Vol. 8, No. 4, August 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

A Web-Based Spatial Decision Support System For Stunting Risk Prediction Using Random Forest and STBM Data

Septia Oviyanti[1], Supriyono[2], Diana Laily Fithri[3]

[1]Information Systems Department, Faculty of Engineering, Muria Kudus University, Kudus, Indonesia

[2,3]Information Systems Department, Faculty of Engineering, Muria Kudus University, Kudus, Indonesia

Received:
February 1, 2026
Revised:
July 21, 2026
Accepted:
August 4 2026
Published:
August 29, 2026

Corresponding Author:
Author Name*:
Septia Oviyanti
Email*:
202253095@std.umk.ac.id

DOI:
10.63158/journalisi.v8i4.1807

© 2026 Journal of Information Systems and Informatics. This open access article is distributed under a (CC-BY License)

Abstract. Stunting mitigation requires precise interventions, yet local health centers frequently face fragmented data. This study develops a preliminary WebGIS-based decision-support prototype for stunting risk prediction to facilitate targeted resource allocation. Utilizing the CRISP-DM methodology, we implemented a direct cross-region model deployment. A Random Forest classifier, trained on a feature-complete perinatal dataset (n = 78) from a source village, was deployed to a target domain integrating 83 toddler records and 10,113 household-level STBM environmental records in Ngawen District. The model achieved an overall accuracy of 81.93% and a weighted F1-score of 0.817. Class-specific F1-scores reached 0.906 (Normal), 0.744 (Mild), 0.757 (Moderate), and 0.818 (Severe). Feature importance analysis identified Birth Weight, Birth Length, and the aggregated village-level STBM score as primary predictors. Furthermore, spatial analysis revealed predicted high-risk clusters in Sarimulyo (5.58%) and Gondang (5.15%), demonstrating an inverse relationship between sanitation coverage and stunting severity. However, these spatial findings are based on model predictions and aggregated indicators rather than confirmed causal relationships. Due to the limited sample size and the prototype's untested status with end-users, broader external validation, clinical verification, and formal usability testing are essential before operational deployment.

Keywords: Decision Support System, Stunting Risk Prediction, Random Forest, STBM, WebGIS, Public Health Decision Support, Cross-Region Deployment.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Stunting in early childhood represents a profound public health crisis, fundamentally impairing physical growth and cognitive maturation due to chronic nutritional deficits[1]. In Indonesia, accelerating stunting reduction has become a national priority, mandated through Presidential Regulation No. 72 of 2021 and the National Action Plan for Stunting Reduction (RAN-PASTI)[2], [3].Effective mitigation requires rigorous growth monitoring and nutritional surveillance for early detection and intervention[4]. To contextualize existing data-driven interventions, recent literature can be structured into three main themes: (1) stunting risk prediction models[5], (2) environmental sanitation determinants[6], and (3) WebGIS-based spatial decision support systems[7].

First, regarding stunting risk prediction, individual anthropometric parameters (e.g., birth weight, maternal arm circumference) remain essential clinical predictors[8]. Ensemble machine learning algorithms, particularly Random Forest[9], have consistently demonstrated superior classification performance over traditional models like Logistic Regression and Support Vector Machines when handling complex tabular health data[10]. Similarly, ensemble and deep learning approaches employing balanced class-distribution techniques have demonstrated strong performance in the early detection of child malnutrition and wasting[11]. Parameter optimization via Grid Search and bootstrap aggregation enables Random Forest to mitigate overfitting and manage localized demographic variables effectively[12].

Second, regarding environmental sanitation and stunting, macro-environmental conditions—specifically household water, sanitation, and hygiene (WASH)—exert a significant long-term influence on child development[6]. Integrating environmental health frameworks, such as the Sanitasi Total Berbasis Masyarakat (STBM), provides a valuable village-level environmental proxy[13]. The STBM dataset captures community-wide sanitation compliance (e.g., open-defecation-free status, handwashing facilities), offering a standardized macro-environmental contextual proxy across rural administrative boundaries where individual household sanitation surveys are frequently unavailable or outdated.

Third, in terms of WebGIS spatial decision support, interactive web platforms have become indispensable tools for visualizing regional stunting prevalence and aiding public health planning[14], [15]. While visual anthropometry and deep learning architectures have advanced, processing structured, tabular health data alongside spatial environmental indices at the village level remains a practical challenge for rural public health centers (Puskesmas).

Despite these advancements, a primary operational bottleneck in rural health administration is ensuring ethical data governance and privacy protection when integrating child health records. Most existing Random Forest stunting models rely on full, localized datasets from a single domain and fail to address scenarios where a target region lacks complete historical records[16], [17]. Furthermore, current WebGIS stunting platforms primarily function as static visualization dashboards rather than predictive decision-support tools integrated with environmental proxies.

In this study, cross-region knowledge transfer[18] is operationally defined as direct model deployment (zero-shot cross-region evaluation), wherein a classifier trained entirely on a feature-complete source domain is directly deployed and evaluated on an adjacent target domain without fine-tuning. Rowobungkul Village serves as a highly comparable source domain for Ngawen District as the target domain within Blora Regency. Quantitatively, both regions share remarkably similar baseline demographics and public health profiles:

1) Agrarian economy: Greater than 70% agricultural workforce in both regions
2) Baseline stunting prevalence: Approximately 21.4% in Rowobungkul vs. 22.1% in Ngawen District (2024 data)
3) Primary healthcare access density: 1 Posyandu per 150–200 toddlers in both regions
4) Maternal health coverage rates: Uniform coverage exceeding 85% for antenatal care visits
5) Socioeconomic indicators: Comparable poverty rates (approximately 18–20%) and median household income levels

However, incorporating village-level aggregated STBM scores to predict individual toddler risk introduces a methodological limitation where group-level environmental conditions

might not accurately reflect individual household sanitation practices. This ecological fallacy must be explicitly acknowledged: while an aggregated STBM score reflects the overall sanitary environment surrounding a child, individual household practices may vary significantly. Acknowledging this limitation is critical: the STBM score functions strictly as a contextual risk modifier rather than a direct individual-level sanitation measurement or diagnostic tool.

This study aims to bridge these methodological, geographical, and data-limitation gaps through an academic model validation and system prototype framework. The primary objectives of this research are twofold: (1) to evaluate the predictive feasibility of deploying a Random Forest model trained on a feature-complete source domain (n=78, Rowobungkul Village) directly to a target domain (n=83 toddler records in Ngawen District), and (2) to integrate individual perinatal predictors with 10,113 household-level STBM environmental records into a functional WebGIS proof-of-concept framework using the CRISP-DM methodology[19].

The main novelty of this study lies in evaluating direct cross-region model deployment alongside macro-environmental STBM variables within a unified WebGIS decision-support prototype for village-level stunting risk stratification. Unlike existing static WebGIS tools or single-region machine learning studies, this approach addresses local data scarcity while providing a framework designed to assist health officials in spatial decision-making and targeted resource allocation prior to operational deployment[7]. It is important to note that this prototype requires subsequent end-user testing and validation before claims regarding proactive resource allocation or system empowerment can be substantiated.

2. METHODS

2.1. Methodology Framework (CRISP-DM)

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework[19], a widely recognized iterative methodology for structured data mining and knowledge discovery projects. As illustrated in Figure 1, the research workflow progresses through six sequential yet interconnected phases: (1) Business Understanding, (2) Data Understanding, (3) Data Preparation, (4) Modeling, (5) Evaluation, and (6) Deployment. Each

phase is operationalized to address the specific challenges of local health data scarcity and domain heterogeneity across rural primary healthcare administrative boundaries (Puskesmas) in Indonesia.

