

Article

Clinical and Sociodemographic Predictors of Tuberculosis and Malnutrition Among Under-Five Migrant Children in Urban Pakistan: A Community-Based Cross-Sectional Study

Javeria Saleem ¹, Afia Zafar ², Abida Tehreem ³, Farooq Manzoor ⁴, Zulfiqar Ali ⁵ , Asad Ullah Mahmood ⁶, Fatima Ali ³, Enrique Roche ^{7,8,9,*} , Ana M. Celorrio San Miguel ¹⁰, Elena Jiménez-Callejo ¹¹, Juan Mielgo-Ayuso ^{12,†}  and Diego Fernández-Lázaro ^{10,13,14,*} 

¹ Department of Public Health, Green International University, Lahore 53000, Pakistan

² Sir Ganga Ram Hospital, Lahore 54000, Pakistan

³ Department of Public Health, University of the Punjab, Lahore 54590, Pakistan

⁴ Additional Secretary Services & General Administration Department, University of the Punjab, Lahore 54000, Pakistan

⁵ College of Statistical Sciences, University of the Punjab, Lahore 54590, Pakistan

⁶ Jinnah Hospital, Lahore 54030, Pakistan

⁷ Department of Applied Biology-Nutrition, Institute of Bioengineering, Miguel Hernández University, 03202 Elche, Spain

⁸ Alicante Institute for Health and Biomedical Research (ISABIAL), 03010 Alicante, Spain

⁹ CIBER Physiopathology of Obesity and Nutrition (CIBEROBN), Carlos III Health Institute (ISCIII), 28029 Madrid, Spain

¹⁰ Neurobiology Research Group, Faculty of Medicine, University of Valladolid, 47005 Valladolid, Spain

¹¹ Research Support Unit, Healthcare Management of Soria, Santa Bárbara Hospital, Soria University Hospital Complex (CAUSO), 42005 Soria, Spain

¹² Department of Health Sciences, Faculty of Health Sciences, University of Burgos, 09001 Burgos, Spain

¹³ Histology Area, Faculty of Health Sciences, University of Valladolid, Soria Campus, 42004 Soria, Spain

¹⁴ Biomedical Research Institute of León (IBIOLEÓN), University Hospital Complex of León, 24071 León, Spain

* Correspondence: eroche@umh.es (E.R.); diego.fernandez.lazaro@uva.es (D.F.-L.)

† Senior Authors.

Abstract

Tuberculosis (TB) and malnutrition remain major health challenges among migrant children living in low-resource urban settings. Identifying clinically relevant patterns associated with these conditions may support standardized clinical assessment and public health planning. The present study aimed to identify clinical, anthropometric and sociodemographic factors associated with TB and malnutrition among migrant children under five years of age in urban Pakistan. A community-based cross-sectional study was conducted in migrant settlements of Rachna Town, Lahore, Pakistan. Clinical, demographic and anthropometric data were collected from 300 children under five years of age. A Genetic Algorithm (GA) combined with Random Forest analysis was applied to identify the most informative variables associated with TB and malnutrition. Model performance was assessed using internal cross-validation. The analytical approach identified a concise subset of variables showing stable internal performance. Key factors associated with malnutrition included Bacillus Calmette-Guérin vaccination status, formula feeding, chest X-ray findings and gastrointestinal symptoms. The most relevant variables associated with TB were Bacillus Calmette-Guérin vaccination score, erythrocyte sedimentation rate, gastrointestinal indicators and parental occupation. The final models demonstrated good internal discrimination and substantial agreement. These findings suggest that routinely collected clinical, anthropometric, laboratory, radiological, and sociodemographic variables may support the identification of clinically relevant patterns associated with TB and malnutrition during standardized community-based assessments. The proposed data-driven approach should



Academic Editor: Salem Al-Tamemi

Received: 13 June 2026

Revised: 30 July 2026

Accepted: 4 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Published by MDPI on behalf of the Oman Medical Association. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

be interpreted as an analytical and decision-support framework for prioritizing informative variables and supporting clinical assessment, rather than as a pre-diagnostic or early risk prediction tool.

Keywords: tuberculosis; malnutrition; migrant children; risk factors; public health; urban health; pediatric health; cross-sectional study

1. Introduction

Tuberculosis (TB) and malnutrition remain two of the most important public health challenges affecting children under five years of age in low-resource settings. Their co-existence exerts a synergistic effect that substantially increases morbidity and mortality, particularly among displaced or migrant populations [1–4]. Slum environments, commonly characterized by poverty, overcrowding and inadequate sanitation, provide favorable conditions for TB transmission and the persistence of chronic undernutrition [1,5].

Globally, TB continues to be one of the leading causes of infectious disease-related mortality [2]. According to the World Health Organization (WHO), nearly one million pediatric TB cases occur annually, representing approximately 10% of the global burden [6–8]. Pakistan ranks fifth worldwide in TB incidence and bears the highest burden in South Asia [9,10]. Children account for ~11% of the national TB caseload [11], and the prevalence of malnutrition in this population remains alarmingly high, contributing to the country's fragile public health situation.

The relationship between malnutrition and TB is bidirectional. Malnutrition weakens immune defenses and increases susceptibility to infection, whereas TB exacerbates wasting and micronutrient deficiencies [3,4,12–16]. Studies conducted in South Asia have reported that the prevalence of TB among severely malnourished children may reach 30–40% [6,16]. This complex interaction is further aggravated by displacement and poverty. Internally displaced populations in Pakistan, many of whom originate from the conflict-affected Federally Administered Tribal Areas, frequently settle in urban slums characterized by poor housing conditions, limited access to healthcare and food insecurity [10]. These socio-environmental factors intensify vulnerability to communicable diseases and nutritional deficiencies, particularly in young children.

Although global initiatives aligned with the sustainable development goals aim to reduce TB incidence and mortality by 2030 [7], surveillance systems in low-income countries remain constrained by limited diagnostic and analytical capacity. Traditional epidemiological approaches are often insufficient to identify the complex and multidimensional determinants of disease risk. In this context, computational analytical approaches have emerged as promising tools for identifying key predictors of health outcomes in pediatric populations [17]. However, their application to combined TB-malnutrition risk assessment remains limited, particularly in community-based settings.

Moreover, most previous studies have examined tuberculosis and malnutrition as independent conditions using conventional epidemiological or statistical approaches [12,16,17]. Consequently, limited evidence is available regarding integrated analytical models capable of simultaneously identifying the multidimensional clinical, anthropometric, and sociodemographic factors underlying both conditions in highly vulnerable pediatric populations [17]. Addressing this knowledge gap is essential for improving the identification of clinically relevant patterns associated with TB and malnutrition and supporting targeted public health interventions in resource-constrained settings [7,17].

Based on these considerations, the present study was designed to address this important knowledge gap by integrating routinely collected clinical, anthropometric, and sociodemographic information with a hybrid machine-learning framework. We hypothesized that combining feature-selection optimization with supervised classification could improve the identification of clinically relevant clinical and sociodemographic patterns associated with TB and malnutrition. To address this gap, a genetic algorithm was applied to identify the most informative clinical, demographic and anthropometric variables associated with TB and malnutrition among under-five migrant children living in an urban slum in Lahore, Pakistan. Despite its name, the genetic algorithm used in the present study refers to a computational optimization technique inspired by evolutionary processes rather than genetic or molecular data. By efficiently selecting a subset of relevant features, this approach may help us to identify context-specific clinical and sociodemographic patterns associated with tuberculosis and malnutrition during standardized community-based clinical assessments. Beyond prioritizing informative variables, the proposed analytical framework provides an interpretable and scalable approach that may support clinical decision-making and evidence-based public health planning in vulnerable pediatric populations.

2. Materials and Methods

2.1. Study Design

A community-based analytical cross-sectional study with embedded structured field observations was conducted over a six-month period (March–August 2024). This design was selected to assess a defined population at a single point in time and to characterize the interaction between socioeconomic conditions and child health outcomes. Quantitative assessments, including clinical, anthropometric and sociodemographic data, were combined with an explanatory triangulation approach, whereby field observations were used to contextualize quantitative patterns of malnutrition and TB among children under five years of age.

Field observations followed a pre-specified checklist covering key environmental health domains, including water access and sanitation, solid waste disposal, household crowding and ventilation, and hygiene behaviors. Observers received standardized training and pilot instruction before data collection. Due to the cross-sectional nature of the study, temporal relationships and causality could not be established. As recruitment was conducted between March and August in a setting characterized by seasonal migration and disease variation, potential seasonal effects were minimized by distributing enrolment evenly across all study months and recording the month of assessment for adjustment in analyses when applicable.

2.2. Study Setting

The study was conducted in Rachna Town, an informal urban settlement located on the outskirts of Lahore, Pakistan, characterized by high population density, limited primary healthcare coverage, suboptimal sanitation and widespread poverty. According to government records, a substantial proportion of residents are internally displaced individuals originating from the former Federally Administered Tribal Areas, with seasonal mobility associated with informal employment.

The site was purposefully selected based on the following operational criteria: (i) Reports from nearby primary healthcare units indicating high notification rates of presumed and confirmed TB among children and adults; (ii) extensive household crowding and poor ventilation; (iii) limited access to safe water and sanitation infrastructure; and (iv) feasibility of sustained community engagement. Within the accessible catchment area, the study included the majority of children under five years of age attending community screening

activities during the study period, thereby enhancing the internal validity of the findings within this setting. Although external validity is naturally limited to comparable informal urban settlements with similar structural vulnerabilities, the mechanisms examined in the present study, including overcrowding, hygiene and healthcare access, are common across similar environments, supporting analytical transferability.

2.3. Sample Selection

The study population comprised children under five years of age residing in Pathan Colony, Rachna Town, a settlement predominantly inhabited by migrant families who temporarily relocate during winter. Although household mobility is common, most families remain in the area between March and August, allowing stable recruitment throughout the study period.

