NOVA MEDICA INFORMATICA

Volume 01, Issue 2, 2026, pp. 15 of 27

Article

Machine Learning-Driven Multi-Class Prediction of Vitamin Deficiency Diseases: A Clinically Interpretable Approach Integrating Socioeconomic Determinants and Biomarker Profiles

Louise Kartika Indah[1*], Danny E.J. Luhulima[2], Vanessa Mutiara[3]

Received: 25-06-2026

Revised: 29-06-2026

Accepted: 30-07-2026

Published: 31-07-2026

Citation: L. K. Indah, D. E. J. Luhulima, and V. Mutiara, "Machine Learning-Driven Multi-Class Prediction of Vitamin Deficiency Diseases: A Clinically Interpretable Approach Integrating Socioeconomic Determinants and Biomarker Profiles," Nova Medica Informatica, vol. 1, no. 2, pp. 15-27, Jul. 2026.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Correspondence (*):
louise.indah@uki.ac.id

[1] Department of Medical Community, Faculty of Medicine, Universitas Kristen Indonesia (Christian University of Indonesia), Jakarta, Indonesia; louise.indah@uki.ac.id

[2] Department of Clinical Pathology, Faculty of Medicine, Universitas Kristen Indonesia (Christian University of Indonesia), Jakarta, Indonesia; danny.luhulima@uki.ac.id

[3] Department of Clinical Pathology, Faculty of Medicine, Universitas Kristen Indonesia (Christian University of Indonesia), Jakarta, Indonesia; vanessa.mutiara@uki.ac.id

Highlights

Multi-class proof-of-concept machine learning system was analyzed using parameterized cohort (N = 4,000) through five different vitamin deficiencies.

What are the main findings?

- Sequential Gradient Boosting gave best multi-class discrimination result (accuracy: 94.6%, macro F1: 91.5%, MCC: 0.926) with stratified cross-validation controls.
- Serum vitamin D, night blindness, and serum B12 were most important global features that showed internal validity in line with physiological pathways.
- Probabilistic risk ranking maintained good discrimination results in case of rare Scurvy cohort (ROC-AUC: 0.9737).

What are the implications of the main findings?

- The application of machine learning-based screening using few biological markers and pathognomonic symptoms constitutes a scalable decision support system suitable for capacity-challenged primary health-care institutions.
- Prospective registration-based validation and adaptive threshold calibration are necessary prerequisites before clinical implementation.
- Biological markers and pathognomonic symptoms have the main predictive value, while diet-related and socio-economic parameters provide additional context.

Abstract

Deficiency disease conditions due to micronutrients represent serious public health issues in low-resource settings wherein central laboratory facilities are not accessible. Even though earlier studies using digital health approaches have concentrated on detection of single diseases in a binary fashion, multi-class prediction involving multiple domains like clinical signs and symptoms, biomarkers and socio-economic factors is relatively unexplored. A proof-of-concept dataset using parameterization (N = 4,000) with 33 parameters from demographic, lifestyle, nutritional, biological markers and pathognomonic sign and symptom domains was used for evaluation. Random Forest, Gradient Boosting and SVM classifiers were trained using stratified splitting in 80/20 ratios and validated by 5-fold cross-validation. Gradient Boosting gave the best multi-class discrimination, with 94.6% global accuracy, a macro-averaged F1-score of 91.5%, and a Matthews Correlation Coefficient of 0.926. Class-specific F1-scores were 0.989 for healthy controls, 0.959 for anemia, 0.937 for night blindness, 0.950 for rickets or osteomalacia, and 0.742 for the minority scurvy cohort. Discrete classification of scurvy was limited by low sample support (4.2% prevalence), but probabilistic risk ranking still performed well (ROC-AUC: 0.9737). The most important features were serum vitamin D (weight: 0.246), night blindness symptom (0.187), serum vitamin B12 (0.149), composite symptom burden (0.112), and hemoglobin (0.089); these ranking matches known physiological pathways. Sequential gradient boosting gives a robust, interpretable screening model for multi-class vitamin deficiency risk. These are proof-of-concept results with internal construct validity; prospective registry validation and threshold calibration are needed before clinical use.

Keywords: clinical decision support; gradient boosting; machine learning; multi-class classification; precision nutrition; socioeconomic determinants; vitamin deficiency prediction

1. Introduction

Deficiency of micronutrients is estimated to be affecting up to two billion people around the world and causes impairment of the immune system, cognitive and physical growth [1]. Causes of these deficiencies include a wide variety of conditions [1]. Iron, cobalamin (vitamin B12) or folate deficiency lead to megaloblastic and microcytic anemias [2], deficiency in vitamin D results in rickets and osteomalacia [3], [4], vitamin A deficiency causes night blindness [1] and deficiency in ascorbic acid results in scurvy. Differentiating between these disorders is challenging for clinicians due to their nonspecific clinical presentation that includes fatigue, weakness and pallor irrespective of different causes of hematomal and metabolic disorders [2], [5]. Anemia is estimated to affect 1.62 billion people by the WHO [6].

In developing countries, the first step that must be overcome in order to manage such deficiencies is the absence of reliable tests [7]. Usually, it is necessary to conduct assays of serum biomarkers, complete blood count tests, and pathological analysis conducted by experts, but unfortunately, all those procedures are not affordable or available in primary health care clinics in low- and middle-income countries [7]. This delay causes the progression of the disease which otherwise could be diagnosed at an early stage.

Machine learning has become one way to ease these diagnostic constraints, combining different types of patient data into a single clinical decision-support tool [8]. These screening models pull together symptom data, dietary intake, and targeted biomarkers to make diagnosis more accessible. Predictive algorithms already work across several metabolic conditions, from forecasting individual blood sugar responses after meals [9] to stratifying risk for diabetes and cardiovascular disease across large cohorts [10], [11]. More recently, researchers have built machine learning models for single micronutrients: predicting cobalamin depletion [12], creening for vitamin D insufficiency at the community level using gradient boosting [13], and diagnosing pediatric vitamin A deficiency in an interpretable way [14].

However, the majority of existing literature on health informatics research is still focused on each nutritional disorder individually, as a single binary classification task [13]. Moreover, a lot of these models neglect some additional factors such as family income, eating habits, and sunlight exposure, which contribute to nutrition-related risks [8], [15], [16]. Explainable machine learning has been used for pediatric undernutrition detection based on population survey data [17], however, so far, there is no framework that predicts the coexistence of different vitamin deficiency syndromes by clinical signs, biomarkers, and social determinants simultaneously [18]. It is essential to distinguish between these disorders when they coexist through modeling overlapping clinical symptoms with various types of features [5].

This study addresses that gap with a proof-of-concept machine learning framework, tested on a parameterized clinical cohort (N = 4,000) with 33 features across five domains: demographics, lifestyle, dietary intake, biomarkers, and symptoms. This serves as an initial feasibility study prior to any possible registry-based

evaluation, and the methods compared include bagging, boosting, and maximum-margin methods. All analyses make use of stratified partitions and cross-validation. The goals of the project include: (1) comparison of the prediction accuracy of Random Forest, Gradient Boosting, and Support Vector Machine classifiers on balanced and imbalanced data sets; (2) identification of features from multiple domains which serve to distinguish between overlapping clinical cases; (3) examination of interactions between socioeconomic conditions, nutritional behavior, and deficiencies; and (4) potential value of automation to clinical decision-making.

