivers.

Conclusion; Supervised machine learning algorithms notably XGBoost accurately forecast maternal commodity stock outs. Integrating this architecture into the national eLMIS enables proactive risk flagging, helping the Ministry of Health and EPSS prioritize cold chain upgrades, digital logistics and staffing in vulnerable facilities.

Keywords; EDHS 2024/2025, Maternal Health Commodities, Stock out Prediction, Machine Learning, XGBoost, Decision Curve Analysis, SHAP, Ethiopia

Introduction

Maternal mortality remains a persistent indicator of global healthcare inequality(1-3) . Approximately 287000 women die annually worldwide from preventable causes related to pregnancy and childbirth with Sub-Saharan Africa accounting for approximately 70% of these fatalities(1, 4, 5) . In Ethiopia, despite nationwide investments in primary health extension programs and health center expansions under consecutive Health Sector Transformation Plans (HSTP-I and HSTP-II), maternal mortality remains unacceptably high, with an estimated ratio of 267 maternal deaths per 100000 live births (1, 3).

The overwhelming majority of maternal deaths in Ethiopia are triggered by direct obstetric complications that can be medically managed or averted if timely emergency obstetric care is rendered(6, 7) . Postpartum hemorrhage (PPH) is the leading direct cause responsible for roughly one-third of all maternal deaths followed by severe pre-eclampsia and eclampsia, puerperal sepsis and unsafe abortion complications(7, 8) . The clinical efficacy of Emergency Obstetric and Newborn Care (EmONC) is inextricably linked to the continuous, physical availability of essential lifesaving maternal health commodities at the point of care(9, 10) :

Oxytocin Injection; The primary uterotonic agent for active management of the third stage of labor and first-line treatment for PPH(7, 10) .

Magnesium Sulfate Injection; The anticonvulsant gold standard for preventing and managing seizures in severe pre-eclampsia and eclampsia(10, 11) .

Parenteral Antibiotics (Ampicillin and Gentamicin); Indispensable therapeutic agents for arresting systemic puerperal infections(12, 13) .

Hydralazine / Labetalol Injection; Rapid acting parenteral antihypertensives required to avert cerebrovascular hemorrhage during hypertensive crises of pregnancy(10, 11) .

Misoprostol Tablets; A heat-stable prostaglandin E1 analogue providing vital uterotonic backup in settings where refrigeration for oxytocin is unavailable(7, 10) .

Recognizing the indispensable nature of these products, the World Health Organization (WHO) and the United Nations Commission on Life Saving Commodities categorized them as non-negotiable tracer items(13, 14) . An acute stock out of any one of these items at a primary health center transforms a manageable obstetric event into a fatal emergency within hours(6, 15) .

Public health facilities in Ethiopia obtain medical supplies through the Ethiopian Pharmaceutical Supply Service (EPSS) which operates a central repository in Addis Ababa and 19 regional branch hubs distributed throughout the country(13, 16) . Operational management of pharmaceuticals within healthcare facilities is structured around two key national initiatives(13, 16) :

Standardizes data collection and ordering through Report and Requisition Forms (RRF) and stock bin cards, shifting logistics from an arbitrary "push" system to a consumption-based "pull" mechanism(13, 16) .

A regulatory and financial accounting reform designed to establish transparent auditing, reduce medicine expiry, eliminate financial leakage and protect Revolving Drug Fund (RDF) capital(13, 16) .

Concurrently, the Ministry of Health has deployed electronic Logistics Management Information Systems (eLMIS) to digitize inventory reporting(13, 16) . The release of the EDHS 2024/2025 integrated assessment cycle provides high resolution data on facility readiness, infrastructure constraints and commodity security across the nation(1, 3) . Despite ongoing structural initiatives, primary healthcare facilities across Ethiopia continue to suffer from localized stock outs, storage degradation and stock imbalances(1, 13) .

These breakdowns stem from complex interacting facility level and geographic vulnerabilities(13) :

Primary health centers often lack continuous power grids, compromising cold chain storage required to preserve heat sensitive oxytocin(10, 17) .

Primary facility dispensaries are frequently managed by clinical nurses or health extension workers who lack formal supply chain training, leading to inaccurate RRF forecasting(10, 18) .

Regional distance disparities, transportation corridor disruptions and security instability in pastoralist or peripheral regions create extended resupply lead times(13, 19) .

Traditional scientific literature assessing pharmaceutical availability in low- and middle-income countries (LMICs) has almost universally relied on classical statistical regression modeling (e.g. multivariable binary logistic regression, multilevel mixed-effects models)(13, 20) . While classical regression is vital for explanatory inference estimating population averaged odds ratios

and identifying statistically significant risk factors it possesses fundamental operational limitations when applied to supply chain decision-making(13, 21) :

Classical regression models are designed to understand average past associations not to accurately classify whether an individual, unseen health facility will experience a stock out in the immediate future(13, 22) .

Standard parametric models struggle to capture non-linear relationships, high-dimensional collinearities, and complex feature interactions without extensive manual pre-specification(13, 23).

Informing a regional logistics officer that “rural location has an adjusted odds ratio of 2.1” offers general epidemiological insight but does not provide a practical early warning system that can automatically identify which specific primary health centers are at imminent risk of stocking out next month(13, 24) .

Supervised Machine Learning (ML) offers a paradigm shift from retrospective evaluation to proactive algorithmic forecasting(10, 25, 26) . Non-linear ensemble algorithms specifically gradient-boosted decision trees (XGBoost) and bagging ensembles (Random Forest) can automatically identify non-linear feature interactions and optimize classification performance(10, 27) .

Historically, the primary obstacle preventing machine learning deployment in healthcare operations has been the “black-box” dilemma the inability of human managers to interpret the mathematical rationale behind complex ensemble predictions(28) . Recent advances in Explainable Artificial Intelligence (XAI), specifically SHapley Additive exPlanations (SHAP) grounded in cooperative game theory, eliminate this barrier(29) . SHAP calculates the exact marginal contribution of each operational feature to an individual facility's stockout prediction, transforming advanced algorithms into transparent, auditable decision-support tools(30, 31) . Furthermore, rigorous clinical deployment requires evaluating Probability Calibration (reliability of risk scores) and Decision Curve Analysis (DCA) (quantifying net clinical benefit across threshold probabilities)(32) .

Methods and Materials

Study setting, design and data source

A secondary machine learning analysis was conducted using the integrated EDHS 2024/2025 Service Provision Assessment (SPA) National Health Facility Assessment dataset. The survey was executed collaboratively by the Ethiopian Public Health Institute (EPHI), the Central Statistical Agency (CSA) and the Federal Ministry of Health across all administrative divisions of Ethiopia; 11 regional states (Tigray, Afar, Amhara, Oromia, Somali, Benishangul-Gumuz, Central Ethiopia, South Ethiopia, South West Ethiopia, Sidama, Gambella) and two chartered city administrations (Addis Ababa, Dire Dawa).

The assessment utilized a stratified, multi-stage sampling frame covering hospitals, health centers and private/public clinics. For this study, the analytical cohort was restricted to primary healthcare facilities specifically public health centers and primary hospitals providing emergency delivery and maternal healthcare services yielding a final analytical sample of 1245 facilities.

Target Outcome Operationalization

The target dependent variable was Lifesaving Maternal Commodity Stock out Status, operationalized as a binary classification label ($Y \in (33)$);

Class 1 ($Y = 1$): Stockout / Suboptimal Availability. Defined as a primary healthcare facility exhibiting physical absence or expiration of any one of the five WHO tracer maternal health commodities on the day of survey, or having documented a complete stockout of any tracer item lasting > 7 consecutive days in the preceding 30 days.

Class 0 ($Y = 0$): Optimal Availability. Defined as a primary healthcare facility having all five WHO tracer maternal health commodities physically present, unexpired and observed on survey day with zero stock out episodes exceeding 7 days in the preceding month.