A total of 300 children were enrolled. The sample size was determined pragmatically from a preliminary community census identifying ~360 eligible children under five years of age; therefore, the final sample represented >80% of the accessible population. Although a formal power calculation was not required due to the census-based design, a posteriori estimation indicated that 200 participants would provide 80% power ($\alpha = 0.05$) to detect moderate effects (Cohen's $h = 0.5$); therefore, the final sample size exceeded this threshold.

As probability sampling was not feasible due to the transient nature of the population, participants were recruited consecutively from all households meeting the inclusion criteria: (i) Age < 5 years; and (ii) residence in Rachna Town for at least one month before enrolment. Exclusion criteria included age ≥ 5 years and lack of parental or guardian consent. The one-month residence criterion ensured recent exposure to local environmental conditions without excluding short-term settlers, who are common within this community. Although sampling was non-probabilistic, the approach achieved broad population coverage and minimized selection bias within the study context.

2.4. Data Collection Tools

2.4.1. Community Health Profiling

To characterize the environmental and socioeconomic context, a structured community health checklist was developed based on WHO and United Nations Children's Fund field-assessment frameworks and reviewed by three senior public health experts to ensure content validity. The instrument included information on household composition, access to clean water, sanitation facilities, housing materials, waste disposal and household crowding. The checklist was piloted in 10 households to ensure clarity and inter-observer reliability ($\kappa > 0.80$).

Data were obtained through direct non-participant observation and brief interviews with household members, with immediate recording in structured forms to minimize observer bias. Field data collection was performed by trained community health workers under continuous supervision by a senior investigator to ensure procedural consistency and quality control.

2.4.2. Screening and Clinical Examination

All children under five years of age residing in the study area underwent screening and clinical examination after written informed consent had been obtained from parents or legal guardians. Examinations were performed at temporary community health posts within the settlement by qualified pediatric healthcare professionals following the WHO Integrated Management of Childhood Illness guidelines and national pediatric assessment standards.

Clinical evaluation included body temperature, respiratory rate, signs of dehydration, lymphadenopathy, cough, respiratory distress, skin and hair changes, and anthropometric indicators, including weight, height and mid-upper arm circumference. All findings were

recorded using pre-tested case report forms translated into Urdu and Pashto to ensure comprehension and completeness.

2.5. Data Collection

2.5.1. Disease Frequency and Nutritional Status Assessment

The prevalence of common morbidities, including respiratory infections, diarrheal disease and dermatological conditions, together with nutritional status, was assessed through standardized clinical and anthropometric evaluations. Anthropometric measurements included weight, height or length, mid-upper arm circumference and head circumference using calibrated equipment.

All anthropometric data were converted into Z-scores (weight-for-age, height-for-age and weight-for-height) according to the WHO Child Growth Standards (2006) [18]. Children presenting any anthropometric index below -2 standard deviations were classified as malnourished.

Clinical examinations documented nutritional signs and morbidity patterns. Laboratory investigations included complete blood count, erythrocyte sedimentation rate (ESR), chest radiography and stool examination. Bacillus Calmette-Guérin (BCG) vaccination status was confirmed through inspection of vaccination scars and verification of medical records, with medical records prioritized in cases of discrepancy. All procedures followed established pediatric field guidelines [11] and WHO malnutrition assessment protocols and were performed by trained healthcare professionals under pediatric supervision.

2.5.2. Diagnosis of Tuberculosis

TB screening was performed using the Kenneth John Scoring System (KJSS), which evaluates ten clinical features commonly associated with pediatric TB, adapted from Kenneth John et al. [19] and previously validated in pediatric screening settings. Each feature was scored according to standardized criteria, whereby scores of 1–2 indicated TB unlikely, scores of 3–4 required clinical observation and scores ≥ 5 indicated a high likelihood of TB. Children with scores ≥ 3 were referred for confirmatory evaluation, including chest radiography and tuberculin skin testing.

Chest radiographs were independently interpreted by two radiologists blinded to the clinical data, and discordant findings were adjudicated by a pediatric TB specialist. Final TB diagnosis was established through comprehensive clinical evaluation and specialist review according to WHO recommendations for childhood TB management.

2.5.3. Anthropometric Measurements and Classification

Anthropometric measurements were obtained according to standardized WHO procedures [18]. Measurements included weight, height or length, mid-upper arm circumference and occipital-frontal circumference, the latter being assessed only in children under 24 months of age. Weight and height measurements were obtained in duplicate using calibrated instruments, and average values were recorded. Technical error of measurement thresholds remained below accepted standards.

Anthropometric data were converted into Z-scores according to the WHO Child Growth Standards (2006) [20]. Children were subsequently classified as normal, moderately malnourished or severely malnourished according to the lowest Z-score among the evaluated indices.

2.6. Training of Data Collectors

All data collectors participated in a three-day intensive training workshop covering study objectives, ethical procedures, standardized anthropometric techniques and TB screening procedures. Pilot testing and supervised role-playing sessions were con-

ducted, and competence was verified through post-training evaluation requiring at least 90% concordance with reference measurements.

2.7. Quality Control and Assurance

Quality assurance procedures included daily review of completed forms, random repeat measurements in 10% of participants and double data entry with cross-validation. Any discrepancies were resolved through comparison with the original records, and datasets were finalized only after complete verification.

2.8. Data Analysis

Descriptive statistics were used to summarize sociodemographic characteristics, nutritional indicators and TB prevalence. Group comparisons were performed using chi-square or Fisher's exact tests, as appropriate.

Predictive modelling was performed using a two-stage machine-learning framework. First, a Genetic Algorithm (GA) was applied as an optimization and feature-selection technique to identify the subset of variables with the greatest predictive relevance. Subsequently, the selected variables were used to train a supervised Random Forest (RF) classifier, in which the predefined outcome variables (nutritional status and tuberculosis status) served as class labels. Thus, the RF represented the supervised learning component of the analytical framework, whereas the GA was used exclusively for feature selection and optimization before model construction.

Because the outcome variables were known *a priori*, the Random Forest classifier was trained and internally validated using a supervised learning approach. Model performance was assessed by 10-fold cross-validation using accuracy, sensitivity, specificity, Cohen's kappa coefficient, and the area under the receiver operating characteristic curve (ROC-AUC).

The objective of the predictive models was to identify the combination of routinely collected clinical, anthropometric, laboratory, radiological, and sociodemographic variables providing the highest discriminatory performance within a standardized community-based clinical assessment, rather than to develop a pre-diagnostic screening algorithm based exclusively on baseline information available before clinical evaluation. Accordingly, the proposed framework should be interpreted as an analytical tool for identifying clinically relevant patterns within the overall assessment process, rather than as an early screening model intended for use before diagnostic investigations. The complete analytical workflow is summarized in Figure 1.

2.9. Ethical Considerations

The study protocol was reviewed and approved by the Institutional Research Committee (IRC) of Green International University (GIU), Lahore, Pakistan (Approval No. IRC-GIU-160-09-2024; 5 September 2024). All study procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki and complied with institutional ethical standards for research involving human participants. Written informed consent was obtained from the parents or legal guardians of all participating children before enrolment. Participant privacy and confidentiality were strictly maintained throughout the study by anonymizing all collected data and restricting access to authorized research personnel only. Community feedback sessions were conducted following completion of data collection to communicate the main findings and promote community engagement.

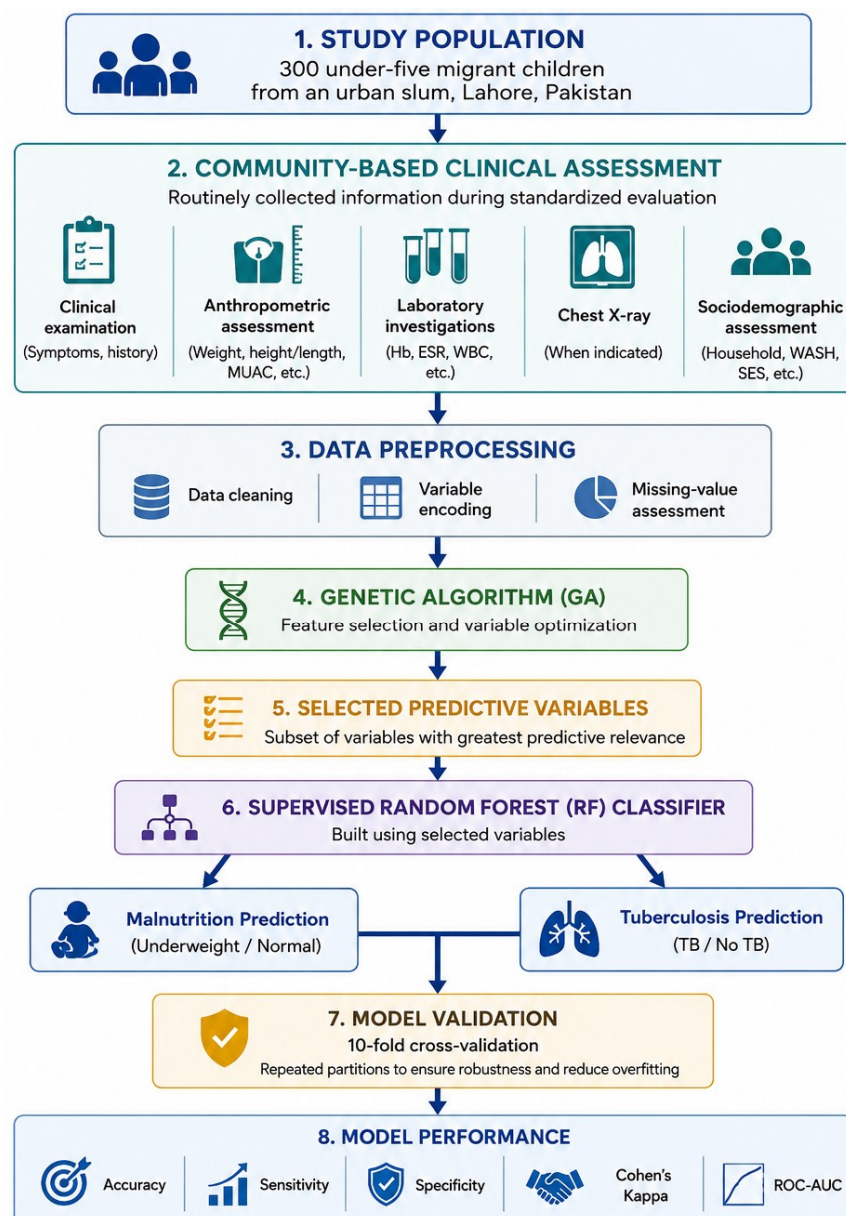


Figure 1. Study workflow and machine-learning framework for predicting malnutrition and tuberculosis among under-five migrant children. The analytical pipeline comprised community-based clinical assessment, data preprocessing, feature selection using a Genetic Algorithm (GA), supervised classification with a Random Forest (RF) algorithm, and internal model validation by 10-fold cross-validation. Model performance was evaluated using accuracy, sensitivity, specificity, Cohen’s kappa coefficient, and the area under the receiver operating characteristic curve (ROC-AUC). The workflow illustrates that the predictive models were developed using routinely collected information obtained during a standardized community-based clinical assessment, rather than solely baseline variables available before clinical evaluation.

