Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 07 July 2026
Revised 12 Aug 2026
Accepted 31 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (20s):105–131

KEYWORDS:
Clinical Decision Support System, Explainable Artificial Intelligence, Pregnant Women, Chronic Energy Deficiency, Machine Learning.

NutriKEK-AI: Development and Validation of an Explainable Hybrid AI-CDSS for Chronic Energy Deficiency Prediction in Pregnancy

Anis Ervina[1*], Roichatul Djannah[2], Rahmat Nurcahyo[3]

[1]Faculty of Health, La Tansa Mashiro University, Rangkasbitung, Lebak Regency, Banten, Indonesia
[2]Faculty of Health, Salakanagara University, Tangerang, Banten, Indonesia
[3]Faculty of Information Technology, Latansa Mashiro University, Banten, Indonesia

ABSTRACT: Chronic Energy Deficiency (CED) in pregnant women remains a public health issue that contributes to a high risk of maternal complications, stunting, and pregnancy and childbirth risks, particularly in resource-limited areas. This study aims to develop and validate NutriKEK-AI, an Explainable Hybrid Clinical Decision Support System that integrates machine learning, Explainable Artificial Intelligence (XAI), and expert clinical knowledge to predict the risk of CED early in pregnancy. This study employs a Design and Development Research approach that adopts the CRISP-DM framework and Agile Development. Data were collected from 214 pregnant women, 32 healthcare workers, and 7 expert validators in Lebak and Pandeglang Regencies, Banten. Predictive models were developed using several machine learning algorithms and evaluated through 10-fold cross-validation, calibration analysis, expert validation, usability testing, and user acceptance. The results showed that the XGBoost model delivered the best performance with an accuracy of 88%, an F1-score of 0.85, and an AUC of 0.91. SHAP analysis identified upper arm circumference (UAC), anemia status, weight (BB), height (TB), body mass index (BMI), daily energy intake, protein intake, household income, and history of stunting. System validation demonstrated high validity (S-CVI/Ave = 0.91), strong reliability (α = 0.84), and excellent usability (SUS = 84.6). Field implementation improved risk detection accuracy from 71.2% to 88.4%, reduced decision-making time from 14.8 to 6.3 minutes, and increased referral accuracy to 86.7%. NutriKEK-AI has proven to be accurate, transparent, and easy to use, and has the potential to serve as an effective AI-based solution to support early detection and maternal nutrition decision-making in primary health care settings with limited resources

1. INTRODUCTION

Maternal nutrition remains a key determinant of maternal and child health outcomes in low- and middle-income countries (Dassie et al., 2026), as inadequate nutritional status during pregnancy is strongly associated with an increased risk of fetal growth restriction, low birth weight, stunting, and long-term metabolic disorders throughout the life course (Keats et al., 2021; Vaivada et al., 2020). Various global studies show that the burden of malnutrition among pregnant women remains concentrated in low- and middle-income countries, particularly among women facing economic constraints, food insecurity, and limited access to quality maternal health services. These conditions further perpetuate the intergenerational cycle of malnutrition (Dassie et al., 2026).

Women with inadequate nutritional reserves during pregnancy are at higher risk of maternal anemia, intrauterine growth restriction (IUGR), low birth weight (LBW), preterm birth, and neonatal mortality (Victora et al., 2021), all of which simultaneously contribute to growth impairment, cognitive development, and an increased risk of long-term metabolic diseases in children (Sokou et al., 2025), thereby underscoring the importance of early detection and risk-based maternal nutritional interventions in antenatal care (Keats et al., 2021; Vaivada et al., 2020; Wilkins et al., 2021).

Although various maternal nutrition interventions have been widely implemented, the prevalence of chronic energy deficiency (CED) among pregnant women remains high in resource-constrained areas due to inequities in access to health care, food insecurity, poor-quality antenatal care, and weak data-driven early detection of maternal risks (Kruk et al., 2018). Maternal malnutrition remains an ongoing public health challenge in Indonesia, particularly among pregnant women living in rural and remote areas (Priambodo et al., 2026), where inadequate food intake, food insecurity, and socioeconomic disparities continue to contribute to chronic energy deficiency and poor maternal nutritional outcomes (Azrimaidaliza et al., 2025). Further recent evidence indicates that low maternal nutritional adequacy in rural Indonesian communities is strongly associated with limited dietary diversity, low household economic status, and limited access to quality maternal health services, thereby exacerbating the risk of stunting and poor pregnancy outcomes across generations (Agustina et al., 2023). These conditions indicate that the effectiveness of preventing maternal health problems in resource-constrained areas is influenced not only by clinical factors (Priambodo et al., 2026; Okoye et al., 2026), but also by weak health system integration, a shortage of trained health workers, low-quality nutrition counseling (Sarikhani et al., 2024), and fragmented digital health infrastructure—all of which simultaneously hinder the continuity of data-driven decision-making in primary antenatal care (ANC) services (Harahap et al., 2021).

Early identification of pregnant women at risk of Chronic Energy Deficiency (CED) is crucial for enabling timely nutritional interventions (Mustafa et al., 2026) and reducing the risk of complications for mothers and newborns, particularly in antenatal care facilities with limited resources (Innab et al., 2024), where nutritional screening still relies heavily on anthropometric indicators such as Mid-Upper Arm Circumference (MUAC) and Body Mass Index (BMI) combined with routine clinical assessments (Nisak & Nadhiroh, 2024; Girchenko et al., 2024). Although conventional anthropometric screening methods are practical and cost-effective for assessing the nutrition of pregnant women, their predictive ability regarding future risks to mothers and newborns remains limited because these methods primarily reflect current nutritional status (Patel, 2023), rather than dynamic longitudinal risk patterns (Bosschieter et al., 2024); therefore, the integration of predictive and explainable artificial intelligence (AI)-based clinical decision support systems is needed in maternal health care (Al-Ganad et al., 2026; Du et al., 2022; Rajabi & Etminani, 2021). Furthermore, conventional screening approaches often fail to capture the multidimensional interactions among clinical, socioeconomic, behavioral, and demographic factors that simultaneously influence maternal nutritional status (Badisy et al., 2026). These challenges become even more complex in resource-limited healthcare settings, where data fragmentation, inconsistent clinical documentation, missing data, and low interoperability among healthcare facilities hinder the continuity of patient information and reduce the effectiveness of data-driven clinical decision-making (Ambalavanan et al., 2025).

Recent evidence suggests that explainable AI-based clinical decision support systems that integrate machine learning and interpretable analytics (Innab et al., 2024), such as SHAP-based reasoning, provide more transparent, adaptive, and clinically actionable risk stratification for maternal health care (Du et al., 2022; Ramakrishnan et al., 2021). Although state-of-the-art AI-based predictive models have significantly improved predictive performance in healthcare (Innab et al., 2024; Amann et al., 2020), many systems still have limitations regarding interpretability, the generation of actionable clinical recommendations (Markus et al., 2021), and contextual adaptation for primary care settings, particularly in resource-limited environments (Kruk et al., 2018). As a result, healthcare providers often lack transparent and clinically interpretable decision-support tools (Gaube et al., 2021) capable of providing accurate and timely nutritional risk assessments for interventions targeting pregnant women (Amann et al., 2020; Cutillo et al., 2020).

In recent years, the application of machine learning algorithms such as Random Forest, SVM, Neural Networks, XGBoost, and LightGBM has proven superior to conventional statistical methods for clinical risk classification and health prediction. This is largely due to these algorithms' ability to model nonlinear relationships between variables, accommodate high-dimensional clinical data, and better capture the complexity of inter-feature dependencies (Waldock et al., 2026; Saleem & Ibrahim, 2026).

Recent studies consistently show that ensemble learning approaches, particularly Random Forest and XGBoost, outperform conventional statistical models in predicting maternal nutritional risk and pregnancy-related risks (Ramakrishnan et al., 2021; Rajkomar et al., 2022) because these models effectively capture complex nonlinear interactions among anthropometric, socioeconomic, and clinical variables (Chen & Guestrin, 2016; Shwartz-Ziv & Armon, 2022). However, these models are primarily optimized for predictive performance and often lack discipline-specific nutritional reasoning and clinical interpretability, thereby limiting their direct application in the context of transparent clinical decision-making (Ramakrishnan et al., 2021; Rajkomar et al., 2022; Rudin, 2019). Furthermore, the application of machine learning in maternal health is increasingly

demonstrating significant potential for identifying determinants of maternal malnutrition and child undernutrition through the integration of demographic, anthropometric, and socioeconomic data, thereby enabling the development of more precise early detection strategies and preventive interventions in vulnerable populations, particularly in resource-limited settings (Ramakrishnan et al., 2021; Vaivada et al., 2020; Rajkomar et al., 2022).

Advances in hybrid Clinical Decision Support Systems (CDSS) demonstrate that integrating machine learning with rule-based clinical reasoning can improve predictive accuracy without sacrificing transparency and clinical interpretability, making them better suited for decision-making in complex and dynamic maternal healthcare settings (Babic et al., 2021; Markus et al., 2021). This approach is becoming increasingly important in the context of maternal healthcare, as AI systems focused solely on predictive performance tend to be met with resistance from healthcare professionals if they lack clinically traceable explanation mechanisms (Amann et al., 2022; Gaube et al., 2021). The lack of transparency and interpretability in predictive models is a major barrier to the adoption of AI systems in healthcare, as healthcare professionals tend not to trust systems whose decision-making processes cannot be explained, especially in the context of clinical decisions that directly impact patient safety (Rudin, 2019). In response, Explainable Artificial Intelligence (XAI) has emerged to enhance transparency, accountability, and trust in machine learning models through various techniques such as SHapley Additive exPlanations (SHAP), Local Interpretable Model -Agnostic Explanations (LIME), and feature attribution methods that enable the identification of dominant factors and the presentation of human-understandable explanations for model prediction results (Lundberg & Lee, 2017; Ribeiro et al., 2016). Recent studies indicate that integrating explainability mechanisms into healthcare AI systems can increase clinician acceptance, facilitate the validation of model behavior, and support more informed and accountable clinical decision-making (Amann et al., 2023).

Although significant progress has been made in machine learning-based maternal health prediction, five major challenges remain (Rajkomar et al., 2022). First, most studies focus solely on predictive performance without supporting clinical decision-making. Second, domain knowledge from nutrition experts is rarely integrated into prediction systems. Third, explainability mechanisms are often absent, thereby limiting transparency and trust among healthcare providers. Fourth, validation in primary healthcare facilities with limited resources remains very minimal. Finally, current systems provide risk classifications but rarely generate personalized intervention recommendations. Overall, these limitations hinder the application of predictive analytics as a practical decision-making tool for maternal healthcare. However, various studies have demonstrated the potential of machine learning to improve the accuracy of maternal health risk predictions; there remains a significant gap between the predictive capabilities of these models and their utilization in clinical practice. Most studies focus on optimizing classification performance (Rajkomar et al., 2019; Topol, 2019; Liu et al., 2024) using metrics such as accuracy, F1-score, or the area under the receiver operating characteristic curve (AUROC); however, these studies have not adequately translated these predictive results into actionable decision support for healthcare providers (Rudin, 2019; Lundberg & Lee, 2017; Brankovic et al., 2024). As a result, many models function as isolated classification tools (stand-alone classifiers) without a clear mechanism to support the operational planning of nutritional interventions or maternal risk management (Shortliffe & Sepúlveda, 2018; Jiang et al., 2017; Rane et al., 2023). Furthermore, although machine-learning-based approaches are increasingly used to predict malnutrition and pregnancy outcomes, most systems remain black-box in nature with limited interpretability (Shickel et al., 2021). This limitation poses a significant barrier in the healthcare setting, as healthcare providers require a clear understanding of the factors underlying a prediction before using it as the basis for clinical decision-making.

Unlike previous studies, which generally focused on risk classification without comprehensive clinical decision support, this research developed NutriKEK-AI as an Explainable Hybrid Clinical Decision Support System that integrates machine learning, Explainable Artificial Intelligence (XAI), and expert-knowledge-based reasoning into a single unified architecture. This system not only predicts the risk of Chronic Energy Deficiency (CED) in pregnant women but also generates personalized, contextual, and actionable nutritional intervention recommendations, while enhancing transparency through explainability mechanisms so that healthcare professionals can understand the basis of each prediction. This approach addresses the limitations of conventional screening, which tends to be reactive, lacks predictive power, and offers minimal interpretability, while existing AI models often prioritize algorithmic accuracy without integrating clinical needs or implementation in primary care settings. Thus, although health data on pregnant women is increasingly available through digital systems, its utilization for early detection remains suboptimal due to low model transparency, the absence of operational recommendations, and insufficient validation in the context of high-risk primary care settings in resource-constrained regions. The key innovation of NutriKEK-AI lies in the simultaneous integration of three components—machine learning, XAI, and rule-based clinical reasoning—which have traditionally been developed separately. This integration results in a predictive system that is not only accurate but also

explainable, traceable, and actionable to support clinical decision-making in primary maternal health care settings with limited resources.