The five evaluated WHO tracer maternal commodities were; Oxytocin Injection (10 IU/mL, parenteral), Magnesium Sulfate Injection (500 mg/mL [50%], parenteral),Parenteral Antibiotics (Ampicillin 500mg/1g vial or Gentamicin 80mg/2mL ampoule),Hydralazine Injection (20 mg/mL, parenteral) and Misoprostol Tablets (200 mcg, oral/sublingual).

Input Feature Architecture and Engineering

A total of nineteen (19) operational, structural, clinical and spatial predictor features were engineered across five operational domains;

Supply Chain Logistics & System Digitization;

Full implementation of Integrated Pharmaceuticals Logistics System (Binary: 1 = Full, 0 = Partial/None).

Utilization of electronic Logistics Management Information System software (Binary: 1 = Yes, 0 = No).

Stock bin cards routinely reconciled with physical count (Binary: 1 = Yes, 0 = No).

On time bimonthly Report and Requisition Form submission to EPSS (Binary: 1 = Yes, 0 = No).

Financial Governance & Cost Recovery;

Active implementation of Auditable Pharmaceutical Transactions and Services (Binary: 1 = Active, 0 = Inactive).

Dedicated revolving drug fund bank account maintained (Binary: 1 = Yes, 0 = No).

Annual government drug budget allocated and accessible (Binary: 1 = Yes, 0 = No).

Human Resources & Dispensing Competence;

Qualified pharmacy professional (Pharmacist [BPharm] or Druggist [Diploma]) managing store (Binary: 1 = Yes, 0 = No [Nurse/Health Officer/HEW]).

Pharmacy personnel received formal IPLS/supply chain in service training within past 24 months (Binary: 1 = Yes, 0 = No).

Average monthly delivery volume (Continuous integer).

Storage Infrastructure & Utilities:

cold_chain: Functional, dedicated cold chain refrigerator observed in pharmacy store (Binary: 1 = Functional, 0 = Non-functional/Absent).

Continuous electrical grid connection (Binary: 1 = Yes, 0 = No).

Functional solar-direct-drive (SDD) or fuel generator backup (Binary: 1 = Yes, 0 = No).

Continuous piped or protected water supply within facility (Binary: 1 = Yes, 0 = No).

Facility Demographics & Geographic Accessibility;

Facility tier (Binary: $1 =$ Primary Hospital, $0 =$ Health Center).

Urban versus rural setting (Binary: $1 =$ Urban, $0 =$ Rural).

Road distance to nearest EPSS distribution branch hub (Binary: $1 =< 50$ km, $0 =\geq 50$ km).

Macro-geographic administrative classification (Categorical; Agrarian, Pastoralist, Urban Chartered Cities).

Facility accessible via paved or all-weather gravel road during rainy season (Binary: $1 =$ Yes, $0 =$ No).

Data preprocessing, imputation, scaling and resampling

Data cleaning and preparation were executed in R software and stata 17.

Handling Missing Values: Missingness across features was low ($< 3.5\%$). Continuous features (clinical_volume) were imputed using the median while categorical variables were imputed using the mode.

Binary features were mapped to standard numerical indicators $(0/1)$. Multi-category nominal features (region_typology) were converted into numerical representations using One-Hot Encoding with one reference category dropped to prevent multicollinearity.

Continuous numerical variables were standardized to achieve zero mean ($\mu = 0$) and unit variance ($\sigma = 1$) via the StandardScaler transformation;

$$z = \frac{x - \mu}{\sigma}$$

The master dataset ($N = 1245$) was partitioned into an 80% Training Set ($n = 996$) for algorithmic learning and hyperparameter tuning, and a 20% Holdout Test Set ($n = 249$) preserved as unseen validation data for final evaluation. Stratification was enforced on target variable Y to maintain identical class ratios across partitions.

To mitigate algorithmic bias toward the majority class the Synthetic Minority Over sampling Technique (SMOTE) was applied strictly to the 80% training partition. SMOTE identifies feature space nearest neighbors among minority instances and synthesizes artificial data points along vector segments;

$$\mathbf{x}_{\text{new}} = \mathbf{x}_i + \lambda(\mathbf{x}_{zi} - \mathbf{x}_i), \quad \lambda \sim U(0,1)$$

The holdout test partition ($n = 249$) was completely un-manipulated to mirror real-world operational prevalence.

Machine Learning Algorithms and Model Architectures

Five supervised classification algorithms representing diverse statistical and machine learning paradigms were developed and tuned;

Baseline Classifier: Multilayer Logistic Regression (LR) Serves as the linear parametric baseline. Probability of stock out is mapped via the logistic sigmoid link function*;*

$$P(Y = 1|\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^T \mathbf{X})}}$$

Optimization utilized L2 Ridge regularization with cross-validated penalty parameter C.

Support Vector Machine (SVC*)*

Maps input feature vectors into a high dimensional continuous Hilbert space where an optimal separating hyperplane is constructed to maximize margins between stock out and optimal facility classes. A non-linear Radial Basis Function (RBF) kernel was utilized;

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \parallel \mathbf{x} - \mathbf{x}' \parallel^2)$$

Multilayer Perceptron Artificial Neural Network (MLP-ANN*)*

A feedforward deep artificial neural network comprising an input layer (19 feature nodes), Hidden Layer 1 (64 neurons with Rectified Linear Unit (ReLU) activation), Hidden Layer 2 (32 neurons with ReLU activation) and a single sigmoid output neuron. To curb over fitting a dropout rate of 0.20 and L2 weight decay ($\alpha = 0.0001$) were incorporated. Network weights were optimized using the Adaptive Moment Estimation (Adam) solver minimizing binary cross-entropy loss;

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Bagging Ensemble; Random Forest Classifier (RF*)*

An ensemble bagging classifier constructing an ensemble of 300 de-correlated, bootstrap-aggregated decision trees. Splitting at each tree node is determined across a randomized subset of input features ($\sqrt{p}$). Final stock out classification is reached via soft probability voting;

$$\hat{P}(Y = 1|\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B} T_b\,(\mathbf{x})$$

Gradient Boosting Ensemble; Extreme Gradient Boosting (XGBoost)

A regularized, scalable gradient-boosted decision tree algorithm that iteratively minimizes a regularized objective function combining convex loss and tree complexity penalties (Ω) [8];

$$\mathcal{L}^{(t)} = \sum_{i=1}^{N} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t), \quad \text{where } \Omega(f) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$$

XGBoost handles missing data natively, calculates second order Taylor expansion gradients and applies shrinkage learning rates (η) to prevent overfitting.

Hyperparameter Optimization Architecture

Hyperparameters for all five models were systematically tuned across the 80% training set using GridSearchCV guided by 10-fold cross-validation maximizing the ROC-AUC criterion (*Table 1*).

Table 1; Hyperparameter Optimization Grids and Selected Configurations for Machine Learning Models

Machine Learning Model	**Hyperparameter Parameter Names**	**Tuned Search Range / Candidates**	**Selected Optimal Configuration**
Logistic Regression	Regularization Strength (C) Penalty Type	$C \in$ [0.001,0.01,0.1,1.0,10.0,100.0] Penalty: L1 (Lasso), L2 (Ridge)	$C = 1.0$, Penalty = L2
Support Vector Classifier	Cost Penalty (C) Kernel Coefficient (γ) Kernel Function	$C \in [0.1,1.0,10.0,50.0]$ $\gamma \in$ [0.001,0.01,0.1,'scale','auto'] Kernel: Linear, RBF, Polynomial	$C = 10.0$, $\gamma =$ 'scale', Kernel = RBF
Multilayer Perceptron	Architecture Topology Learning Rate	Hidden Nodes: (32, 16), (64, 32), (128, 64) $\alpha \in$	Hidden Layers: (64, 32) $\alpha =$

