Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 4, October 2026, pp: 1665-1686 e-ISSN: 2656-8632

REVIEW ARTICLE OPEN ACCESS

Machine Learning Models Using Gut Microbiome Data for Obesity-Relevant Classification, Dietary Response Prediction, and Personalised Nutrition: A PRISMA-Guided Systematic Review

Poppy Fransisca Amelia[1], Fitra Adi Prayogo[1]*, Isy Royhanaty[1], Heni Wijayanti[1], Egy Sunanda Putra[2,3]

[1] Study Program of Biomedical Sciences, Universitas Karya Husada, Semarang, Indonesia
[2] Department of Health Promotion, Poltekkes Kemenkes Jambi, Jambi, Indonesia
[3] Centre of Excellence for Science and Technology, Poltekkes Kemenkes Jambi, Jambi, Indonesia

Corresponding author: Fitra Adi Prayogo (e-mail: fitraadi@unkaha.ac.id). **Author(s) Email**: Poppy Fransisca Amelia (e-mail: poppy@stikesyahoedsmg.ac.id), Isy Royhanaty (e-mail: isy@stikesyahoedsmg.ac.id), Heni Wijayanti (heniwijayanti55@gmail.com), Egy Sunanda Putra (egyputra93@poltekkesjambi.ac.id).

Abstract **Obesity is a major non-communicable disease burden, yet population-level dietary guidance does not capture interindividual metabolic variability related to gut microbiome composition. Machine learning (ML) may support microbiome-informed precision nutrition, but evidence remains heterogeneous across classification, response-prediction, and intervention-adjacent studies. This systematic review evaluated ML models that use gut microbiome data for obesity-relevant classification, dietary response prediction, and personalised nutrition outcomes. A PRISMA 2020-guided search of PubMed/MEDLINE, Scopus, Web of Science, and IEEE Xplore identified English-language peer-reviewed articles published from January 2015 to February 2026. Eligible studies included adult cohorts with overweight/obesity participants, BMI-stratified microbiome data, or dietary-response outcomes relevant to obesity-related precision nutrition. Risk of bias was assessed using PROBAST, Newcastle-Ottawa Scale, and RoB 2.0 as appropriate. Twelve studies met the inclusion criteria and were assigned to dominant evidence families: seven classification or biomarker-modelling studies, two dietary or postprandial-response prediction studies, one small, blinded, personalised-diet validation within a PPGR cohort, one prospective, controlled-feeding cohort, and one longitudinal weight-loss prediction study. XGBoost-based classifiers showed the strongest comparative support for static obesity/BMI classification (peak internal AUC 0.90), with SHAP improving interpretability in the best-performing study. Random Forest/MetAML performance was variable, ranging from cross-study AUCs around 0.60 to internal or robust-signature AUCs of 0.83-0.85. The main translational finding was a generalisation envelope from a peak internal AUC of 0.90 to approximately 0.60 across cross-study/meta-analytic benchmarks. Specificity was not reported in any classification-capable study, and calibration metrics were absent across all included studies. Microbiome-based ML shows early promise for precision-nutrition CDSS, but clinical translation requires external validation, compositional-data-aware preprocessing, complete metric reporting, and prospective dietary trials.**

Keywords **Gut microbiome; Machine learning; Obesity; Personalised nutrition; Systematic review**

I. Introduction

Obesity is among the most pressing non-communicable diseases of the twenty-first century. The World Health Organization reports that more than one billion adults live with obesity, and obesity substantially increases the risk of type 2 diabetes, cardiovascular disease, non-alcoholic fatty liver disease, obstructive sleep apnoea, and multiple cancers [1], [2]. Long-term weight management remains difficult despite behavioural, nutritional, pharmacological, and surgical approaches; this limited durability is partly explained by interindividual differences in postprandial and metabolic responses that are not captured by population-average dietary guidelines [3]-[6]. Precision nutrition seeks to individualise dietary guidance using biological, behavioural, and environmental data. Gut microbiota is central to this effort because microbial communities contribute to energy harvest, short-chain

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

fatty-acid and bile-acid metabolism, immune signalling, low-grade inflammation, and satiety pathways [7]-[13]. However, these relationships are difficult to model. The data are non-linear, high-dimensional, sparse, and compositional, and they are strongly affected by technical and population heterogeneity [14]-[20].