2. Materials and Methods

2.1. Dataset and Synthetic Cohort Design

This study randomly generated a synthetic patient cohort (N = 4,000) to match realistic epidemiological distributions of multi-class vitamin deficiency, using a fixed random seed (42) for full replicability. The dataset had 33 features across five categories. Demographic features were age, sex, and body mass index (BMI). The category of lifestyle and behavior included information about smoking habits, indices of physical activities, and alcohol drinking, with a missing at random (MAR) value of 30%. The dietary intake category included diet types along with the percentages of Recommended Dietary Allowance (%RDA) of eight crucial micronutrients, such as vitamins A, C, D, E, and B12, folic acid, iron, and calcium. Physiological markers were modeled using the continuous laboratory marker domain which involved levels of hemoglobin (*g/dL*), serum vitamin D (*ng/mL*), serum vitamin B12 (*pg/mL*), and serum folate (*ng/mL*). Finally, the category of socioeconomic and environment included information on income level, indices of frequency of sun exposure, and regions of latitude.

Clinical symptom profiles were captured through eight binary pathognomonic indicators (fatigue, pale complexion, bone pain, night blindness, bleeding gums, joint pain, alopecia, and dyspnea) along with a continuous composite symptom score. To establish a controlled computational benchmark for evaluating algorithmic feature recovery prior to prospective empirical registry testing, clinical symptoms were encoded using conditional probability distributions derived from established clinical diagnostic guidelines [8], [18]. Specifically, the conditional probability of presenting with bone pain was parameterized at p = 0.80 within the Rickets/Osteomalacia class, night blindness at p = 0.94 within the Vitamin A deficiency cohort, capillary bleeding gums at p = 0.82 within the Scurvy class, and generalized fatigue at p = 0.82 among Anemia cases. Continuous physiological biomarkers were modeled according to standard laboratory reference intervals using localized mean and standard deviation (Mean $\pm$ SD) distributions. For instance, serum vitamin D concentration was defined by a localized distribution of 12 ng/mL (*SD* = 5) in the Rickets/Osteomalacia cohort compared to 35 *ng/mL* (*SD* = 8) in healthy controls.

The dependent target variable encompassed five mutually exclusive clinical diagnostic categories displaying an epidemiologically skewed multi-class distribution: Healthy (11.5%, n = 460), Anemia (27.5%, n = 1,100), Night Blindness (18.9%, n = 756), Rickets/Osteomalacia (37.9%, n = 1,516), and Scurvy (4.2%, n = 168). To resolve the 30% missingness introduced into the categorical alcohol consumption vector under a missing at random (MAR) mechanism, simple mode imputation was executed to establish a low-overhead operational baseline suitable for resource-constrained clinical screening environments [19]. Post the above preprocessing step for dealing with missingness, categorical variables of alcohol consumption, exercise level, diet category, income category, sun exposure, latitude region, sex, and smoking were converted into numerical variables using one-hot encoding approach along with drop-first.

2.2. Data Preprocessing

This study one-hot encoded categorical features (alcohol consumption, physical activity, dietary pattern, income category, sun exposure index, geographic latitude, sex, and smoking status) with a drop-first setup to avoid the dummy variable trap. This expanded the feature space from 33 to 42 dimensions. Continuous variables were standardized with a z-score transform (Equation 1).

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

Here x is the raw value, μ is the feature mean, and σ is the standard deviation, both estimated only from the training partition to avoid data leakage. We split the data into an 80% training set (n = 3,200) and a 20% held-out validation set (n = 800) using stratified random sampling, which kept class proportions consistent across both sets and prevented the rare scurvy class (4.2%, test n = 34) from being underrepresented or dropped [20], [21].

2.3. Machine Learning Classifiers

We evaluated three machine learning approaches for the multi-class prediction baseline. The Random Forest (RF) classifier used bagging over independent decision trees trained on bootstrap samples with randomized feature subspace splitting (100 estimators, max depth 15) to reduce variance [22]. Conversely, the Gradient Boosting (GB) framework applied sequential boosting to minimize differentiable multi-class loss by fitting weak base learners to negative pseudo-residuals [23]. To ensure that the superior discriminative performance of Gradient Boosting was not an artifact of localized overfitting to parameterized generative structures, explicit structural regularization constraints were enforced during compilation: limiting maximum tree depth to $d = 5$, capping the ensemble at 100 base estimators, and imposing a learning rate shrinkage coefficient of $\eta = 0.10$ [23].

Additionally, a Support Vector Machine (SVM) constructed maximum-margin hyperplanes within a high-dimensional feature space via a radial basis function (RBF) kernel, with posterior class probabilities calibrated using cross-validated Platt scaling [24]. Comparison of these three different machine learning algorithms will allow for a systematic evaluation of bagging, sequential boosting, and maximum margin learning methods for tabular health informatics data. All of the prediction models were developed under a Python 3.11 framework employing the Scikit-Learn machine learning library (v1.3), requiring that a common seed condition be imposed (random_state = 42).

2.4. Evaluation Protocol

The performance of algorithmic discrimination was assessed based on the reserved test dataset (n=800) by employing four statistical measures in order to provide multi-dimensional evaluation under the condition of asymmetric class distribution. Classification accuracy was considered as a measure representing overall correctness of diagnosis. In order to take class size variations into account, Weighted F1-score was determined as the average of F1-scores of individual classes multiplied by their support. In order to avoid bias towards major class distributions during evaluation, the Macro-averaged F1-score method was applied with each medical condition receiving equal weight. Finally, the Matthews correlation coefficient (MCC) was used as a metric for balanced assessment even under the condition of severe multi-class imbalance, with a score close to one meaning perfect correlation between outputs [20], [21].

Precision, recall, and F1-scores for each class along with multi-class confusion matrices were also obtained with multi-class ROC curves for each class with one-vs-rest macro averaging of AUC along with precision-recall (PR) curves for estimating average precision. Feature importance analysis was done based on obtaining global importance measures provided by the model, specifically using MDI (Mean Impurity Decrease) for RF model and impurity decrease through gradient descent method for GB models [18]. Generalization ability of each classifier was examined by conducting 5-fold stratified cross validation test over the whole dataset (N=4,000), giving us the mean and respective standard deviation for accuracy, F1-macro, and F1-weighted measures of performance for each algorithm. Importantly, all the three machine learning models have been tested using their pre-set parameters without tuning; such an approach was deliberately adopted to provide a benchmark without any grid search.

3. Results

3.1. Overall Classifier Performance

Overall classification results of the three examined algorithms on the withheld test set (n = 800) are shown in Table I. The Gradient Boosting was the algorithm that reached the best performance across all four criteria, reaching 94.6% accuracy, 94.6% weighted F1-score, 91.5% macro-averaged F1-score and 0.926 MCC. Although the SVM had similar performance compared to the Random Forest in terms of global accuracy, which was 93.9%, the kernelized hyperplanes of the SVM achieved a slightly better result for macro-averaged F1-score (88.8% vs 87.9%).