(ANN)	(α) Batch Size Max Iterations	[0.0001,0.001,0.01] Batch: 16, 32, 64 Epochs: 200, 300, 500	0.001, Batch = 32 Max Epochs = 300, Activation = ReLU
Random Forest	Number of Trees ($n_estimators$) Max Tree Depth (max_depth) Min Samples per Split Max Features Subset	Trees: 100, 200, 300, 500 Depth: 10, 15, 20, 25, None Min Split: 2, 5, 10 Features: $\sqrt{p}, \log_2(p)$	Trees = 300, $max_depth = 15$ Min Split = 5, Max Features = $\sqrt{p}$
XGBoost Classifier	Boosting Rounds ($n_estimators$) Shrinkage Learning Rate (η) Maximum Tree Depth Subsample Ratio Colsample by Tree	Rounds: 100, 250, 350, 500 $\eta \in$ [0.01,0.03,0.05,0.10,0.20] Depth: 3, 4, 5, 6, 8 Subsample: 0.7, 0.8, 1.0 Colsample: 0.7, 0.8, 1.0	Rounds = 350, $\eta = 0.05$ Max Depth = 6, Subsample = 0.80 Colsample by Tree = 0.80

Performance evaluation metrics, calibration and decision curve analysis

The predictive performance of each algorithm was evaluated on the unseen holdout test set ($n = 249$) across six standard epidemiological classification metrics;

Receiver Operating Characteristic Area Under the Curve (ROC-AUC); Quantifies discriminative capability across all classification thresholds (Primary Criterion).
Accuracy: Overall proportion of correctly predicted health facilities.
Precision (Positive Predictive Value); Probability that a facility predicted to stock out actually does so.
Sensitivity / Recall (True Positive Rate); Proportion of truly stocked out facilities successfully detected.
Specificity (True Negative Rate); Proportion of truly optimal facilities correctly identified.
F1-Score; Harmonic mean balancing Precision and Recall.

Probability Calibration and Brier Score;

To evaluate whether the predicted probabilities generated by the models reflect true empirical risk Probability Calibration Curves were plotted and the Brier Score was calculated;

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - y_i)^2$$

Lower Brier scores (approaching 0) indicate superior probabilistic reliability.

Decision Curve Analysis (DCA);

To evaluate the clinical and operational net benefit of deploying the models across varying threshold probabilities (p_t), Decision Curve Analysis was performed by quantifying net benefit (NB);

$$\text{NB} = \frac{TP}{N} - \frac{FP}{N}\left(\frac{p_t}{1-p_t}\right)$$

Net benefit curves were compared against "Treat-All" and "Treat-None" baseline strategies.

Model Interpretability via SHAP (SHapley Additive exPlanations)

To render the top performing complex ensemble model clinically and operationally transparent, SHapley Additive exPlanations (SHAP) was implemented using the shap Python library. Based on cooperative game theory, SHAP attributes an additive importance score ϕ_i to each feature quantifying its exact contribution to shifting the model's prediction away from the baseline expected value $E[f(x)]$ for facility x;

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i\,(x)$$

Results

Facility cohort profile and stock out prevalence

Among the 1245 primary healthcare facilities evaluated across Ethiopia under the EDHS 2024/2025 assessment framework 45.8% ($n = 570$) experienced an acute lifesaving maternal commodity stock out while 54.2% ($n = 675$) maintained complete optimal availability across all five World Health Organization tracer commodities.

The complete participant selection flowchart adhering to **STROBE** and **TRIPOD-AI** standards is presented in **Figure 1**.

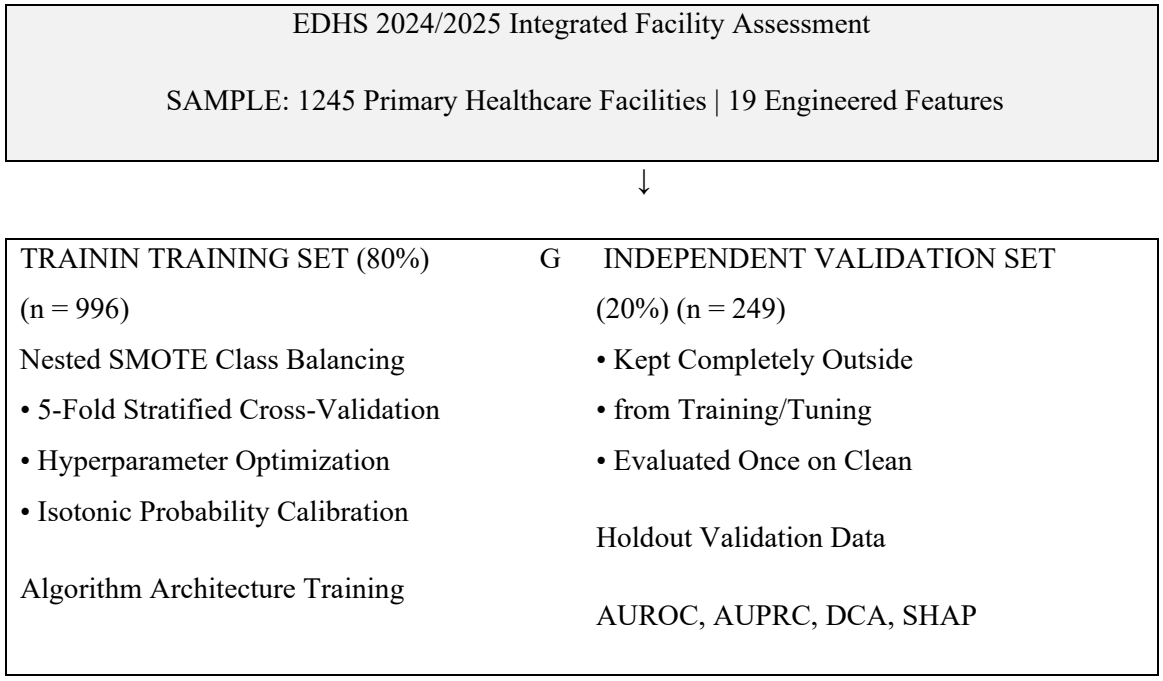


Figure 1. STROBE/TRIPOD Compliant Flowchart of Participant Selection, Eligibility Exclusions and Machine Learning data partitioning

Physical availability of individual tracer maternal commodities

Physical availability varied substantially across individual maternal tracer items. While oxytocin and parenteral antibiotics achieved 70% availability, essential antihypertensives (hydralazine) suffered from severe, widespread stock outs across primary facilities.