ML methods can learn non-linear structure from microbiome feature spaces and integrate multi-omics inputs, including 16S rRNA, whole-genome shotgun metagenomics, metabolomics, host genetics, and continuous glucose monitoring. Classical ML algorithms and deep-learning architectures have therefore been applied to obesity classification, postprandial-response prediction, microbiome biomarker discovery, and dietary-recommendation generation [21]-[32]. In Medical Informatics, however, high apparent performance is insufficient unless models are externally validated, interpretable, calibrated, and reported in a way that allows CDSS safety assessment [33]-[35]. Despite the expanding literature, prior reviews have generally focused on gut dysbiosis biology, precision nutrition broadly, or ML in healthcare rather than on the specific evidence chain linking gut microbiome-based ML to obesity classification, dietary-response prediction, and personalised-nutrition translation [36]-[42]. Three specific gaps remain unaddressed. First, existing reviews conflate computational classification benchmarks with clinical dietary efficacy, treating a high internal AUC as evidence of clinical readiness even when no prospective dietary outcome was tested. Second, no prior review has paired a quantitative performance synthesis with a domain-level risk-of-bias audit and an explicit metric-coverage matrix, so the field lacks a clear account of which discrimination, calibration, and specificity metrics are actually reported across the obesity-specific evidence base. Third, prior syntheses have not translated their findings into a CDSS-readiness framework that states the external-validation, calibration, and specificity thresholds a microbiome-ML model must meet before it can be deployed for dietary guidance. This review is therefore different because it restricts the evidence pool to adult obesity-relevant microbiome-ML studies, separates classification-only evidence from response-prediction and intervention evidence, and audits both reporting completeness and bioinformatics-pipeline risk rather than summarising headline performance alone.

Following PRISMA 2020 reporting guidance, this systematic review evaluates ML models applied to gut microbiome data across three connected task families: (i) obesity/BMI classification and obesity-relevant microbiome stratification from microbial features; (ii) dietary and postprandial metabolic-response prediction; and (iii) prospective personalised-nutrition or weight-management evidence. The evidence synthesis is based on 12 eligible studies, with supporting literature used only for background, methodological, bioinformatics, and translational context. The specific contributions of this review are:

1. A quantitative performance synthesis of ML algorithms (AUC-ROC, Accuracy, F1-Score, Recall, Pearson R) across 12 eligible studies, identifying the most consistently high-performing algorithm class for gut microbiome-based obesity/BMI classification and obesity-relevant microbiome stratification within the limits of available evidence.
2. A structured separation of classification/biomarker-modelling, dietary-response prediction, intervention-adjacent, and weight-management evidence to avoid conflating computational benchmarks with clinical dietary efficacy.
3. A combined metric-reporting and risk-of-bias audit, including PROBAST domain-level ratings and a standardised metric-coverage matrix, to identify the evidence gaps that constrain clinical interpretation.
4. A bioinformatics and CDSS reporting framework summarising recurrent pipeline challenges, compositional-data requirements, validation expectations, and TRIPOD+AI-aligned reporting standards for future microbiome-ML studies.
5. A translation-oriented synthesis that maps each evidence family to a Clinical Decision Support System (CDSS) readiness level, distinguishing models suitable only for upstream microbiome stratification from those with prospective dietary-response evidence, and specifying the minimum external-validation, calibration, and specificity thresholds required before clinical deployment can be considered.