TABLE I
Overall Classification Performance on the Held-Out Test Partition (n = 800).

Model	Accuracy	F1-Weighted	F1-Macro	MCC
Random Forest	0.939	0.935	0.879	0.916
Gradient Boosting	0.946	0.946	0.915	0.926
SVM	0.939	0.936	0.888	0.916

This implies that hyperplanes obtained from the kernel SVM perform well in separating rare groups such as scurvy [24]. The Random Forest performed poorly in macro-averaged F1 score (87.9%), among the three, implying it did a poor job in misclassifying minority classes in the presence of class imbalance, despite maintaining its high accuracy [21]. Fig. 1 shows all baseline performances.

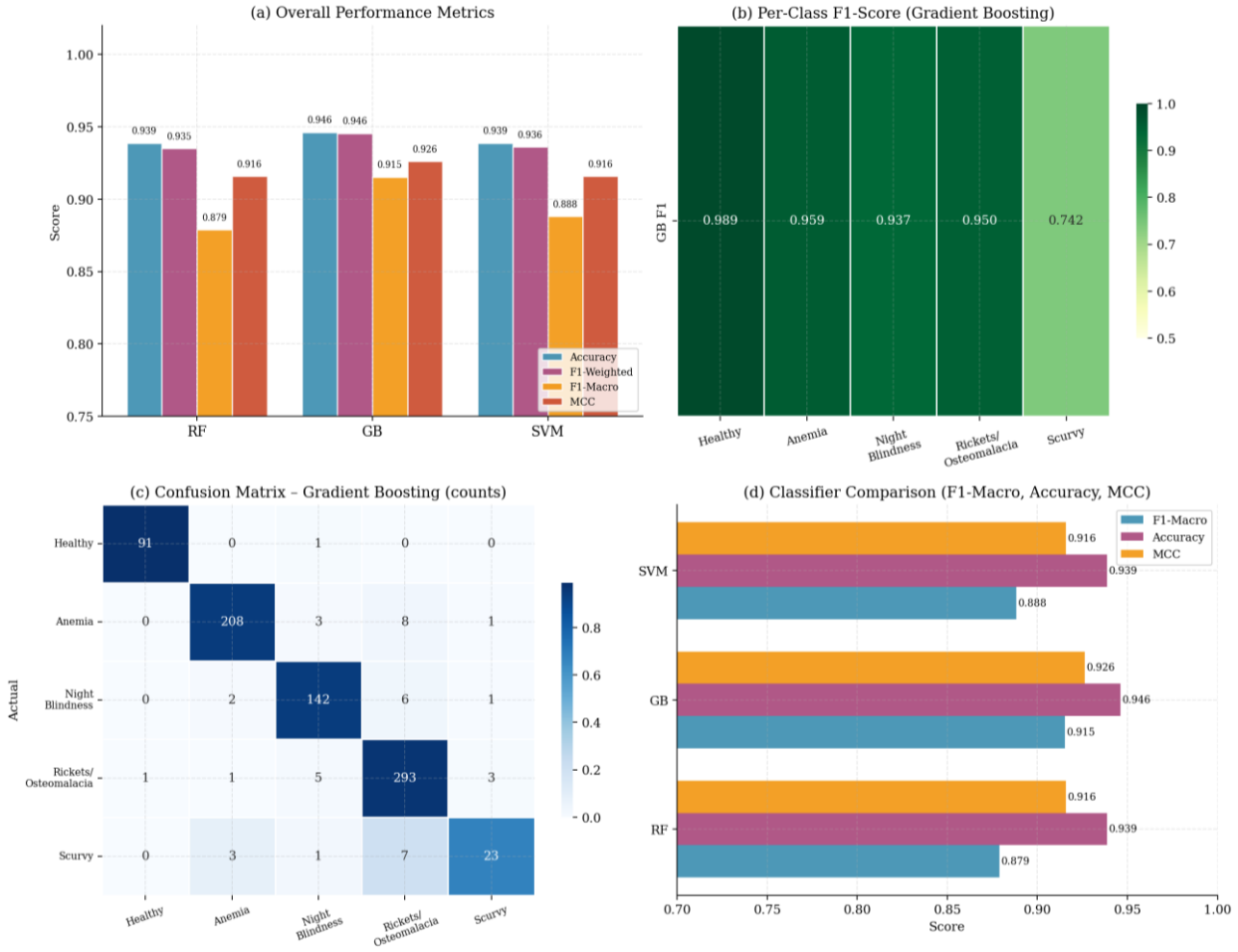


Fig. 1. Performance overview: (a) multi-metric classifier comparison; (b) per-class F1-score heatmap for GB; (c) row-normalized GB confusion matrix; (d) horizontal bar comparison of F1-macro, accuracy, and MCC across classifiers.

Table II presents the results of 5-fold stratified cross validation performance measures of the algorithms on the entire parameterized cohort (N = 4,000). Gradient Boosting maintained high cross validated mean accuracy (94.7% ± 0.6%) and macro F1 score (89.2% ± 1.6%) suggesting that its higher discriminatory power is consistent across folds and not due to any local over fitting of synthetic structure [23]. SVM had similar mean accuracy (94.8% ± 0.7%) but lower macro F1 score (88.4% ± 1.1%) proving that its high accuracy was mainly due to the majority class population. Random Forest had highest macro F1 score variance (± 1.2%), reflecting sensitivity to minority class sample distribution across folds [21]. The close numerical convergence between held-out test metrics (Table I) and cross-validated estimates (Table II) confirms the absence of partition bias in the stratified 80/20 split.

TABLE II
5-Fold Stratified Cross-Validation Summary Across the Full Dataset (N = 4,000).

Model	Accuracy	F1-Macro	F1-Weighted
Random Forest	0.942 ± 0.004	0.867 ± 0.012	0.937 ± 0.005
Gradient Boosting	0.947 ± 0.006	0.892 ± 0.016	0.945 ± 0.007
SVM	0.948 ± 0.007	0.884 ± 0.011	0.943 ± 0.007

3.2. Per-Class Performance

Class-specific precision (P), recall (R), and F1-scores are depicted in Table III together with the respective test-set support (n) for each class for the Random Forest and Gradient Boosting classifiers. Among all tested structural baselines, the Healthy control class exhibited the highest discrimination capabilities, having attained F1-scores of 0.989 for both classifiers as well as F1-score value of 1.000 for the SVM classifier in the hyperplane configuration. This outstanding discriminative ability can be attributed to the unique physiology of the healthy control sample, defined by the balanced levels of all homeostatic biomarkers along with virtually no weightings of clinical symptoms. Rickets or Osteomalacia diagnostic group demonstrated the high prediction stability with F1-scores of 0.950 for both Random Forest and Gradient Boosting classifiers as well as the F1-score of 0.942 for SVM in light of powerful multi-class discrimination based on depressed serum vitamin D concentration levels combined with a prevalence of bone pain (78%).