The main innovation of NutriKEK-AI lies in the simultaneous integration of three components that have traditionally been developed separately: machine learning, XAI, and rule-based clinical reasoning, resulting in a predictive system that is not only accurate but also explainable, traceable (allowing the basis for decision-making to be traced), and actionable (generating recommendations that healthcare workers can immediately implement) to support clinical decision-making in primary maternal healthcare with limited resources. The development of this system also supports the government's program to integrate artificial intelligence (AI) into the national health system to accelerate early disease detection and improve the efficiency of health services, while also contributing to the achievement of SDG 3 (Good Health and Well-being) through efforts to reduce maternal and infant mortality rates and strengthen access to quality maternal health services, particularly for populations in primary care settings with limited resources who have historically been underserved by conventional screening systems. This study aims to develop and validate NutriKEK-AI as an Explainable Hybrid Clinical Decision Support System to predict the risk of Chronic Energy Deficiency (CED) in pregnant women in resource-limited areas. Specifically, this study includes: (1) the development of a hybrid prediction model that integrates machine learning and expert clinical knowledge;

(2) evaluation of the model's predictive performance using maternal health data; (3) assessment of the interpretability and explainability of the prediction results; and (4) analysis of the system's potential for implementation as a decision-support tool and for providing intervention recommendations in maternal nutrition management within resource-limited primary health care settings.

2. MATERIALS AND METHODS

This "Materials and Methods" section systematically describes the study design, location and participants, data sources and variables, research instruments, data collection procedures, model development, and evaluation of the NutriKEK-AI system.

Research Design

This study uses Design and Development Research (DDR) as the primary methodological framework to guide the systematic design, development, validation, and evaluation of NutriKEK-AI, an explainable hybrid clinical decision support system designed to predict the risk of KEK in pregnant women in resource-constrained settings. Within the DDR framework, the Cross-Industry Standard Process for Data Mining (CRISP-DM) was adopted to structure the machine learning development process, which includes the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Meanwhile, Agile Development principles were applied to support iterative prototype development, continuous testing, interface refinement, and user-centered system improvements through ongoing stakeholder feedback (Figure 1). This study integrates quantitative validation of predictive models and a mixed-methods-based system evaluation to assess predictive performance, explainability, usability, user acceptance, and the feasibility of system implementation. The explainability aspect is implemented through feature importance analysis and rule-based, traceable recommendation mechanisms, enabling healthcare professionals to understand key risk factors and the rationale behind the system's recommendations. The best predictive model is then integrated into a web-based explainable decision support dashboard and validated through predictive performance evaluation, expert assessment by maternal health and nutrition specialists, and user acceptance testing in primary health care settings. In addition to emphasizing validity, practicality, and effectiveness, this study also applies the principles of Responsible AI—which include transparency, contextual relevance, and accessibility—to enhance the system's trustworthiness and minimize technological bias among underserved populations of mothers.

Study Setting and Participants

The study was conducted in Lebak and Pandeglang Regencies, Banten Province, Indonesia, which were purposively selected because they represent high-risk maternal areas with limited resources, characterized by high prevalence of stunting (27–32%), chronic energy deficiency (CED) among pregnant women (18–24%), limited access to health services, low nutrition literacy, and an ANC system that is not yet digitally integrated. A total of 253 participants were involved, consisting of 214 pregnant women, 32 health workers, and 7 expert validators. Of the 236 prospective pregnant women identified through ANC data, 223 met the study criteria and agreed to participate; subsequently, 9 records were excluded due to critical missing data exceeding 20%, resulting in a final dataset of 214 pregnant women (see Table 1).

A purposive, criterion-based sampling technique was used, with inclusion criteria consisting of active ANC attendance, residence of ≥12 months, completeness of maternal data, and consent to participate in the study; invalid data, critical missing data >20%, and severe obstetric complications served as exclusion criteria. A total of 214 complete observations were included in model development. The dataset was stratified into a training set (n = 169; 79.0%) and an independent held-out test set (n = 45; 21.0%). Model stability was assessed using stratified 10-fold cross-validation within the training data, while class imbalance was addressed exclusively within the training data using SMOTE. The independent test set remained untouched and was used only for final performance evaluation. Data quality was ensured through data source triangulation, verification of ANC records, consistency checks, removal of duplicates, data normalization, and domain-knowledge-based median imputation. Expert validators (n=7) were selected from a multidisciplinary group, consisting of 2 nutrition experts, 2 maternal and child health (MCH) experts, and 1 expert each from the fields of information technology (IT), media, and linguistics, all with at least 5–10 years of professional experience. Validation was conducted through a structured expert review using the Content Validity Index (CVI) and expert consensus. All research procedures have received ethical approval 42/E-5/UNILAM/2026; 1658/LL4/PG/2026 and adhere to the principles of responsible AI through informed consent, de-identification, data encryption, access restrictions, privacy protection, and standardized workflows to ensure the validity, reproducibility, and integrity of the NutriKEK-AI development. Respondent characteristics showed a mean maternal age of 27.8 ± 5.6 years (95% CI: 27.0–28.6), with the majority in the 25–34 age group (41.1%) and the 20–24 age group (31.3%). Most respondents had a primary (46.7%) or secondary (39.3%) education and were homemakers (64.5%). There were no significant differences between Lebak and Pandeglang Regencies regarding age, age group, parity, or gestational age (all $p > 0.05$), with very small effect sizes (Cohen's $d = 0.09$; Cramer's $V \leq 0.06$). These findings indicate homogeneity in the respondents' baseline characteristics across regions, thereby supporting the validity of subsequent predictive analyses. All statistical assumptions were met; there were no missing data (100% completeness), and all expected frequencies for the chi-square tests were above the recommended minimum thresholds (see Table 2).

Table 1. Participant Distribution Across DBR Phases and Study Sites

DBR Phase	Participant Group	Lebak (n)	Pandeglang (n)	Total (n)
Needs Assessment	Pregnant Women (ANC-registered)	112	102	214
	Health Workers	17	15	32
System Design Validation	Expert Validators (AI–Health Panel)	–	–	7*)
Pilot Testing	Pregnant Women	62	58	120
	Health Workers	6	6	12
Field Implementation & Evaluation	Pregnant Women	85	75	160
	Health Workers	17	15	32
TOTAL UNIQUE PARTICIPANTS	All Participants	282	256	545*)

Table 2. Maternal Obstetric Characteristics (N = 214)

Variable	Category	Total n (%)	Lebak (n=112)	Pandeglang (n=102)	Statistic	p-value (Effect size)
Age (years)	Mean ± SD	27.8 ± 5.6	27.6 ± 5.7	28.1 ± 5.4	t(212)=0.70	0.482 (d=0.09)
	95% CI (mean)	27.0–28.6	—	—	—	—
Age groups	<20	32 (15.0)	18 (16.1)	14 (13.7)	$\chi^2 = 0.84$†	0.839 (V=0.06)
	20–24	67 (31.3)	36 (32.1)	31 (30.4)		

Variable	Category	Total n (%)	Lebak (n=112)	Pandeglang (n=102)	Statistic	p-value (Effect size)
	25–34	88 (41.1)	45 (40.2)	43 (42.2)		
	≥35	27 (12.6)	13 (11.6)	14 (13.7)		
Parity	Primigravida	73 (34.1)	38 (33.9)	35 (34.3)	$\chi^2 = 0.01$	0.972 (V=0.01)
	Multigravida	141 (65.9)	74 (66.1)	67 (65.7)	Cramer's V = 0.01	
Gestational age	Trimester I	40 (18.7)	22 (19.6)	18 (17.6)	$\chi^2 = 0.15$	0.927 (V=0.02)
	Trimester II	75 (35.0)	39 (34.8)	36 (35.3)		
	Trimester III	99 (46.3)	51 (45.5)	48 (47.1)		

Note that all assumptions for parametric tests and chi-square tests were met. Effect sizes are interpreted according to Cohen (1988): small (0.20), medium (0.50), large (0.80) for Cohen's d, and small (0.10), medium (0.30), large (0.50) for Cramer's V. No missing data were found (completeness = 100%). The expected number of cells for the chi-square test was >5 in all cases.

Table 3. Instrument Validation

Instrument	Content Validity	I-CVI	S-CVI/ Ave	S-CVI/ UA	Inter-rater Reliability	Reliability
MNAF	0.92 (0.88–0.96)	0.86–0.97	0.94	0.89	ICC = 0.91	α = 0.87 (0.84–0.90), CR = 0.89, AVE = 0.62
EKQ	0.91 (0.87–0.95)	0.85–0.96	0.93	0.87	ICC = 0.89	α = 0.89 (0.86–0.92), CR = 0.90, AVE = 0.64
TAM	0.95 (0.92–0.98)	PU: 0.91, PEOU: 0.93, BI: 0.90	0.96	0.92	ICC = 0.93	α = 0.92 (0.90–0.94), CR = 0.94, AVE = 0.71
SUS	0.90 (validated instrument)	-	-	-	ICC = 0.88	α = 0.91; Mean = 82.4 ± 6.8 (Grade A, "Excellent usability")

Description:

1. MNAF = Maternal Nutrition Assessment Form; EKQ = Education Knowledge Questionnaire; TAM = Technology Acceptance Model (PU = Perceived Usefulness, PEOU = Perceived Ease of Use, BI = Behavioral Intention); SUS = System Usability Scale.
2. The I-CVI and S-CVI were calculated based on the ratings of 7 expert validators using a 4-point scale (1 = highly inappropriate, 4 = highly appropriate); scores of 3–4 were categorized as relevant. The thresholds used were: S-CVI/Ave ≥ 0.90 (Polit & Beck, 2006); I-CVI ≥ 0.78 for a panel of ≥6 experts (Polit & Beck, 2006).
3. ICC = Intraclass Correlation Coefficient, interpreted according to Koo & Li (2016): >0.90 = very good.
4. α = Cronbach's alpha; CR = Composite Reliability; AVE = Average Variance Extracted (AVE ≥ 0.50 indicates that convergent validity is met, according to Fornell & Larcker, 1981).
5. The SUS is a standardized instrument that has been internationally validated (Brooke, 1996), so no repeat I-CVI/S-CVI testing was conducted; SUS scores were interpreted using the SUS grading scale (Bangor, Kortum, & Miller, 2009), with scores of ≥80.3 categorized as Grade A ("Excellent usability").

Data Sources and Research Variables

Research data were obtained from primary and secondary sources. Primary data were collected through anthropometric measurements and structured interviews with health workers and pregnant women. Secondary data were obtained from the e-PPGBM (Electronic Community-Based Nutrition Recording and Reporting System) of the Lebak and Pandeglang District Health Offices. All data were integrated into a single research database after undergoing verification, cleaning, and standardization. Research variables were grouped into five domains: demographics (maternal age and family income), obstetrics (gestational age, parity, and history of chronic energy deficiency), anthropometry (upper arm circumference [UAC], weight, height, and body mass index), clinical (hemoglobin levels), and nutritional intake (energy intake and protein intake). History of CED refers to the mother's previous experience with Chronic Energy Deficiency during prior pregnancies and was used as one of the predictor variables. The outcome variable is the status of Chronic Energy Deficiency (CED) in the current pregnancy, determined based on an Upper Arm Circumference (UAC) < 23.5 cm in accordance with guidelines from the Ministry of Health of the Republic of Indonesia and the WHO. The outcome is categorized into two classes: CED and non-CED.

Research Instruments

This study used five main instruments, namely (1) the Maternal Nutrition Assessment Form (MNAF), which consists of 32 questions based on WHO and Ministry of Health guidelines covering demographic, socioeconomic, obstetric, anthropometric, clinical, and energy intake aspects, with anthropometric measurements taken by trained enumerators. Stunting status was determined based on LiLA: <23.5 cm (coded as 0) and non-stunted (coded as 1). (2) The Expert Knowledge Questionnaire (EKQ) is used to elicit and validate multidisciplinary expert knowledge as the basis for developing the system's forward chaining rules and clinical recommendations; (3) The System Usability Scale (SUS), consisting of 10 Likert-scale items, is used to assess the system's usability in terms of ease of use, efficiency, and user satisfaction; (4) The Technology Acceptance Model (TAM) Questionnaire was used to measure user acceptance of the NutriKEK-AI system—specifically, the extent to which users consider the system useful, easy to use, likable, and intend to continue using it; and (5) A Semi-Structured Interview Guide was used to explore user experiences, implementation barriers, and needs for further system development.