II. Method

A. Study Design

This systematic review was designed according to the PRISMA 2020 statement [43]. The protocol was structured before data extraction. The PICO framework, eligibility criteria, and extraction schema were fixed before screening. The review was not prospectively registered in PROSPERO, OSF, or an equivalent registry. This is a material limitation for protocol transparency and is acknowledged here and in Section IV.C. To mitigate the absence of registration, the PICO, full eligibility criteria, extraction schema, and risk-of-bias instruments were defined a priori and reported in full (Algorithm 1; Section II). Complete search strings, screening counts, and exclusion reasons are provided as supplementary files so that the process can be independently audited and reproduced. No primary patient data were collected, and all analyses were based on previously published peer-reviewed literature.

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

B. PICO Framework

Population: Adult human cohorts (age ≥18 years) that included confirmed overweight/obesity participants (BMI ≥25 kg/m²), BMI-stratified microbiome data, or dietary/postprandial response outcomes directly relevant to obesity-related precision nutrition. Gut microbiota had to be profiled via 16S rRNA amplicon sequencing or whole-genome shotgun (WGS) metagenomics. Studies utilising public microbiome databases (e.g., GMrepo, Human Microbiome Project 2, European Genome-phenome Archive) were eligible when adult BMI metadata, obesity/BMI stratification, or relevant dietary-response outcomes were accessible for subgroup or model interpretation.

Intervention: Application of ≥1 supervised, unsupervised, or semi-supervised ML or deep learning (DL) algorithm (including Random Forest, Support Vector Machine, XGBoost, Gradient Boosting, Artificial Neural Network, CNN, LSTM, Autoencoder, Contrastive Learning, Transformer, or Reinforcement Learning architectures) to classify obesity status, cluster microbiome phenotypes, predict dietary responses, or recommend personalised dietary interventions from gut microbiota feature profiles.

Comparison: Conventional statistical methods (ANOVA, logistic regression, PCoA, PERMANOVA, linear mixed-effects models), non-ML clinical scoring systems, or alternative ML architectures used as within-study performance benchmarks.

Outcome: Primary outcomes: quantitative ML performance metrics (AUC-ROC, Accuracy, F1-Score, Precision, Recall (Sensitivity), Specificity, Matthews Correlation Coefficient (MCC), RMSE, MAE, Pearson Correlation (R)). Secondary outcomes: dietary recommendation clinical efficacy (postprandial glycaemic response, weight loss, macronutrient optimisation), microbiome biomarker identification and feature importance rankings, and model interpretability assessments.

C. Search Strategy

Systematic searches were conducted in PubMed/MEDLINE, Scopus, Web of Science Core Collection, and IEEE Xplore between 1 and 14 February 2026. Searches combined microbiome, obesity/overweight, ML/deep-learning, and diet/precision-nutrition terms using database-specific field tags. Database-level filters were English language, publication year 2015-2026, and peer-reviewed article type. Review-type records captured for search sensitivity were excluded during screening unless they contained original extractable data. IEEE Xplore records were retained only when full methods and performance metrics were available. Deduplication used EndNote 21 with DOI-first matching followed by title/first-author similarity and manual adjudication by two reviewers. Full search strings and hit counts are provided in Supplementary Table S1.

D. Inclusion and Exclusion Criteria

Studies were included if all of the following criteria were satisfied: (I-1) original peer-reviewed research article applying ML or DL to gut microbiota data; (I-2) human adult subjects ≥18 years in overweight/obesity cohorts, BMI-stratified datasets, or dietary-response cohorts directly relevant to obesity-related precision nutrition; (I-3) explicit deployment and evaluation of ≥1 ML algorithm with described cross-validation or train-test split methodology; (I-4) gut microbiota profiled by 16S rRNA sequencing or WGS metagenomics with reportable OTU/ASV abundance tables, metagenomic profiles, or equivalent microbiome features; (I-5) reporting of ≥1 quantifiable ML performance metric or dietary-response outcome; (I-6) linkage of ML microbiome predictions to dietary recommendation output, postprandial glycaemic/lipemic response modelling, obesity/BMI stratification, or weight-management outcome; (I-7) English-language, peer-reviewed publication in an indexed journal, 2015–2026.