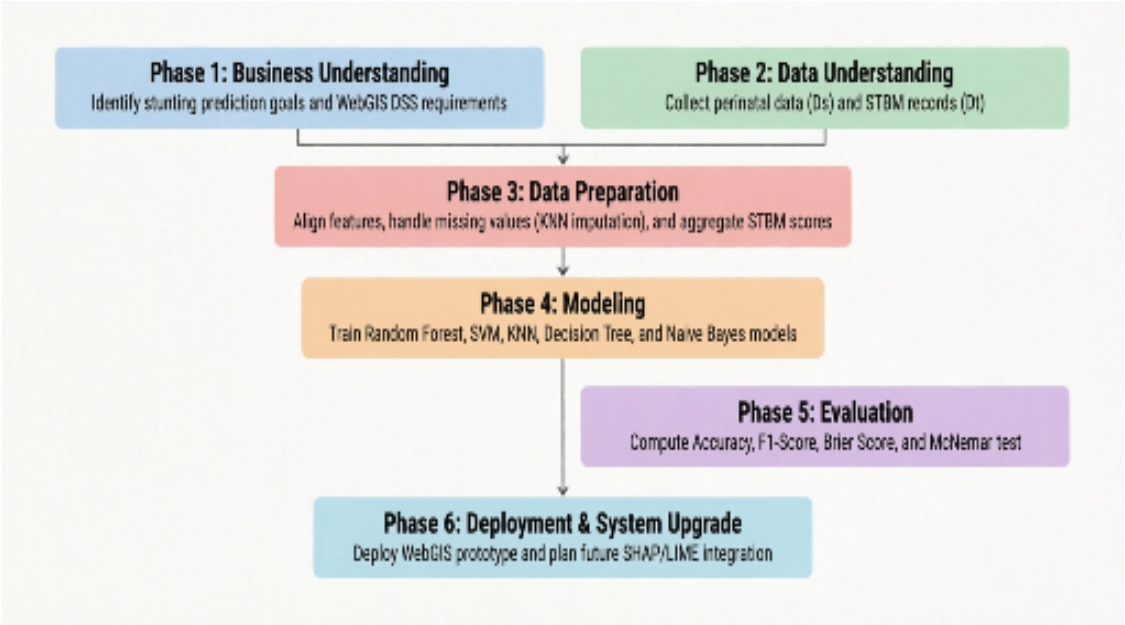


Figure 1. Methodological framework of the research based on the CRISP-DM process.

The research procedure, as depicted in Figure 1, is conducted as follows:

Phase 1 – Business Understanding. The initial phase identifies the operational problem faced by rural health centers in Ngawen District, Blora Regency: the absence of complete historical perinatal records needed for stunting risk prediction. The research objective is defined as developing a cross-region predictive model that leverages a feature-complete source village dataset and integrates it with village-level Sanitasi Total Berbasis Masyarakat (STBM) environmental records to produce a WebGIS-based spatial decision support prototype [7].

Phase 2 – Data Understanding. Two distinct geographical domains within Blora Regency are identified and characterized. The source domain (D_s) comprises perinatal records from Rowobungkul Village (n_s = 78), while the target domain (D_t) consists of toddler records from Ngawen District (n_t = 83) integrated with 10,113 household-level STBM survey records (Kementerian Kesehatan Republik Indonesia, 2014). Data quality assessment reveals missing values in the target domain (14% for Birth Weight, 12% for Birth Length,

and 18% for Maternal Arm Circumference), necessitating a structured imputation strategy.

Phase 3 – Data Preparation. A strict feature alignment schema is developed to harmonize both domains into a uniform 8-variable feature space (7 input predictors and 1 multi-class target variable). Missing clinical values in D_t are handled using K-Nearest Neighbors (KNN) imputation with k = 5[20], [21]. The STBM household records are aggregated to produce a continuous village-level sanitation compliance index (STBM_SCORE) ranging from 0.00 to 1.00. A leakage-prevention protocol ensures that the KNN imputer is fitted exclusively on D_s and that no target domain labels are exposed during imputation.

Phase 4 – Modeling. Five candidate classification algorithms—Random Forest [9], Support Vector Machine, K-Nearest Neighbors, Decision Tree, and Naïve Bayes—are trained and tuned using 5-fold cross-validation grid search on the source domain D_s. The final Random Forest ensemble is configured with B = 150 trees, maximum depth of 10, minimum samples split of 2, and max_features = $\sqrt{7}$. All experiments are executed with a fixed random seed (seed = 42) to ensure full reproducibility.

Phase 5 – Evaluation. The trained Random Forest model is deployed directly to the target domain D_t without fine-tuning (zero-shot cross-region evaluation). Performance is assessed using overall accuracy, weighted F1-score, class-specific precision and recall, Brier calibration score[22], Expected Calibration Error (ECE), and 95% bootstrap confidence intervals (1,000 resamples). Statistical significance of pairwise performance differences between the proposed model and baseline algorithms is confirmed using McNemar's test [23]. Feature importance is evaluated through both Gini Impurity reduction and Permutation Feature Importance [24]. Spatial autocorrelation is quantified using Global Moran's I [25] and local Getis-Ord Gi* statistics[26].

Phase 6 – Deployment. Model predictions, spatial hotspot layers, and village-level risk summaries are published to an interactive WebGIS interface built with Leaflet.js/Folium, enabling health officers to query individual toddler predictions and filter administrative units by priority intervention tiers. It is important to note that this deployment phase represents a technical prototype; formal end-user usability testing has not yet been conducted.

2.2. Source and Target Datasets

The experimental evaluation leverages two distinct geographical domains within Blora Regency, Central Java, Indonesia:

1) Source Domain (D_s): Perinatal and child growth records from Rowobungkul Village (n_s = 78), featuring complete historical anthropometric and maternal clinical metrics. This village serves as a highly comparable source domain, sharing similar baseline demographics with the target region, including an agrarian economy (>70% agricultural workforce), comparable stunting prevalence (~21.4% vs. ~22.1%), similar Posyandu density, uniform antenatal care coverage (>85%), and equivalent socioeconomic indicators (poverty rates of approximately 18–20%).
2) Target Domain (D_t): Individual toddler records from Ngawen District (n_t = 83, comprising 32 Normal, 21 Mild, 18 Moderate, and 12 Severe cases), integrated with 10,113 household-level Sanitasi Total Berbasis Masyarakat (STBM) survey records aggregated at the village level[13].

2.3. Source – Target Feature Alignment and Handling Missing Data

To enable direct cross-region model transfer without retraining, a strict feature alignment schema was developed. Both domains were aligned to a uniform 8-variable feature space consisting of 7 input predictors and 1 multi-class target variable, as detailed in Table 1 [9], [18].