3. Results

3.1. Observational Findings

Field observations documented marked environmental deprivation and precarious living conditions in Rachna Town. Most households lacked reliable access to clean water, adequate sanitation and organized waste disposal systems. Open drainage channels and accumulated solid waste were frequently observed near residential areas. Household overcrowding was

common, with multiple family units often sharing a single poorly ventilated room, conditions known to facilitate the transmission of respiratory and gastrointestinal infections.

Observational data were collected using structured field checklists and independently verified by two trained observers to improve consistency. Children frequently presented with respiratory symptoms, diarrheal episodes, skin lesions and visible clinical signs of malnutrition, including muscle wasting and stunting. Caregivers, primarily mothers, commonly reported chronic fatigue, anemia and limited access to preventive healthcare services, including routine childhood immunization.

Seasonal migration patterns were also reported to disrupt continuity of healthcare and clinical follow-up, contributing to incomplete vaccination schedules and delayed recognition and management of infectious diseases. These qualitative observations were consistent with the quantitative clinical and laboratory findings and provided contextual support for the subsequent analyses of malnutrition and TB presented in the following sections.

3.2. Sociodemographic Characteristics

A total of 300 children under five years of age were included in the analysis, of whom 269 (89.7%) were classified as malnourished and 31 (10.3%) as well nourished (Tables 1 and 2). No eligible participants were excluded due to missing data.

Table 1. Sociodemographic characteristics of malnourished children under 5 years ($n = 269$).

Characteristic	Frequency (n)	Percentage (%)	Statistical Comparison (p)
Age (months)			
<12 months	56	20.8	$\chi^2 = 1.56; p = 0.21$
12–36 months	107	39.8	
36–60 months	106	39.4	
Gender			
Male	136	50.6	$\chi^2 = 0.22; p = 0.64$
Female	133	49.4	
Father’s Education			
Not educated	264	98.5	Fisher = 0.33
Educated	4	1.5	
Mother’s Education			
Not Educated	269	100	— (all not educated)
Educated	0	0	
Father’s Occupation			
Unskilled worker	4	1.7	Fisher = 0.27
Skilled worker	225	97.4	
Not alive	2	0.9	
Mother’s Occupation			
Housewife	267	99.3	— (all housewives)
Unskilled worker	2	0.7	

Table 1. Cont.

Characteristic	Frequency (n)	Percentage (%)	Statistical Comparison (p)
Socio-economic Status			
Lower	132	49.1	Fisher = 0.78
Middle	130	48.3	
Upper middle	6	2.2	
Upper	1	0.4	

Abbreviations = χ^2 : Chi-square; <: less than.

Table 2. Sociodemographic characteristics of well-nourished children under 5 years ($n = 31$).

Characteristic	Frequency (n)	Percentage (%)	Statistical Comparison (p)
Age (months)			
<12 months	4	12.9	$\chi^2 = 1.56; p = 0.21$
12–36 months	13	41.9	
36–60 months	14	45.2	
Gender			
Male	17	54.8	$\chi^2 = 0.22; p = 0.64$
Female	14	45.2	
Father’s Education			
Not educated	29	93.5	Fisher = 0.33
Educated	1	3.2	
Mother’s Education			
Not Educated	31	100	— (all not educated)
Educated	0	0	
Father’s Occupation			
Unskilled worker	2	7.4	Fisher = 0.27
Skilled worker	25	92.6	
Mother’s Occupation			
Housewife	31	100	— (all housewives)
Unskilled worker	0	0	
Socio-economic Status			
Lower	16	51.6	Fisher = 0.78
Middle	14	45.2	
Upper middle	1	3.2	
Upper	0	0.0	

Abbreviations = χ^2 : Chi-square; <: less than.

Age distribution was comparable between groups. Most children were aged ≥ 12 months (malnourished, 79.2% vs. well nourished, 87.1%), with no statistically significant differences between groups ($\chi^2 = 1.56$; $p = 0.21$). Sex distribution was also balanced, with males representing 50.6% of the malnourished group and 54.8% of the well-nourished group ($\chi^2 = 0.22$; $p = 0.64$).

Parental education levels were uniformly low across both groups. Nearly all fathers were uneducated (malnourished, 98.5% vs. well nourished, 93.5%; Fisher's exact test, $p = 0.33$), and all mothers in both groups reported no formal education. Fathers were predominantly skilled laborers (malnourished, 97.4% vs. well nourished, 92.6%; Fisher's exact test, $p = 0.27$), whereas almost all mothers identified as housewives.

Socioeconomic status was similarly distributed between groups, with >95% of families classified within the lower or middle socioeconomic strata (Fisher's exact test, $p = 0.78$). No statistically significant differences were observed for any sociodemographic variable examined (all $p > 0.05$).

Overall, the malnourished and well-nourished cohorts were demographically comparable, reducing the likelihood that subsequent findings were confounded by baseline sociodemographic differences.

3.3. Feature Selection Using the Genetic Algorithm

A Genetic Algorithm (GA) was implemented to evaluate 48 clinical, 48 anthropometric, laboratory, radiological, and sociodemographic variables according to their discriminatory value for nutritional status classification. The algorithm evolved over 100 generations, with each generation producing 10 candidate models. Crossover and mutation probabilities were set at 0.8 and 0.1, respectively, to balance exploratory diversity and convergence stability.

The fitness function was designed to maximize overall classification accuracy using 10-fold cross-validation. To reduce the risk of overfitting, the evolutionary search process was terminated after five consecutive generations without performance improvement. The GA was combined with a Random Forest (RF) classifier consisting of 500 trees, which served as the supervised classification algorithm.

Model performance was primarily assessed using accuracy and Cohen's kappa, complemented by sensitivity, specificity and the area under the receiver operating characteristic curve (ROC-AUC) to evaluate discrimination and class balance. Across 1000 candidate models, the GA consistently retained a subset of clinical, anthropometric, laboratory, radiological, and sociodemographic informative variables that provided the highest classification performance for nutritional status classification. The most consistently selected informative variables included BCG vaccination status, formula feeding, chest X-ray findings, gastrointestinal symptoms, and anthropometric indicators, which were subsequently used to train the RF classifier.

The final GA-selected model comprised 16 informative variables spanning clinical signs, systemic conditions, anthropometric indices, and selected sociodemographic characteristics. This dimensionality reduction, from 48 candidate variables to 16 informative variables, substantially improved model interpretability while maintaining stable internal classification performance. A detailed description of the selected informative variables and their clinical relevance is presented in Table 3.

Table 3. Detailed description of Genetic Algorithm selected predictors for nutritional status and their clinical relevance.

Variable	Description	Relevance to Nutritional Status
Intermittent Symptoms	Periodic or recurrent manifestations of illness or feeding disruption.	Recurring infections or inconsistent food intake can directly affect a child's capacity to maintain adequate nutritional status, making this a key early predictor.
Dehydration	Clinical signs of fluid loss, often observed during diarrhoeal or febrile episodes.	Dehydration during illness can rapidly deteriorate nutritional status, especially in malnourished children, increasing the risk of severe clinical outcomes.

Table 3. Cont.

Variable	Description	Relevance to Nutritional Status
Cyanosis	Bluish discoloration of the skin caused by insufficient oxygenation.	Indicates impaired oxygen transport; may signal severe systemic disease or complications frequently associated with advanced malnutrition.
Dyspigmentation	Abnormal changes in skin pigmentation.	Serves as a visible marker of nutritional deficiencies, particularly in essential micronutrients such as zinc and vitamin A.
Gastrointestinal Symptoms	Includes vomiting, diarrhoea, or abdominal pain.	Reflects impaired nutrient absorption and chronic gastrointestinal disturbances that can lead to or worsen malnutrition.
Respiratory Symptoms	Cough, shortness of breath, or difficulty breathing.	Frequent respiratory infections are common in malnourished children and may further compromise growth and overall health.
ESR	Laboratory test measuring systemic inflammation or infection.	Elevated ESR reflects inflammatory stress and infection burden, both of which exacerbate nutritional decline.
Chest X-ray Findings	Radiological assessment of pulmonary and systemic pathology.	Identifies respiratory or systemic diseases such as TB or pneumonia, which are closely associated with malnutrition in children.
Kenneth John Scoring System	Clinical scoring tool for pediatric malnutrition.	Quantifies malnutrition severity based on observable physical signs and symptoms; useful for standardized field assessment.
Age Category	Classification of the child's age group.	Younger children are more susceptible to malnutrition due to higher metabolic needs and greater vulnerability to infection.
Weight Category	Classification of weight-for-age or weight-for-height percentile.	Direct anthropometric indicator of nutritional status; underweight values reflect chronic or acute malnutrition.
MUAC	Anthropometric measure of muscle and fat stores.	Sensitive indicator of acute malnutrition; low MUAC values strongly correlate with wasting and increased morbidity.
Father's Education	Educational attainment of the child's father.	Higher paternal education is associated with better awareness of nutritional needs and access to healthcare services.
Mother's Education	Educational attainment of the child's mother.	Maternal education is a key determinant of feeding practices, hygiene behaviors, and early healthcare-seeking actions.
Father's Occupation	Employment type or economic activity of the father.	Economic stability derived from the father's occupation influences food availability, living conditions, and child well-being.