Studies were excluded if any of the following applied: (E-1) conference abstracts, editorials, commentary, preprints, or narrative reviews without data extraction tables; (E-2) exclusively paediatric populations (age <18 years); (E-3) no ML component or ML applied exclusively to non-gut omics data; (E-4) oral, vaginal, or skin microbiota profiling without a gut component; (E-5) no reportable quantifiable ML performance metrics; (E-6) pharmacological or bariatric surgical interventions without a dietary component; (E-7) non-English publications or studies predating January 2015.

Criterion I-6 was applied by assigning each included study to a dominant evidence family to avoid conflating evidence types: (a) classification or biomarker-modelling studies, where obesity/BMI classifiers, obesity-relevant microbial signatures, SHAP/feature-importance outputs, or health-stratification outputs provide upstream microbiome stratification relevant to diet-modulable taxa [44]-[50]; (b) dietary-response prediction studies, where ML predicts postprandial glycaemic, lipemic, or metabolite responses that directly inform dietary ranking [51], [52]; and (c) intervention or intervention-adjacent studies, including one small blinded personalised-diet validation within a PPGR cohort [4], one prospective controlled-feeding cohort [5], and one longitudinal weight-loss cohort [53].

E. Data Extraction and Quality Appraisal

A pre-specified, standardised data extraction form was independently applied to each included study by two reviewers, capturing: author(s) and year; study design; sample size and demographic characteristics; ML algorithm(s) evaluated; input data features and sequencing methodology; bioinformatics preprocessing pipeline and reference database; clinical or nutritional outcome; and all reported performance metrics including AUC-ROC, Accuracy,

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Algorithm 1. Reproducible PRISMA 2020-guided systematic-review workflow covering database searching, deduplication, screening, eligibility assessment, data extraction, risk-of-bias appraisal, evidence-family assignment, and narrative synthesis.

Input: Databases D = {PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore}
Output: Set S of included studies and their narrative-synthesis groups
Define PICO, eligibility criteria (I-1..I-7, E-1..E-7) and extraction schema before screening
R ← empty_set *// raw records*
for each database d **in** D **do**
    R ← set_union(R, Search(d, microbiome AND obesity AND ML AND diet; 2015–2026, English, peer-reviewed))
UniqueRecords ← Deduplicate(R) by DOI first, then title/first-author similarity + manual adjudication
T ← TitleAbstractScreen(UniqueRecords) by two reviewers *// exclude off-scope records*
S ← empty_set
for each full-text record f **in** T **do** *// eligibility assessment*
    if f satisfies all I-1..I-7 and no E-1..E-7 **then**
        extract data with the pre-specified form (design, n, algorithm, sequencing, pipeline, outcome, metrics in Eqs. (1)–(11))
        appraise risk of bias: PROBAST (Eq. (12)), NOS, or RoB 2.0 by design
        assign f to a dominant evidence family (classification / response / intervention)
        S ← set_union(S, {f})
    else record structured exclusion reason (Supplementary Table S2)
resolve all reviewer disagreements by consensus + third-reviewer adjudication
narrative synthesis: group S by algorithm class and by primary clinical outcome
return S and synthesis groups

F1-Score, Precision, Recall, Specificity, MCC, RMSE, and Pearson Correlation. Where a study compared multiple ML algorithms, data were extracted for all reported models and their respective benchmarks. Disagreements between reviewers were resolved by consensus discussion, with a third senior reviewer adjudicating unresolved cases.