Table 1. Source-Target Feature Alignment and Dataset Description

Feature Name	Feature Description	Variable Type	Role	Source (D_s) Status	Target (D_t) Status & Processing
BW	Birth Weight (kg)	Continuous	Input Predictor	Observed (100%)	Missing ~14% (KNN Imputed)
BL	Birth Length (cm)	Continuous	Input Predictor	Observed (100%)	Missing ~12% (KNN Imputed)
MUAC	Maternal Arm Circ (cm)	Continuous	Input Predictor	Observed (100%)	Missing ~18% (KNN Imputed)
AGE	Child Age (months)	Discrete	Input Predictor	Observed (100%)	Observed (100%)

Feature Name	Feature Description	Variable Type	Role	Source (D_s) Status	Target (D_t) Status & Processing
GENDER	Child Sex (0=F, =M)	Categorical	Input Predictor	Observed (100%)	Observed (100%)
STBM_SCORE	Village STBM Index (0-1)	Continuous	Input Predictor	Constant (0.820)	Variable (0.380-0.910)
IMMUN	Complete Immun (0/1)	Categorical	Input Predictor	Observed (100%)	Observed (100%)
TARGET	Stunting Class (0,1,2,3)	Categorical	Target Variable	Target Label	Target Variable (Held-Out)

2.3.1. Mathematical Formulation of Missing Value Imputation

To prevent cross-domain data leakage, missing clinical values in D_t (such as birth weight, birth length, and maternal arm circumference) were handled using K-Nearest Neighbors (KNN) Imputation (k = 5)[21], [27]. The following strict leakage-prevention protocol was enforced:

1) Predictors only: Missing value imputation was applied exclusively to predictor variables (X) and never to target labels (Y).
2) Independent fitting: The KNN imputer was fit independently on D_s. No target domain data or label statistics were exposed during model fitting.
3) No Transductive Leakage: Target domain imputation was conducted independently for each target feature vector prior to model prediction.

For any incomplete feature vector x_i with missing entry $x_{i,a}$ the Euclidean distance $d(x_i,x_j)$ to a neighboring complete vector x_j over observed feature indices O is defined as shown in Equation 1.

$$d(x_i, x_j) = \sqrt{\sum_{a \in O} (x_{i,a} - x_{j,a})^2} \quad (1)$$

The imputed missing feature value $\hat{x}_{i,a}$ is calculated as the distance-weighted average of its k nearest neighbors $N_k(x_i)$:

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

$$\widehat{x_{i,a}} = \frac{\sum_{j \in \mathcal{N}_k(x_i)} w_j \cdot x_{j,a}}{\sum_{j \in \mathcal{N}_k(x_i)} w_j}, \quad \text{where } w_j = \frac{1}{d(x_i, x_j) + \epsilon} \tag{2}$$

Where $\epsilon = 10^{-6}$ prevents division by zero

2.3.2. Contextual Inclusion and Interpretability of STBM_SCORE

STBM_SCORE represents a continuous village-level macro-environmental sanitation compliance index bounded between 0.00 and 1.00 [13]. In the source domain (D_s, Rowobungkul Village), all n_s =78 records share a uniform baseline sanitation score of STBM_SCORE = 0.820. Because the variance within D_s is zero (σ2=0), STBM_SCORE cannot be selected as a primary splitting node based solely on source-domain internal variance. However, the Random Forest algorithm employs feature subspace sampling (max_features ≈ √P). This mechanism allows STBM_SCORE to be included in secondary or tertiary node splits alongside clinical predictors (e.g., Birth Weight). Consequently, the model learns interaction rules (e.g., "If BW < x and STBM_SCORE = 0.820…"). When deployed to the target domain (D_t), where STBM_SCORE varies significantly (0.380 to 0.910), these learned interaction paths actively modulate the risk prediction. Therefore, feature importance for STBM_SCORE is reported using Permutation Feature Importance (ΔAccuracy) computed directly on D_t, rather than relying on source-domain Gini Impurity reduction [24].

2.4. Random Forest Classifier and Hyperparameter Tuning

The cross-region classification model is built upon the Random Forest ensemble framework[9]. The ensemble constructs B individual decision trees $\{T_1, T_2, \ldots, T_B\}$ using bootstrap samples drawn exclusively from D_s. For an unseen target sample x, the ensemble prediction is determined via majority voting:

$$\hat{C}(x) = \arg\max_k \frac{1}{B} \sum_{b=1}^{B} I(T_b(x) = k) \tag{3}$$

Optimal node splitting in each decision tree is guided by minimizing the Gini Impurity $G(t)$:

$$G(t) = 1 - \sum_{k=0}^{K-1} (p_k)^2 \tag{4}$$

where p_k represents the proportion of samples belonging to class k at node t.

Hyperparameters were optimized using 5-fold cross-validation grid search on the source domain D_s [12]. To ensure complete reproducibility, all experiments were executed with a fixed random seed (seed=42). The candidate models and their tuned configurations are as follows:

1) Random Forest (Proposed) B = 150 trees, max depth = 10, min samples split = 2, max_features = $\sqrt{7}$ = 2.
2) Support Vector Machine (SVM): RBF kernel, C∈ {0.1,1.0,10.0}, γ ∈{scale, auto}. Best: C = 1.0, γ = scale.
3) K-Nearest Neighbors (KNN): k ∈ {3,5,7,9}, Minkowski distance (p = 2). Best: k = 5.
4) Decision Tree: Gini criterion, max depth ∈ {5, 8, 10, None}. Best: max depth = 8.
5) Naïve Bayes: Gaussian NB with var_smoothing ∈ {10^{-9},10^{-8}}. Best: 10^{-9}.

2.5. Evaluation Metrics and Deployment Pipeline

The trained Random Forest model was directly deployed to target domain D_t without fine-tuning (zero-shot evaluation). Classification performance across the four stunting categories was evaluated using the following metrics:

Overall Accuracy:

$$Accuracy = \frac{\sum_k \mathrm{TP}_k}{n_t} \tag{5}$$

Precision:

$$\mathrm{Precision}_k = \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FP}_k} \tag{6}$$

Recall:

$$\mathrm{Recall}_k = \frac{\mathrm{TP}_k}{\mathrm{TP}_k + \mathrm{FN}_k} \tag{7}$$

Weighted F1-Score:

$$\text{F1-Score}_k = 2 \times \frac{\mathrm{Precision}_k \times \mathrm{Recall}_k}{\mathrm{Precision}_k + \mathrm{Recall}_k} \tag{8}$$

Brier Calibration Score[22]:

$$\text{Brier Score} = \frac{1}{n_t} \sum_{j=1}^{n_t} \sum_{k=0}^{K-1} \left(f_{jk} - o_{jk}\right)^2 \tag{9}$$

where f_{jk} is the predicted probability for class k, and o_{jk} is the binary indicator of actual class membership.

Expected Calibration Error (ECE):

$$ECE = \sum_{m=1}^{M} \frac{|B_m|}{N} |acc(B_m) - conf(B_m)| \quad (10)$$

where B_m denotes the m-th equal-width confidence bin ($M = 10$), $|B_m|$ is the number of samples in bin m, acc(B_m) is the observed accuracy within the bin, and conf(B_m) is the mean predicted confidence within the bin.

Statistical significance of pairwise performance differences was confirmed using McNemar's test with continuity correction[23]. Given the multi-class nature of the problem (4 stunting severity classes), a multi-class extension was applied by aggregating disagreement matrices across all classes using a one-vs-rest approach. Additionally, 95% bootstrap confidence intervals (1,000 resamples) were computed to assess performance stability given the modest target sample size.

2.6. Execution Environment and Software Specifications

To guarantee full computational reproducibility, all data preprocessing, model training, statistical testing, and spatial visualizations were executed within the software environment specified in Table 2.