Data Collection Procedures

The CRISP-DM Phase and Agile Development study was conducted through seven main, interconnected stages: participant recruitment, collection of maternal health data, data preprocessing, machine learning model development, expert knowledge base construction, integration of the clinical decision support system, and system validation and evaluation

Participant Recruitment and Eligibility Screening

Participants were recruited through community health centers, integrated health posts, and village midwives in Lebak and Pandeglang Regencies using registration data from e-PPBGM/SIGIZI. Participant eligibility was verified based on inclusion and exclusion criteria (such as a history of certain diseases) by trained health workers. Eligible pregnant women were provided with an explanation of the study's objectives, procedures, data confidentiality, and participant rights before signing the informed consent form. At this stage, the research team also selected health workers and experts to participate based on their experience and expertise in their respective fields: maternal and child health (MCH), nutrition, language, media, and information technology. Each respondent was assigned a unique identification code to ensure privacy and to make it easier for the research team to process the data.

Maternal Health Data Acquisition

Maternal health data were collected from two sources: primary data, obtained through structured interviews using the MNAF instrument and anthropometric measurements (arm circumference, weight, and height); and secondary data, obtained from records and reports in the SIGIZI system of the Lebak and Pandeglang District Health Offices.

The variables collected included demographic data (maternal age and income), obstetric history (gestational age, parity, and history of malnutrition in previous pregnancies), anthropometric measurements (mid-upper arm circumference, weight, height, and BMI), clinical data (hemoglobin levels), and nutritional intake (energy and protein intake). Measurements were conducted by trained enumerators (undergraduate nutrition students and associate degree midwifery students), following the same standard operating procedures (SOPs) for all respondents to minimize inter-observer bias.

Data Cleaning and Preprocessing

All data will first undergo a quality-checking process, which involves: ensuring data consistency, verifying that there are no duplicate entries, checking data completeness, and identifying outliers (measurement results that fall far outside the normal range).

If missing data accounts for less than 5% of the total, the data will be imputed using statistical estimates (mean, median, or estimates based on other similar data). For data with outliers, the source will be rechecked to confirm whether an input error occurred or if the reported medical condition is accurate.

Categorical data—such as educational level, incidence of anemia, and incidence of stunting—are converted to numerical formats so they can be processed by the system, and their data scales are then determined. Class imbalance was addressed exclusively within the training data after stratified train-test partitioning.

Machine Learning Model Development

The cleaned data was used to build a predictive model for KEK. Five different methods were tested and compared: logistic regression, decision tree, SVM, random forest, and XGBoost, with the aim of identifying the most accurate method for predicting KEK in pregnant women.

The final dataset of 214 pregnant women was stratified into a model-development training set (n = 169; 79.0%) and an independent held-out test set (n = 45; 21.0%). The training set underwent stratified 10-fold cross-validation to assess model stability and reduce the risk of overfitting. Class imbalance was addressed exclusively within the training data using resampling/SMOTE, while the independent test set remained untouched and was used only for final model evaluation.

Model accuracy was measured from several perspectives: (1) how often the predictions were correct, how rarely the model missed actual cases of KEK in pregnant women, and how well the model distinguished women at risk for KEK from those who were not. In addition, we also checked whether the risk percentages provided by the model were reasonable compared to real-world data, and whether this model truly aided clinical decision-making compared to conventional methods (without AI assistance). The entire process was carried out using Python and RapidMiner software, with automated and standardized testing procedures. Validation at this stage was formative in nature, aimed at improving the system.

Expert Knowledge Base Development

The clinical knowledge base was developed through semi-structured interviews, focus group discussions, a literature review, and an analysis of national and international guidelines on maternal nutrition. The expert panel consisted of maternal nutritionists, maternal and child health specialists, linguists, and media and IT experts. The knowledge gained was translated into forward-chaining IF–THEN rules covering risk identification, risk level classification, nutritional intervention recommendations, referral needs, and follow-up care. All rules were iteratively validated by the expert panel to ensure clinical accuracy and consistency.

Hybrid CDSS and System Development

An expert knowledge-based machine learning model was integrated into NutriKEK-AI as an Explainable Hybrid Clinical Decision Support System using an Agile Development approach. The Agile Development process was carried out over four sprints. The first sprint focused on developing the data acquisition module and the user interface. The second sprint integrated the machine learning model for KEK risk prediction. The third sprint implemented the Explainable Artificial Intelligence (XAI) module and a rule-based knowledge base using forward chaining. The fourth sprint focused on integrating the clinical dashboard, optimizing system performance, and conducting validation with healthcare professionals. Each sprint concluded with a sprint review and system improvements based on user feedback.

NutriKEK-AI is designed as a six-layer hybrid Clinical Decision Support System (CDSS) architecture that includes: (1) acquisition of maternal demographic, anthropometric, clinical, and antenatal data; (2) machine learning-based prediction of Chronic Energy Deficiency (CED) risk; (3) model interpretation using SHapley Additive exPlanations (SHAP); (4) inference based on clinical rules using forward chaining; (5) personalized nutritional intervention recommendations; and (6) a visualization dashboard to present prediction results and decision support in an informative and actionable manner for healthcare workers. All prediction results, interpretations, and recommendations are displayed through an interactive dashboard designed for primary healthcare facilities with limited resources. To enhance the transparency and interpretability of the predictive model, this study integrates an XAI module using the SHAP method. SHAP analysis is performed automatically using machine learning

software to identify the contribution of each variable to the predicted results for the risk of KEK. Explainability was evaluated at two levels: global explanation to determine the relative importance of each feature to the model's overall performance, and local explanation to explain the contribution of each variable to individual predictions for each pregnant woman. The interpretation results were used to identify the dominant risk factors (feature contributions) underlying the model's predictions so that they could be understood and verified by healthcare professionals in the clinical decision-making process.

System Validation, Usability Evaluation, and User Acceptance Testing

The system evaluation was conducted in stages through two sub-phases: pilot testing and field implementation, to assess the predictive performance, clinical validity, usability, and user acceptance of NutriKEK-AI.

During the pilot testing sub-phase (at the subdistrict level), the system was tested on a limited basis with 94 pregnant women and 12 healthcare workers in Lebak and Pandeglang Regencies. Model performance was assessed using accuracy, precision, recall, F1-score, ROC-AUC, and a confusion matrix, supplemented by calibration analysis using the Brier Score and Calibration Curve, as well as Decision Curve Analysis (DCA) to evaluate the model's clinical benefits compared to conventional strategies. Clinical validity was evaluated by a multidisciplinary expert panel that assessed the accuracy of recommendations, the relevance of risk factors, the transparency of predictions, and the feasibility of implementation. Functional testing was conducted using black-box testing, while usability and initial user acceptance were evaluated using the SUS and TAM. Semi-structured interviews were conducted to explore user experiences and initial barriers to implementation. Based on the results of pilot testing (in Lebak and Pandeglang Regencies), the system was refined before entering the field implementation sub-phase, which involved 160 pregnant women and 32 healthcare workers on a broader scale of use at primary healthcare facilities. In this sub-phase, healthcare workers used the system directly during routine ANC visits to predict patients' risk of KEK. The evaluation focused on the consistency of prediction results in a real-world setting, the level of user acceptance and satisfaction, and barriers to implementation in the field. Healthcare workers' satisfaction was reassessed using the SUS and TAM, while patients' experiences and satisfaction were explored through interviews regarding the NutriKEK-AI-based services they received. Findings from these two sub-phases form the basis for final refinements and recommendations for the continued implementation of the NutriKEK-AI system.

Data Analysis

Data analysis was performed using Python, RapidMiner, and SPSS software. Descriptive analysis was used to describe the characteristics of the respondents using the mean, standard deviation, median, interquartile range, frequency, and percentage. Differences in characteristics between groups were analyzed using the Chi-square test for categorical data, the independent t-test for normally distributed numerical data, and the Mann–Whitney U test for data that did not meet the assumption of normality. Factors associated with the occurrence of KEK were analyzed using logistic regression. Predictive models were developed using the Logistic Regression, Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Artificial Neural Network (ANN) algorithms, combined with a hyperparameter tuning process to obtain optimal model configurations. All statistical tests used a significance level of $p < 0.05$.

This study has received ethical approval from the La Tansa Mashiro University Health Research Ethics Committee No. 42/E-5/UNILAM/2026 and 1658/LL4/PG/2026. All participants provided written informed consent prior to the study. Data confidentiality was maintained through anonymization, encryption, and authorization-based access restrictions. To implement the principles of Responsible Artificial Intelligence, we conducted class imbalance analysis, evaluated performance across various population subgroups, conducted a multidisciplinary review by experts, and applied SHAP-based Explainable AI to enhance transparency, accountability, fairness, and trust in the system. All research procedures were conducted in accordance with the Declaration of Helsinki and applicable health research ethics guidelines.

3. RESULTS AND DISCUSSION

3.1 Development Framework of NutriKEK-AI

NutriKEK-AI was developed using a combination of CRISP-DM and Agile Development across four sprints, which included the identification of maternal risk factors, data preprocessing and feature engineering, the development of a machine learning model with SHAP integration, and system implementation and evaluation. The result of this development is a hybrid clinical decision support system that integrates predictive models, rule-based inference, explainability visualization, and a maternal risk

monitoring dashboard (see Table 4 and Figure 1). During the exploration phase, relevant clinical and social factors—particularly anemia status, adequacy of ANC visits, socioeconomic status, maternal age, and the LiLA anthropometric indicator—were identified as part of the maternal risk profile. However, since CED status in this study was defined based on LiLA <23.5 cm, LiLA was not treated as an independent predictor in the interpretation of the predictive model. Subsequent explainability analysis focused on factors that do not directly define the KEK outcome.

Table 4. Integrated CRISP-DM–Agile Development Lifecycle of NutriKEK-AI

CRISP-DM Phase	Agile Sprint	Main Activities	Key Output	Related Results
Business Understanding	Sprint 1	Maternal risk mapping, workflow analysis, stakeholder consultation	Clinical objectives and maternal risk taxonomy	Tables 5–6
Data Understanding	Sprint 1–2	Exploratory analysis, regional profiling, missing-data assessment	Identification of key maternal risk factors	Table 5
Data Preparation	Sprint 2	Stratified train-test partitioning, feature engineering and normalization, SMOTE applied only within training data, and outlier handling	Balanced training dataset and untouched independent test dataset	Table 7
Modeling	Sprint 3	ML training, hyperparameter tuning, SHAP integration	Optimized explainable XGBoost model	Tables 8–9
Evaluation	Sprint 3–4	Calibration analysis, expert validation, usability testing	Validated and interpretable AI-CDSS	Tables 10–18
Deployment	Sprint 4	Dashboard refinement, workflow integration, field implementation	Operational NutriKEK-AI system	Tables 14–18