Methodological quality was assessed using design-appropriate instruments: PROBAST for ML prediction models (four domains: Participants, Predictors, Outcome, Analysis) [54]; Newcastle-Ottawa Scale for observational cohort and cross-sectional studies [55]; and Cochrane RoB 2.0 for RCTs [56]. Two reviewers independently assessed each study; disagreements were resolved by a third. A formal quantitative meta-analysis was not statistically appropriate for five specific reasons: (i) effect sizes not directly comparable across studies (AUC, accuracy, Pearson R, R^2, or variance explained reported variably, with no valid conversion); (ii) confidence intervals absent in 9/12 studies, precluding inverse-variance weighting; (iii) classification thresholds (BMI ≥25, ≥27, ≥30 kg/m²) and obesity case definitions non-harmonisable; (iv) validation schemes heterogeneous (5-fold CV, 10-fold CV, 80/20 split, Leave-One-Dataset-Out, nested CV, external test cohorts) used non-uniformly; (v) outcome definitions ranged from binary obesity status, BMI regression, postprandial glycaemic response, to long-term weight loss with no shared common effect. A structured narrative synthesis was therefore conducted, grouping findings by (a) ML algorithm class (ensemble, deep learning, classical ML) and (b) primary clinical outcome (obesity classification, postprandial-response prediction, weight loss prediction, biomarker identification).

F. Review Workflow

The end-to-end review procedure, from database search to narrative synthesis, is summarised as a reproducible workflow in Algorithm 1. This complements the PRISMA 2020 flow diagram (Fig. 1) by making the deduplication rule, screening order, eligibility logic, extraction schema, quality-appraisal instruments, and synthesis grouping explicit and repeatable.

G. Performance-Metric Definitions and Risk-of-Bias Scoring

To support transparent and reproducible comparison across the heterogeneous metrics reported by the included studies, the performance measures extracted during data extraction are defined below [33], [83]. For a binary obesity/non-obesity classifier, true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are derived from the confusion matrix. Accuracy, Eq. (1) [33], is the proportion of correctly classified cases, but it is unreliable under class imbalance, which is common in obesity cohorts.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Precision, Eq. (2) [33], quantifies the reliability of a positive obesity prediction, while Recall, equivalently Sensitivity or the true-positive rate, Eq. (3) [33], quantifies the proportion of true obese cases that are detected. In a dietary-recommendation context, low precision implies false-positive recommendations and low recall implies missed at-risk individuals [33].

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \text{Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

Specificity, Eq. (4) [33], is the true-negative rate and is essential for evaluating false-positive dietary recommendations; its absence across the included classification-capable studies is a key reporting deficit identified in this review [33]. The F1-score, Eq. (5) [33], is the harmonic mean of precision and recall and is preferred over accuracy for imbalanced obesity datasets [33].

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

The area under the receiver-operating-characteristic curve (AUC-ROC) summarises discrimination across all decision thresholds and equals the probability that a randomly chosen obese case receives a higher predicted score than a randomly chosen non-obese case, Eq. (6) [33], where TPR and FPR denote the true-positive and false-positive rates. An AUC of 0.5 indicates chance-level discrimination and 1.0 indicates perfect discrimination; values around 0.90 reflect strong internal performance, whereas values near 0.60 reflect weak cross-cohort transfer.

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR}) \quad (6)$$

The Matthews correlation coefficient (MCC), Eq. (7) [33], is a balanced single-value summary of the confusion matrix that remains informative under class imbalance [33], ranging from −1 (total disagreement) through 0 (chance) to +1 (perfect agreement).

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

For studies reporting continuous outcomes, such as postprandial glycaemic or metabolite response, regression error was summarised by the root-mean-square error (RMSE), Eq. (8) [83], and the mean absolute error (MAE), Eq. (9) [83], where y_i is the observed value, $\hat{y}_i$ the predicted value, and n the number of observations. RMSE penalises large deviations more heavily than MAE.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (8)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (9)$$

Agreement between predicted and observed continuous responses was reported by several studies as the Pearson correlation coefficient, Eq. (10) [83], where x_i and y_i are paired observed and predicted values with means $\bar{x}$ and $\bar{y}$.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (10)$$

Probabilistic calibration was operationalised through the Brier score, Eq. (11) [33], the mean squared difference between predicted probabilities p_i and observed binary outcomes o_i, where lower values indicate better-calibrated risk estimates. No included study reported a Brier score or calibration curve; this metric is included because TRIPOD+AI lists calibration as a required reporting item for clinical prediction models [33].