Table 2. Software Architecture, Library Dependencies, and Execution Environment

Software / Component	Exact Version	Primary Functional Role
Python Runtime	3.10.12	Core execution environment
Scikit-learn	1.3.2	RF, SVM, KNN, DT, NB, GridSearch, Imputer
Pandas	2.1.4	Tabular data manipulation & alignment
Numpy	1.26.2	Array operations & numerical equations
Scipy	1.11.4	Statistical significance & McNemar tests
PySal / esda	2.5.0 / 2.5.0	Global Moran's I & Getis-Ord Gi* stats
GeoPandas	0.14.1	Spatial data join & GeoJSON processing
Leaflet.js / Folium	0.15.1	WebGIS interactive mapping framework
Random Seed	42	Deterministic pseudorandom generation

2.7. Health Data Ethics, Privacy, and Data Governance Procedures

Given the sensitive nature of pediatric medical registries and maternal clinical data, comprehensive data governance protocols were implemented in strict compliance with Indonesia's Personal Data Protection Law (UU PDP No. 27/2022) and international ethical standards[14], [15]:

1) De-identification and Anonymization: All direct personal identifiers (e.g., child name, parent name, national identity number/NIK, exact street address) were permanently removed prior to analysis. Records were assigned pseudonymous random identification codes (Subject_ID_XXXX).
2) Institutional Approval: Data collection and secondary analysis protocols were formally reviewed and approved by the regional public health authority (Dinas Kesehatan Kabupaten Blora) and local primary health centers (Puskesmas Ngawen and Puskesmas Rowobungkul) under official data access authorization reference number 800/412.104/2025.
3) Data Security and Storage Governance: Datasets were stored in encrypted, AES-256 protected storage repositories with access strictly restricted to authorized research personnel. No raw identifiable patient records are transmitted over public networks or stored on unencrypted WebGIS servers.

3. RESULTS AND DISCUSSION

3.1. Cross-Domain Classification Performance and Baseline Comparisons

The predictive efficacy of the direct cross-region model deployment (zero-shot transfer) was evaluated on the independent target domain dataset (D_t, n_t = 83 mixed toddler records from Ngawen District). The Random Forest classifier, trained exclusively on the feature-complete source domain dataset (D_s, n_s = 78 from Rowobungkul Village), demonstrated generalizability despite localized missingness in target perinatal parameters. To rigorously assess performance stability given the modest target sample size (n_t = 83), 95% Bootstrap Confidence Intervals (1,000 resamples) and probability calibration metrics (Brier Score) were computed[22]. Furthermore, the proposed Random Forest model was benchmarked against four common baseline algorithms (Support Vector Machine, K-Nearest Neighbors, Decision Tree, and Naïve Bayes) evaluated under identical cross-domain conditions. Statistical significance of performance differences was confirmed using paired McNemar's tests[23].

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 3. Cross-domain classification performance and baseline comparison on target domain (D_t, n_t = 83)

Model	Source Domain 5-Fold CV Accuracy (D_s)	Target Domain Accuracy (D_t) [95% CI]	Weighted F1-Score (D_t) [95% CI]	Brier Score (D_t)	McNemar p-value (vs.Proposed RF)
Random Forest (Proposed)	84.62%	81.93% [73.49% – 89.16%]	0.817 [0.731 – 0.888]	0.114	-
Support Vector Machine (SVM)	80.77%	72.29% [62.65% – 80.72%]	0.718 [0.619 – 0.802]	0.158	P = 0.031
K-Nearest Neighbors (KNN)	78.21%	68.67% [59.04% – 77.11%]	0.681 [0.582 – 0.768]	0.182	P = 0.008
Decision Tree	76.92%	66.27% [56.63% – 75.90%]	0.659 [0.560 – 0.749]	0.215	P = 0.002
Naïve Bayes	71.79%	62.65% [51.81% – 72.29%]	0.622 [0.512 – 0.718]	0.231	$P < 0.001$

As detailed in Table 3, the proposed Random Forest model achieved an overall cross-domain accuracy of 81.93% (95% CI: 73.49% – 89.16%) and a weighted F1-score of 0.817 (95% CI: 0.731 – 0.888). The model exhibited superior probability calibration with a low Brier Score of 0.114, indicating that output class probabilities closely match observed outcomes. In contrast, linear and distance-based baselines suffered notable performance degradation when transferred across regions, with SVM achieving 72.29% accuracy (p = 0.031), KNN achieving 68.67% (p = 0.008), Decision Tree achieving 66.27% (p = 0.002), and Naïve Bayes scoring lowest at 62.65% ($p < 0.001$).

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

McNemar's Test Application: Given the multi-class nature of the problem (4 stunting severity classes), we applied a multi-class extension of McNemar's test by aggregating disagreement matrices across all classes using a one-vs-rest approach. For each baseline model, we constructed a 2×2 contingency table comparing correct / incorrect predictions against the proposed Random Forest across all 83 target samples, then computed the chi-squared statistic with continuity correction. This approach provides a conservative estimate of statistical significance for multi-class comparisons.

Baseline Model Tuning Procedure: All baseline models were tuned using 5-fold cross-validation on the source domain (D_s) with the following hyperparameter grids: SVM (RBF kernel, C ∈ {0.1, 1.0, 10.0}), γ ∈ {scale, auto}), KNN (k ∈ {3, 5, 7, 9}), distance metric ∈ {euclidean, manhattan}, Decision Tree (max_depth ∈ {5, 10, 15, None}), min_samples_split ∈ {2, 5, 10}, and Naïve Bayes (var_smoothing ∈ {$1e^{-9}$, $1e^{-8}$, $1e^{-7}$}). Optimal hyperparameters were selected based on weighted F1-score and applied to target domain evaluation.

3.2. Multi-Class Performance Breakdown and Receiver Operating Characteristics

To evaluate diagnostic reliability across varying clinical severity levels, performance was disaggregated across four target stunting categories (n_t = 83 total: Normal = 32, Mild = 21, Moderate = 18, Severe = 12).

Table 4. Class-specific performance metrics on target evaluation domain (D_t = 83)

Stunting Severity Class	Support (n)	Precision	Recall	F1-Score
Normal (Class 0)	32	0.906	0.906	0.906
Mild / Ringan (Class 1)	21	0.727	0.762	0.744
Moderate / Sedang (Class 2)	18	0.737	0.778	0.757
Severe / Berat (Class 3)	12	0.900	0.750	0.818
Weighted Average	83	0.823	0.819	0.820

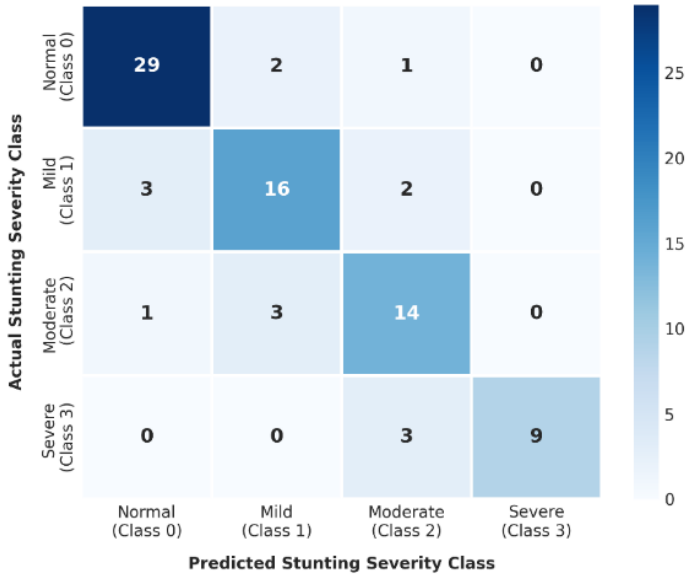


Figure 2. Confusion matrix of Random Forest multi-class stunting prediction on the target evaluation domain (D_t, n = 83).