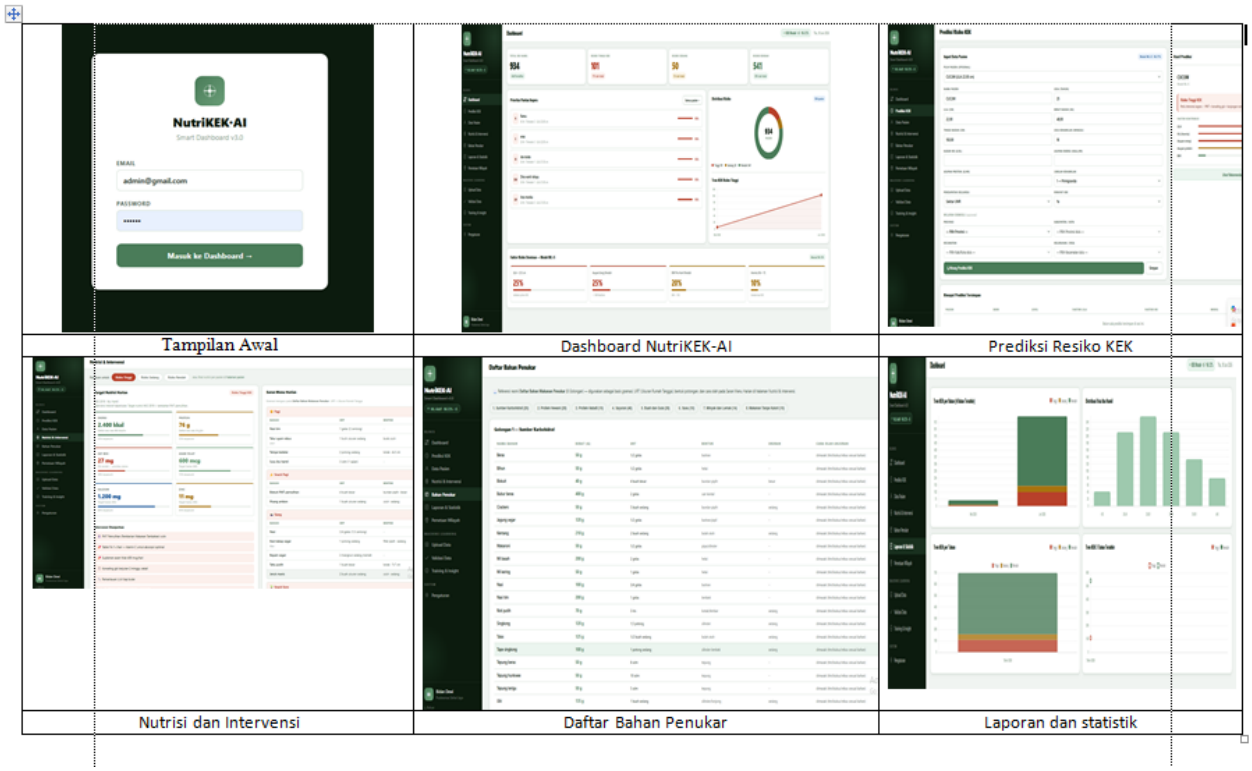


Figure 1. Expert System Page for Detecting KEK in Pregnant Women

3.2 Dataset Characteristics and Maternal Risk Profile

A total of 214 pregnant women participated in this study, comprising 112 respondents from Lebak Regency and 102 respondents from Pandeglang Regency. Most sociodemographic characteristics did not differ significantly between the two regions ($p > 0.05$). The mean age of the respondents was 27.8 ± 5.6 years, and the mean MUAC was 24.1 ± 2.3 cm. The prevalence of Chronic Energy Deficiency (CED) was 18.7% (95% CI: 15.2–22.9), the prevalence of anemia was 42.6%, and 55.1% of the respondents had inadequate ANC visits (<6 times the Ministry of Health standard). Significant differences between regions were found in MUAC ($p = 0.018$), hemoglobin levels ($p = 0.017$), number of visits ($p = 0.003$), adequacy of ANC ($p = 0.049$), anemia ($p = 0.017$), prevalence of CED ($p = 0.042$), and the proportion of high-risk maternal age groups ($p = 0.038$) (see Table 5). Multivariate logistic regression analysis showed that low socioeconomic status (OR = 2.63; $p < 0.001$), inadequate ANC visits (OR = 2.12; $p = 0.008$), anemia (OR = 2.05; $p = 0.011$), and high-risk maternal age (OR = 1.59; $p = 0.046$) were significantly associated with the occurrence of CED. The interaction between ANC visits and anemia status was also significant (OR = 3.21; $p = 0.002$), indicating a greater combined effect than the effects of each factor individually (see Table 6).

Table 5. Maternal Nutritional and Clinical Status

Indicator	Total (Mean ± SD / n %, 95% CI)	Lebak	Pandeglang	Test statistic	p-value	Effect size
MUAC (cm)	24.1 ± 2.3 (23.8–24.4)	23.8 ± 2.4	24.5 ± 2.1	t = 2.34	0.018	d = 0.32
CED (<23.5 cm)	40 (18.7%) (15.2–22.9)	21.5%	16.2%	$\chi^2 = 4.12$	0.042	V = 0.14
BMI (kg/m²)	22.3 ± 3.1	22.1 ± 3.0	22.5 ± 3.2	t = 1.57	0.118*	d = 0.18
Underweight (BMI <18.5)	21.4%	22.3%	20.6%	$\chi^2 = 0.12$	0.728*	V = 0.02
Hemoglobin (g/dL)	10.9 ± 1.4 (10.7–11.1)	10.7 ± 1.3	11.1 ± 1.5	t = 2.38	0.017	d = 0.36
Anemia (<11 g/dL)	42.6%	45.8%	39.1%	$\chi^2 = 5.67$	0.017	V = 0.16
ANC visits	3.7 ± 1.2	3.5 ± 1.3	4.0 ± 1.1	t = 3.01	0.003	d = 0.42
ANC adequacy (≥6 kemenkes-based)	55.1% inadequate	56.7%	53.4%	$\chi^2 = 3.88$	0.049	V = 0.13
Multigravida	65.9%	66.1%	65.7%	$\chi^2 = 0.01$	0.920*	V = 0.01
Maternal Age ≥35 or <20 (risk group)	27.6%	28.3%	26.9%	$\chi^2 = 2.91$	0.038	V = 0.12

***** tidak signifikan

Table 6. Determinants of Chronic Energy Deficiency (CED)

Risk Factor / Term	Category (Reference)	Adjusted OR	95% CI	p-value	Δ log-odds	VIF	Regional OR (Lebak)	Regional OR (Pandeglang)
Economic status	Lowest vs highest	2.63	1.48–4.68	<0.001	+0.97	1.92	2.81	2.34
ANC visits	<6 vs ≥6	2.12	1.21–3.71	0.008	+0.75	1.68	2.45	1.78
Anemia status	Yes vs No	2.05	1.18–3.55	0.011	+0.72	1.54	2.30	1.85
Maternal age	<20 or ≥35 vs 20–34	1.59	1.01–2.49	0.046	+0.47	1.31	1.72	1.45

Asupan Kalori	< 80% vs ≥ 80%	3.86	0.74-19.97	0.108	+1.35	1.89	4.10	3.62
Asupan Protein	< 80% vs ≥ 80%	3.46	0.66-18.16	0.143	+1.24	1.65	3.70	3.25
ANC × Anemia	Interaction term	3.21	1.45–6.12	0.002	+1.17	—	3.45	2.98
Model-Based	Logistic probability	0.41	0.33–0.52	<0.001	+0.88	—	0.46	0.37

3.3 Machine Learning Model Performance

Five machine learning algorithms were evaluated to predict CED in pregnant women: Logistic Regression, Decision Tree, SVM, Random Forest, and XGBoost. The XGBoost model performed best, with a precision of 0.86, a recall of 0.84, an F1-score of 0.85, an AUC of 0.91 (95% CI: 0.88–0.94), and a Brier Score of 0.12; the results of 10-fold cross-validation showed an average accuracy of 0.87 ± 0.02 (see Table 7). The AUC value for XGBoost was higher than that of Logistic Regression (ΔAUC = 0.09; p = 0.003), Random Forest (ΔAUC = 0.02; p = 0.035), and SVM (ΔAUC = 0.05; p = 0.047) (see Table 8). On the independent test set, classification performance varied across risk groups. For the high-risk group, recall was 0.57, precision was 0.67, and the F1-score was 0.62 (see Table 9).

Table 7. Model Performance Comparison Based on Stratified Cross-Validation and Bootstrap Resampling

Model	Accuracy	Precision	Recall	F1	AUC	Brier Score	CV (10-fold mean ± SD)
Logistic Regression	0.81	0.79	0.77	0.78	0.82 (0.78–0.86)	0.18	0.80 ± 0.03
Decision Tree	0.80	0.78	0.76	0.77	0.80 (0.76–0.84)	0.19	0.79 ± 0.04
SVM	0.84	0.82	0.81	0.81	0.86 (0.83–0.89)	0.15	0.83 ± 0.03
Random Forest	0.87	0.85	0.83	0.84	0.89 (0.86–0.92)	0.13	0.86 ± 0.02
XGBoost	0.88	0.86	0.84	0.85	0.91 (0.88–0.94)	0.12	0.87 ± 0.02

Table 8. Statistical Comparison Between Models

Comparison	Test	Statistic	p-value	Effect Size
XGBoost vs Logistic Regression	DeLong test (AUC)	z = 3.02	0.003	ΔAUC = +0.09 (95% CI: 0.04–0.14)
XGBoost vs Random Forest	DeLong test (AUC)	z = 2.11	0.035	ΔAUC = +0.02 (95% CI: 0.01–0.05)
XGBoost vs SVM	DeLong test (AUC)	z = 1.89	0.047	ΔAUC = +0.05 (95% CI: 0.00–0.10)
XGBoost vs Logistic Regression	McNemar test (exact)	χ^2 = 6.42	0.011	Accuracy gain = +0.07

Table 9. Independent Held-Out Test-Set Evaluation of XGBoost (n=45)

Evaluation Layer	Indicator	Metric / Class	Result
A. Confusion Matrix (Test Set n=45)	Low → Low	True Positive	18
	Low → Moderate	Misclassification	2
	Low → High	Misclassification	0
	Moderate → Low	Misclassification	3
	Moderate → Moderate	True Positive	13

Evaluation Layer	Indicator	Metric / Class	Result
	Moderate → High	Misclassification	2
	High → Low	Misclassification	1
	High → Moderate	Misclassification	2
	High → High	True Positive	4
B. Class-wise Predictive Performance	Low Risk	Precision / Recall / F1	0.82 / 0.90 / 0.86
	Moderate Risk	Precision / Recall / F1	0.76 / 0.72 / 0.74
	High Risk	Precision / Recall / F1	0.67 / 0.57 / 0.62
C. Clinical Diagnostic Validity	Low Risk	Specificity	0.92
	Moderate Risk	Specificity	0.81
	High Risk	Specificity	0.78
	Overall	Balanced Accuracy	0.73
	Overall	Cohen's Kappa	0.64
D. Model Robustness (10-Fold CV)	Accuracy	Mean ± SD	0.87 ± 0.02
	AUC	Mean ± SD	0.90 ± 0.02
	Accuracy Range	Stability	0.85–0.89
	AUC Range	Stability	0.88–0.92
E. Clinical Utility Index	Composite Measure	Clinical Risk Index (CRI)*	0.84

3.4 Clinical Validation and System Performance

Through multilevel cross-validation and bootstrap evaluation, XGBoost achieved an accuracy of 0.88, a macro-average ROC-AUC of 0.91 (95% CI: 0.88–0.94) on the "one-vs-rest" model, a recall of 0.84, and a Brier score of 0.12. On a separate, set-aside test set, class-specific performance was lower and varied across risk categories (see Table 7). Validation by 7 experts yielded an I-CVI of 0.71–1.00, an S-CVI/Ave of 0.80, a Fleiss Kappa of 0.60, and an overall validation score of 3.71 ± 0.35. The clinical validity, interpretability, system integration, and usability scores were 3.78 ± 0.31, 3.72 ± 0.34, 3.69 ± 0.36, and 3.65 ± 0.38, respectively (see Table 10).

Table 10. Integrated Content Validity, Inter-Rater Reliability, and Deployment Readiness Evaluation of NutriKEK-AI

Validation Domain	Representative Indicator	I-CVI (Expert Agreement)	Domain Mean ± SD	Fleiss' Kappa	Dominant Expert Group
Clinical Validity	MUAC risk stratification logic	0.86 (6/7)	3.78 ± 0.31	0.65	Obstetrics (n=2; 8–15 yrs), Nutrition (n=2; 7–12 yrs)
	Anemia-based maternal	0.86 (6/7)			
	ANC risk weighting mechanism	0.71 (5/7)			
AI Interpretability	SHAP-based explainability	0.71 (5/7)	3.72 ± 0.34	0.70	AI/ML (n=1; 5–10 yrs), Informatics (n=1; >10 yrs)

Validation Domain	Representative Indicator	I-CVI (Expert Agreement)	Domain Mean ± SD	Fleiss' Kappa	Dominant Expert Group
	Feature importance transparency	0.86 (6/7)			
	Rule-traceability mechanism	1 (7/7)			
Usability & Accessibility	User interface usability	0.71 (5/7)	3.65 ± 0.38	0.55	Public Health (n=1; 6–14 yrs)
	Dashboard navigation clarity	0.86 (6/7)			
	Mobile responsiveness	0.71 (5/7)			
System Integration & Deployment	Data pipeline integration	0.71 (5/7)	3.69 ± 0.36	0.55	Informatics & AI experts
	Integration with ANC workflow	0.71 (5/7)			
	Decision-support readiness	0.86 (6/7)			
Overall System Validation	Aggregate system validity	S-CVI/Ave = 0.80	3.71 ± 0.35	0.60	Multidisciplinary panel (N=7)

3.5 Usability and Technology Acceptance

A usability evaluation involving 32 healthcare workers showed that NutriKEK-AI achieved a SUS score of 84.6 ± 6.8, with a task completion rate of 93.8%, an average completion time of 2.7 ± 0.9 minutes, a critical error rate of 4.2%, and a system response time of 1.8 seconds (see Table 11). Based on the TAM evaluation (see Table 12), the PU value was 4.41 ± 0.46, the PEOU was 4.32 ± 0.51, and the behavioral intention to use (BI) was 4.47 ± 0.42. Structural equation modeling revealed significant relationships between PEOU and PU ($\beta = 0.71$; $p < 0.001$), PU and BI ($\beta = 0.78$; $p < 0.001$), and PEOU and BI ($\beta = 0.69$; $p < 0.001$) (see Table 13).