$$\text{BS} = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2 \quad (11)$$

Risk-of-bias scoring followed each instrument's published logic rather than an arbitrary numeric cut-off. Under PROBAST, each of the four domains (Participants, Predictors, Outcome, Analysis) was rated low, high, or unclear from its signalling questions; a study was rated low overall only when every domain was low, high overall if any domain was high, and unclear otherwise [54]. The overall PROBAST judgement can therefore be expressed as the worst domain rating, as in Eq. (12) [54], where the ordering is low < unclear < high. Newcastle-Ottawa quality was scored additively across the Selection, Comparability, and Outcome domains to a maximum of nine stars [55], and Cochrane RoB 2.0 judgements were derived domain-wise for the single randomised dietary-validation arm [56].

$$\text{RoB}_{\text{overall}} = \max(d_{\text{Part}}, d_{\text{Pred}}, d_{\text{Out}}, d_{\text{Anal}}) \quad (12)$$

Eqs. (1)–(12) were not used to pool studies, since meta-analysis was not appropriate for the reasons given above; they define the metric vocabulary used to extract, tabulate, and critically appraise the heterogeneous performance values reported by the included studies, and they underpin the metric-coverage matrix presented in Table 4.

III. Results

A. Study Selection and PRISMA Flow

The database search yielded 1,240 raw records (PubMed/MEDLINE 412, Scopus 391, Web of Science 289, IEEE Xplore 148). After removing 434 duplicates, 806 records underwent title/abstract screening; 564 were excluded, leaving 242 full texts assessed for eligibility. Full-text screening excluded 230 records: no ML component despite computational terminology (n = 67), no reportable ML performance metric (n = 58), no dietary, metabolic, BMI, or obesity-relevant linkage (n = 49), review-only design without extractable primary data (n = 43), and other reasons including non-adult, non-gut, non-English, pre-2015, or incomplete-scope records (n = 13). Twelve studies met all criteria and