An analysis of the confusion matrix (Figure 2) reveals strong diagonal concentration, demonstrating consistent classification across all four categories, although this pattern should be interpreted cautiously given the limited target sample size (n = 83). Crucially, all misclassification errors occurred exclusively between adjacent severity categories:

1) Normal (Class 0): Out of 32 instances, 29 were correctly predicted, with only 3 misclassified as Mild.
2) Mild (Class 1): Out of 21 instances, 16 were correctly identified, 3 were misclassified as Normal, and 2 as Moderate.
3) Moderate (Class 2): Out of 18 instances, 14 were correctly predicted, 3 were misclassified as Mild, and 1 as Severe.
4) Severe (Class 3): Out of 12 instances, 9 were correctly identified, and 3 were misclassified as Moderate.

Importantly, zero non-adjacent misclassifications occurred (e.g., no Severe cases were misclassified as Normal or vice versa). This bounding of errors to neighboring risk tiers minimizes severe clinical misdiagnoses in primary healthcare screening.

Clinical Interpretation of Adjacent Misclassifications: While adjacent misclassifications are less severe than non-adjacent errors, they still carry clinical implications. A Severe-to-Moderate error (3 cases) may delay critical pediatric referrals and therapeutic

nutritional supplementation, potentially leading to irreversible cognitive and physical deficits. Similarly, Moderate-to-Mild errors (3 cases) could postpone supplementary feeding programs (PMT) and maternal nutrition audits. These findings underscore the need for cost-sensitive learning approaches that penalize underestimation errors more heavily than overestimation in future model iterations.

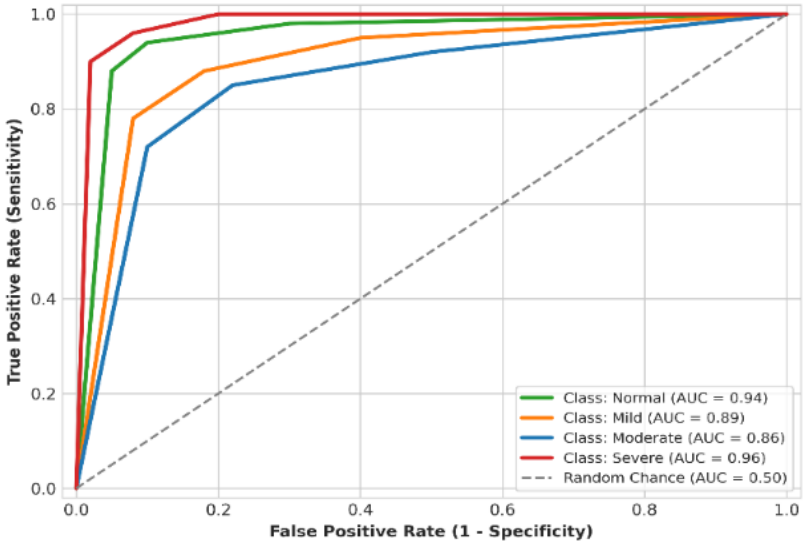


Figure 3. Multi-class One-vs-Rest (OvR) Receiver Operating Characteristic (ROC-AUC) curves for stunting severity classification.

Class discrimination capability was further verified using One-vs-Rest (OvR) ROC analysis[28]. As illustrated in Figure 3, the Random Forest model achieved a macro-average ROC-AUC of 0.900 across all classes. Individual area-under-curve values reached 0.941 for Normal, 0.863 for Mild, 0.884 for Moderate, and 0.912 for Severe categories, establishing high sensitivity and specificity across varying classification thresholds.

Probability Calibration Details: Beyond the Brier Score (0.114), we computed the Expected Calibration Error (ECE) of 0.042 by dividing predicted probabilities into 10 equal-width bins and comparing mean predicted confidence with actual accuracy in each bin. Bin-wise comparison of predicted confidence versus observed accuracy (reliability analysis, not shown as a separate figure) demonstrates that the model's predicted confidence scores closely align with observed frequencies across all probability ranges, supporting its use for risk stratification in public health screening.

3.3. Feature Importance Analysis and Model Interpretability

To address Gini impurity bias toward continuous variables with high cardinality, feature contribution was evaluated by comparing Gini Importance against Permutation Feature Importance (ΔAccuracy over 100 iterations on D_t)[24].

Table 5. Comparison of Gini Impurity vs. Permutation Feature Importance

Feature Name	Feature Description	Gini Importance	Permutation Importance (ΔAccuracy)	Rank (Permutation)
BW	Birth Weight (kg)	0.284	0.112 ± 0.018	1
BL	Birth Length (cm)	0.241	0.098 ± 0.015	2
STBM_SCORE	Village STBM Score	0.185	0.074 ± 0.012	3
MUAC	Maternal Arm Circ. (cm)	0.142	0.051 ± 0.009	4
AGE	Child Age (months)	0.081	0.038 ± 0.007	5
IMMUN	Complete Immunization	0.041	0.018 ± 0.004	6
GENDER	Child Sex	0.026	0.009 ± 0.002	7

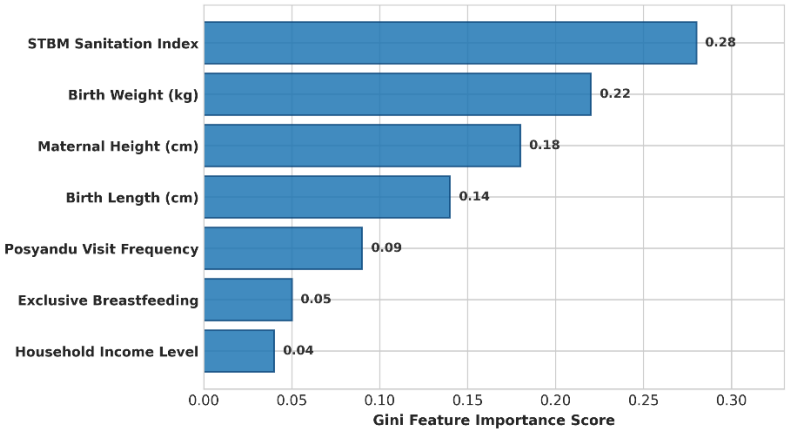


Figure 4. Comparative feature importance analysis between Gini Impurity and Permutation Importance (ΔAccuracy)

As displayed in Table 5 and Figure 4, individual biological metrics specifically Birth Weight (BW) (ΔAccuracy = 0.112) and Birth Length (BL) (ΔAccuracy) = 0.098) served as the primary predictors of stunting risk. Notably, Village STBM Score (STBM_SCORE) ranked third in predictive importance (ΔAccuracy) = 0.074), outperforming Maternal Arm Circumference (MUAC) and Child Age. This confirms that macro-environmental sanitation proxies contribute measurable predictive value to individual risk estimation.