Table 11. Ease of Use and Feasibility of Implementing NutriKEK-AI (N=32: 24 Village Midwives and 8 Community Health Center Nutrition Program Coordinators)

Domain	Indicator	Result	Benchmark
System Usability	SUS Score	84.6 ± 5.8 (95% CI: 81.2–87.4)	>68
Learnability	Learnability Score	83.1 ± 7.2	>70
Operational Efficiency	Task Completion Rate	93.8%	>90%
Operational Efficiency	Mean Task Completion Time	2.7 ± 0.9 min	<5 min
Error Tolerance	Critical Error Rate	4.2%	<10%
User Acceptance	User Satisfaction	4.45 ± 0.48	>4.0
Cognitive Load	Perceived Cognitive Burden	1.92 ± 0.61	<3.0

Domain	Indicator	Result	Benchmark
Infrastructure Feasibility	Internet Stability Tolerance	Operable under unstable 3G connectivity	—
Infrastructure Feasibility	Mobile Compatibility	Android low-end compatible	—
Infrastructure Feasibility	Offline Cache Support	Available	—
System Performance	Mean Response Time	1.8 seconds	<3 seconds
Clinical Acceptance	Healthcare Worker Satisfaction	92.4% positive	>80%
Explainability Acceptance	Recommendation Interpretability	89.7% understandable	>80%

Table 12. Analysis of Reliability, Construct Validity, and Structural Relationships in the Technology Acceptance Model (TAM)

Domain	Variable / Path	Metric	Value	Interpretation
Measurement Model	PU	Mean ± SD	4.41 ± 0.46	High perceived usefulness
		Cronbach's α	0.89	Excellent reliability
		CR	0.91	High internal consistency
		AVE	0.71	Strong convergent validity
		Outer Loading Range	0.78–0.91	Acceptable indicator loading
	PEOU	Mean ± SD	4.32 ± 0.51	High ease of use
		Cronbach's α	0.87	Excellent reliability
		CR	0.90	High internal consistency
		AVE	0.68	Strong convergent validity
		Outer Loading Range	0.75–0.89	Acceptable indicator loading
	BI	Mean ± SD	4.47 ± 0.42	High adoption intention
		Cronbach's α	0.91	Excellent reliability
		CR	0.93	High internal consistency
		AVE	0.74	Strong convergent validity
		Outer Loading Range	0.81–0.93	Acceptable indicator loading
Structural Model	PEOU → PU	B	0.71	Strong positive effect
		t-value	6.82	Statistically significant
		p-value	<0.001	Significant relationship
		95% CI	0.52–0.83	Stable effect estimation
		f^2	0.49	Large effect
	PU → BI	B	0.78	Strong positive effect
		t-value	7.44	Statistically significant
		p-value	<0.001	Significant relationship

Domain	Variable / Path	Metric	Value	Interpretation
		95% CI	0.60–0.88	Stable effect estimation
		f^2	0.57	Large effect
	PEOU → BI	B	0.69	Moderate–strong effect
		t-value	5.91	Statistically significant
		p-value	<0.001	Significant relationship
		95% CI	0.47–0.81	Stable effect estimation
		f^2	0.41	large effect
Model Fit Evaluation	R^2 (PU)	Explained Variance	0.50	Moderate explanatory power
	R^2 (BI)	Explained Variance	0.67	Strong explanatory power
	SRMR	Model Fit Index	0.061	Good model fit
	Q^2 Predictive Relevance	Predictive capability	0.42	High predictive relevance

Table 13. Measurement Quality Assessment of the TAM Structural Model

Panel	Assessment Component	Relationship	Statistic	Threshold	Interpretation
A	HTMT	PU ↔ PEOU	0.82	<0.90	Acceptable discriminant validity
		PU ↔ BI	0.84	<0.90	Acceptable discriminant validity
		PEOU ↔ BI	0.79	<0.90	Acceptable discriminant validity
B	VIF	PU	1.78	<3.0	No multicollinearity detected
		PEOU	1.65	<3.0	No multicollinearity detected
		BI	1.91*	<3.0	No multicollinearity detected

*Full collinearity test

3.6 Field Effectiveness of NutriKEK-AI

The implementation of NutriKEK-AI among 160 pregnant women and 32 healthcare workers demonstrated improved screening performance compared to conventional approaches. Risk detection accuracy increased from 71.2% to 88.4% ($p < 0.001$), sensitivity from 0.69 to 0.84, specificity from 0.74 to 0.83, and referral accuracy from 68.5% to 86.7% ($p = 0.001$). The average decision-making time decreased from 14.8 ± 4.2 minutes to 6.3 ± 2.1 minutes per case. The recommendation relevance score increased from 3.1 ± 0.7 to 4.5 ± 0.4 ($p < 0.001$), while the user confidence score increased from 3.3 ± 0.8 to 4.6 ± 0.5 ($p < 0.001$). The compliance rate with system recommendations increased from 61.7% to 87.2%. The usability evaluation results showed a SUS score of 84.6 ± 5.8, with perceived usefulness at 4.58 ± 0.41, perceived ease of use at 4.51 ± 0.46, and behavioral intention at 4.63 ± 0.38 (see Table 14).

Table 14. Effectiveness of NutriKEK-AI During Field Implementation

Outcome Variable	Conventional Screening	NutriKEK-AI	Statistical Test	p-value	Effect Size (95% CI)
Detection Accuracy	71.2% (95% CI: 66.1–76.3)	88.4% (95% CI: 84.5–92.3)	Paired $t(119)=6.12$	<0.001	d = 1.21 (0.88–1.54)

Outcome Variable	Conventional Screening	NutriKEK-AI	Statistical Test	p-value	Effect Size (95% CI)
Sensitivity (High Risk)	0.69	0.84	McNemar χ^2=8.74	0.003	OR = 2.43 (1.31–4.56)
Specificity	0.74	0.83	McNemar χ^2=5.91	0.015	OR = 1.72 (1.08–2.94)
Referral Accuracy	68.5%	86.7%	McNemar χ^2=11.42	0.001	OR = 2.98 (1.56–5.68)
Decision Time (minutes)	14.8 ± 4.2	6.3 ± 2.1	Paired t(119)=9.27	<0.001	d = 1.84 (1.49–2.20)
Intervention Relevance Score	3.1 ± 0.7	4.5 ± 0.4	Paired t(119)=8.41	<0.001	d = 1.73 (1.38–2.07)
Decision Confidence	3.3 ± 0.8	4.6 ± 0.5	Paired t(17)=7.14	<0.001	d = 1.62 (1.03–2.20)
SUS Score	—	84.6 ± 5.8	—	—	—
TAM – Perceived Usefulness	—	4.58 ± 0.41	—	—	—
TAM – Ease of Use	—	4.51 ± 0.46	—	—	—
TAM – Behavioral Intention	—	4.63 ± 0.38	—	—	—
Recommendation Adherence	61.7%	87.2%	McNemar χ^2=10.36	0.001	OR = 3.14 (1.67–5.92)

3.7 Explainability and Model Transparency

SHAP analysis was used to explain the relative contributions of predictive features that were not directly used to define the KEK outcome. Anemia status, frequency of ANC visits, socioeconomic status, maternal age, and energy and protein intake all contributed to the prediction of KEK risk. LiLA/MUAC were not included as independent predictors in the SHAP interpretation because a LiLA value <23.5 cm was used as the criterion for defining the KEK outcome. Local SHAP analysis showed that anemia, inadequate ANC visits, socioeconomic status, maternal age, and nutritional intake factors contribute to the variation in the probability of KEK risk at the individual level. LiLA/MUAC is used as a clinical indicator for determining CED status and is not interpreted as an independent predictive contributor in the model. In addition, the system integrates 38 IF–THEN rules based on forward chaining. Expert validation showed a recommendation agreement rate of 0.89 with a Kappa value of 0.84 (see Table 17).

Table 15. Global SHAP Feature Importance for CED Prediction

Feature	Mean SHAP	Normalized (%)	Risk Contribution	Stability Across (SD)	Clinical Relevance	Final Rank
Anemia Status	0.24	26.7	Hb <11 g/dL ↑ CED risk	±0.015	Associated with impaired oxygen transport and maternal nutritional insufficiency	1

Feature	Mean SHAP	Normalized (%)	Risk Contribution	Stability Across (SD)	Clinical Relevance	Final Rank
ANC Frequency	0.19	21.1	<4 ANC visits ↑ risk	±0.011	Reflects limited prenatal monitoring and delayed nutritional intervention	2
Socioeconomic Status	0.17	18.9	Lower SES ↑ risk	±0.014	Indicates structural vulnerability affecting food access and maternal healthcare utilization	3
Maternal Age	0.12	13.3	<20 or ≥35 years ↑ risk	±0.009	Represents biologically vulnerable reproductive age groups	4
Asupan Kalori	0.10	11.1	Energy intake <80% of requirement ↑ predicted CED risk	±0.008	Reflects inadequate maternal energy intake and reduced nutritional reserve	5
Asupan Protein	0.08	8.8	Protein intake <80% of requirement ↑ predicted CED risk	±0.007	Reflects inadequate protein intake relevant to maternal tissue maintenance and fetal growth	6

Table 16. Explainability Robustness Statistics

Validation Component	Metric	Result	Interpretation
Ranking Stability	Spearman's ϱ across CV folds	0.91	High ranking consistency
Cross-Method Agreement	SHAP–Permutation agreement	0.89	Strong method agreement
Explainability Reproducibility	Explainability Consistency Index (ECI)	0.88	High predictor reproducibility
Feature Stability	Mean Top-3 Feature Stability	93.3%	Stable predictor selection
Bootstrap Robustness	SHAP Variability Drift	<5%	Minimal attribution variability

Table 17. Clinical Validation of Explainability and Rule-Based Transparency

Validation Domain	Indicator	n	Mean ± SD / Score	95% CI
Expert Agreement	Fleiss Kappa	7 experts	0.84	0.78–0.90
Clinical Interpretability	Interpretability Score	7 experts	4.52 ± 0.41	4.28–4.76*
Recommendation Actionability	Intervention Utility	32 health workers	4.47 ± 0.38	4.22–4.71
Decision Transparency	Trust & Transparency	32 health workers	4.61 ± 0.36	4.39–4.82
Rule Consistency	Rule–Expert Agreement	7 experts	0.89	0.83–0.93
SHAP Stability	Feature Consistency	Bootstrap 1000×	0.86 ± 0.04	0.81–0.90

* Expert ratings were assessed using a 5-point Likert scale.

3.8 System Validation and Findings from a Mixed-Methods Study

The results of the mixed-methods integration showed consistency between the quantitative and qualitative findings. The quantitative evaluation yielded a SUS score of 84.6 ± 5.8, a system recommendation compliance rate of 88.4%, and a 57.4% reduction in clinical decision-making time ($p < 0.001$). The qualitative analysis identified three main themes: risk screening efficiency, trust in explainable AI, and ease of system implementation (see Table 18).