Abbreviations = ESR = Erythrocyte Sedimentation Rate; MUAC = Mid-Upper Arm Circumference; TB: Tuberculosis. *Note* = Variables are ordered by clinical relevance and category (clinical, systemic, anthropometric, sociodemographic).

3.4. Model Performance and Variable Importance

The GA-RF hybrid model demonstrated stable internal classification performance. Preliminary tuning across pilot runs ranging from 50 to 200 generations indicated stable convergence at 100 generations, with optimal exploration achieved using a crossover probability of 0.8 and a mutation probability of 0.1. The fitness function jointly maximized accuracy, Cohen's kappa and the area under the receiver operating characteristic curve (ROC-AUC) under 10-fold cross-validation, thereby reducing potential bias associated with class imbalance (malnourished, ~90%; well nourished, ~10%).

The RF classifier, consisting of 500 trees, was optimized by tuning the *mtry* parameter, corresponding to the number of informative variables randomly sampled at each node split, across a range of values from 2 to 12. The configuration minimizing the out-of-bag error was selected for the final model. This hybrid GA-RF approach has previously been applied in biomedical feature-selection studies and provides an appropriate balance between interpretability and classification performance.

Permutation-based variable importance analysis confirmed the stability of feature rankings ($p < 0.01$). Variable importance was quantified using both Mean Decrease Accuracy and Mean Decrease Gini indices and is summarized in Table 4. The Kenneth John Scoring System (KJSS) emerged as the most influential predictor (Mean Decrease Gini = 7.32), followed by chest X-ray findings, erythrocyte sedimentation rate (ESR), and weight category. Together, these variables highlight the importance of clinical assessment, systemic inflammation, radiological findings, and growth impairment in the classification of nutritional status within the proposed analytical framework.

Table 4. Variable importance for the nutritional-status Random Forest (RF) model (Mean Decrease Accuracy and Mean Decrease Gini).

Variable	Mean Decrease Accuracy	Mean Decrease Gini	Interpretation
Kenneth John Scoring System	0.0313	7.3233	Highly relevant composite score summarizing disease severity; markedly enhances model accuracy.
Erythrocyte Sedimentation Rate	−0.0122	5.9415	Inflammatory biomarker indicating nutritional depletion and infection burden.
Weight Category	0.0212	4.9672	Anthropometric indicator reflecting underweight and growth faltering; moderate predictive contribution.
Chest X-ray Findings	0.0269	4.2972	Reflects pulmonary or systemic infection; strongly associated with morbidity linked to malnutrition.
Age Category	0.0073	2.7512	Captures age-related vulnerability to nutritional deficits; modest but relevant effect.
Mid-Upper Arm Circumference	0.0013	2.5279	Sensitive anthropometric marker of acute malnutrition; consistent positive association.
Father's Education	−0.0007	0.5238	Slight negative effect; limited predictive value due to low variability.
Gastrointestinal Issues	−0.0000	0.4074	Weak, context-dependent effect; contributes minimally to classification accuracy.
Respiratory Signs	0.0010	1.3422	Marginal impact on model performance; may add noise to predictive power.
Father's Occupation	−0.0026	1.1094	Negative contribution, suggesting low relevance for nutritional prediction.
Intermittent Symptoms	−0.0005	0.3046	Minimal contribution to discrimination; low prediction.
Dehydration	0.0003	0.1522	Represents acute physiological stress and inadequate fluid intake; minor positive influence.
Cyanosis	0.0003	0.1441	Indicator of hypoxia and severe systemic compromise; limited but physiologically meaningful relevance.

Table 4. Cont.

Variable	Mean Decrease Accuracy	Mean Decrease Gini	Interpretation
Mother's Education	0.0000	0.0244	No measurable impact on classification; reflects homogeneity in educational background.
Dyspigmentation	0.0000	0.0000	Not a significant predictor; negligible influence on classification.

Note = Importance values were normalized to facilitate comparison across predictors. Variables are ordered by overall importance (Mean Decrease Gini).

By contrast, sociodemographic variables, including parental education and occupation, exhibited negligible or negative importance values, suggesting limited contribution to classification performance after incorporation of clinical, anthropometric, laboratory, and radiological indicators. Negative Mean Decrease Accuracy values reflected unstable or redundant contributions rather than true inverse associations. Collectively, these findings support the internal consistency of the GA-RF framework for identifying clinically relevant patterns associated with malnutrition and suggest that biological and functional indicators exert a stronger influence than sociodemographic factors within this highly deprived population.

3.5. Summary Interpretation

The GA-RF framework identified a concise and clinically interpretable subset of predictors that demonstrated high internal validity, with an overall accuracy of 0.956, Cohen's kappa of 0.882 and a ROC-AUC of 0.97. These performance metrics indicated strong internal discrimination and agreement despite the marked class imbalance within the study population.

Clinical indicators such as dehydration, cyanosis, ESR and abnormal chest X-ray findings emerged as relevant markers of physiological deterioration associated with malnutrition. Together with the KJSS, these variables reflected the multidimensional interaction between infection, inflammation, and nutritional depletion within the standardized clinical assessment.

Although internal classification performance was high, the absence of external validation requires cautious interpretation of the findings. Therefore, the results should be considered exploratory and hypothesis-generating rather than diagnostic. Nevertheless, the analysis demonstrated that routinely collected clinical and anthropometric variables, without dependence on advanced laboratory infrastructure, may effectively discriminate nutritional status in community-based settings with limited resources.

Overall, these findings highlight the potential utility of the GA-RF approach as an analytical and decision-support framework for identifying clinically relevant patterns in vulnerable populations while acknowledging the methodological limitations inherent to internal validation.

3.6. Predictive Modeling for Tuberculosis Status

A second GA-RF model was developed to classify TB status using the same computational framework described for nutritional status. The algorithm evaluated 52 candidate variables encompassing clinical signs, anthropometric indicators, sociodemographic characteristics and laboratory parameters (Table 5). After 100 generations, with 10 candidate models generated per generation, 18 informative variables were retained in >50% of iterations, yielding stable internal discrimination.

Table 5. Detailed description of the predictive variables identified by the Genetic Algorithm and their relevance to tuberculosis status.

Variable	Description	Relevance to TB Status
Mother's Occupation	Type of maternal employment or domestic role.	The mother's occupation can influence household income and access to healthcare, which in turn affects TB prevention and treatment adherence.
Shortness of Breath	Clinical symptoms reflecting respiratory distress or difficulty breathing.	May indicate respiratory complications or infections such as TB, highlighting pulmonary involvement and potential disease severity.
Temperature	Body temperature measurement recorded during clinical evaluation.	Elevated temperature can indicate active infection, including TB, and is critical for early diagnosis and monitoring of treatment response.
Dehydration	Clinical sign of fluid loss, commonly associated with diarrhoeal or febrile conditions.	Dehydration can complicate TB treatment and recovery, particularly when coexisting with other infections or systemic inflammation.
Skin Condition	Observations of skin integrity, pigmentation, or lesions.	Reflects general nutritional and immune status, which are important for TB prognosis and response to therapy.
Clubbing	Enlargement of fingertips or toes due to chronic hypoxia.	May be associated with long-term pulmonary disease, including advanced TB, and indicates severity of respiratory compromise.
Dyspigmentation	Abnormal changes in skin pigmentation.	Can suggest underlying metabolic or nutritional disturbances that may co-occur with TB or impact its progression.
Respiratory Symptoms	Presence of cough, sputum production, or breathing difficulty.	Core indicator of pulmonary TB; frequency and duration of symptoms can support clinical diagnosis.
ESR	Laboratory test indicating systemic inflammation or infection.	Elevated ESR is a strong indicator of active TB and reflects inflammatory activity related to disease burden.
Platelet Count	Quantitative measure of blood platelets.	Altered platelet levels can occur in TB due to immune activation or chronic inflammation, reflecting disease progression or severity.
Chest X-ray Findings	Radiological observations from thoracic imaging.	Critical diagnostic indicator for pulmonary TB, revealing infiltrates, lesions, or cavitations typical of infection.
BCG Vaccination Score	Record of Bacillus Calmette–Guérin vaccination status.	Indicates prior immunization against TB; helps interpret susceptibility and potential protection level.
Kenneth John Scoring System	Clinical scoring system for pediatric TB and malnutrition.	Assigns points based on ten clinical features commonly associated with TB; useful for case identification in resource-limited settings.
Age Category	Classification of the child's age into standard developmental groups.	Younger children show higher vulnerability to TB due to immature immune defenses and increased exposure risks.
Weight Category	Classification of the child's weight-for-age or weight-for-height percentile.	Weight loss or underweight status may reflect chronic infection and nutritional depletion linked to TB.

Table 5. Cont.

Variable	Description	Relevance to TB Status
Height Category	Classification of height-for-age percentile.	Growth impairment is a common consequence of chronic TB and indicates prolonged disease impact.
MUAC	Anthropometric measure of muscle and fat mass.	Low MUAC identifies malnutrition, which worsens TB outcomes by impairing immune function and recovery capacity.
Father's Education	Educational level of the child's father.	Higher paternal education may be associated with better awareness of TB symptoms and improved treatment compliance.
Mother's Education	Educational level of the child's mother.	Maternal education influences health-seeking behavior and awareness of TB prevention and management practices.
Father's Occupation	Type of employment or labor performed by the father.	Economic stability associated with the father's occupation may affect household exposure, living conditions, and access to care.
Haemoglobin Records	Blood test result indicating hemoglobin levels.	Low hemoglobin levels (anemia) are frequently observed in TB patients and correlate with disease severity and poor prognosis.