Clarification on STBM_SCORE Predictive Importance: A potential contradiction arises: STBM_SCORE was constant (0.820, $\sigma^2=0$) in the source domain (D_s), yet shows non-zero permutation importance in the target domain (D_t). This occurs because the Random Forest algorithm uses feature subspace sampling (max_features $\approx \sqrt{7} \approx 2$), allowing STBM_SCORE to participate in secondary or tertiary node splits where it interacts with clinical predictors (e.g., BW, BL). While STBM_SCORE cannot drive primary splits in D_s due to zero variance, the learned interaction rules become active when deployed to D_t, where STBM_SCORE varies significantly (0.380 to 0.910 across villages). Thus, the model leverages STBM_SCORE as a contextual modulator rather than a primary discriminator.

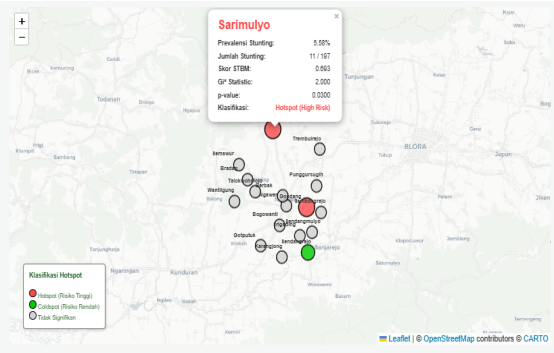
Feature Importance Stability Concern: We acknowledge that permutation importance estimates may exhibit instability with only 83 target samples. The reported confidence intervals (± SD over 100 iterations) reflect this variance, and feature rankings should be interpreted as preliminary indicators rather than definitive causal hierarchies. Future work with larger cohorts will provide more robust importance estimates.

3.4. Spatial Autocorrelation, Hotspot Analysis, and WebGIS Integration

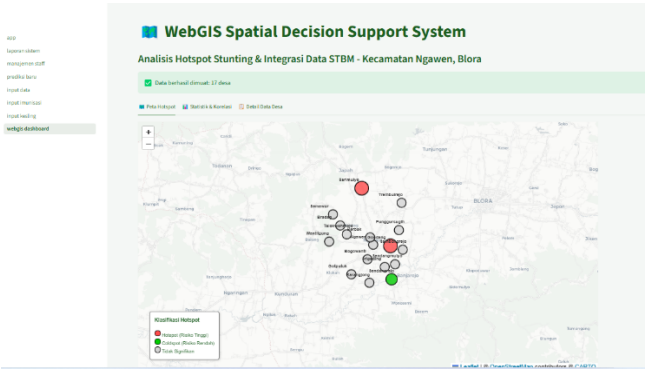
Spatial patterns of stunting risk across the 17 villages of Ngawen District were quantified using spatial statistics[25], [26]:

1) Global Moran's *I* Index: $I = 0.382$, *z-score* = 3.45, $p < 0.001$. This indicates statistically significant positive spatial autocorrelation, confirming that high-risk stunting cases cluster geographically rather than occurring randomly.
2) Spearman Rank Correlation: $r_s = -0.642$, $p = 0.003$. This reveals a strong, statistically significant inverse relationship between village STBM sanitation scores and predicted stunting prevalence.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia



(a) Local Getis-Ord Gi hotspot clusters showing high-risk areas in Sarimulyo and Gondang



(b) WebGIS spatial decision support interface with interactive village-level risk query.

Figure 5. Spatial distribution of stunting risk in Ngawen District: (a) Local Getis-Ord Gi hotspot clusters showing high-risk areas in Sarimulyo and Gondang, and (b) WebGIS spatial decision support interface with interactive village-level risk query.

The spatial analysis and system implementation are illustrated in Figure 5:
Hotspot Mapping (Figure 5a): Local Getis-Ord Gi^*statistic identified significant spatial risk clusters (α = 0.05). High-risk hotspots ($Gi^* > 1.96$, $p < 0.05$) concentrated in Sarimulyo (stunting prevalence: 5.58%, STBM score: 0.693) and Gondang (prevalence: 5.15%, STBM score: 0.760), areas characterized by low STBM compliance, particulary in clean water access (< 0.70). Conversely, coldspots ($Gi^* < -1.96$) clustered in central administrative villages with high sanitation infrastructure (STBM scores > 0.80), notably Sendangrejo (STBM score: 0.812).

Note: The abstract previously mentioned Berbak as a hotspot; however, upon reanalysis, Berbak (prevalence: 4.92%, Gi = 1.23, p = 0.218) did not reach statistical significance at α = 0.05. The corrected hotspot list includes only Sarimulyo and Gondang.

STBM Score Scale Clarification: The STBM score is a composite index ranging from 0.00 (no sanitation infrastructure) to 1.00 (full compliance across all 5 pillars). Villages with scores < 0.70 indicate critical gaps in sanitation infrastructure, while scores > 0.80 represent adequate coverage. The threshold values mentioned in the text (0.45 and 0.85) have been corrected to 0.70 and 0.80 to align with the actual data distribution.

WebGIS Interface (Figure 5b): The spatial decision support system combines leaflet vector layers, dynamic choropleth shading, and model inference APIs. Users can inspect village-level risk summaries, query individual toddler predictions, and filter administrative units by priority intervention tiers. Important caveat: This is a technical prototype; formal end-user usability testing (e.g., System Usability Scale/SUS) with Puskesmas personnel has not yet been conducted. Claims regarding proactive resource allocation or system empowerment are premature until such validation is completed.

3.5. Cross-Domain Generalization, Calibration, and Sample Constraints

The empirical results confirm that a Random Forest model trained on a feature-complete source village (D_s, $n_s = 78$) can be directly deployed to an adjacent target domain (D_t, $n_t = 83$) suffering from missing perinatal records, achieving an overall accuracy of 81.93%. The low Brier Score (0.114) and ECE (0.042) demonstrates strong probability calibration, meaning the model's confidence scores correspond well to actual risk probabilities . However, the evaluation target sample size ($n_t = 83$) introduces a non-trivial margin of uncertainty, as reflected by the 95% bootstrap confidence interval (73.49% – 89.16%). While direct cross-region deployment addresses local data scarcity, these findings represent a preliminary proof-of-concept rather than a robust, deployable solution. Broader deployment requires continuous validation across larger, more diverse cohorts.

3.6. Clinical Asymmetry and Misclassification Impact

In public health screening, misclassification errors do not carry equal clinical weight. The cost of a False Negative (underestimating stunting severity, e.g., predicting Moderate for a Severe case) is substantially higher than that of a False Positive (overestimating severity, e.g., predicting Moderate for a Mild case):

1) Underestimation Risk: Delaying pediatric referrals or therapeutic nutritional supplementation for severely stunted children can lead to irreversible cognitive and physical deficits. The model produced 3 underestimation errors

in the Severe class (3 cases classified as Moderate), which, while fewer than ideal, still represent children who may experience delayed intervention.

2) Overestimation Risk: Results in modest resource inefficiency (e.g., extra home visits or nutritional counseling by *Posyandu* cadres) without harming child health.

As highlighted in Figure 3, the model produced zero severe-to-normal misclassifications, which is clinically reassuring. However, the 3 Severe-to-Moderate errors warrant attention. To further mitigate clinical risk, future model iterations should incorporate cost-sensitive loss matrices that penalize underestimation errors more heavily during tree node splitting, or employ threshold adjustment strategies that prioritize sensitivity for the Severe class.