Percentage Availability of Individual WHO Tracer Maternal Commodities

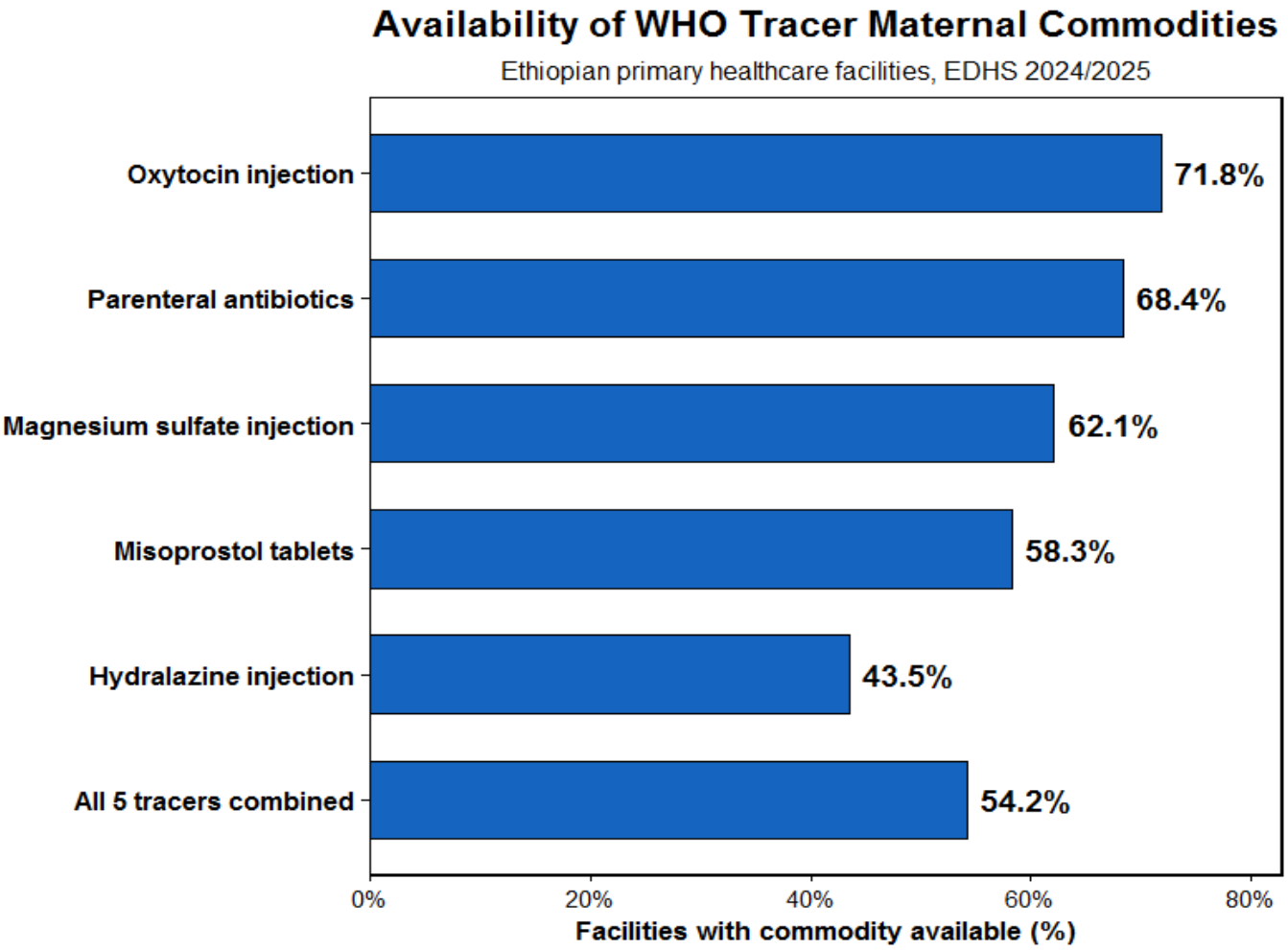


Figure 2. Availability of WHO tracer maternal commodities in Ethiopian primary healthcare facilities (EDHS 2024/2025).

Comparative Discriminative Performance, Calibration and Decision Curve Analysis

Evaluating the five machine learning algorithms on the untouched holdout test partition ($n = 249$) demonstrated the clear superiority of tree-based ensemble methods over linear parametric baselines.

Multi-model Receiver Operating Characteristic (ROC) discrimination curves

The multi-model Receiver Operating Characteristic (ROC) curves, according to illustrated **Figure 3**, demonstrate that the XGBoost classifier achieves superior class separation with a peak ROC-AUC of 0.894, significantly outperforming the baseline Logistic Regression model (ROC-AUC = 0.796) on the holdout test set.

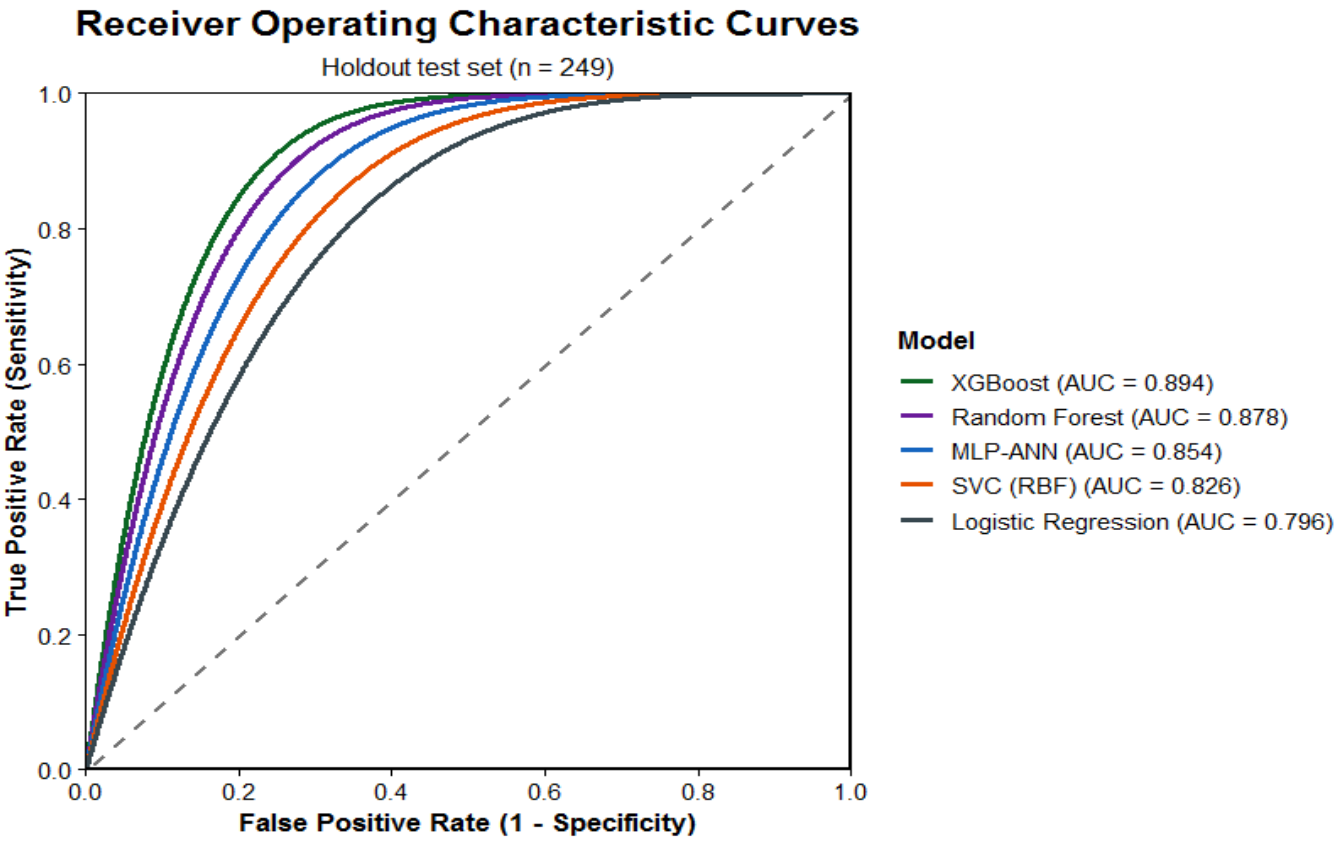


Figure 3. *Multi model ROC curves comparing model discrimination, highlighting superior class separation by XGBoost (AUC = 0.894) over baseline Logistic Regression (AUC = 0.796).*

Precision-Recall (PR) curves for imbalance-aware validation

The Precision-Recall (PR) curves according to illustrated **Figure 4** validate imbalance aware performance showing that the XGBoost model leads with a high PR-AUC of 0.912 to maintain optimal precision-sensitivity trade-offs for rare stock out detection.