Gradient Boosting reached an F1-score of 0.959 for the anemia cohort by combining reduced hemoglobin, serum B12, and serum folate with high fatigue prevalence (82% of cases). Night blindness classification (GB F1: 0.937) relied mainly on the ocular symptom itself, present in 94% of cases. Scurvy performed worse across all three models (RF F1: 0.577, recall 44.1%; SVM F1: 0.618; GB F1: 0.742, recall 67.6%). This ceiling comes from the small sample size (4.2% prevalence, test n = 34) combined with symptom overlap with related sub-acute deficiencies [20], [21].

TABLE III
Class-Specific Discrimination Metrics Between Random Forest and Gradient Boosting Ensembles.

Condition	RF-P	RF-R	GB-P	GB-R	GB-F1	*n*
Healthy	0.989	0.989	0.989	0.989	0.989	92
Anemia	0.950	0.959	0.972	0.945	0.959	220
Night Blindness	0.927	0.921	0.934	0.940	0.937	151
Rickets or Osteomalacia	0.928	0.974	0.933	0.967	0.950	303
Scurvy	0.833	0.441	0.821	0.676	0.742	34

3.3. Feature Importance and Clinical Interpretability

Table IV shows the ranking of the feature importance scores for the Gradient Boosting and Random Forest models. The serum vitamin D was identified as the most important predictor in both models (Gradient Boosting: 0.246; Random Forest: 0.156). This is understandable since vitamin D plays an important role in separating rickets and osteomalacia from other metabolic diseases. The binary measure of night blindness as a symptom indicator took the second place among both Gradient Boosting (0.187) and Random Forest (0.139) frameworks, corresponding to its pathognomonic role in detecting serious vitamin A deficiency. The continuous measure of serum vitamin B12 concentration represented another major axis of predictive relevance in the Gradient Boosting algorithm (0.149) to distinguish macrocytic anemia caused by deficiencies of B complex vitamins. The continuous composite measure of symptom count offered another strong multi-class discrimination power within overlapping conditions (Gradient Boosting: 0.112). Objectively measured hemoglobin parameters were the fifth predictor within the Gradient Boosting hierarchy (0.089) and third among the Random Forest predictors (0.095).

TABLE IV
Predictive Feature Importance Weights and Associated Clinical Annotations.

Feature	GB Importance	RF Importance	Clinical Relevance
Serum vitamin D	0.246	0.156	Biomarker indicating vitamin D deficiency/bone disease
Night blindness symptom	0.187	0.139	Pathognomic feature of vitamin A deficiency
Serum B12	0.149	0.084	Indicator separating B12 deficiency anemia from others
Symptom count	0.112	—	Deficiency score; classifier cutting across classes
Hemoglobin	0.089	0.095	Marker of anemia; levels lowered by deficiencies of iron, folate, and B12
Bone pain	—	0.085	Sign of rickets or osteomalacia; No. 4 in RF list
Serum folate	—	0.068	Differentiates folate deficiency anemia from other types

As per the Random Forest algorithm, the symptom of localized bone pain was the fourth most significant attribute (0.085), which supports clinical decision-making on bone tissue matrix pathology. The most important aspect is the presence of behavior-related dietary predictors and sociostructural variables concentrated mostly in the lower quartile of the algorithmic importance measure. These results suggest that biological laboratory values and clinical symptomology bear the main computational responsibility at the moment. Graphs demonstrating feature importance are presented in Fig. 2.

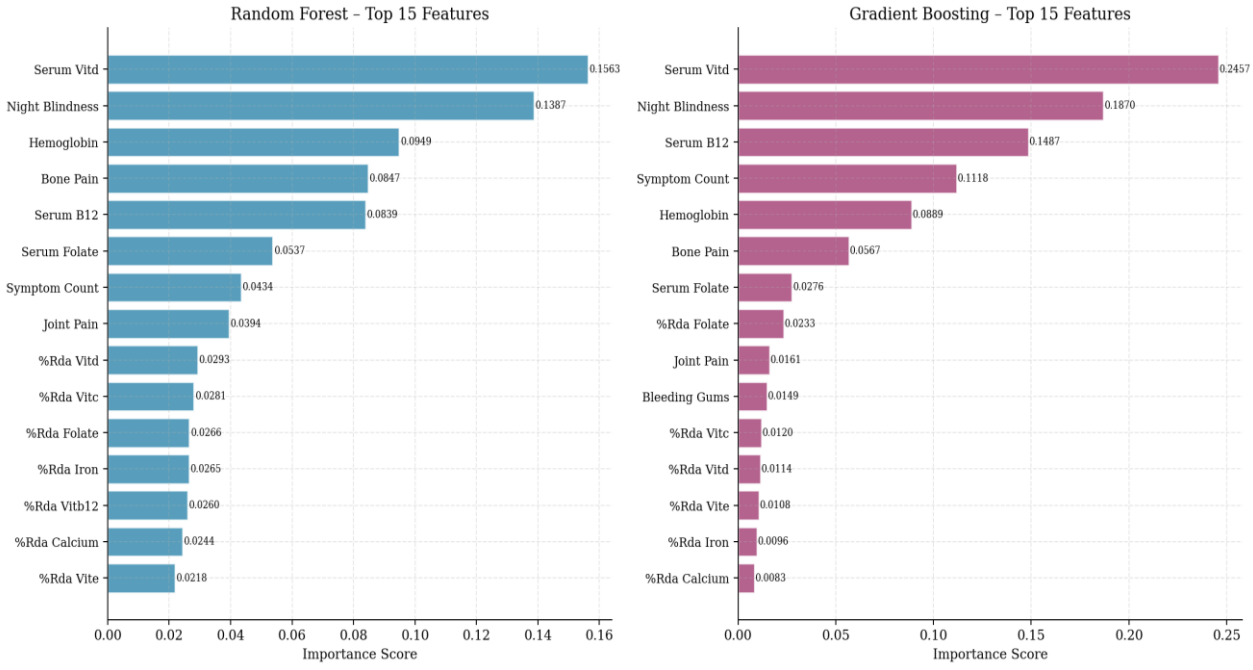


Fig. 2. Comparison of Feature Importance by Mean Decrease Impurity Weights: Random Forest (left) and Gradient Boosting (right).

3.4. Socioeconomic and Dietary Determinants

Fig. 3 shows the analysis of disease class distributions according to behavior type (multidomain), income level, and environmental photobiological factors.