Abbreviations = BCG = *Bacillus Calmette–Guérin*; ESR = Erythrocyte Sedimentation Rate; MUAC = Mid-Upper Arm Circumference; TB = Tuberculosis. Note = Variables are ordered by their clinical category (sociodemographic, clinical, anthropometric, laboratory) and their relative importance within the GA–RF model.

This dimensionality reduction, from 52 to 18 informative variables, substantially improved model interpretability while maintaining strong internal classification performance (accuracy = 0.94; Cohen's kappa = 0.86; ROC-AUC = 0.96). The retained informative variables primarily reflected the infection-inflammation profile characteristic of active TB.

Key informative variables included chest X-ray findings, ESR, body temperature, shortness of breath and haemoglobin (HB) levels, indicating the central role of respiratory involvement, systemic inflammation and haematological impairment in TB classification. Anthropometric indicators, including weight, height and mid-upper arm circumference categories, retained moderate importance, supporting the close relationship between nutritional status and susceptibility to TB.

By contrast, sociodemographic variables, including parental education and maternal occupation, exhibited limited contribution to classification performance after incorporation of clinical and physiological indicators. Permutation testing confirmed the stability of the retained informative variables ($p < 0.01$).

As with the nutritional-status model, the TB classification framework should be interpreted as exploratory and internally validated. Nevertheless, these findings highlight the potential of low-cost analytical approaches integrating routine clinical and anthropometric information to support standardized community-based clinical assessment in resource-constrained settings.

3.7. Variable Importance in the TB Model

The variable-importance rankings for the TB RF model are presented in Table 6. Importance estimates were averaged across 10-fold cross-validation runs and normalized to a maximum value of 1.0 to facilitate direct comparison among informative variables.

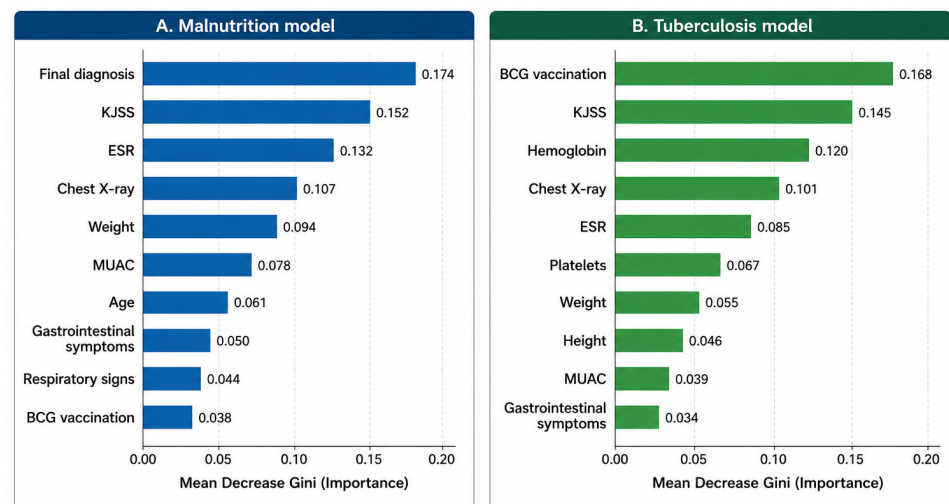
Table 6. Variable importance for the tuberculosis-status Random Forest model (Mean Decrease Accuracy and Mean Decrease Gini).

Variable	Mean Decrease Accuracy	Mean Decrease Gini	Interpretation
BCG Vaccination Score	0.0833	30.5700	Strongest predictor; highlights the importance of vaccination in TB susceptibility.
Kenneth John Scoring System	0.0420	17.1530	Composite clinical score integrating multiple features; robust predictor of TB severity.
Haemoglobin Records	0.0090	13.0033	Relevant biomarker; anaemia is a common comorbidity in TB.
Chest X-ray Findings	0.0080	9.9530	Crucial diagnostic indicator confirming pulmonary TB.
ESR	0.0200	8.7601	Elevated ESR denotes systemic inflammation typical of active TB.
Platelet Count	0.0040	2.6986	Reflects immune activation and chronic infection status.
Age Category	0.0010	0.5612	Slight positive effect; younger age increases susceptibility.
Mid-Upper Arm Circumference	0.0000	0.5032	Indicates nutritional depletion influencing infection risk.
Father's Occupation	0.0000	0.2766	Slight indirect effect through socioeconomic pathways.
Father's & Mother's Education	0.0000	0.0000	Minimal relevance; low variability within the dataset.
Temperature	0.0000	0.0520	Suggests febrile response but limited predictive power.
Respiratory Symptoms	0.0000	0.0506	Associated with TB but redundant due to overlap with other clinical variables.
Dehydration	0.0000	0.0100	Minor influence; reflects disease burden rather than diagnostic specificity.

Abbreviations = BCG: *Bacillus Calmette–Guérin*; ESR = Erythrocyte Sedimentation Rate. Note = Importance values were normalized to facilitate comparison across predictors. Variables are ordered by overall importance (Mean Decrease Gini).

The BCG vaccination score emerged as the most influential informative variable associated with TB classification (Mean Decrease Gini = 30.57; Mean Decrease Accuracy = 0.083), followed by the KJSS (Mean Decrease Gini = 17.15; Mean Decrease Accuracy = 0.042), HB records (13.00; 0.009), chest X-ray findings (9.95; 0.008) and ESR (8.76; 0.020). Collectively, these five variables accounted for >80% of the total model importance, reflecting the biological interaction between infection, inflammation and immune-nutritional impairment in pediatric TB.

The relative importance of the variables selected by the Genetic Algorithm for both Random Forest classification models is graphically presented in Figure 2. This visualization complements the numerical information reported in Tables 4–6 and facilitates the interpretation of the relative contribution of the most influential informative variables for malnutrition and tuberculosis classification.



Note. Variables are ranked according to their relative importance (Mean Decrease Gini) in the Random Forest models.
 KJSS: King's Jeddah Score; ESR: Erythrocyte Sedimentation Rate; BCG: Bacillus Calmette–Guérin; MUAC: Mid–Upper Arm Circumference.

Figure 2. Relative importance of the variables selected by the Genetic Algorithm for the Random Forest prediction models. Relative variable importance was estimated using the Mean Decrease Gini index. Panel (A) shows the variables contributing to the prediction of nutritional status, whereas Panel (B) presents the variables contributing to the prediction of tuberculosis. Higher Mean Decrease Gini values indicate a greater contribution of the corresponding predictor to the classification performance of the Random Forest models.

Lower-ranking informative variables, including platelet count, mid-upper arm circumference and age category, contributed modestly to model discrimination. By contrast, sociodemographic variables such as parental education and occupation exhibited near-zero or negative importance scores, indicating limited contribution to classification performance after incorporation of direct clinical and laboratory indicators. The consistency of rankings across both Mean Decrease Gini and Mean Decrease Accuracy metrics supports the internal stability of feature weighting within the RF framework. Overall, the variable-importance profile delineated a multivariate TB signature predominantly characterized by immune-inflammatory and respiratory indicators, whereas socioeconomic factors showed comparatively limited contribution within this highly homogeneous population.

3.8. Data Completeness and Class Balance

Missing data were minimal across all variables, with <3% missingness per variable. Incomplete records were handled through listwise exclusion, resulting in a final analytical sample of 300 children. Exclusion of incomplete cases did not materially alter class proportions or the distribution of key variables.

The nutritional-status dataset exhibited moderate class imbalance, with 269 children classified as malnourished (89.7%) and 31 classified as well nourished (10.3%). To address this imbalance, internal validation procedures incorporated stratified resampling, and Cohen's kappa coefficient was prioritized as the principal performance metric to account for agreement beyond chance.

For the exploratory TB classification model, performance evaluation focused on internal discrimination and feature-importance metrics. The GA-RF model achieved high internal accuracy (0.94), substantial agreement (kappa = 0.86) and strong discriminatory capacity (ROC-AUC = 0.96). Given the absence of an external validation cohort, these estimates are presented as indicators of internal classification performance rather than externally generalizable model performance.

Overall, the procedures used to address data completeness and class imbalance supported the internal validity of the analytical framework and provided an appropriate basis for the subsequent robustness and diagnostic analyses.

3.9. Model Diagnostics and Robustness

A comprehensive series of diagnostic analyses was conducted to evaluate the internal robustness and stability of the GA-RF models.

Permutation importance analysis was used to assess the consistency of variable-importance rankings in both the nutritional and TB frameworks. Across 1000 permutations, the five most influential informative variables [FD, KJSS, chest X-ray findings, ESR and weight category for nutritional status; BCG vaccination score, chest X-ray findings, KJSS, ESR and HB records for TB] maintained >85% concordance with the original rankings (randomization test, $p < 0.01$), supporting the reproducibility of the feature-selection process.

Feature stability was further evaluated using bootstrap resampling (1000 replicates). In total, 12 of the 16 variables retained in the nutritional model and 13 of the 18 variables retained in the TB model were selected in >75% of bootstrap iterations, indicating a high degree of internal consistency and low selection variability within the GA optimization process.

To assess potential multicollinearity, variance inflation factors (VIFs) were calculated for all continuous and ordinal predictors. No variable exceeded a VIF threshold of 5, suggesting that inter-variable correlations did not materially influence model estimates or variable-importance rankings.