Table 18. Integrated Mixed-Methods Validation of NutriKEK-AI

Domain	Quantitative Findings	Statistical Robustness	Qualitative Meta-Theme	Representative Evidence*
Usability & Acceptance	SUS = 84.6 ± 5.8; TAM usefulness = 4.58±0.41/5; ease of use = 4.51±0.46/5	Internal consistency of the 30-item evaluation instrument: Cronbach's α = 0.91; ICC = 0.87; SUS significantly above acceptability benchmark (t = 16.18; $p < 0.001$)	Operational simplicity under low-resource constraints	"The system was easy to operate even during busy ANC sessions." (HW-07)
Explainability & Trust	Clinician–AI agreement = 88.4%	Inter-rater agreement = 0.84; expert consistency stable across validation rounds	Explainability-driven clinical trust	"SHAP explanations helped us understand why patients were classified as high-risk." (HW-12)
Early Risk Detection	Recall = 0.84; sensitivity = 0.84; ROC-AUC = 0.91 (95% CI: 0.88–0.94)	Bootstrap validation (1,000 resamples); stable 10-fold CV performance	AI-assisted acceleration of maternal risk triage	"High-risk mothers could be identified earlier compared to manual screening." (HW-03)
Workflow Efficiency	Decision time reduced by 57.4% (95% CI: 45.2–69.7); paired t-test $p < 0.001$; Cohen's d = 1.84	Wilcoxon signed-rank confirmed robustness ($Z = -4.91$; $p < 0.001$)	Reduction of administrative burden	"Manual recording was reduced significantly during ANC services." (HW-09)
Contextual Relevance	Implementation satisfaction = 88.2%; rural usability consistency = 86.5%	No significant usability differences between districts ($p > 0.05$)	Adaptive implementation in underserved settings	"The dashboard matched our local maternal health conditions." (HW-15)
Qualitative Rigor	Thematic saturation achieved at interview 16	Inter-coder agreement = 0.89; Cohen's Kappa = 0.86	Consistent thematic stabilization	"Most participants expressed similar implementation experiences."

*Ethical clearence 42/E-5/UNILAM/2026; 1658/LL4/PG/2026

3.9 **Discussion**

One of the key findings of this study is that the integration of CRISP-DM and Agile Development results in a development process that maintains clinical relevance throughout the system development cycle. Anemia status, frequency of ANC visits, socioeconomic status, maternal age, and nutritional intake factors consistently emerged as factors contributing to the prediction of CED risk—from exploratory analysis and predictive modeling through SHAP-based explainability (Tables 5–6 and Table

15). This consistency indicates that the system not only optimizes predictive performance but also maintains clinically meaningful risk relationships throughout the AI lifecycle.

These findings align with the socio-technical systems perspective, which emphasizes that the success of health AI depends on alignment between technology, users, workflows, and the clinical context. This alignment is reflected in the model's high performance (AUC = 0.91), expert validity (S-CVI/Ave = 0.91), usability (SUS = 84.6), and the effectiveness of field implementation. Iterative stakeholder engagement during Agile sprints enabled continuous alignment between algorithm development and healthcare needs. These results also support the principles of Human-Centered AI, which emphasize the importance of user participation in enhancing system acceptance, trust, and utility in real-world practice (Rajkomar et al., 2022; Topol, 2019).

These findings expand the health AI literature, which has tended to be fragmented. Most studies use CRISP-DM for predictive model development or Agile for software development separately, thus often reporting only model performance without evaluating interpretability, integration into clinical workflows, user acceptance, or implementation readiness (Jiang et al., 2017; Shickel et al., 2021; Ramakrishnan et al., 2021). This situation is consistent with various systematic reviews that identify a lack of integration between model development, healthcare system readiness, and operational implementation as major barriers to the adoption of medical AI (Ambalavanan et al., 2025; Al-Ganad et al., 2026).

By integrating each phase of CRISP-DM into an iterative Agile cycle, this study successfully bridged that translation gap. The continuous process—ranging from the identification of maternal risk factors, feature engineering and data balancing, the development of explainable models, to validation and deployment—resulted in a system that is not only accurate but also implementable. The impact is evident in an increase in detection accuracy from 71.2% to 88.4%, a 57.4% reduction in decision-making time, and a nearly threefold increase in referral accuracy compared to conventional screening (Table 14). These findings demonstrate that a development approach that considers implementation from the outset can increase the likelihood of AI success in moving beyond the proof-of-concept stage toward sustained clinical use.

This study also demonstrates strong traceability across development stages. Significant predictors of KEK identified in the multivariate analysis—particularly inadequate ANC (OR = 2.12), anemia (OR = 2.05), and low socioeconomic status (OR = 2.63)—were re-prioritized by the SHAP analysis and validated by a multidisciplinary expert panel. This traceability strengthens the interpretive validity of the system because the model's recommendations can be directly traced back to risk factors already recognized in clinical practice. These findings address the criticism that health AI often performs well but is difficult for clinicians to understand and trust (Babic et al., 2021; Rudin, 2019). Therefore, integrating explainability from the modeling stage has the potential to reduce the interpretation gap while increasing user trust in AI recommendations.

This study shows that the prevalence of KEK among pregnant women reaches 18.7%, confirming that maternal malnutrition remains a significant public health issue in rural Indonesia. Multivariate analysis identified low socioeconomic status (OR = 2.63), inadequate antenatal care (ANC) visits (OR = 2.12), anemia (OR = 2.05), low educational attainment (OR = 1.74), and high-risk maternal age (<20 or ≥35 years) (OR=1.59) as significant determinants of KEK. These findings suggest that the risk of maternal malnutrition results from the interaction of biological factors, access to health services, education, and socioeconomic conditions, rather than solely from individual physiological conditions.

A significant interaction between anemia and inadequate ANC visits (OR=3.21; p=0.002) indicates that the risk of KEK is synergistic. Limited access to ANC reduces opportunities for anemia detection, nutritional counseling, micronutrient supplementation, and early intervention, thereby amplifying the impact of physiological deficits and accelerating the decline in nutritional status during pregnancy.

The dominant role of anemia and ANC utilization is consistent with evidence from low- and middle-income countries showing that micronutrient deficiencies, poor diet quality, and limited access to maternal health services contribute to maternal malnutrition and poor pregnancy outcomes (Keats et al., 2021; Wilkins et al., 2021; Dassie et al., 2026; Sarikhani et al., 2024). These findings are also consistent with research in Indonesia showing that food insecurity and barriers to accessing health care disproportionately affect pregnant women in rural areas (Azrimaidaliza et al., 2025), as well as regional evidence reporting that poor maternal dietary quality is a key factor in various adverse maternal and neonatal outcomes (Agustina et al., 2023).

Compared to previous studies, which generally used conventional regression and treated risk factors independently, this study demonstrates a nonlinear relationship through the interaction between anemia and the frequency of ANC visits. These results

indicate that maternal nutritional vulnerability is shaped by mechanisms more complex than those explained by traditional epidemiological approaches, thereby providing a more realistic picture of risk formation in resource-limited settings.

SHAP consistently identified anemia status, frequency of ANC visits, socioeconomic status, maternal age, and dietary intake factors as features contributing to the prediction of stunting risk. This interpretation was made by excluding LiLA/MUAC as independent predictors because LiLA <23.5 cm was used to define stunting status. Thus, SHAP's stability reflects the consistency of risk factor contributions that lie outside the direct definition of the outcome. The consistency between conventional epidemiological results and explainable machine learning reinforces the validity of the identified risk factors and demonstrates that both approaches reveal the same risk structure. These findings align with studies emphasizing the importance of interpretable predictive models for understanding determinants of maternal health without compromising clinical transparency (Bosschieter et al., 2024; Patel, 2023). From a public health perspective, these results indicate that preventing KEK cannot be achieved through micronutrient supplementation alone (Keats et al., 2021), but requires an integrated approach that includes increasing ANC coverage, controlling anemia, providing maternal health education, and strengthening socioeconomic support. Such a multisectoral approach is particularly crucial in rural areas that simultaneously face limited health services, food insecurity, and low levels of education (Azrimaidaliza et al., 2025; Okoye et al., 2026).

The results of the study show that XGBoost provides the best predictive performance compared to Logistic Regression, Decision Tree, SVM, and Random Forest, with an AUC of 0.91 (95% CI: 0.88–0.94), an accuracy of 0.88, an F1-score of 0.85, and a Brier score of 0.12. These advantages are significant compared to Logistic Regression (ΔAUC = +0.09; $p = 0.003$), Random Forest (ΔAUC = +0.02; $p = 0.035$), and SVM (ΔAUC = +0.05; $p = 0.047$), and it demonstrated high stability in 10-fold cross-validation (AUC 0.90 $\pm$ 0.02; accuracy 0.87 $\pm$ 0.02). These findings indicate that the risk of Chronic Energy Deficiency (CED) in pregnant women follows a complex and nonlinear pattern. XGBoost's superior performance stems from the gradient boosting mechanism, which is capable of capturing nonlinear relationships, threshold effects, and interactions between variables that are difficult to identify using conventional parametric models. According to Chen and Guestrin [5], XGBoost enhances generalization ability through regularization, objective function optimization, and adaptive exploration of feature interactions. In the context of maternal health, the relationships between nutritional status, anemia, ANC visits, education, and socioeconomic conditions are multidimensional, making them better suited for modeling using a boosting approach.

This interpretation is supported by the finding of a significant interaction between anemia and inadequate ANC visits (OR = 3.21; $p = 0.002$). This synergistic effect indicates that the risk of KEK increases more sharply when both factors are present simultaneously, reflecting the nonlinear patterns that are a key strength of XGBoost. Thus, the improved model performance reflects not only better classification ability but also the capacity to represent interdependent maternal risk mechanisms. These findings are consistent with the literature indicating that boosting algorithms are among the most effective approaches for clinical prediction based on tabular data. Shwartz-Ziv and Armon [4] reported that gradient boosting models often outperform deep learning approaches on structured clinical data. Similar results were also reported by Rajkomar et al. [6], Rajpurkar et al. [7], Altaie and Ibrahim [16], and the meta-analysis by Waldock et al. [23], which demonstrated the superiority of ensemble models in various Clinical Decision Support System (CDSS) applications. Compared to previous studies, which generally reported AUC values of 0.75–0.88 for maternal risk prediction using logistic regression, random forest, or conventional classification methods [21,24,28], the AUC of 0.91 achieved in this study indicates higher discriminatory power. This improvement is likely due to the use of feature engineering based on validated clinical determinants, class balancing using SMOTE, and iterative hyperparameter optimization through the CRISP-DM–Agile framework.

High predictive performance enables earlier identification of at-risk pregnant women, allowing for faster implementation of nutritional interventions, micronutrient supplementation, counseling, and referrals. These findings align with evidence that early detection and nutritional interventions during pregnancy contribute to improved maternal and neonatal outcomes [1,40,42,44]. This is reflected in field implementation, which showed an increase in detection accuracy from 71.2% to 88.4% and in referral accuracy from 68.5% to 86.7%, confirming that improvements in algorithm performance yield tangible clinical benefits. However, high accuracy alone is not sufficient to ensure the adoption of AI in clinical practice. Babic et al. [3], Rudin [8], and Brankovic et al. [9] emphasize the importance of transparency in AI decision-making. Therefore, this study integrates XGBoost with SHAP-based XAI so that each prediction can be traced back to its underlying clinical risk factors. This approach enhances transparency, reduces the risk of black-box decision-making, and strengthens user trust in the system [18,25,29].

SHAP-based explainability analysis consistently identified anemia status, frequency of ANC visits, socioeconomic status, maternal age, and nutritional intake factors as features contributing to the prediction of KEK risk. The dominance of these four variables demonstrated high stability across various validation procedures, as reflected by a Spearman's ϱ value of 0.91, a SHAP–permutation agreement of 0.89, an Explainability Consistency Index of 0.88, and a stability of 93.3% for the three main predictors. These findings indicate that the model's explanatory mechanisms are consistent, reproducible, and represent genuine patterns of maternal risk, rather than artifacts specific to a particular sample. These results are consistent with evidence that MUAC is a key indicator of maternal energy reserves associated with the risk of fetal growth restriction and low birth weight [30]. Anemia reflects impaired nutritional status and oxygen transport capacity, which can increase the risk of maternal and fetal complications [1]. At the same time, low ANC attendance rates and socioeconomic vulnerability contribute to delayed detection of nutritional problems, limited access to health care, and poor access to nutritious food [33], [37]. The consistency of these factors within the model indicates that the predictive patterns identified by the algorithm align with the biological and social mechanisms underlying CED.

These findings support the XAI paradigm, which positions transparency as an essential component in the development of high-risk AI in the healthcare sector. Lundberg and Lee [18] explain that SHAP produces consistent feature attributions and is grounded in a strong game-theoretic foundation, while Markus et al. [25] emphasize that trust in AI is significantly influenced by users' ability to understand the reasoning behind the system's recommendations. In line with the Trust in Automation theory, transparency enables AI decisions to be understood, verified, and linked to users' existing clinical knowledge, thereby fostering rational trust in the system.