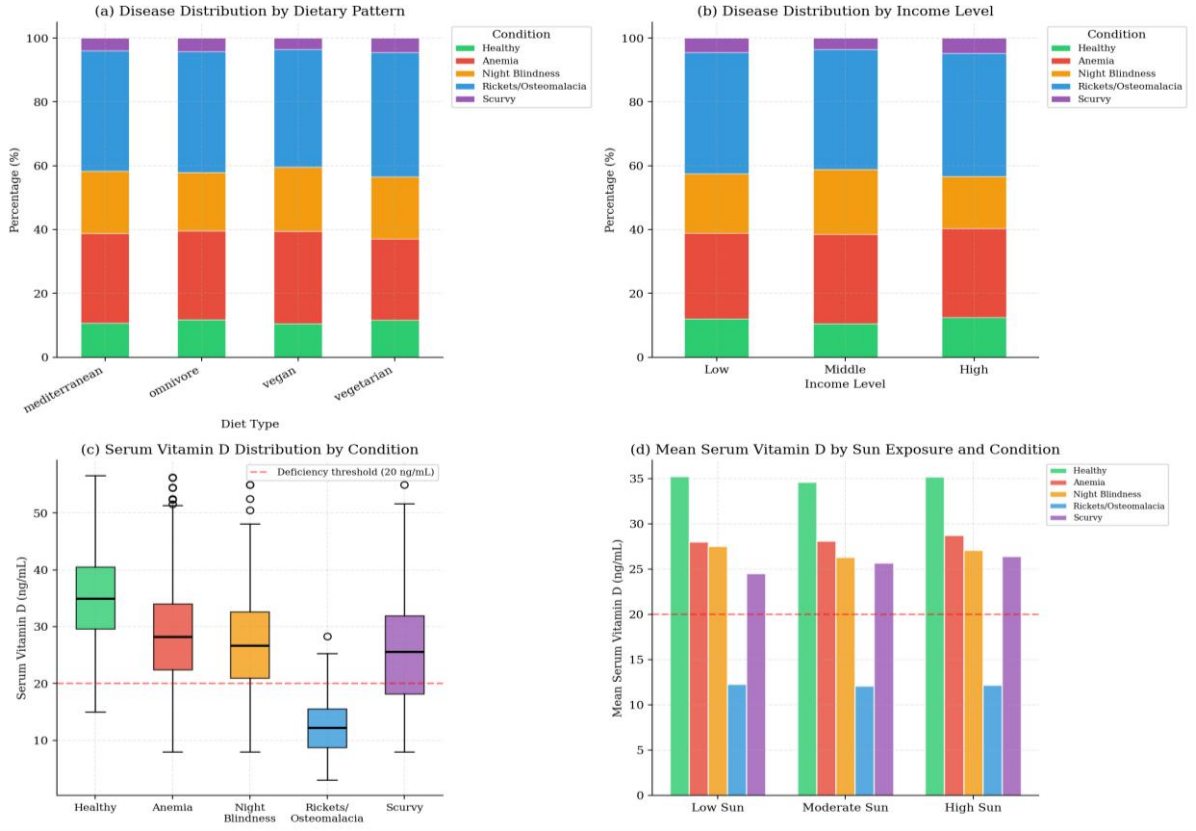


Fig. 3. Socioeconomic and dietary factors: (a) distribution of diseases according to dietary groupings; (b) distribution of diseases according to income levels; (c) serum vitamin D box plots for each condition; (d) mean serum vitamin D according to sun exposure groups and conditions.

Within modeled behavioral groups, those whose diets were purely vegan showed the greatest relative occurrence of Rickets or Osteomalacia (36.8% of all vegans calculated within the entire dataset, $N = 4{,}000$) and associated Anemia conditions (28.9%). Such patterns are in line with physiological mechanisms of restricted absorption capabilities related to vitamin D, heme iron, and B12 from purely vegetarian meals [25], [26]. To compare the baseline figures, the omnivorous group had a local Rickets or Osteomalacia risk of 37.8% and Anemia incidence rate of 27.8%. At the same time, Mediterranean diets showed respective values of 37.8% and 28.1%. Minor variations between such groups could be considered a result of the stochastic nature of the data generation process used for this analysis. Alignment between Mediterranean and omnivorous groups on vegan deficient levels of diseases confirms that these variations should be viewed as behavioral predictors rather than disease determinants. Household income markers showed no significant over-dominance in relation to absolute disease class occurrence in this particular mathematical cohort.

The fraction of subjects showing evidence of systemic elevation in symptom burden (as defined by the simultaneous presentation of ≥ 3 symptoms) was similarly distributed among economic subpopulations, being observed at 41.0%, 42.1%, and 41.2% among low-income, middle-income, and high-income groups respectively. The uniformity of this pattern implies weakening of the linear relationship between economic factors and disease as produced by the artificial model, compared to natural populations which show a clear correlation between lower socioeconomic status and poorer nutrition and food insecurity [15]. The serum vitamin D levels were shown to demonstrate powerful condition-based distributions based on environmental factors (Fig. 3c). Specifically, while the Rickets/Osteomalacia category showed consistently low serum levels for every sun exposure frequency, high latitudes were associated with lower serum vitamin D levels irrespective of condition. High frequencies of sun exposure located in low-latitude environments, on the other hand, were shown to lead to the highest serum levels for every condition tested. This pattern is consistent with biological processes of cholecalciferol production in skin under UVB light at different latitudes [3].

3.5. Symptom Profiles and Biomarker Patterns

Fig. 4 provides the clinical symptom signature for the disease in question based on the anticipated symptom presentation in the medical world.