Model calibration was evaluated by comparing predicted and observed probabilities and by calculating the Brier score. The nutritional model achieved a Brier score of 0.06, indicating good calibration with slight overestimation at predicted probabilities >0.9. The TB model demonstrated a comparable calibration pattern (Brier score = 0.07), suggesting acceptable probability scaling despite the absence of external validation.

Sensitivity analyses were performed to evaluate model robustness following removal of influential predictors. Exclusion of the FD variable from the nutritional model resulted in a moderate reduction in performance (accuracy = 0.91; Cohen's kappa = 0.76; ROC-AUC = 0.93), indicating that FD contributed substantial, although not exclusive, predictive information. Similarly, removal of the KJSS variable from the TB model reduced the ROC-AUC by 0.04 without altering the overall hierarchy of the remaining predictors, supporting the complementary contribution of individual features.

Overall, these diagnostic analyses indicated that the GA-RF models demonstrated stable internal performance, consistent feature selection and adequate calibration. Although external validation remains necessary, the internal findings suggest that the analytical framework was methodologically coherent and not substantially overfitted to the available data.

3.10. Comparative Interpretation and Clinical Implications

Comparative evaluation of the two GA-RF classification frameworks revealed convergent patterns in feature selection and relative variable importance, highlighting the close biological interdependence between malnutrition and TB within this population.

In both frameworks, markers of systemic inflammation such as erythrocyte sedimentation rate (ESR), composite clinical scores such as the Kenneth John Scoring System (KJSS), and radiological findings, particularly chest X-ray abnormalities, consistently ranked among the most influential informative variables. These variables reflect overlapping pathophysiological mechanisms, including chronic infection, immune activation, and tissue catabolism, which are central to both nutritional deterioration and TB progression.

The additional contribution of haemoglobin (HB) levels further emphasizes the role of haematological impairment within this infection–nutrition continuum.

By contrast, sociodemographic variables, including parental education and occupation, contributed minimally to both classification frameworks. This finding is consistent with the relative homogeneity of socioeconomic deprivation within the study setting, where limited variability may reduce the contribution of demographic factors to classification performance. Under these conditions, clinical and biological indicators appear to provide a more direct representation of individual health status.

From a translational perspective, these findings illustrate the capacity of hybrid analytical approaches such as the GA-RF framework to identify clinically relevant patterns within small community-based datasets. Identification of a concise subset of highly informative variables, including the KJSS, ESR, chest X-ray findings, and BCG vaccination status, suggests that routinely collected, low-cost clinical measures may support standardized community-based clinical assessment in resource-limited settings.

Taken together, this comparative interpretation reinforces the bidirectional relationship between infection and nutritional status and provides an analytical basis for integrated clinical assessment strategies. The findings also support the potential feasibility of incorporating computational feature-selection methods into primary healthcare practice to assist clinical decision-making and facilitate timely clinical assessment and appropriate management of vulnerable children under five years of age.

4. Discussion

4.1. Overview of the Analytical Framework and Key Findings

The present study provides a community-based analytical framework integrating clinical, anthropometric and sociodemographic variables to characterize clinically relevant patterns associated with malnutrition and TB among children under five years of age living in urban slum environments. By applying a hybrid GA-RF pipeline, reproducible subsets of informative variables were identified that jointly captured the infection–inflammation–nutrition continuum underlying child morbidity in resource-limited settings. The convergence of key variables, particularly ESR, chest X-ray findings, HB levels and the KJSS, supports their biological plausibility and clinical relevance [3,12,13,16,21–25].

In contrast to conventional epidemiological models that generally require large sample sizes and predefined covariate structures, the present approach used data-driven feature selection to identify physiologically meaningful patterns from relatively small field-derived datasets. The resulting dimensionality reduction, from 48 and 52 candidate variables to 16 and 18 informative variables for malnutrition and TB, respectively, substantially improved interpretability while maintaining stable internal discriminatory performance (area under the curve, ~0.96). These findings support the feasibility of applying computational analytical approaches within standardized community-based clinical assessments and suggest that routinely collected clinical variables may contribute to identifying clinically relevant patterns associated with malnutrition and TB. The analytical strategy combined a Genetic Algorithm for feature optimization with a supervised Random Forest classifier, allowing robust classification while reducing model complexity.

From a methodological perspective, the principal contribution of the GA-RF framework was not the identification of previously unknown clinical predictors, but rather the objective prioritization and optimization of routinely collected variables according to their combined discriminatory value. Unlike conventional epidemiological approaches, which generally evaluate predefined associations based on a priori hypotheses, the GA-RF framework simultaneously considered many interacting variables, identified the most informative subset, and reduced model dimensionality while maintaining high classi-

fication performance. This characteristic may facilitate the development of simplified decision-support tools requiring fewer routinely collected variables, making the approach particularly suitable for standardized community-based clinical assessments in resource-constrained settings.

Although the high internal accuracy and stability of the selected informative variables support the methodological consistency of the analytical framework, the absence of external validation requires cautious interpretation. Therefore, the proposed models should be considered proof-of-concept analytical frameworks identifying clinically relevant patterns rather than deployable diagnostic tools. Accordingly, the proposed GA-RF framework should be interpreted as a complementary analytical and decision-support framework rather than as a replacement for conventional epidemiological analyses or clinical judgment. Its principal strength lies in supporting the data-driven prioritization of routinely available clinical, anthropometric, laboratory, radiological, and sociodemographic variables that may facilitate targeted clinical assessment and resource allocation in vulnerable populations. The primary objective of the present study was not to develop a deployable diagnostic model, but to identify clinically and socially relevant patterns associated with TB and malnutrition in vulnerable pediatric populations. Nevertheless, the consistency observed across both classification models provides a rationale for extending this analytical approach to larger prospective cohorts to refine classification thresholds and evaluate applicability under real-world conditions. Accordingly, the present findings should be considered hypothesis-generating rather than immediately applicable to routine clinical implementation or public health surveillance until validated in independent populations.

4.2. Clinical and Epidemiological Interpretation

The findings of the present study indicate that several informative clinical, laboratory, radiological, anthropometric, and sociodemographic variables were associated with TB prevalence, with malnutrition emerging as a central determinant among preschool-aged children. These observations are consistent with previous reports demonstrating a strong association between undernutrition and increased susceptibility to TB during early childhood [16,21–24]. Children under five years of age constitute a biologically vulnerable population due to the immaturity of the immune system, which increases susceptibility to infectious diseases and nutritional deterioration [8,26,27]. In addition, TB in this age group frequently presents with atypical or extrapulmonary manifestations, contributing to diagnostic delays and poorer clinical outcomes [7,28,29].

The present analysis also highlights the influence of parental education on child health outcomes. Paternal illiteracy emerged as an important sociodemographic factor associated with both TB and malnutrition, in agreement with previous epidemiological studies linking low parental education with adverse pediatric health indicators [30]. By contrast, maternal education demonstrated minimal variability within the cohort, likely reflecting the uniformly low educational attainment among women in this community. These findings underscore the importance of incorporating parental education and health literacy components into community-based TB and nutritional intervention programs.

In addition, diarrheal illness and dehydration emerged as frequent and clinically relevant variables associated with both TB and malnutrition, corroborating previous reports [25,31]. Gastrointestinal infections and associated mucosal inflammation may impair nutrient absorption and accelerate wasting in vulnerable children. This bidirectional and cyclical relationship between diarrheal disease and undernutrition further emphasizes the importance of timely recognition and appropriate management of gastrointestinal disorders to prevent progressive deterioration of nutritional status and immune function [8].

4.3. Immunological and Pathophysiological Correlates

Anemia, elevated ESR, and low HB levels emerged among the most informative biomarkers prioritized by the GA-RF framework, underscoring the central role of chronic inflammation and iron dysregulation in the pathogenesis of both TB and malnutrition. In TB, anemia is largely mediated by inflammatory cytokine-induced upregulation of hepcidin, which suppresses ferroportin expression and restricts iron availability, leading to the classical anemia of chronic disease [32,33]. Malnutrition may further exacerbate these mechanisms through deficiencies in essential micronutrients, including iron, zinc, and vitamin A, as well as through impaired erythropoiesis [34].

Similar mechanistic interactions have been described by Schaible and Kaufmann and by Cegielski and McMurray, who highlighted a reciprocal feedback loop in which chronic infection induces oxidative stress, proteolysis and growth impairment, whereas nutritional deprivation compromises host immune competence [12,35]. Within this framework, infection and undernutrition may act synergistically to sustain a state of persistent immunometabolic stress.

Mechanistically, the present findings reinforce the immunometabolic model linking infection-driven inflammation with nutritional decline. TB-associated cytokine activation, particularly involving interleukin-6, tumor necrosis factor- α and interferon- γ , increases resting energy expenditure and promotes skeletal muscle catabolism, whereas reduced dietary intake and impaired nutrient absorption limit immune-cell proliferation and phagocytic capacity. This interaction may explain the concurrent presence of low HB concentrations, elevated ESR values and reduced mid-upper arm circumference observed in the present cohort.

These findings are consistent with previous studies reporting similar patterns of immunometabolic deterioration among malnourished children with chronic infectious diseases [13,36], thereby supporting the biological plausibility of the variables prioritized by the analytical framework. Collectively, these observations support the concept that TB and malnutrition represent interdependent immunopathological processes rather than independent or sequential conditions, particularly in vulnerable pediatric populations living in resource-constrained settings.

4.4. Role of Vaccination and Respiratory Morbidity

In the present analysis, BCG vaccination emerged as one of the most informative variables associated with TB outcomes, consistent with previous evidence demonstrating that BCG confers partial protection against severe forms of TB, including miliary disease and tuberculous meningitis, in young children [15,28,37]. Although overall immunization coverage did not differ significantly between groups, the presence of a BCG scar was inversely associated with TB status. These findings reinforce the role of BCG as a modifiable preventive factor and underscore the importance of maintaining high vaccination coverage in socioeconomically deprived populations [15,37,38].