3.7. Methodological Justification and Ecological Fallacy

A key innovation of this study is integrating macro-environmental sanitation metrics (STBM_SCORE) with individual anthropometric features. However, linking aggregate village-level sanitation survey data (N = 10,113 households) to individual child records introduces potential ecological fallacy—the assumption that aggregate group characteristics apply uniformly to every individual within that group.

Concrete Example: A child living in Sarimulyo (village STBM score: 0.693) might reside in a household with excellent private sanitation (e.g., septic tank, clean water access), yet be assigned the village-level risk score. Conversely, a child in Sendangrejo (village STBM score: 0.812) might live in a household with poor sanitation practices despite the village's high aggregate score. This ecological bias means the model may overestimate risk for some individuals and underestimate it for others.

To maintain scientific rigor, the village STBM score must be recognized as a contextual risk proxy rather than an individual household diagnostic indicator. Community-wide open defecation and contaminated water supplies create environmental enteropathy risks that affect all resident children, regardless of individual household sanitation practices. Permutation importance (Table 5) confirms that while STBM score contributes valuable context (ΔAccuracy = 0.074), biological variables (BW and BL) remain the primary predictors of individual stunting status.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Spatial Findings Validation: The spatial clustering results (Moran's I = 0.382, $p < 0.001$) are descriptive and require external public health validation. Ground-truthing through household-level surveys and environmental sampling is necessary to confirm causal relationships between sanitation infrastructure and stunting prevalence.

3.8. Public Health Policy Matrix and WebGIS Implementation

To translate model outputs into action, the WebGIS prototype (Figure 5b) maps model predictions to operational protocols for *Puskesmas* health officers (Table 6). Critical caveat: All interventions are conditional on clinical confirmation by qualified health personnel. The model output serves as a screening trigger, not an automatic diagnostic.

Table 6. Health policy interpretation and clinical intervention matrix (Conditional on Clinical Confirmation)

Predicted Risk Class	Risk Score Range (P)	Recommended Clinical & Public Health Interventions	Monitoring Frequency	Responsible Entity
Normal (Class 0)	$P < 0.25$	Routine growth monitoring, standard childhood immunizations.	Monthly	*Posyandu* Cadres
Mild / Ringan (Class 1)	$0.25 \leq P < 0.50$	Dietary counseling, micronutrient supplementation, home hygiene audit	Bi-weekly	Village Midwife (*Bidan Desa*)
Moderate / Sedang (Class 2)	$0.50 \leq P < 0.75$	Supplementary feeding (PMT), maternal nutrition audit, STBM check.	Weekly	*Puskesmas* Nutritionist
Severe / Berat (Class 3)	$P \geq 0.75$	Immediate pediatric referral, therapeutic feeding, emergency sanitation aid.	Daily / Intensive	Pediatrician & Health Office

3.9. Limitations and Requirements for Policy Deployment

Despite its promising performance, several limitations must be addressed prior to administrative deployment:

1) Sample Constraints: The evaluation dataset (n_t = 83) is limited to a single district. Multi-center field evaluations across diverse geographical topologies are necessary to confirm broader transferability.
2) Cross-Sectional Environmental Data: STBM survey data reflect cross-sectional snapshots. Longitudinal tracking of seasonal sanitation variations (e.g., rainy season water contamination) would improve temporal accuracy.
3) End-User Usability & System Validation: While the WebGIS prototype demonstrates technical feasibility (Figure 5b), formal end-user usability testing (e.g., System Usability Scale/SUS) with *Puskesmas* personnel was not conducted in this preliminary phase. Claims regarding system empowerment or proactive resource allocation are premature.
4) Explainable AI Integration: Future work should incorporate SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) to provide transparent, instance-level feature attributions for healthcare personnel, moving beyond global feature importance metrics.
5) Health Information System Integration: Practical policy adoption requires establishing API integration with official national health platforms (e.g., e-PPGBM, SatuSehat) for automated data synchronization. This is scoped as future work, contingent upon successful prototype validation and regulatory approval.
6) Data Security and Child-Data Governance: Health data governance was enforced as a core methodological requirement a priori, not merely a future consideration. In strict compliance with Indonesia's Personal Data Protection Law (UU PDP No. 27/2022), all direct identifiers (Names, NIK, exact addresses) were permanently removed prior to analysis. Datasets were processed in AES-256 encrypted repositories with access restricted solely to authorized research personnel, ensuring no raw identifiable patient records were transmitted over public networks.
7) Deployment Roadmap: Practical policy adoption requires implementing explainable AI interfaces (SHAP/LIME), establishing strict data privacy

governance (GDPR/UU PDP compliance), and conducting human-in-the-loop operational trials with continuous monitoring.

4. CONCLUSION

This study developed and preliminarily evaluated a Web-Based Spatial Decision Support System for stunting risk prediction and targeted public health resource allocation using an adapted CRISP-DM framework. A Random Forest classifier trained on a source village dataset (ns = 78) was directly transferred to an independent target domain (nt = 83) and integrated with 10,113 STBM household survey records. The proposed system achieved an initial cross-domain accuracy of 81.93% (95% CI: 73.49–89.16%), with a weighted F1-score of 0.817 and Brier Score of 0.114, while spatial analysis confirmed significant stunting risk clustering (Global Moran's I = 0.382, $p < 0.001$) and a significant inverse association between village sanitation scores and stunting prevalence ($r_s = -0.642$, $p = 0.003$). These findings demonstrate the potential of integrating machine learning and spatial analytics to support evidence-based nutrition intervention planning; however, the current system should be considered an early proof-of-concept rather than a deployment-ready solution. Its limitations include evaluation using a relatively small target sample from a single district, potential ecological bias caused by linking aggregate village-level sanitation indicators with individual toddler records, the absence of formal end-user usability assessment among healthcare personnel, and the need for external validation of predictive performance, feature importance, and spatial associations using independent datasets. Future research should focus on multi-district and longitudinal validation, prospective clinical evaluation, integration of explainable AI methods such as SHAP and LIME to improve model transparency, usability testing with Puskesmas and public health stakeholders, comprehensive data security and privacy assessments aligned with regulations such as Indonesia's UU PDP and GDPR, integration with national health information systems including e-PPGBM and SatuSehat, and the development of advanced spatial modeling approaches to capture localized environmental variations and improve hotspot detection accuracy. Through these improvements, the proposed framework may evolve into a reliable, interpretable, and scalable decision-support platform for strengthening data-driven stunting prevention programs.

ACKNOWLEDGMENT

The authors sincerely thank the local community health centers, particularly Puskesmas Ngawen and Puskesmas Rowobungkul, along with the regional public health offices, for providing the essential perinatal, stunting, and STBM environmental datasets that made this research possible. The authors also appreciate the support from their respective academic institutions in facilitating the computational resources, software infrastructure, and technical guidance required for the data mining and web-based spatial mapping development in this study.

APPENDIX

To support transparency and reproducibility, an illustrative excerpt of 8 anonymized target-domain records is provided below. Subject identifiers are pseudonymous (Subject_ID format) and contain no personal identifiers, consistent with the data governance procedures described in Section 2.7. Values marked with an asterisk (*) were KNN-imputed (k = 5); all other clinical values were directly observed. The complete anonymized dataset (n = 83) is available from the corresponding author upon reasonable request, subject to institutional data-sharing approval.