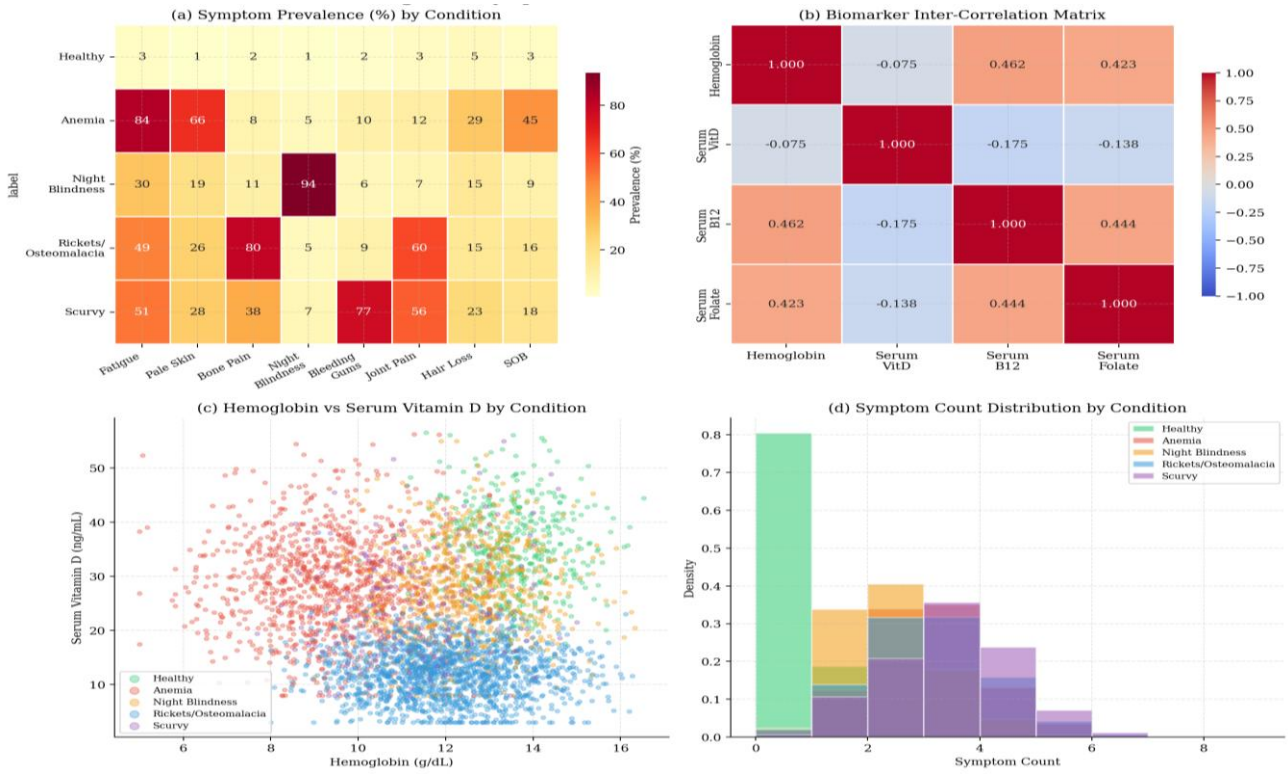


Fig. 4. Symptom profile and biomarker pattern: (a) heatmap of symptom probability under given conditions; (b) biomarker correlation matrix (Pearson's r); (c) scatter plot of hemoglobin against vitamin D levels; (d) distribution of symptom count per condition.

Conditional prevalence of night blindness among individuals with ocular night blindness deficiency was 94%, while somatic bone pain had a prevalence rate of 78% for Rickets or Osteomalacia. Capillary bleeding of gums had a probability of 82% for Scurvy, whereas generalized fatigue showed a prevalence rate of 82% in the case of Anemia. On the other hand, symptom prevalence rates in the Healthy control group were less than 5% throughout the laboratory investigations conducted for each individual indicator. Correlation coefficients (r) of the objective laboratory biomarkers showed low to moderate linear intercorrelations. Primary linear intercorrelation was recorded between continuous hemoglobin levels and serum vitamin B12 concentration (r=0.462) followed by hemoglobin and serum folate (r=0.423). The relatively moderate values are indicative of the intertwined biological functions of cobalamin and folate in normal erythropoiesis in humans. The continuous vitamin D concentrations in blood serum showed relatively weak correlations with hemoglobin levels (r = -0.076), serum vitamin B12 (r = -0.175), and serum vitamin folate (r = -0.138). It is evident that the relatively poor correlation profile demonstrates that the diagnostic model of bone diseases based on vitamin D is mostly independent of the anemia model based on B vitamins in the feature space.

3.6. ROC and Precision-Recall Performance

Quantitative representation of the exact Area Under the Curve (ROC-AUC) and Average Precision (AP) values obtained by applying the one-vs-rest approach on the Gradient Boosting technique is provided in Table V below. Based on the performance matrix of the benchmark, the Healthy control group demonstrated a 1.0000 AUC curve value, indicating that there was absolute differentiation of the homeostatic biomarkers and symptomatology from any other disease states. The Anemia and Night Blindness target states recorded near-perfect AUC discrimination curves of 0.9956 and 0.9957, respectively, whereas the AUC optimum for Rickets and Osteomalacia was 0.9914.

TABLE V
Exact Multi-Class ROC-AUC and Average Precision (AP) Metrics for Gradient Boosting.

Condition	AUC (ROC)	AP (PR-AUC)
Healthy	1.0000	0.9998
Anemia	0.9956	0.9931
Night Blindness	0.9957	0.9852
Rickets or Osteomalacia	0.9914	0.9812
Scurvy	0.9737	0.8376

Notably, despite registering a reduced hard-threshold F1-score of 0.742 under default decision boundaries (0.50 cutoff), the minority Scurvy class achieved a high area under the receiver operating characteristic curve (ROC-AUC: 0.9737). This shows that even though hard classification boundaries suffer from low sample support, the model still ranks probabilistic risk well across continuous outputs [21]. Point-of-care screening tools generally favor continuous probability scores over fixed binary labels anyway, since scores let clinicians adjust the decision threshold to avoid missing severe ascorbic acid depletion [21], [27]. Precision-recall was high for the majority classes; scurvy scored 0.8376 average precision, reflecting both its low prevalence and its symptom overlap with other conditions. Fig. 5 shows discrimination and precision-recall curves for all target states.

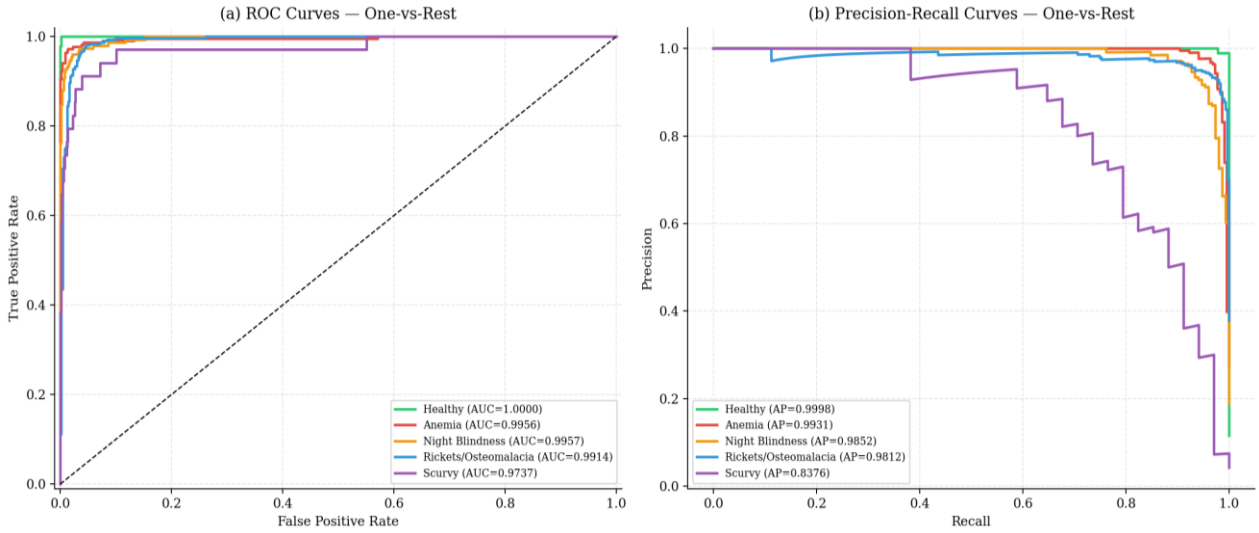


Fig. 5. ROC and precision-recall curves for Gradient Boosting (one-vs-rest): (a) class-specific ROC curves with AUC values; (b) precision-recall curves with average precision (AP).

4. Discussion

Our findings have demonstrated that the usage of the ensemble methods, particularly sequential gradient boosting, is possible to build up robust dividing lines for vitamin deficiencies using an adequate set of features. The model includes laboratory parameters, symptoms associated with the specific diseases, eating habits, and social economic status. All the cut-off points obtained in this study were derived from the testing set.