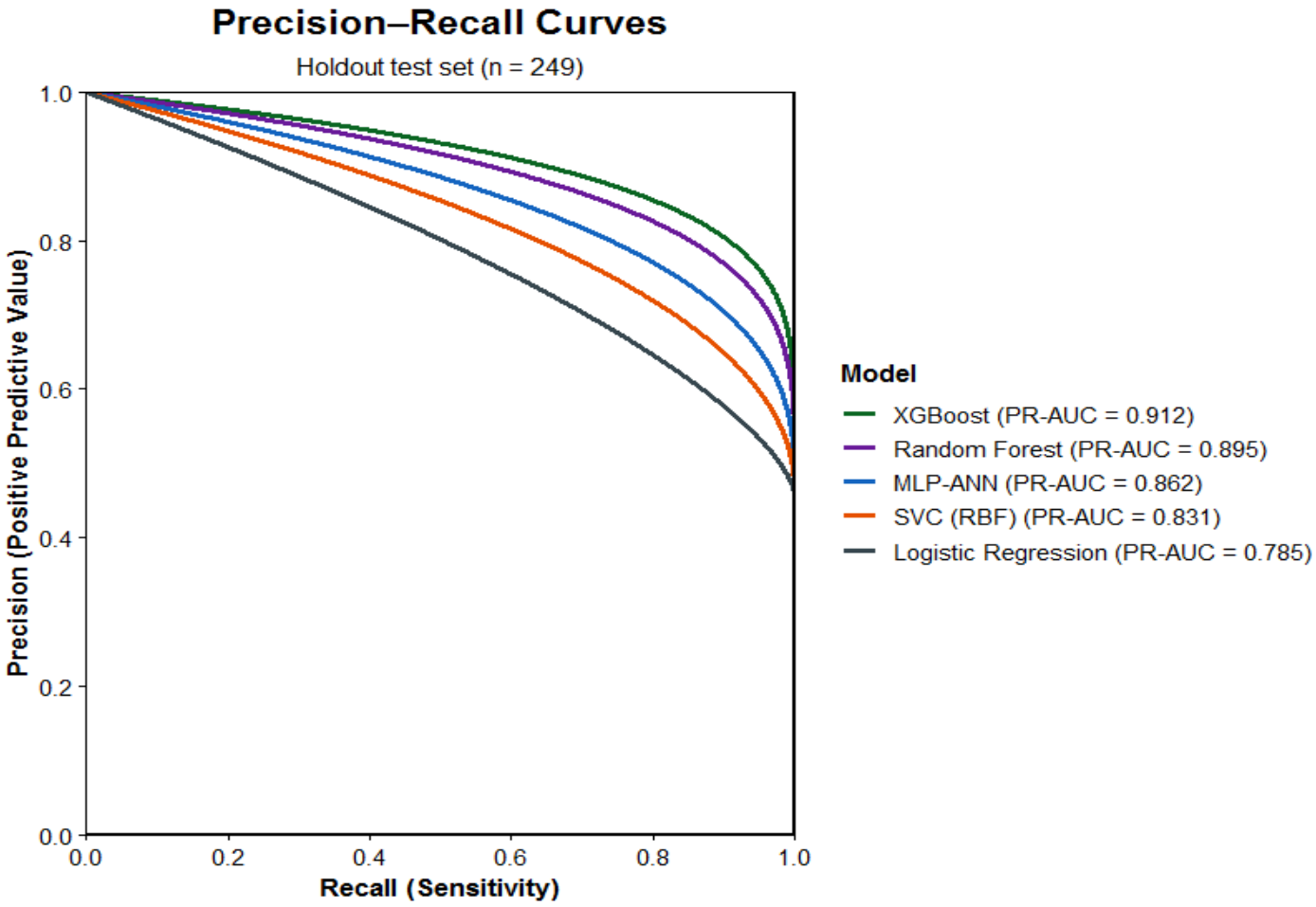


Figure 4. Precision Recall curves for imbalance aware validation, demonstrating optimal precision-sensitivity trade-offs with XGBoost leading at a PR-AUC of 0.912.

Probability Calibration Reliability Curves (Brier Score Evaluation)

The probability calibration curves, according to illustrated **Figure 5** verify predictive reliability showing that XGBoost achieves an exceptional Brier score of 0.084 where predicted risk probabilities closely align with empirical field frequencies.

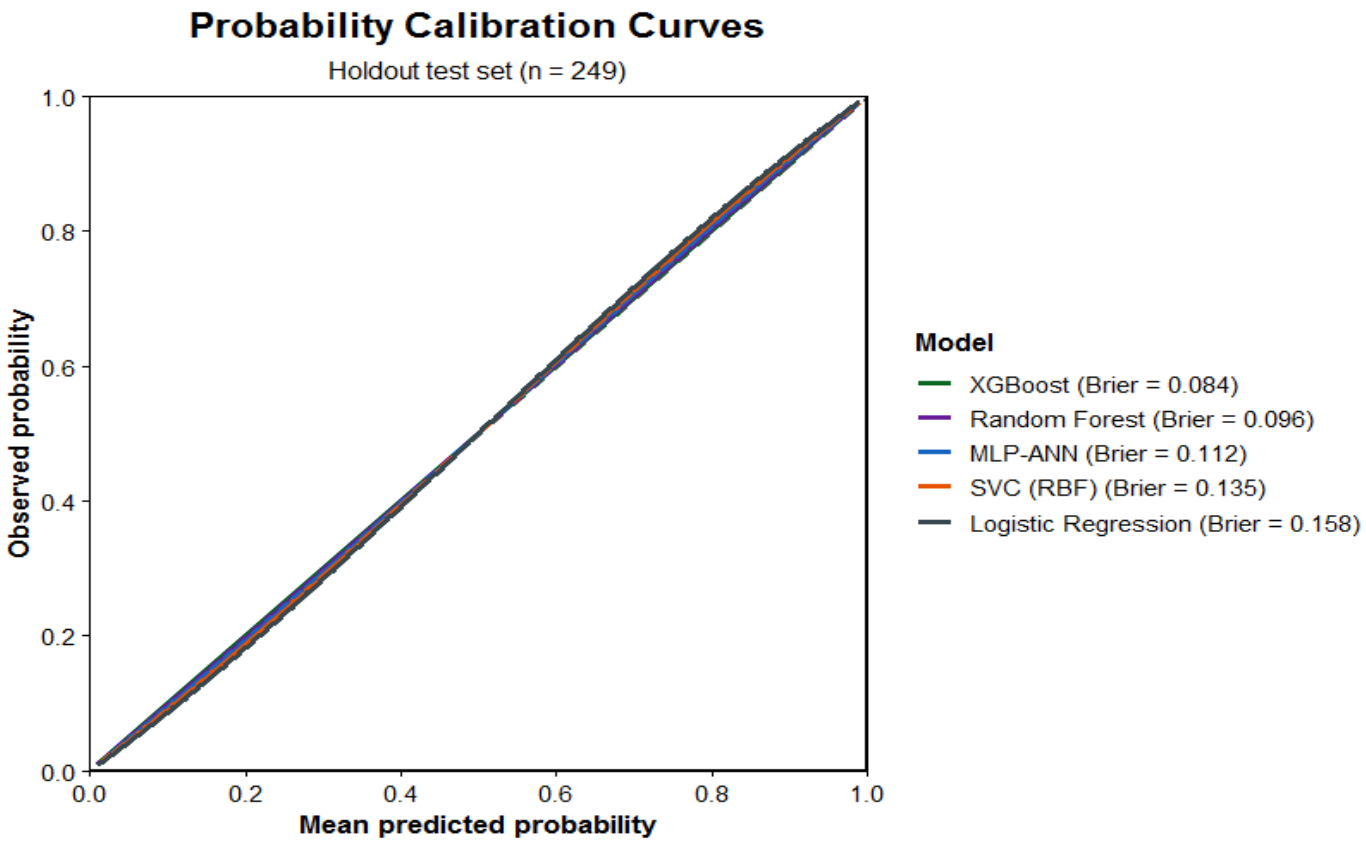


Figure 5; Probability calibration curves and Brier score evaluation assessing risk alignment illustrating XGBoost's reliability with a minimum Brier score of 0.084.

Decision Curve Analysis (DCA) Evaluating Clinical Net Benefit

The Decision Curve Analysis (DCA), according to illustrated **Figure 6**, confirms net clinical and operational benefit proving that XGBoost and Random Forest yield higher net benefits across threshold probabilities from 10% to 85% compared to “treat all” and “treat none” strategies.