Respiratory symptoms, particularly productive cough, dyspnea and wheezing, were consistently associated with both TB and malnutrition. This observation is consistent with the pathophysiology of TB as a chronic pulmonary disease characterized by alveolar injury, impaired gas exchange, hypoxia and sustained inflammatory activation [28]. Similar to previous reports [5,36,39], the present findings support a bidirectional interaction in which respiratory infections accelerate nutritional decline through increased metabolic demand and reduced dietary intake, whereas malnutrition compromises mucosal integrity and immune defense mechanisms.

These interactions highlight the importance of integrating systematic respiratory assessment into nutritional assessment protocols, particularly in high-risk pediatric populations. Timely recognition of respiratory symptoms in malnourished children may facilitate

appropriate TB investigation and improve case detection, particularly in children presenting with nonspecific or atypical clinical manifestations [3,4,7].

4.5. Clinical and Public Health Implications

From a translational perspective, the present findings highlight the feasibility of integrating low-cost, data-driven analytical approaches such as the GA-RF framework into standardized primary healthcare assessment workflows in resource-limited settings. This analytical strategy may complement conventional regression-based models by capturing non-linear relationships and complex interactions that are often overlooked in traditional epidemiological analyses.

Identification of a concise and clinically interpretable subset of informative variables, including ESR, KJSS, chest X-ray findings, and HB levels, suggests that routinely collected clinical, laboratory, radiological, anthropometric, and sociodemographic data may support the identification of clinically relevant patterns associated with TB and malnutrition during standardized clinical assessment, thereby facilitating prioritization for further diagnostic evaluation and timely intervention.

Given the bidirectional relationship between infection and malnutrition, integrated clinical assessment of both conditions may facilitate comprehensive clinical evaluation, support clinical decision-making, and enhance management efficiency in high-burden communities. These observations are consistent with previous recommendations advocating integrated models of care combining TB assessment, nutritional support, and community-based education to interrupt the infection-malnutrition cycle [4,5].

Strengthening nutritional supplementation within TB control programs, ensuring continuity of vaccination coverage, and improving access to essential healthcare services remain important strategies for reducing morbidity and mortality among vulnerable pediatric populations in comparable high-burden settings [1,5,10,13,35].

4.6. Strengths, Limitations and Future Research Directions

The present study has several methodological and conceptual strengths. First, it represents one of the few community-based investigations integrating clinical, anthropometric and laboratory data to examine the dual burden of TB and malnutrition among children under five years of age living in deprived urban environments. Use of the hybrid GA-RF framework enabled objective feature selection and reduced the likelihood of model overfitting despite the relatively modest sample size. In contrast to conventional regression approaches, this computational strategy accommodates non-linear relationships, multicollinearity and high-dimensional data structures, thereby allowing a broader representation of biological and epidemiological complexity.

Second, the study demonstrated strong internal validity through repeated cross-validation, stratified resampling, permutation testing and bootstrap stability analyses, supporting the robustness and reproducibility of the selected informative variables. Third, because the models relied on routinely collected clinical, anthropometric, laboratory, radiological, and sociodemographic variables obtained during standardized community-based clinical assessments, the analytical framework may be feasible and potentially scalable within resource-constrained primary healthcare systems. Finally, integration of structured field observations with clinical and laboratory assessments enhanced the ecological validity of the findings by ensuring that the analytical models reflected real-world environmental and healthcare conditions within this vulnerable population.

Several limitations should nevertheless be acknowledged. The cross-sectional design precludes causal inference and limits temporal interpretation of the observed associations between malnutrition, infection and clinical indicators. The absence of universal

microbiological confirmation for TB introduces the possibility of diagnostic misclassification, particularly in extrapulmonary or paucibacillary cases commonly observed in young children. Seasonal migration and population mobility may also have influenced representativeness, whereas reliance on maternal recall for age estimation may have introduced minor anthropometric imprecision. Although the sample included >80% of the accessible population, it remained insufficient for external validation or detailed stratified analyses according to age, sex or disease phenotype. Furthermore, despite the internal consistency of the GA-RF framework, the absence of an independent validation cohort requires cautious interpretation of predictive performance as exploratory rather than definitive.

An additional methodological consideration is that several of the variables identified as highly informative by the GA-RF framework, including chest radiographic findings, the Kenneth John Scoring System, and anthropometric measurements, are obtained during the routine clinical assessment rather than before diagnostic evaluation. Consequently, the present models were designed to identify the combination of routinely collected variables providing the greatest discriminatory performance within a comprehensive community-based clinical assessment, rather than to develop a pre-diagnostic screening tool based exclusively on baseline information. Therefore, the reported classification performance should be interpreted within this clinical context. Future studies should evaluate simplified analytical or classification models using only variables available before diagnostic investigations to determine their suitability for early community-based clinical assessment and public health implementation.

An additional limitation is that tuberculosis and malnutrition are complex multifactorial conditions influenced by numerous environmental, social, biological, and behavioral determinants that were not comprehensively captured in the present analytical framework. Although the models incorporated a broad range of routinely collected clinical, anthropometric, laboratory, radiological, and sociodemographic variables, other potentially relevant factors, including household income, parental education, dietary diversity, food security, environmental pollution, genetic susceptibility, microbiome composition, and detailed immunological characteristics, were not available for analysis. Consequently, the identified informative variables should be interpreted as representing the most informative variables within the available dataset rather than an exhaustive characterization of all determinants contributing to TB and malnutrition.

Future research should address these limitations by incorporating a broader range of environmental, socioeconomic, behavioral, and biological determinants into longitudinal multicenter studies, thereby enabling a more comprehensive characterization of the complex pathways underlying tuberculosis and malnutrition in vulnerable pediatric populations. Larger cohorts incorporating molecular and immunological biomarkers, including ferritin, hepcidin, soluble transferrin receptor, and cytokine profiles, may further clarify the mechanistic pathways linking inflammation, anemia, and nutritional deterioration. Following external validation, the incorporation of machine learning algorithms into routine community health surveillance systems may facilitate real-time monitoring of nutritional and infectious health profiles, particularly when integrated with mobile health or telemedicine platforms. Finally, intervention studies combining nutritional rehabilitation, TB assessment, and caregiver education are warranted to determine whether data-driven prioritization of clinically relevant variables can translate into measurable improvements in early diagnosis, treatment adherence, and child survival.

Taken together, these considerations position the present study as a proof-of-concept analytical foundation for the development of scalable, integrated, and precision-oriented child health strategies in low-resource settings.

5. Conclusions

In conclusion, the present study demonstrates that hybrid data-driven analytical models may identify clinically relevant patterns associated with TB and malnutrition using relatively small community-based datasets. The convergence of inflammatory, hematological, radiological, and anthropometric indicators highlights the close interrelationship between infection and nutritional status during early childhood, particularly among vulnerable urban migrant populations.

These findings reinforce the importance of integrated community-based approaches combining nutritional support, early TB control strategies, and health education as complementary strategies to improve child health outcomes in deprived settings. Although the proposed GA-RF framework demonstrated promising internal performance, its findings should be considered hypothesis-generating and require external validation in larger and more heterogeneous populations before routine implementation in clinical practice or public health programs can be recommended.

Nevertheless, the proposed GA-RF framework should be interpreted as an analytical and decision-support framework rather than as a pre-diagnostic screening or early risk prediction tool. Within standardized community-based clinical assessments, it may support the identification of clinically relevant patterns and the prioritization of informative clinical, anthropometric, laboratory, radiological, and sociodemographic variables, thereby informing future research and evidence-based decision-making in resource-limited environments.

Author Contributions: Conceptualization, D.F.-L., J.M.-A. and J.S.; methodology, D.F.-L., J.M.-A., E.R. and A.M.C.S.M.; software, D.F.-L. and A.U.M.; validation, D.F.-L., E.R. and J.M.-A.; formal analysis, D.F.-L., A.U.M. and F.M.; investigation, J.S., A.Z., A.T., F.M., Z.A., A.U.M. and F.A.; resources, J.S., A.Z. and A.T.; data curation, J.S., A.U.M. and D.F.-L.; writing—original draft preparation, J.S., D.F.-L. and A.U.M.; writing—review and editing, A.Z., A.T., F.M., Z.A., F.A., E.R., A.M.C.S.M., E.J.-C., J.M.-A. and D.F.-L.; visualization, A.U.M. and D.F.-L.; supervision, D.F.-L., J.M.-A. and E.R.; project administration, D.F.-L.; funding acquisition, D.F.-L., E.R. and J.M.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Department of Education of the Junta of Castile and León and the European Regional Development Fund (FEDER) under the TCUE Plan 2024–2027, approved by the Order of 16 September 2024, Grant No. 067/230003.

Institutional Review Board Statement: This research was reviewed and approved by Institutional Research Committee (IRC) of Green International University (GIU), Lahore, Pakistan (Approval No. IRC-GIU-160-09-2024; 5 September 2024). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and complied with the ethical standards of the Institutional Research Committee. Written informed consent was obtained from the parents or legal guardians of all participating children before enrolment. Participants' privacy, dignity, and confidentiality were strictly protected throughout the study by anonymizing all collected data and restricting access to authorized research personnel only.