4.1. Performance Contextualization

In comparison with the baseline of 20% due to pure chance in a five-class classification problem, the results for gradient boosting show high levels of accuracy of 94.6% and F1-score of 91.5%. The high level of performance is provided mostly by continuous serum biomarkers such as vitamin D and vitamin B12, which discriminate well for their diseases. The models that use just subjective symptoms or subjective evaluation of one's diet without using biomarkers will probably show lower results due to similarity in symptoms in different types of deficiency. Biomarker availability seems to be the key limitation factor. The global accuracy of SVM and Gradient Boosting was the same (93.9%) but macro-average F1 score of SVM was lower (88.8%), which indicates that SVM was less sensitive in the case of scurvy. The largest difference between global accuracy and macro-averaged F1 was shown by Random Forest (93.9% and 87.9%) with minority-class recall as low as 44.1%.

The best F-score of Scurvy was found to be 0.742 for all three models due to the high complexity of the problem caused by extreme imbalance [20], [28]. With just 34 cases tested (4.2%), models that aim at maximizing precision cannot help but sacrifice recall [21]. Any future clinical implementation of this model requires implementing safety mechanisms: SMOTE for mixed table data [28], cost-sensitive loss functions [20], [27], and threshold probability adjustment to not miss severe ascorbic acid deficiency [21].

4.2. Clinical Interpretation of Feature Importance

Vitamin D was the best predictor in the Gradient Boosting Model (0.246) and in the Random Forest (0.156) models as well due to its significance in metabolic bone diseases (rickets or osteomalacia comprised 37.9% of the total sample population). The next best predictors were night blindness (GB 0.187) which is unique to vitamin A deficiency. Serum B12 (0.149), folate, and hemoglobin (0.089) aided in the differentiation between overlapping types of anemias [2], [5], [12]. Both cobalamin/folate and iron deficiency anemia present symptoms such as fatigue and shortness of breath, but the addition of serum helped differentiate between them [5].

The alignment of the feature ranking with physiological knowledge adds to the internal validity of the construct of the model, which is relevant to develop explainable clinical decision support models [5], [18], [29]. In light of the fact that most of the predictive strength resides within the laboratory parameters and physical symptoms instead of dietary intake information, there is room to develop a diagnostic model based on three laboratory biomarkers and some selected physical parameters to save time in clinical work flow.

4.3. Socioeconomic Determinants and Health Equity

It was discovered that there is increased prevalence of rickets or osteomalacia (36.8%) and anemia (28.9%) in strict vegans due to the low absorption of non-heme iron, lack of vitamin B12 in plant products, and less cholecalciferol intake in strict vegans, based on literature on nutrition research [25], [26]. This is not an argument against plant-based diets since the positive effects of such diets on cardiovascular and metabolic systems have been proven.

The levels of vitamin D in serum were different depending on latitude and exposure to sunlight, as expected due to synthesis of cholecalciferol via UV-B rays [3], [4]. One-hot encoding expands the number of features from 33 to 42 dimensions, raising the common problem of collinearity, but tree-based algorithms, such as Gradient Boosting and Random Forest, are not affected by it as linear models are [22], [23]. Decision trees use splitting based on impurity, not coefficients, which allows using these algorithms in the presence of correlation between socioeconomic and environmental factors [18], [22]. The factor of income was not found to have a pattern in the experiment since the synthetic data used for analysis assigned it randomly; however, in real life socioeconomic and food deprivation is associated with risk of micronutrients deficiency [15], [16].

4.4. Limitations

These findings come with some real limitations. The synthetic cohort (N = 4,000) was necessary for a controlled proof-of-concept, but it can't capture the confounders, measurement noise, or physiological variation you'd see in real patients [8]. Since our probability distribution between conditions and symptoms are set based on clinically proven evidence (e.g., night blindness at 94% prevalence within the vitamin A deficiency category), the feature importance values obtained from the models show that they are internally consistent rather than being effective in practical applications. Prospective studies to validate the performance of such models on actual clinical registries would be necessary before applying them in clinical practice [8], [12]. Second, regarding data preprocessing topology, resolving the 30% missingness in categorical alcohol consumption via univariate mode imputation represented a low-overhead operational baseline. However, mode replacement artificially truncates variable entropy and attenuates bivariate covariance structures with downstream socioeconomic attributes [19].

Future empirical investigations should implement multivariate imputation frameworks, such as Multiple Imputation by Chained Equations (MICE) or MissForest, to preserve feature covariance topology [19]. Third, baseline model training did not incorporate active resampling or cost-sensitive re-weighting during compilation, constraining hard-threshold classification performance for the minority Scurvy class (n = 34 test examples) [27], [28]. Fourth, local explainability frameworks, such as SHAP or LIME vector attributions, were omitted from the primary evaluation protocol, representing a critical component for future regulatory approval and clinician trust [18], [29]. Finally, the cross-sectional design precludes longitudinal tracking of deficiency progression or therapeutic response.

4.5. Future Directions

The future directions for primary research vectors going forward are as follows:

1. Conducting prospective validation using actual electronic medical records of communities and applying multi-micronutrient screening programs to LMIC regions where the prevalence of the dual burden of multi-micronutrient malnutrition is concentrated, including the regions within Sub-Saharan Africa and South or Southeast Asia [1], [17].
2. Implementing effective multi-class imbalance methods with an emphasis on the use of hybrid resampling methods and cost-effective threshold setting to minimize false positives for minority classes while not adversely affecting global specificity [5], [28].
3. Creating and evaluating localized visualization systems such as SHAP/LIME for individual vector analysis, ensuring that each risk prediction generated by the automated model can be effectively evaluated by field clinicians for accuracy and reliability [18].
4. Development of federated learning systems for training across different organizations and updating weights iteratively without breaching any restrictions on patient data privacy or the centralized storage of critical medical registries.
5. Combining the weights from the optimized gradient boosting algorithm into existing point-of-care microfluidic platforms, creating a mobile and field-ready clinical decision-making aid for rapid non-invasive nutritional risk screening [8].

5. Conclusions

This study highlights the computational viability of the sequential gradient boosting approach to perform multi-class predictions simultaneously on five different vitamin deficiencies using a proof-of-concept framework with parameters. The best Gradient Boosting model was able to provide an accuracy of 94.6%, macro-averaged F1-score of 91.5%, and a Matthews Correlation Coefficient value of 0.926. Symmetric feature importance among ensemble models established that the continuous level of vitamin D in the blood, night blindness in the eyes, and blood level of vitamin B12 are the main physiological markers.

While hard-threshold classification of the uncommon Scurvy population was limited due to lack of sample size support under default thresholds (F1 score: 0.742), the continuous probabilistic risk scoring maintained high discrimination capability (ROC-AUC: 0.9737). The above result highlights the importance of using cost-sensitive loss tuning and adaptive thresholds in any future implementation to avoid potential false negative triaging. Importantly, all metrics are based on reproducible experimental workflow with strict split management. Empirical testing on various patient populations is a must prior to actual deployment. In sum, this work lays down the groundwork for a scalable digital health platform focused on medicine-driven screening.