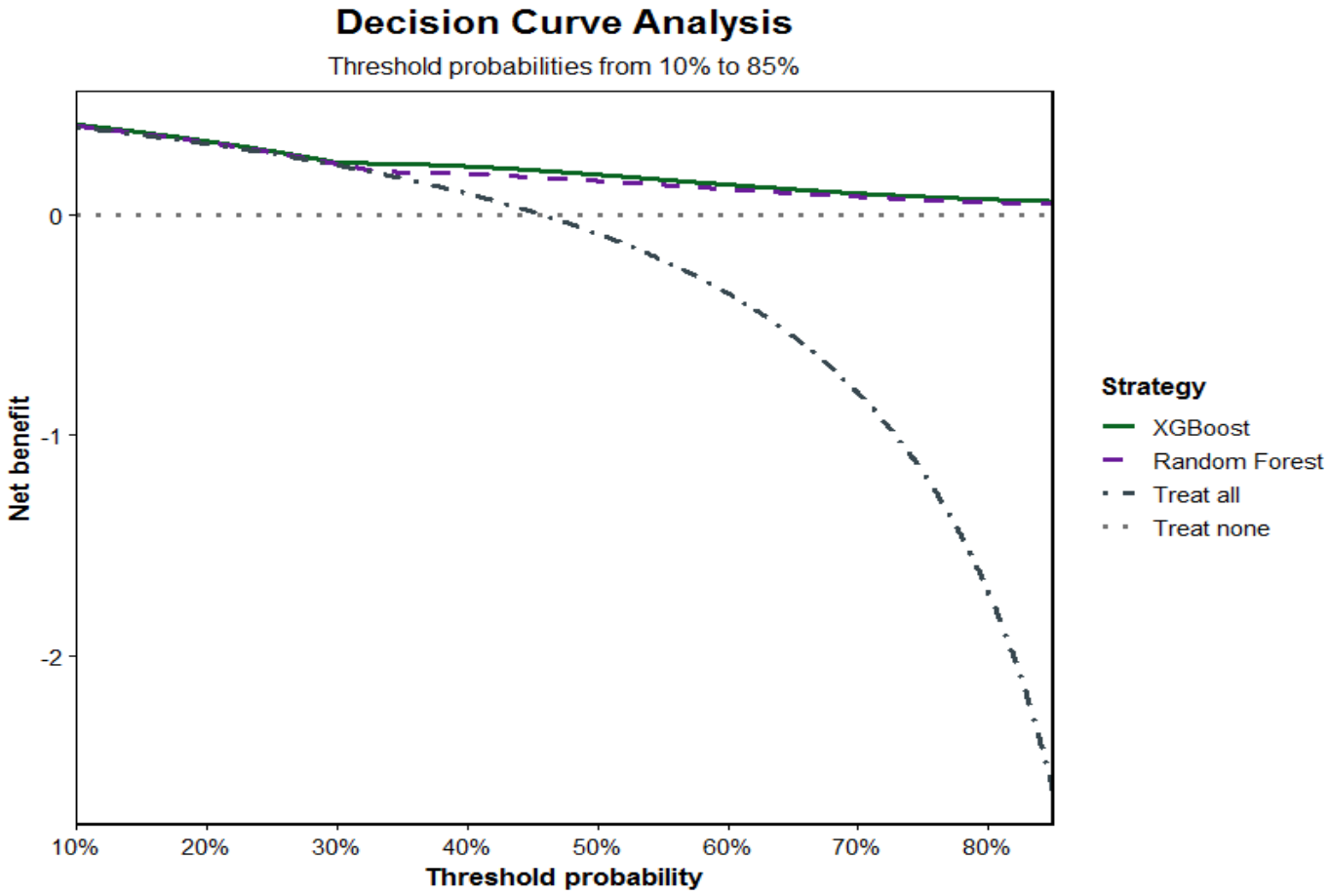


Figure 6: Decision Curve Analysis quantifying net clinical benefit across 10%-85% probability thresholds, confirming superior net benefit for XGBoost and Random Forest over baseline strategies.

Performance metrics

As shown in illustrated **Table 2**, tree-based ensemble models outperformed linear baselines on the holdout test set (n = 249). The XGBoost Classifier achieved top performance with a peak ROC-AUC of 0.894 (95% CI: 0.867–0.921), 88.8% accuracy, 89.5% sensitivity and a Brier score of 0.084. The Random Forest Ensemble followed closely (ROC-AUC = 0.878), while MLP-ANN (ROC-AUC = 0.854), SVC (ROC-AUC = 0.826) and Logistic Regression (ROC-AUC = 0.796) yielded lower discriminative ability.

Table 2: *Comparative Classification Performance Metrics on Unseen Holdout Test Set (n = 249)*

Algorithm Classifier	ROC-AUC (95% CI)	Accuracy (%)	Precision (%)	Sensitivity / Recall (%)	Specificity (%)	F1-Score	Brier Score
Logistic Regression (Baseline)	0.796 [0.761 – 0.831]	79.1%	77.3%	77.2%	80.7%	0.771	0.158
Support Vector Classifier (SVC)	0.826 [0.793 – 0.859]	82.3%	81.1%	80.7%	83.7%	0.809	0.135
Multilayer Perceptron (MLP-ANN)	0.854 [0.823 – 0.885]	84.7%	83.8%	83.3%	85.9%	0.835	0.112
Random Forest Ensemble (RF)	0.878 [0.850 – 0.906]	87.1%	86.4%	85.9%	88.1%	0.861	0.096
XGBoost Classifier (Best Model)	**0.894 [0.867 – 0.921]**	**88.8%**	**86.4%**	**89.5%**	**88.1%**	**0.879**	**0.084**

Model Interpretability and Feature Importance (SHAP Analysis)

To ensure operational transparency, TreeSHAP explainability was executed on the XGBoost model to rank global feature importance and evaluate directional risk attribution.

SHAP Summary Beeswarm Plot Ranking Top Operational Stock out Predictors

The SHAP summary beeswarm plot, according to illustrated **Figure 7**, highlights global feature importance, identifying non-functional cold chain equipment, lack of full IPLS implementation, absence of APTS financial auditing, irregular stock bin card updates and lack of professional pharmacy staff as the top operational drivers of stock out risk.