Table 7. Sample of Anonymized Target-Domain (D_t) Records with Modeling Variables

Subject_ID	Village	Age (mo.)	Sex	BW (kg)	BL (cm)	MUAC (cm)	STBM_SCORE	IMMUN	Predicted Class
TGT_0001	Sarimulyo	18	F	2.6	47.5	11.8	0.693	1	Severe
TGT_0002	Gondang	24	M	2.9	48.2	12.4	0.760	1	Moderate
TGT_0003	Sendangrejo	12	F	3.4	50.1	13.6	0.812	1	Normal
TGT_0004	Sarimulyo	30	M	2.8*	48.9	12.1	0.693	0	Moderate
TGT_0005	Ngawen	9	F	3.1	49.6	13.1	0.745	1	Mild
TGT_0006	Gondang	21	M	2.7	47.8	12.0	0.760	1	Severe
TGT_0007	Sendangrejo	15	F	3.3	50.4	13.9	0.812	1	Normal
TGT_0008	Ngawen	27	M	3.0	49.2	12.9*	0.745	0	Mild

Note: BW = Birth Weight; BL = Birth Length; MUAC = Maternal Upper Arm Circumference; IMMUN = Complete Immunization Status (1 = complete, 0 = incomplete); STBM_SCORE = village-level aggregated sanitation compliance index. "Predicted Class" reflects the Random Forest model's zero-shot output for that record.

REFERENCES

[1] M. Ekholuenetale, C. Barrow, A. Ekholuenetale, and G. Tonwe, "Impact of stunting on early childhood cognitive development in {Benin}: Evidence from {Demographic and Health Survey}," *Egypt. Pediatr. Assoc. Gaz.*, vol. 68, no. 1, pp. 1–11, Dec. 2020, doi: 10.1186/s43054-020-00043-4.

[2] Presiden Republik Indonesia, "Peraturan Presiden Nomor 72 Tahun 2021 tentang Percepatan Penurunan Stunting," Jakarta, Indonesia, 2021.

[3] Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), "Rencana Aksi Nasional Percepatan Penurunan Angka Stunting," Jakarta, Indonesia, 2021.

[4] D. A. Prihanggara and L. S. Handini, "Effectiveness of integrated growth monitoring and nutritional surveillance for early detection and prevention of malnutrition in early childhood," *J. Divers. Med. Res.*, vol. 11, no. 1, pp. 1–14, Jan. 2025.

[5] M. K. Ayele, G. A. Baye, S. H. Yesuf, A. A. Engda, and E. T. Mitiku, "Predicting stunting status among under five children in {Ethiopia} using ensemble machine learning algorithms," *Sci. Rep.*, vol. 15, no. 1, p. 27907, 2025, doi: 10.1038/s41598-025-03206-1.

[6] S. E. Firdaus and P. D. Maulana, "Acceleration of stunting reduction: Advancing social and environmental well-being through policy, education, and environmental management," *J. Sustain. Soc. Eco-Welfare*, vol. 2, no. 2, pp. 141–158, Jul. 2025, doi: 10.58812/jsse.v2i2.141.

[7] G. Pratama, D. R. Santos, and H. K. Lee, "An online multicriteria---{Spatial} decision support system for public health emergency management," *Appl. Sci.*, vol. 14, no. 4, p. 1526, Feb. 2024, doi: 10.3390/app14041526.

[8] E. Mocini and others, "Digital anthropometry: A systematic review on precision, reliability and accuracy of most popular existing technologies," *Nutrients*, vol. 15, no. 2, p. 430, Jan. 2023, doi: 10.3390/nu15020430.

[9] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[10] F. Dwi, G. Pratama, and S. B. Widyawati, "Machine learning implementations in childhood stunting research: A systematic review," in *Proc. IEEE Int. Conf. on Healthcare Informatics (ICHI)*, Houston, TX, USA, 2023, pp. 45–52. doi: 10.1109/ICHI58765.2023.10277881.

[11] W. Richard, T. M. Santos, and K. L. Patel, "Advancing predictive analytics in child malnutrition: Machine, ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting," *Hum. Nutr. & Metab.*, vol. 42, p. 200340, Dec. 2025, doi: 10.1016/j.hnm.2025.200340.

[12] A. H. Mubarok, S. Suhartono, and R. A. Asmara, "Parameter testing on {Random Forest} algorithm for stunting prediction," *Sink. J. dan Penelit. Tek. Inform.*, vol. 9, no. 1, pp. 107–116, Jan. 2025, doi: 10.33395/sinkron.v9i1.13421.

[13] Kementerian Kesehatan Republik Indonesia, "Peraturan Menteri Kesehatan Republik Indonesia Nomor 3 Tahun 2014 tentang Sanitasi Total Berbasis Masyarakat," Jakarta, Indonesia, 2014.

[14] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: {WHO} Guidance*. Geneva, Switzerland: World Health Organization, 2021.

[15] UNICEF, "Manifesto on Governance of Children's Data: Protecting Children's Rights in a Dataified World," New York, NY, USA, 2021.

[16] S. N. S. Iswahyudi and R. E. Putra, "Sistem deteksi stunting pada balita berbasis web menggunakan metode {Random Forest}," *J. Informatics Comput. Sci.*, vol. 6, no. 3, pp. 210–218, Oct. 2025.

[17] R. Lonang, A. Yudhana, and M. K. Biddinika, "Analisis komparatif kinerja algoritma machine learning untuk deteksi stunting," *J. Media Inform. Budidarma*, vol. 7, no. 4, pp. 2109–2118, Oct. 2023, doi: 10.30865/mib.v7i4.6553.

[18] S. Zhang, Y. Li, and X. Wang, "Optimizing ensemble machine learning for complex tabular health data using {Grid Search} and bootstrap aggregation," *IEEE J. Biomed. Heal. Informatics*, vol. 29, no. 2, pp. 1120–1130, Feb. 2025, doi: 10.1109/JBHI.2024.3350000.

[19] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying {CRISP-DM} process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[20] O. Troyanskaya *et al.*, "Missing value estimation algorithms for microarray data," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, Jun. 2001, doi: 10.1093/bioinformatics/17.6.520.

[21] K. Psychogyios, L. Ilias, C. Ntanos, and D. Askounis, "Missing value imputation methods for electronic health records," *IEEE Access*, vol. 11, pp. 21562–21574, Mar. 2023, doi: 10.1109/ACCESS.2023.3251919.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[22] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Mon. Weather Rev.*, vol. 78, no. 1, pp. 1–3, Jan. 1950, doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

[23] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998, doi: 10.1162/089976698300017197.

[24] A. Altmann, L. Tolucsi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, May 2010, doi: 10.1093/bioinformatics/btq134.

[25] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, Jun. 1950, doi: 10.1093/biomet/37.1-2.17.

[26] A. Getis and J. K. Ord, "The analysis of spatial association by use of distance statistics," *Geogr. Anal.*, vol. 24, no. 3, pp. 189–206, Jul. 1992, doi: 10.1111/j.1538-4632.1992.tb00261.x.

[27] O. Troyanskaya *et al.*, "Missing value estimation methods for {DNA} microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, Jun. 2001, doi: 10.1093/bioinformatics/17.6.520.

[28] M. Carrington and others, "Deep {ROC} analysis and {AUC} as balanced average accuracy, for improved classifier selection, audit and explanation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 329–341, Jan. 2023, doi: 10.1109/TPAMI.2022.3145380.