Author Contributions: Conceptualization, L.K. and D.E.J.L.; methodology, L.K.; software, L.K.; validation, L.K., D.E.J.L. and V.M.; formal analysis, L.K.; investigation, L.K.; resources, L.K.; data curation, L.K.; writing original draft preparation, L.K.; writing review and editing, L.K., D.E.J.L. and V.M.; visualization, L.K.; supervision, D.E.J.L.; project administration, L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgements: The authors acknowledge the Python open-source ecosystem, including scikit-learn, pandas, NumPy, matplotlib, and seaborn, used in the experimental pipeline.

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

References

[1] R. L. Bailey, K. P. West Jr., and R. E. Black, "The Epidemiology of Global Micronutrient Deficiencies," *Annals of Nutrition and Metabolism*, vol. 66, no. Suppl. 2, pp. 22–33, 2015, doi: 10.1159/000371618.

[2] L. H. Allen, "Causes of Vitamin B 12 and Folate Deficiency," *Food and Nutrition Bulletin*, vol. 29, no. 2_suppl1, pp. S20–S34, Jun. 2008, doi: 10.1177/15648265080292S105.

[3] K. D. Cashman *et al.*, "Vitamin D deficiency in Europe: pandemic?," *The American Journal of Clinical Nutrition*, vol. 103, no. 4, pp. 1033–1044, Apr. 2016, doi: 10.3945/ajcn.115.120873.

[4] M. F. Holick, "Vitamin D Deficiency," *New England Journal of Medicine*, vol. 357, no. 3, pp. 266–281, Jul. 2007, doi:

10.1056/NEJMra070553.
[5] B. S. D. Darshan *et al.*, "Differential diagnosis of iron deficiency anemia from aplastic anemia using machine learning and explainable Artificial Intelligence utilizing blood attributes," *Scientific Reports*, vol. 15, no. 1, p. 505, Jan. 2025, doi: 10.1038/s41598-024-84120-w.
[6] E. McLean, M. Cogswell, I. Egli, D. Wojdyla, and B. de Benoist, "Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993–2005," *Public Health Nutrition*, vol. 12, no. 4, pp. 444–454, Apr. 2009, doi: 10.1017/S1368980008002401.
[7] N. Hwalla *et al.*, "The Prevalence of Micronutrient Deficiencies and Inadequacies in the Middle East and Approaches to Interventions," *Nutrients*, vol. 9, no. 3, p. 229, Mar. 2017, doi: 10.3390/nu9030229.
[8] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nature Medicine*, vol. 28, no. 1, pp. 31–38, Jan. 2022, doi: 10.1038/s41591-021-01614-0.
[9] D. Zeevi *et al.*, "Personalized Nutrition by Prediction of Glycemic Responses," *Cell*, vol. 163, no. 5, pp. 1079–1094, Nov. 2015, doi: 10.1016/j.cell.2015.11.001.
[10] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," *Computational and Structural Biotechnology Journal*, vol. 15, pp. 104–116, Jan. 2017, doi: 10.1016/j.csbj.2016.12.005.
[11] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," *PLOS ONE*, vol. 14, no. 5, p. e0213653, May 2019, doi: 10.1371/journal.pone.0213653.
[12] F. Demirci, O. Yıldırım, A. Demirci, and P. Akan, "Development and Temporal Validation of Explainable Machine Learning Models for Predicting Vitamin B12 Deficiency Using Routine Laboratory Analytes," *Diagnostics*, vol. 16, no. 4, p. 563, Feb. 2026, doi: 10.3390/diagnostics16040563.
[13] J. Guo, Q. He, and Y. Li, "Machine learning-based prediction of vitamin D deficiency: NHANES 2001-2018," *Frontiers in Endocrinology*, vol. 15, Feb. 2024, doi: 10.3389/fendo.2024.1327058.
[14] J. Ramesh, S. Sabu Thomas, A. H. Khamis, A. Sagahyroon, F. Aloul, and A. Rushood, "Interpretable Machine Learning Models for Vitamin A Deficiency Screening in School-Aged Children," *IEEE Access*, vol. 13, pp. 212999–213015, 2025, doi: 10.1109/ACCESS.2025.3644236.
[15] N. Darmon and A. Drewnowski, "Does social class predict diet quality?," *The American Journal of Clinical Nutrition*, vol. 87, no. 5, pp. 1107–1117, May 2008, doi: 10.1093/ajcn/87.5.1107.
[16] M. Marmot, J. Allen, R. Bell, E. Bloomer, and P. Goldblatt, "WHO European review of social determinants of health and the health divide," *The Lancet*, vol. 380, no. 9846, pp. 1011–1029, Sep. 2012, doi: 10.1016/S0140-6736(12)61228-8.
[17] M. M. Islam, N. M. S. J. Kibria, S. Kumar, D. C. Roy, and M. R. Karim, "Prediction of undernutrition and identification of its influencing predictors among under-five children in Bangladesh using explainable machine learning algorithms," *PLOS ONE*, vol. 19, no. 12, p. e0315393, Dec. 2024, doi: 10.1371/journal.pone.0315393.
[18] F. Di Martino and F. Delmastro, "Explainable AI for clinical and remote health applications: a survey on tabular and time series data," *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5261–5315, Jun. 2023, doi: 10.1007/s10462-022-10304-3.
[19] J. A. C. Sterne *et al.*, "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls," *BMJ*, vol. 338, no. jun29 1, pp. b2393–b2393, Sep. 2009, doi: 10.1136/bmj.b2393.
[20] Haibo He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.
[21] J. X. C. Ke, A. DhakshinaMurthy, R. B. George, and P. Branco, "The effect of resampling techniques on the performances of machine learning clinical risk prediction models in the setting of severe class imbalance: development and internal validation in a retrospective cohort," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 91, Nov. 2024, doi: 10.1007/s44163-024-00199-0.
[22] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[23] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.
[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
[25] V. Melina, W. Craig, and S. Levin, "Position of the Academy of Nutrition and Dietetics: Vegetarian Diets," *Journal of the Academy of Nutrition and Dietetics*, vol. 116, no. 12, pp. 1970–1980, Dec. 2016, doi: 10.1016/j.jand.2016.09.025.
[26] W. J. Craig, "Health effects of vegan diets," *The American Journal of Clinical Nutrition*, vol. 89, no. 5, pp. 1627S-1633S, May 2009, doi: 10.3945/ajcn.2009.26736N.
[27] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, "Handling imbalanced medical datasets: review of a decade of research," *Artificial Intelligence Review*, vol. 57, no. 10, p. 273, Sep. 2024, doi: 10.1007/s10462-024-10884-2.
[28] Y. Yang, H. A. Khorshidi, and U. Aickelin, "A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems," *Frontiers in Digital Health*, vol. 6, Jul. 2024, doi: 10.3389/fdgth.2024.1430245.
[29] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, Jul. 2019, doi: 10.1002/widm.1312.