Informed Consent Statement: Written informed consent was obtained from the parents or legal guardians of all children involved in the study prior to participation.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors want to thank the Neurobiology Research Group, Faculty of Medicine of the University of Valladolid, for their collaboration. CIBEROBN is an initiative of Instituto de Salud Carlos III, Spain. Artificial intelligence-based tools were used exclusively to assist with language editing, stylistic refinement, and improvement of clarity and organization of the manuscript. All scientific content, data interpretation, and conclusions were generated by the authors, who take full responsibility for the integrity and originality of the work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Munthali, T.; Chabala, C.; Chama, E.; Mugode, R.; Kapata, N.; Musonda, P.; Michelo, C. Tuberculosis caseload in children with severe acute malnutrition related with high hospital-based mortality in Lusaka, Zambia. *BMC Res. Notes* **2017**, *10*, 206. [CrossRef] [PubMed]
2. Khalil, B.; Hussain, M.; Taj, W.; Iqbal, S.; Irshad, M.; Khan, M.J.; Ullah, I. Frequency of pulmonary tuberculosis in severely acute malnourished children and its association with inappropriate feeding practices. *J. Med. Sci.* **2020**, *28*, 252–255.
3. Kumar, M.; Kumar, P.; Singh, A. Recent advances in the diagnosis and treatment of childhood tuberculosis. *J. Nat. Sci. Biol. Med.* **2015**, *6*, 314–320. [CrossRef] [PubMed]
4. Swaminathan, S.; Rekha, B. Pediatric tuberculosis: Global overview and challenges. *Clin. Infect. Dis.* **2010**, *50*, S184–S194. [CrossRef] [PubMed]
5. Graham, S.M.; Sismanidis, C.; Menzies, H.J.; Marais, B.J.; Detjen, A.K. Black RE: Importance of tuberculosis control to address child survival. *Lancet* **2014**, *383*, 1605–1607. [CrossRef] [PubMed]
6. Grobusch, M.P.; Kapata, N. Global burden of tuberculosis: Where we are and what to do. *Lancet Infect. Dis.* **2018**, *18*, 1291–1293. [CrossRef] [PubMed]
7. Dodd, P.J.; Gardiner, E.; Coghlan, R.; Seddon, J.A. Burden of childhood tuberculosis in 22 high-burden countries: A mathematical modelling study. *Lancet Glob. Health* **2014**, *2*, e453–e459. [CrossRef] [PubMed]
8. Seddon, J.A. Shingadia D: Epidemiology and disease burden of tuberculosis in children: A global perspective. *Infect. Drug Resist.* **2014**, *7*, 153–165. [CrossRef] [PubMed]
9. Waheed, M.S. Tuberculosis. World Health Organization Regional Office for the Eastern Mediterranean. Available online: <https://www.emro.who.int/tuberculosis/home/index.html> (accessed on 25 June 2025).
10. Ullah, H.; Ahmed, H.; Salman, A.; Iqbal, R.; Hussaini, S.J.; Malikzai, A. The rising tide of tuberculosis in Pakistan: Factors, impact and multifaceted approaches for prevention and control. *Health Sci. Rep.* **2024**, *7*, e70130. [CrossRef] [PubMed]
11. Yaqoob, A.; Alvi, M.R.; Fatima, R.; Najmi, H.; Samad, Z.; Nisar, N.; Haq, A.U.; Javed, B.; Khan, A.W.; Hinderaker, S.G. Geographic accessibility to childhood tuberculosis care in Pakistan. *Glob. Health Action* **2022**, *15*, 2095782. [CrossRef] [PubMed]
12. Cegielski, J.P.; McMurray, D.N. The relationship between malnutrition and tuberculosis: Evidence from studies in humans and experimental animals. *Int. J. Tuberc. Lung Dis.* **2004**, *8*, 286–298. [PubMed]
13. Vasantha, M.; Gopi, P.G.; Subramani, R. Weight gain in patients with tuberculosis treated under directly observed treatment short-course (DOTS). *Indian J. Tuberc.* **2009**, *56*, 5–9. [PubMed]
14. Lamb, G.S.; Starke, J.R. Tuberculosis in infants and children. *Microbiol. Spectr.* **2017**, *5*, TNMI7-0037-2016. [CrossRef] [PubMed]
15. Trunz, B.B. Fine PEM and Dye C: Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: A meta-analysis and assessment of cost-effectiveness. *Lancet* **2006**, *367*, 1173–1180. [CrossRef] [PubMed]
16. Martin, S.J. Sabina EP: Malnutrition and associated disorders in tuberculosis and its therapy. *J. Diet. Suppl.* **2019**, *16*, 602–620. [PubMed]
17. Rao, B.; Rashid, M.; Hasan, M.G.; Thunga, G. Machine learning in predicting child malnutrition: A meta-analysis of demographic and health surveys data. *Int. J. Env. Res. Public Health* **2025**, *22*, 449. [CrossRef] [PubMed]
18. World Health Organization: WHO Child Growth Standards: Methods and Development. Available online: <https://www.who.int/publications/i/item/924154693X> (accessed on 5 December 2025).
19. Farid, A.; Iqbal, S.M.J.; Hanif, A.; Sultan, M.A. Diagnostic accuracy of modified Kenneth Jones scoring criteria in confirmed cases of tuberculosis in children. *Ann. King Edw. Med. Univ.* **2013**, *19*, 160–163.
20. World Health Organization: Child Growth Standards. Available online: <https://www.who.int/tools/child-growth-standards> (accessed on 11 January 2026).
21. Isanaka, S.; Mugusi, F.; Urassa, W.; Willett, W.C.; Bosch, R.J.; Villamor, E.; Spiegelman, D.; Duggan, C.; Fawzi, W.W. Iron deficiency and anemia predict mortality in patients with tuberculosis. *J. Nutr.* **2012**, *142*, 350–357. [CrossRef] [PubMed]
22. Minchella, P.A.; Donkor, S.; Owolabi, O.; Sutherland, J.S.; McDermid, J.M. Complex anemia in tuberculosis: The need to consider causes and timing when designing interventions. *Clin. Infect. Dis.* **2015**, *60*, 764–772. [CrossRef] [PubMed]
23. Cantor, D.; Swartz, J.; Roberts, B.; Abbara, A.; Ager, A.; Bhutta, Z.A.; Blanchet, K.; Bunte, D.M.; Chukwuorji, J.C.; Daoud, N.; et al. Understanding the health needs of internally displaced persons: A scoping review. *J. Migr. Health* **2021**, *4*, 100071. [CrossRef] [PubMed]
24. Merk, H.; Kodmon, C.; van der Werf, M.J. Will we reach the Sustainable Development Goals target for tuberculosis in the European Union/European Economic Area by 2030? *Euro Surveill.* **2019**, *24*, 1900153. [CrossRef] [PubMed]
25. Kinyoki, D.K.; Berkley, J.A.; Moloney, G.M.; Kandala, N.B.; Noor, A.M. Predictors of the risk of malnutrition among children under the age of 5 years in Somalia. *Public Health Nutr.* **2015**, *18*, 3125–3133. [CrossRef] [PubMed]

26. Ramos, J.M.; Perez-Butragueno, M.; Tesfamariam, A.; Reyes, F.; Tiziano, G.; Endirays, J.; Balcha, S.; Elala, T.; Biru, D.; Comeche, B.; et al. Comparing tuberculosis in children aged under 5 versus 5 to 14 years old in a rural hospital in southern Ethiopia. *BMC Public Health* **2019**, *19*, 856. [[CrossRef](#)] [[PubMed](#)]
27. Fernandes, P.; Pereira, R.; Wiedenhof, G. Organizational culture and individuals' discretionary behaviors at work: A cross-cultural analysis. *Front Sociol.* **2023**, *8*, 1190488. [[CrossRef](#)] [[PubMed](#)]
28. Marais, B.J. Childhood tuberculosis: Reflections from the front line. *Pediatr. Ann.* **2004**, *33*, 695–700. [[CrossRef](#)] [[PubMed](#)]
29. Jenkins, H.E.; Tolman, A.W.; Yuen, C.M.; Parr, J.B.; Keshavjee, S.; Perez-Velez, C.M.; Pagano, M.; Becerra, M.C.; Cohen, T. Incidence of multidrug-resistant tuberculosis disease in children: Systematic review and global estimates. *Lancet* **2014**, *383*, 1572–1579. [[CrossRef](#)] [[PubMed](#)]
30. Shetty, N.; Shemko, M.; Vaz, M.; D'Souza, G. An epidemiological evaluation of risk factors for tuberculosis in South India. *Int. J. Tuberc. Lung Dis.* **2006**, *10*, 80–86. [[PubMed](#)]
31. Becker, P. Measurement of mid-upper arm circumference to screen for childhood malnutrition. *Nutr. Clin. Pract.* **2024**. *Epub ahead of print*. [[CrossRef](#)] [[PubMed](#)]
32. Ganz, T. Systemic iron homeostasis. *Physiol. Rev.* **2013**, *93*, 1721–1741. [[CrossRef](#)] [[PubMed](#)]
33. Cullis, J.O. Anaemia of chronic disease. *Clin. Med.* **2013**, *13*, 193–196. [[CrossRef](#)] [[PubMed](#)]
34. Thakur, N.; Chandra, J.; Pemde, H.; Singh, V. Anemia in severe acute malnutrition. *Nutrition* **2014**, *30*, 440–442. [[CrossRef](#)] [[PubMed](#)]
35. Schaible, U.E.; Kaufmann, S.H.E. Malnutrition and infection: Complex mechanisms and global impacts. *PLoS Med.* **2007**, *4*, e115. [[CrossRef](#)] [[PubMed](#)]
36. Rodriguez, L.; Cervantes, E.; Ortiz, R. Malnutrition and gastrointestinal and respiratory infections in children. *Int. J. Environ. Res. Public Health* **2011**, *8*, 1174–1205. [[CrossRef](#)] [[PubMed](#)]
37. Qu, M.; Zhou, X.; Li, H. BCG vaccination strategies against tuberculosis: Updates and perspectives. *Hum. Vaccin Immunother.* **2021**, *17*, 5284–5295. [[CrossRef](#)] [[PubMed](#)]
38. Gelaw, Y.; Getaneh, Z.; Melku, M. Anemia as a risk factor for tuberculosis: A systematic review and meta-analysis. *Environ. Health Prev. Med.* **2021**, *26*, 13. [[CrossRef](#)] [[PubMed](#)]
39. Zar, H.J.; Hanslo, D.; Apolles, P.; Swingle, G.; Hussey, G. Induced sputum versus gastric lavage for microbiological confirmation of pulmonary tuberculosis in infants and young children. *Lancet* **2005**, *365*, 130–134. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.