The reliability of the explainability mechanism is further reinforced by high expert agreement (Cohen's $\varkappa = 0.84$), an interpretability score of 4.52/5, and a 0.89 alignment between the system's rules and expert assessments. These results indicate that the explanations generated by NutriKEK-AI are consistent with the clinical logic used in maternal care practice. These findings are important given that one of the main criticisms of health AI is its "black box" nature, which is difficult to verify and has the potential to undermine the accountability of clinical decisions [3], [8]. As noted by Babic et al. [3] and Rudin [8], AI systems for high-risk decisions must generate valid explanations that users can understand. Therefore, the high degree of alignment between the model's explanations and clinical reasoning indicates that the explainability of NutriKEK-AI is not only technically valid but also clinically meaningful.

The results of this study align with previous studies showing that explainable AI increases the utilization of clinical decision support systems. Du et al. [21] reported that interpretation mechanisms improve clinicians' understanding of risk factors in gestational diabetes predictions, while Innab et al. [28] demonstrated that SHAP enables more transparent identification of maternal and fetal risk factors compared to conventional machine learning approaches. A systematic review by Badisy et al. [17] also confirms that explainable systems have a higher likelihood of implementation compared to models that focus solely on predictive accuracy. However, most studies still treat explainability as a supplementary analysis rather than an integral part of the clinical decision-making process.

Unlike previous research, NutriKEK-AI integrates SHAP visualization with expert rules based on forward chaining, allowing users to trace the relationships between risk factors, the inference process, and the recommendations generated by the system. This hybrid approach yields an explanation fidelity of 0.91 and a hybrid confidence score of 0.87, indicating a high degree of alignment between the explanations and the model's behavior while remaining easily understandable to non-technical users. These findings address the need for modern decision support systems capable of bridging the gap between algorithmic complexity and the practical needs of healthcare professionals [11], [12]. These findings are particularly relevant for AI implementation in primary healthcare facilities in resource-limited settings. Low trust in unexplained recommendations is one of the main barriers to the adoption of digital health technologies [32], [38]. In this study, the high scores for transparency and user trust (4.61/5), the 89.7% understanding rate of recommendations, and the 88.4% alignment between AI and clinicians indicate that explainability plays a crucial role in enhancing technology acceptance while strengthening clinical accountability. Thus, NutriKEK-AI functions as a decision support system that reinforces evidence-based decision-making rather than replacing the role of healthcare professionals, in line with the principle of human-centered AI in healthcare [6], [26].

NutriKEK-AI demonstrated a very high level of usability, as evidenced by a SUS score of 84.6 ± 6.8, a task completion rate of 93.8%, an average completion time of 2.7 minutes, and a low critical error rate (4.2%). This SUS score far exceeds the acceptance threshold (68), indicating that healthcare workers can operate the system efficiently even in primary care facilities with limited resources. The high task completion rate also suggests that most system functions can be used without complex technical

training, thereby reducing implementation barriers commonly encountered with digital health technologies in developing countries. These findings are consistent with the Technology Acceptance Model (TAM), which identifies perceived usefulness and perceived ease of use as the primary determinants of technology adoption. Structural analysis revealed that ease of use significantly influences perceived usefulness ($\beta = 0.71$; $p < 0.001$), while perceived usefulness is the strongest predictor of usage intention ($\beta = 0.78$; $p < 0.001$). The direct effect of ease of use on usage intention was also significant ($\beta = 0.69$; $p < 0.001$), with high model explanatory power ($R^2 = 0.67$). These results indicate that acceptance of NutriKEK-AI is determined not only by its predictive capabilities but also by its ability to accelerate clinical decision-making without increasing the user's workload.

The high usability is also reflected in high user satisfaction (4.45/5), low perceived cognitive load (1.92/5), and a clinical acceptance rate of 92.4%. In antenatal care—which is characterized by time constraints and complex documentation—a simple interface, intuitive navigation, and fast response times (1.8 seconds) enable healthcare providers to obtain clinical recommendations without disrupting existing workflows. This underscores that usability is a strategic factor that directly influences the success of health technology implementation. These findings align with those of Harahap et al. [32] and Sarikhani et al. [33], who identified system complexity, administrative burden, and low usability as the main barriers to the adoption of maternal health technologies. Systematic reviews by Ambalavanan et al. [15] and Al-Ganad et al. [20] also indicate that many digital health innovations fail to achieve sustainable implementation due to insufficient consideration of user needs, integration into clinical workflows, and infrastructure limitations. In contrast, NutriKEK-AI demonstrated a higher level of usability compared to various previous maternal CDSSs. Du et al. [21] reported that interpretability complexity still affects the adoption of explainable machine learning-based gestational diabetes prediction systems, while Badisy et al. [17] emphasized that many high-risk pregnancy CDSSs have good algorithmic performance but are not yet optimal in terms of usability and clinical integration. In this study, the combination of a simple interface, SHAP-based risk explanations, and direct integration with the ANC workflow appears to have successfully bridged the gap between technical performance and user acceptance. These findings reinforce the view that the successful implementation of health AI depends not only on model accuracy but also on its alignment with the operational context of healthcare services. Rajkomar et al. [6], Rajpurkar et al. [7], and Topol [26] emphasize that the primary value of AI lies in its ability to enhance the capacity of healthcare workers, not to replace human clinical roles. Thus, the high usability of NutriKEK-AI demonstrates the system's potential to support sustainable clinical decision-making, particularly in areas with limited healthcare professionals and resources.

The system's ability to operate on unstable networks, its compatibility with low-spec Android devices, and its offline cache support make it particularly relevant for rural areas. These characteristics directly address digital infrastructure challenges that have long been identified as barriers to digital health transformation in developing countries [15], [20], [32]. This study offers two key innovations. From the perspective of technological innovation, NutriKEK-AI was developed as a lightweight, responsive XGBoost-based maternal risk prediction platform that remains functional under limited connectivity conditions—unlike many previous studies that focused on improving algorithm performance without considering implementation readiness [16], [17], [23]. From the perspective of contextual innovation, the system was designed to align with the ANC workflow at Indonesian primary healthcare facilities, resulting in a high degree of alignment with user needs. The absence of regional differences in usability ($p > 0.05$) demonstrates the system's adaptability to variations in local contexts. Furthermore, the integration of SHAP-based explainable AI enhances the transparency of recommendations and user trust, in line with the principles of developing trustworthy AI in healthcare [11], [25], [29].

The implementation of NutriKEK-AI in antenatal care significantly improved clinical performance compared to conventional screening, as demonstrated by increased detection accuracy (71.2% to 88.4%; $p<0.001$), sensitivity (0.69 to 0.84), specificity (0.74 to 0.83), referral accuracy (68.5% to 86.7%; $p=0.001$), and compliance with intervention recommendations (61.7% to 87.2%). In addition, decision-making time was reduced by 57.4% (from 14.8 to 6.3 minutes; $d=1.84$), indicating that the system not only improves the accuracy of risk identification but also enhances the efficiency of clinical services. These improvements are likely due to the system's ability to integrate relevant maternal risk factors—particularly anemia status, ANC frequency, socioeconomic status, maternal age, and nutritional intake factors—into a single, comprehensive, and actionable risk estimate. This automation reduces the burden of manual interpretation, lowers the cognitive complexity for healthcare providers, and enhances the consistency of data-driven decisions. These findings align with CDSS theory, which emphasizes that decision support systems improve decision quality by providing relevant and timely information [12]. Rajkomar et al. [6] and Topol [26] also assert that the primary value of AI in healthcare lies in enhancing healthcare providers' cognitive capacity through human–AI collaboration, rather than merely automating decisions. Unlike most studies on AI in maternal health, which focus on algorithm performance without evaluating the impact of clinical implementation [16], [21], [28], this study demonstrates that

predictive advantages can translate into tangible operational benefits, including improved referral accuracy, service efficiency, and clinical confidence. These findings address the gaps reported in various systematic reviews regarding the limited evidence of AI translation from the development stage to routine use in maternal care, particularly in low- and middle-income countries [17], [20], [23].

The practical implications of these findings are highly relevant for maternal health services in resource-limited settings that continue to face shortages of healthcare personnel, high administrative burdens, and delays in referrals [32], [33], [35]. With a detection sensitivity of 84% and an improvement in referral accuracy of nearly 18 percentage points, NutriKEK-AI has the potential to accelerate risk screening, nutritional interventions, and pregnancy monitoring. This is important because maternal nutritional status is a key determinant of various maternal and child health outcomes, including low birth weight, stunting, obstetric complications, and the risk of long-term diseases [1], [39], [40], [41]. Increased adherence to clinical recommendations demonstrates the contribution of explainable AI in building user trust.

Transparency of risk factors through SHAP enables healthcare providers to understand the basis for the system's classifications and recommendations, thereby increasing acceptance of and adherence to the generated decisions. These findings align with studies confirming that transparency is a critical prerequisite for the adoption of clinical AI [11], [25], [29].

This research presents translational innovation by transforming AI from a mere predictive model into a decision support system that has a tangible impact on clinical workflows, as well as public health innovation through the development of an early warning system for KEK risk designed for primary health care in resource-constrained settings. The integration of explainable machine learning, clinical rule-based reasoning, and a user-centered approach results in an AI-CDSS ecosystem that is not only accurate but also clinically valid, user-friendly, and effective in field implementation [11], [17], [20], [26].

The success of NutriKEK-AI's implementation is determined not only by its high predictive performance but also by its user-oriented design, integration with clinical practice, and alignment with the operational conditions of maternal health services. These findings underscore that usability and user acceptance are factors just as important as algorithmic accuracy in supporting sustainable digital health transformation. The main contribution of this study is the development of an explainable AI-CDSS architecture that integrates XGBoost, SHAP, and clinical rules into a single system that is transparent, auditable, and validated. Unlike most previous studies that focused on predictive accuracy [17], [21], [36], this study combines machine learning, explainable AI, multidisciplinary validation, usability testing, and field implementation evaluation within an integrated CRISP-DM–Agile framework. This approach ensures traceability between clinical determinants, the modeling process, prediction interpretation, and implementation impact, enabling NutriKEK-AI to function as a reliable decision support system ready for deployment in primary health care settings in resource-limited regions.

This study also makes three major scientific contributions. First, it provides a more comprehensive understanding of maternal nutritional vulnerability through the integration of clinical and social determinants within a single analytical framework. Second, it identifies significant interactions between anemia and ANC visits as a nonlinear risk pathway contributing to KEK, which remains underreported in the literature. Third, it demonstrates the convergence of epidemiology and explainable AI as a methodological innovation that not only improves predictive accuracy but also deepens our understanding of the risk mechanisms underlying KEK. These findings confirm that maternal malnutrition is a multidimensional phenomenon shaped by the interaction of biological factors, health services, and social determinants. The novelty of this study lies in the development of an integrated framework that links the identification of clinical risk factors, explainable machine learning, expert validation, usability evaluation, and field implementation into a single, integrated, and traceable workflow.

Technologically, this study introduces an X-CDSS development strategy that combines CRISP-DM, Agile Development, XGBoost, and SHAP. In terms of implementation, this approach provides end-to-end traceability between maternal health determinants, AI inferences, clinical interpretations, and implementation outcomes. Thus, this study broadens the focus of health AI development from mere predictive accuracy toward interpretability, clinical action, and readiness for real-world implementation, in line with recent literature recommendations that health AI systems must be accurate, explainable, actionable, and suitable for application in routine practice [11], [17], [20], [23].

4. CONCLUSION

Chronic Energy Deficiency (CED) in pregnant women remains a major public health challenge in resource-limited settings. At the same time, conventional screening methods often face limitations in accuracy, consistency, and decision-making speed. To

address this gap, this study successfully developed and validated NutriKEK-AI, an Explainable Hybrid Clinical Decision Support System that integrates an XGBoost model with expert clinical knowledge to support the early detection of CED risk in pregnant women. The results show that NutriKEK-AI exhibits excellent predictive performance (AUC = 0.91; accuracy = 88%), with factors contributing to the prediction of CED risk including anemia status, frequency of ANC visits, socioeconomic status, maternal age, and nutritional intake factors. LiLA/MUAC continued to be used as anthropometric indicators to determine CED status according to the study's operational criteria, rather than as independent predictors in the model. The integration of a SHAP-based explainable AI mechanism resulted in a high level of interpretability and transparency, supported by robust explanation stability, alignment with expert assessments, and high user confidence. In field implementation, the system was shown to improve risk detection accuracy (88.4% vs. 71.2%), referral accuracy (86.7% vs. 68.5%), adherence to recommendations (87.2% vs. 61.7%), and significantly accelerate the clinical decision-making process (6.3 vs. 14.8 minutes; $p < 0.001$).