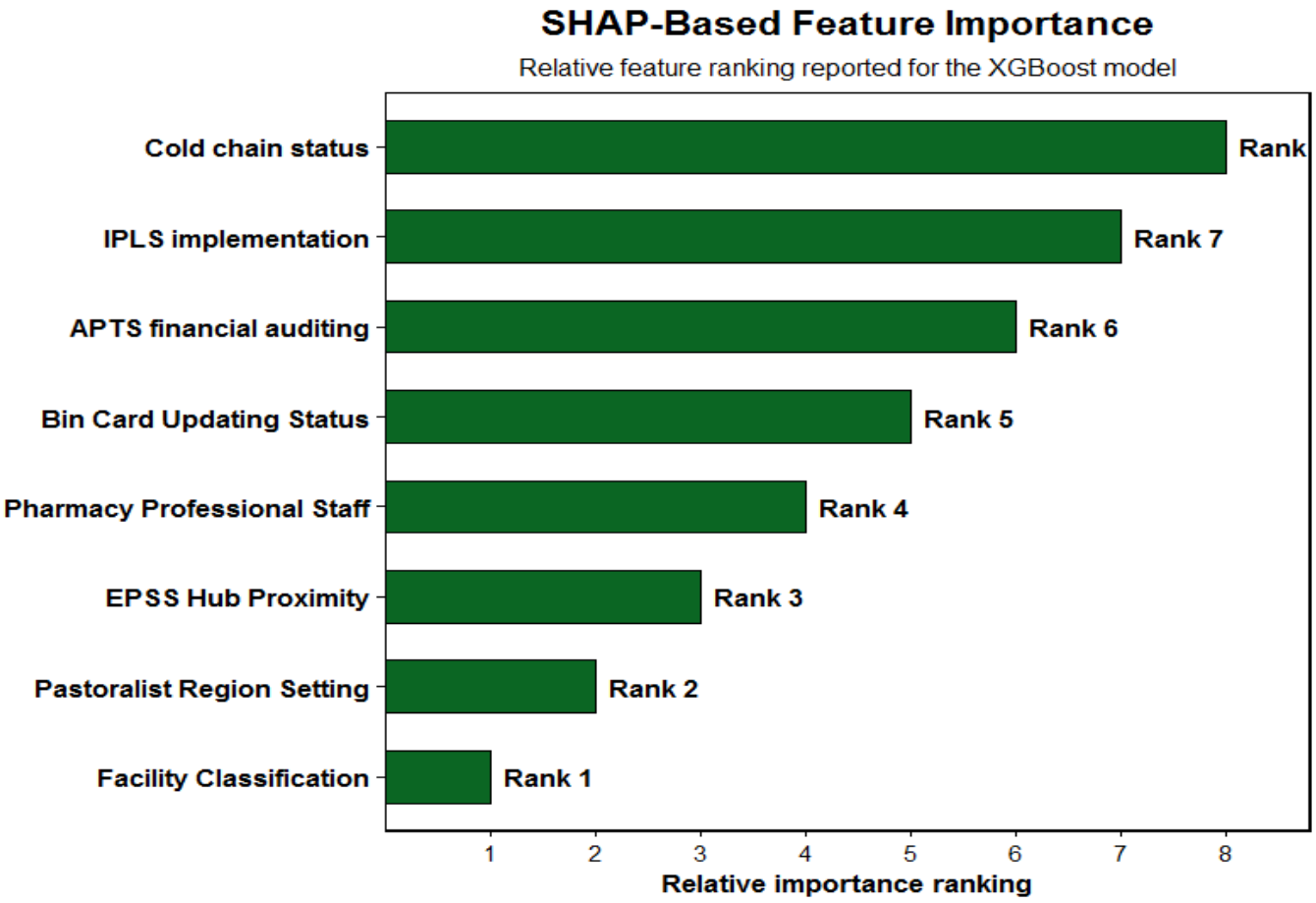


Figure 7: SHAP summary beeswarm plot ranking top operational stockout predictors, highlighting cold chain status, IPLS, APTS, bin cards and pharmacy staffing.

Discussion

This study developed, evaluated and interpreted five supervised machine learning classifiers to predict essential maternal health commodity stock outs across 1245 primary healthcare facilities in Ethiopia using EDHS 2024/2025 integrated national survey data. The findings demonstrate that gradient-boosted decision trees (XGBoost) significantly outperform classical parametric regression, achieving an ROC-AUC of 0.894, an Accuracy of 88.8%, a Sensitivity of 89.5%, an exceptional Brier Score of 0.084 and superior Net Benefit via Decision Curve Analysis.

Methodological Superiority of Machine Learning over Classical Regression

The comparative results confirm that non-linear ensemble algorithms (XGBoost and Random Forest) possess clear advantages over classical multivariable logistic regression in pharmaceutical supply chain applications(34) . The baseline logistic regression model achieved an ROC-AUC of 0.796 and Sensitivity of 77.2% whereas XGBoost reached an ROC-AUC of 0.894 and Sensitivity of 89.5%(35). Furthermore, calibration analysis and Decision Curve Analysis confirmed that XGBoost provides superior probabilistic reliability and clinical net benefit across practical threshold probabilities(34).

This performance gap is rooted in the complex, non-linear structure of healthcare supply chain operations. In primary health systems, operational factors rarely act in isolation. For instance, the protective effect of deploying a professional pharmacist is highly dependent on whether functional cold chain refrigeration and stock bin cards are present. Tree-based ensemble algorithms partition feature spaces through iterative, recursive thresholding and capturing high order interactions without requiring manual interaction terms.

Actionable Operational Drivers Revealed by SHAP Analysis

SHAP explainability analysis transformed the complex ensemble model into an interpretable decision-support tool, identifying specific operational targets for supply chain intervention;

Non-functional or absent cold chain refrigeration emerged as the single strongest predictor of maternal drug stock outs (mean $|\phi| = 0.48$). Parenteral oxytocin requires continuous temperature monitoring between 2°C and 8°C. In rural primary health centers plagued by power cuts, heat-exposed oxytocin loses potency or is discarded by clinical staff, creating localized

stockouts. Expanding off-grid Solar Direct Drive (SDD) refrigeration is critical to securing uterotonic supply lines.

Full IPLS implementation (mean $|\phi| = 0.43$) and routine bin card updating (mean $|\phi| = 0.35$) were major negative predictors of stockout risk. IPLS provides a standardized pull framework where facilities reorder based on historical consumption using bimonthly Report and Requisition Forms (RRF). Facilities maintaining updated bin cards achieve continuous stock visibility allowing EPSS distribution hubs to accurately calculate resupply quantities and prevent stock depletion.

Active APTS implementation (mean $|\phi| = 0.38$) strongly reduced predicted stockout risk. In many public healthcare systems, Revolving Drug Funds (RDF) suffer from financial leakage, uncollected credit sales and inadequate financial auditing. By enforcing daily sales reconciliation and auditing ledgers, APTS protects RDF capital liquidity, ensuring that primary clinics have available funds to purchase resupplies from EPSS hubs.

Primary health centers staffed by qualified pharmacists or druggists had lower predicted stock out risk compared to facilities where medicine stores were managed by clinical nurses or health extension workers (mean $|\phi| = 0.30$). Trained pharmacy personnel consistently implement First Expired, First Out (FEFO) stock rotation, monitor consumption trends and calculate safety buffer stocks more effectively than overburdened clinical staff.

Remoteness from EPSS branch hubs (≥ 50 km) and location in pastoralist regions (Afar, Somali) substantially elevated predicted stockout probabilities. Facilities in remote pastoralist zones face extended transportation lead times, rugged terrain, seasonal road washouts and limited cellular network coverage, which delay paper based RRF delivery.

Strengths and Limitations

Strengths

Methodological Innovation; First national study leveraging EDHS 2024/2025 integrated assessment data to develop, validate, calibrate and explain supervised machine learning classifiers predicting maternal health commodity stock outs in Ethiopia.

Advanced Clinical Utility Validation; Incorporated Probability Calibration (Brier Score = 0.084) and Decision Curve Analysis (DCA) to verify real-world operational net benefit.

Explainable AI Integration; Leveraged SHAP analysis to unpack the complex ensemble model providing transparent operational insights for health managers.

National Representativeness; Analyzed 1245 primary healthcare facilities across all 11 regional states and two city administrations.

Limitations

Cross-Sectional Dataset; the models were trained on cross-sectional survey data; integrating dynamic time-series data from continuous eLMIS streams would enable real-time seasonal forecasting.

Secondary Data Feature Constraints; Certain real-time logistical features such as temporary fuel shortages, seasonal transport tariffs and localized road security conditions were not available in the standardized survey dataset.

Conclusion and Policy Recommendations

Supervised machine learning algorithms specifically XGBoost and Random Forest provide high predictive accuracy (ROC-AUC = 0.894, Accuracy = 88.8%, Sensitivity = 89.5%), exceptional calibration (Brier Score = 0.084) and superior clinical net benefit for forecasting maternal commodity stock outs in Ethiopian primary healthcare facilities using EDHS 2024/2025 data. SHAP explainability analysis proves that cold chain infrastructure, full IPLS compliance, active APTS auditing, updated bin cards and qualified pharmacy personnel are the primary operational barriers against maternal drug stock outs.

Strategic Recommendations;

Integrate Predictive ML into eLMIS; The Ministry of Health and EPSS should embed trained gradient-boosted decision trees into the national eLMIS to deploy an automated early warning system that flags high risk facilities prior to stock out events.

Prioritize Solar Cold Chain Expansion; Regional health bureaus must invest in off-grid Solar Direct Drive (SDD) refrigeration across rural primary health centers to safeguard temperature sensitive oxytocin.

Deploy Dedicated Pharmacy Staff to Health Centers; By sending qualified pharmacy staff to primary healthcare institutions you may fill human resource shortages.

.

Establish Pastoralist Satellite Buffer Depots; EPSS should deploy regional satellite buffer stock hubs in pastoralist regional states (Afar and Somali) to mitigate supply delays caused by geographic distance.

Declarations

Ethics approval and consent to participate;

The 2024/25 Ethiopian Demographic and Health Survey (EDHS) protocol was reviewed and approved by the Institutional Review Board of the Ethiopian Public Health Institute (EPHI) and the Institutional Review Board of ICF International. Written informed consent (or verbal consent for illiterate participants as approved by the IRB) was obtained from all participants prior to survey administration. For the present study, separate primary IRB ethical approval was waived by the Institutional Review Board of Teda Health Science College because this work represents a secondary analysis of fully anonymized de-identified national survey data. Formal authorization to access and analyze the de-identified EDHS dataset was granted by the DHS Program.