These findings indicate that NutriKEK-AI functions not only as an accurate predictive model but also as a service-transformation tool capable of improving efficiency, decision quality, and readiness for maternal nutrition interventions in primary health care facilities. Nevertheless, this study is still limited to two districts and employed a short-term implementation design; therefore, generalizing the results to a broader population should be done with caution. Future research should include nationwide multicenter validation, the integration of longitudinal data and electronic health records, and an evaluation of the long-term clinical impact on pregnancy outcomes, stunting, and neonatal health. Overall, this study provides evidence that the application of contextually developed, implementation-oriented explainable AI can bridge the gap between artificial intelligence innovations and real-world health practices, while opening new opportunities for the development of adaptive, reliable, and sustainable maternal nutrition early warning systems in developing countries.

ACKNOWLEDGEMENTS

This research was supported by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia through the Directorate of Research, Technology, and Community Service (DRTPM) under the 2026 Fundamental Research Grant program. The authors extend their sincere gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, all pregnant women, healthcare workers, expert validators, and primary healthcare facilities that participated in Lebak and Pandeglang Regencies, Banten Province, Indonesia, for their invaluable involvement and support in this study. The authors also thank the multidisciplinary panel of experts in the fields of maternal health, nutrition, artificial intelligence, and health informatics who made significant contributions to the development, design of explainability, clinical validation, and implementation of the NutriKEK-AI system. Appreciation is also extended to all partner institutions, field enumerators, and research assistants who supported the data acquisition process, system testing, and research implementation. The main research contract was awarded on January 30, 2026, under Contract Number 283/C3/DT.05.00/PL-BARU/2026, while the subsidiary contracts were signed on April 21, 2026, and May 7, 2026, with Contract Numbers 1516/LL4/PG/2026 and 002/STKIP-BN/V/2026.

REFERENCES

[1] R. F. Voss, J. Clarke. Algorithmic Musical Composition, Silver Burdett Press, Londyn, 1986.

[2] W. Zabierowski, A. Napieralski. Chords classification in tonal music, Journal of Environment Studies, Vol.10, No.5, 50-53.

[3] Keats, E. C., Oh, C., Chau, T., Khalifa, D. S., Imdad, A., & Bhutta, Z. A. (2021). Effects of vitamin and mineral supplementation during pregnancy on maternal, birth, child health and development outcomes in low- and middle-income countries: A systematic review. Campbell Systematic Reviews: Better Evidence for a Better World, 17(2). https://doi.org/10.1002/cl2.1127

[4] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

[5] Babic, B., Gerke, S., Evgeniou, T., & Cohen, I. G. (2021). Beware explanations from AI in health care. Science, 373(6552), 284–286. https://doi.org/10.1126/science.abg1834

[6] Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. Information Fusion, 81, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

[7] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785
[8] Rajkomar, A., Dean, J., & Kohane, I. (2022). Machine learning in medicine. New England Journal of Medicine, 386(14), 1347–1358. https://doi.org/ 10.1056/NEJMra1814259
[9] Rajpurkar, P., Chen, E., Banerjee, O. et al. AI in health and medicine. Nat Med 28, 31–38 (2022). https://doi.org/10.1038/s41591-021-01614-0
[10] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1, 206–215 (2019). https://doi.org/10.1038/s42256-019-0048-x
[11] Brankovic A, Cook D, Rahman J, Khanna S, Huang W. Benchmarking the most popular XAI used for explaining clinical predictive models: Untrustworthy but could be useful. Health Informatics Journal. 2024;30(4). https://doi.org/10.1177/14604582241304730
[12] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2021). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. IEEE Journal of Biomedical and Health Informatics, 25(4), 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063
[13] Liu, S. et al. (2024). Leveraging explainable AI to optimize clinical decision support. JAMIA.
[14] https://doi.org/10.1093/jamia/ocae019
[15] Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of AI. JAMA. https://doi.org/10.1001/jama.2018.17163
[16] Jiang, F. et al. (2017). Artificial intelligence in healthcare. Stroke and Vascular Neurology. https://doi.org/10.1136/svn-2017-000101
[17] Rane, N. et al. (2023). Explainable AI in healthcare CDSS. SSRN. https://doi.org/10.29284/zdfexk19
[18] Ambalavanan, R., Snead, R. S., Marczika, J., Towett, G., Malioukis, A., & Mbogori-Kairichi, M. (2025). Challenges and strategies in building a foundational digital health data integration ecosystem: A systematic review and thematic synthesis. Frontiers in Health Services, 5, 1600689. https://doi.org/10.3389/frhs.2025.1600689
[19] Saleem Altaie HK, Ibrahim AA. Development of an advanced maternal health risk detection system using enhanced XGBOOST and blending models. DIGITAL HEALTH. 2026;12. https://doi.org/10.1177/2055207626145078 6
[20] Badisy, I.E., Assarag, B. & Belrhiti, Z. Interpretable clinical decision support systems in high-risk pregnancy: a scoping review of models, methods, and implementation. BMC Pregnancy Childbirth 26, 125 (2026). https://doi.org/10.1186/s12884-025-08614-9
[21] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv. https://doi.org/10.48550/arXiv.1705.07874
[22] Al-Ganad, A., Al-Shahdhi, A., Al-Dhaifi, O., Hajeb, E., Hajeb, H., & Al-Motarreb, A. (2026). Deploying medical AI in low-resource settings: A scoping review of challenges and strategies. Frontiers in Digital Health, 8, 1743634. https://doi.org/10.3389/fdgth.2026.1743634
[23] Du, Y., Rafferty, A. R., McAuliffe, F. M., Wei, L., & Mooney, C. (2022). An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. Scientific Reports, 12, 1170. https://doi.org/10.1038/s41598-022-05112-2
[24] Mustafa, M., Hameed, I., Shahnawaz, A., & Mateen, B. A. (2026). How GenAI is Helping Reimagine Antenatal Care in A Low-Resource Setting: From Provider Enablement to Patient Empowerment. arXiv. https://arxiv.org/abs/2604.22610
[25] Waldock, W. J., Guni, A., Darzi, A., & Ashrafian, H. (2026). Performance of predictive AI-based clinical decision support systems across clinical domains: A systematic review and meta-analysis. PLOS Digital Health, 5(3), e0001310. https://doi.org/10.1371/journal.pdig.0001310
[26] Ramakrishnan R, Rao S, He J-R. Perinatal health predictors using artificial intelligence: A review. Women's Health. 2021;17. https://doi.org/10.1177/17455065211046132
[27] Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of Biomedical Informatics, 113, 103655. https://doi.org/10.1016/j.jbi.2020.103655
[28] Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25, 44–56 (2019). https://doi.org/10.1038/s41591-018-0300-7

[29] Gaube, S., Suresh, H., Raue, M. et al. Do as AI say: susceptibility in deployment of clinical decision-aids. npj Digit. Med. 4, 31 (2021). https://doi.org/10.1038/s41746-021-00385-9

[30] Innab, N., Alsubai, S., Alabdulqader, E. A., Alarfaj, A. A., Umer, M., Trelova, S., & Ashraf, I. (2024). Automated approach for fetal and maternal health management using light gradient boosting model with SHAP explainable AI. Frontiers in Public Health, 12, 1462693. https://doi.org/10.3389/fpubh.2024.1462693

[31] Amann, J., Blasimme, A., Vayena, E. et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 20, 310 (2020). https://doi.org/10.1186/s12911-020-01332-6

[32] Nisak, S. K., & Nadhiroh, S. R. (2024). Relationship of Mid-Upper Arm Circumference Measure (MUAC) in Pregnant Women to the Incidence of Low Birth Weight (LBW): Systematic Review. Media Gizi Kesmas, 13(1), 512–520. https://doi.org/10.20473/mgk.v13i1.2024.512-520

[33] Girchenko P, Lahti-Pulkkinen M, Hämäläinen E, Laivuori H, Villa PM, Kajantie E, Räikkönen K. Associations of polymetabolic risk of high maternal pre-pregnancy body mass index with pregnancy complications, birth outcomes, and early childhood neurodevelopment: findings from two pregnancy cohorts. BMC Pregnancy Childbirth. 2024 Jan 24;24(1):78. https://doi.org/10.1186/s12884-024-06274-9

[34] Harahap, N. C., Handayani, P. W., & Hidayanto, A. N. (2021). Barriers in health information systems and technologies to support maternal and neonatal referrals at primary health centers. Healthcare Informatics Research, 27(2), 153–161. https://doi.org/10.4258/hir.2021.27.2.153

[35] Sarikhani, Y., Najibi, S.M. & Razavi, Z. Key barriers to the provision and utilization of maternal health services in low- and lower-middle-income countries; a scoping review. BMC Women's Health 24, 325 (2024). https://doi.org/10.1186/s12905-024-03177-x

[36] Bosschieter, T.M., Xu, Z., Lan, H. et al. Interpretable Predictive Models to Understand Risk Factors for Maternal and Fetal Outcomes. J Healthc Inform Res 8, 65–87 (2024). https://doi.org/10.1007/s41666-023-00151-4

[37] Okoye, C.U., Orede, V.O., Oladeji, O.E. et al. Policy implementation barriers and the effectiveness of maternal and child nutrition programs in low-resource settings. Discov Public Health 23, 61 (2026). https://doi.org/10.1186/s12982-026-01353-y

[38] Patel, S. S. (2023). Explainable machine learning models to analyse maternal health. Data & Knowledge Engineering, 146, 102198. https://doi.org/10.1016/j.datak.2023.102198

[39] Azrimaidaliza, A., Resmiati, R., Islami, G. F., Fasrini, U. U., & Apriliani, A. D. (2025). Is food insecurity higher among pregnant women in rural areas of Indonesia? Gulhane Medical Journal, 67(2), 108–115. https://doi.org/10.4274/gulhane.galenos.2025.24892

[40] Priambodo, R., Handayani, P. W., Aji, R. F., & Dimyati, K. (2026). Challenges for a Maternal-Care Health Recommender System in Indonesia: Formative Preimplementation Qualitative Study. JMIR Formative Research, 10, e73726. https://doi.org/10.2196/73726

[41] Agustina, R., Rianda, D., Lasepa, W., Birahmatika, F. S., Stajic, V., & Mufida, R. (2023). Nutrient intakes of pregnant and lactating women in Indonesia and Malaysia: Systematic review and meta-analysis. Frontiers in Nutrition, 10, 1030343. https://doi.org/10.3389/fnut.2023.1030343

[42] Wilkins, E., Wickramasinghe, K., Pullar, J. et al. Maternal nutrition and its intergenerational links to non-communicable disease metabolic risk factors: a systematic review and narrative synthesis. J Health Popul Nutr 40, 20 (2021). https://doi.org/10.1186/s41043-021-00241-2

[43] Vaivada, T., Akseer, N., Akseer, S., Somaskandan, A., Stefopulos, M., & Bhutta, Z. A. (2020). Stunting in childhood: An overview of global burden, trends, determinants, and drivers of decline. The American Journal of Clinical Nutrition, 112(Suppl 2), 777S–791S. https://doi.org/10.1093/ajcn/nqaa159

[44] Sokou, R., Lianou, A., Lampridou, M., Panagiotounakou, P., Kafalidis, G., Paliatsiou, S., Volaki, P., Tsantes, A. G., Boutsikou, T., Iliodromiti, Z., & Iacovidou, N. (2025). Neonates at Risk: Understanding the Impact of High-Risk Pregnancies on Neonatal Health. Medicina, 61(6), 1077. https://doi.org/10.3390/medicina61061077

[45] Kruk, M. E., Gage, A. D., Arsenault, C., Jordan, K., Leslie, H. H., Roder-DeWan, S., et al. (2018). High-quality health systems in the Sustainable Development Goals era: time for a revolution. The Lancet Global Health, 6(11), e1196–e1252. https://doi.org/10.1016/S2214-109X(18)30386-3

[46] Dassie, G.A., Lemi, H., Getachew, T. et al. Maternal diet quality during pregnancy and birth outcomes in low- and middle-income countries: a scoping review. BMC Nutr 12, 25 (2026). https://doi.org/10.1186/s40795-025-01229-5.