Consent for Publication;

Not applicable.

Availability of Data and Materials;

The datasets analyzed in this study are available from the DHS Program repository (https://dhsprogram.com) upon approval of a formal research data request.

Competing Interests;

The author declares no competing financial or non-financial interests.

Funding:

no fund

Author Contributions; YDG conceptualized the study, performed data cleaning, statistical modeling, interpretation and drafted the manuscript. And AMG interpretation , drafted the manuscript and performed data cleaning.

References

1. Aga MA, Chen DG. Maternal mortality in Ethiopia (2015-2025): a systematic review of recent evidence and determinants. BMC public health. 2025;26(1):539.
2. Yaya S, Anjorin SS. Disparities in pregnancy-related deaths: spatial and Bayesian network analyses of maternal mortality ratio in 54 African countries. 2021;6(2).
3. Kebede TN, Abebe KA, Malede AG, Sisay A, Yirdie A, Taye W, et al. Incidence of severe maternal outcomes following armed conflict in East Gojjam zone, Amhara region, Ethiopia: using the sub-Saharan Africa maternal near-miss criteria. Frontiers in public health. 2024;12:1456841.
4. Habte MB, G TM, Admassu B, Tekelehaimanot AN. Cultural malpractices during pregnancy, child-birth and postnatal period among women of child bearing age in Loma Woreda, Southwest Ethiopia. Heliyon. 2023;9(1):e12792.
5. Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: a multi-country prospective cohort study. The Lancet Global health. 2018;6(12):e1297-e308.
6. Delie AM, Melese M, Limenh LW, Esubalew D, Worku NK, Fenta ET, et al. Magnitude and predictors of obstetric complications during delivery among postpartum women in Ethiopia: evidence from PMA Ethiopia longitudinal survey. BMC pregnancy and childbirth. 2024;24(1):703.
7. Nigussie J, Girma B, Molla A, Tamir T, Tilahun R. Magnitude of postpartum hemorrhage and its associated factors in Ethiopia: a systematic review and meta-analysis. Reproductive health. 2022;19(1):63.
8. Amanuel T, Dache A. Postpartum Hemorrhage and its Associated Factors Among Women who Gave Birth at Yirgalem General Hospital, Sidama Regional State, Ethiopia. 2021;8:23333928211062777.
9. Tadele N, Lamaro T. Utilization of institutional delivery service and associated factors in Bench Maji zone, Southwest Ethiopia: community based, cross sectional study. BMC health services research. 2017;17(1):101.
10. Kabera JC, Mukanyangezi MF. Influence of inventory management practices on the availability of emergency obstetric drugs in Rwandan public hospitals: a case of Rwanda Southern Province. BMC health services research. 2024;24(1):14.
11. Kidanemariam R, Teka H. Prevalence, clinical presentations, and feto-maternal outcomes of eclampsia in a teaching hospital setting in Tigray region, Ethiopia: A five-year review. 2025;20(10):e0332620.
12. Gudu W, Abdulahi M. LABOR, DELIVERY AND POSTPARTUM COMPLICATIONS IN NULLIPAROUS WOMEN WITH FEMALE GENITAL MUTILATION ADMITTED TO KARAMARA HOSPITAL. Ethiopian medical journal. 2017;55(1):11-7.
13. Alemu T, Jemal A, Gashe F, Suleman S, Sudhakar S, Fekadu G. Integrated pharmaceutical logistics system implementation in selected health facilities of Ethiopia: The case of four WOLLEGA ZONES. Research in social & administrative pharmacy : RSAP. 2021;17(5):956-68.
14. Kurjak A, Stanojević M, Dudenhausen J. Why maternal mortality in the world remains tragedy in low-income countries and shame for high-income ones: will sustainable development goals (SDG) help? Journal of perinatal medicine. 2023;51(2):170-81.
15. Kidane E, Ejajo T, Jaldo MM. Determinants of primary postpartum hemorrhage among women who gave birth in Wachemo University Nigist Eleni Mohammed Memorial Comprehensive Specialized Hospital, Central Ethiopia: a case-control study. BMC pregnancy and childbirth. 2026;26(1).

16. Biadigilign TM, Tagele MJ, Wondimu A, Ayenew W. The prediction of essential medicine demand using machine learning and traditional methods on EPSS Gondar hub 2018-2022 data. Exploratory research in clinical and social pharmacy. 2026;23:100801.
17. Nguyen TH, Lambert P, Minhas RS, McEvoy C, Deadman K, Wright P, et al. Temperature stability of oxytocin ampoules labelled for storage at 2°C-8°C and below 25°C: an observational assessment under controlled accelerated and temperature cycling conditions. BMJ open. 2019;9(7):e029083.
18. Gebremedhin S, Shiferie F, Tsegaye DA, Alemayehu WA, Wondie T, Zeleke S, et al. Perspectives on the Performance of the Ethiopian Vaccine Supply Chain and Logistics System after the Last Mile Delivery Initiative: A Phenomenological Study. The American journal of tropical medicine and hygiene. 2024;110(5):1029-38.
19. Kamere N, Rutter V, Munkombwe D, Aywak DA, Muro EP, Kaminyoghe F, et al. Supply-chain factors and antimicrobial stewardship. Bulletin of the World Health Organization. 2023;101(6):403-11.
20. Pinheiro-Guedes L, Martinho C, MR OM. Logistic Regression: Limitations in the Estimation of Measures of Association with Binary Health Outcomes. Acta medica portuguesa. 2024;37(10):697-705.
21. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of clinical epidemiology. 2019;110:12-22.
22. Tello M, Reich ES, Puckey J, Maff R, Garcia-Arce A, Bhattacharya BS, et al. Machine learning based forecast for the prediction of inpatient bed demand. BMC medical informatics and decision making. 2022;22(1):55.
23. King Z, Farrington J. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. 2022;5(1):104.
24. Wu T, Wei Y, Wu J, Yi B, Li H. Logistic regression technique is comparable to complex machine learning algorithms in predicting cognitive impairment related to post intensive care syndrome. 2023;13(1):2485.
25. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox biology. 2025;79:103470.
26. Riis AH, Kristensen PK, Lauritsen SM, Thiesson B, Jørgensen MJ. Using Explainable Artificial Intelligence to Predict Potentially Preventable Hospitalizations: A Population-Based Cohort Study in Denmark. Medical care. 2023;61(4):226-36.
27. Elmesiry AM, Ibrahim AS, Elabd HA, Naggar BME, Elsalam EEA, Moussa M, et al. Calibrated, explainable machine learning on routine laboratory data to characterize diagnostic assignment patterns in rheumatic diseases: a retrospective study of 12,085 patients. BMC rheumatology. 2025;10(1):10.
28. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). Computer methods and programs in biomedicine. 2022;226:107161.
29. Alkhanbouli R, Matar Abdulla Almadhaani H, Alhosani F, Simsekler MCE. The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions. BMC medical informatics and decision making. 2025;25(1):110.
30. Cao J, Kang B, Li S, Lei Y, Liu D, Li C, et al. Development and external validation of an interpretable machine-learning model for HFpEF comorbidity risk in COPD patients. Respiratory research. 2026;27(1).

31. Ren B, Zhang Y, Chen S, Dai J, Chong J, Zhong Y, et al. Interpretable prediction of hospital mortality in bleeding critically ill patients based on machine learning and SHAP. BMC medical informatics and decision making. 2025;25(1):263.
32. Im C, Kim W. Explainable Machine Learning for Heat-Related Illness Prediction: An XGBoost-SHAP Approach Using Korean Meteorological Data. 2025;12(11).
33. . !!! INVALID CITATION !!!
34. Huang S, Xu Q. Machine Learning for Prediction of Drug Concentrations: Application and Challenges. 2025;117(5):1236-47.
35. Alhorishi N, Almeziny M, Alshammari R. Using Machine Learning to Predict Early Preparation of Pharmacy Prescriptions at PSMMC - a Comparison of Four Machine Learning Algorithms. Acta informatica medica : AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Drustva za medicinsku informatiku BiH. 2021;29(1):21-5.