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

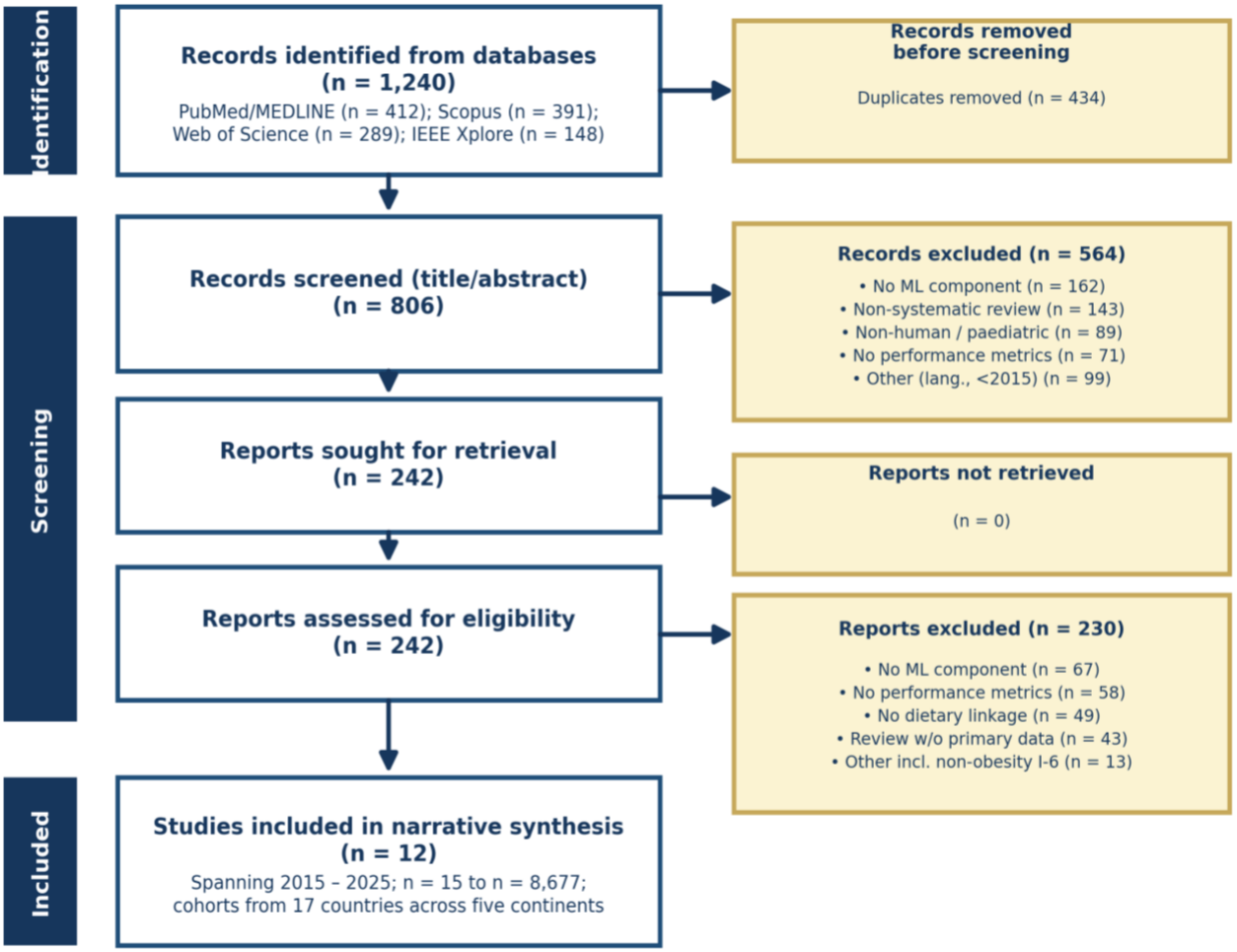


Fig. 1. PRISMA 2020 flow diagram showing study identification, screening, eligibility assessment, and inclusion. From 1,240 records retrieved from four databases, 12 studies were included in the final narrative synthesis.

were included in the final narrative synthesis (Fig. 1). To support transparency and alignment with PRISMA 2020, a study-by-study list of the 230 full-text exclusions, each with its citation and primary exclusion reason categorised into the five exclusion groups reported above (which map to exclusion criteria E-1 to E-7 and to inclusion criterion I-6), is provided in Supplementary Table S2.

B. ML Algorithm Performance Across Included Studies

The 12 included studies are summarised in Table 1. The evidence pool was assigned to dominant evidence families: seven classification or biomarker-modelling studies with obesity or dietary-stratification relevance [44]-[50], two dietary or postprandial-response prediction studies [51], [52], one small blinded personalised-diet validation within the Zeevi PPGR cohort [4], and one prospective controlled-feeding Wang et al. [51] was counted as response-prediction evidence because its primary contribution was metabolite-response modelling, although the model used pooled dietary-intervention source datasets. Some studies span more than one translational function, but each was grouped by its dominant evidentiary role. Sample sizes ranged from n = 15 [53] to n = 8,677 [44]. Profiling methods included 16S rRNA, WGS metagenomics, and metatranscriptomics, which produce distinct feature spaces that require careful harmonisation [57], [58].

1. Ensemble Methods: XGBoost and Random Forest

Ensemble methods appeared in 9/12 studies. XGBoost was the most consistently top-ranked classifier within the studies that compared it, but this should be interpreted as within-study comparative evidence rather than general dominance because datasets, preprocessing pipelines, feature spaces, and validation schemes differed [22], [45], [47], [50]. Wu et al. [45]

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

reported the highest internal discrimination (AUC 0.90; accuracy 0.814; precision 0.819; recall 0.848; F1 0.833) using GMrepo-derived 16S data and SHAP interpretation. SHAP identified negative-BMI taxa including Akkermansia muciniphila, Alistipes onderdonkii, and Bifidobacterium pseudocatenulatum; downstream experiments supported B. pseudocatenulatum as a mechanistically plausible probiotic candidate [45]. Liu Y et al. [47] reported AUC 0.85 across a five-country WGS framework and used deep reinforcement learning to propose minimum-intervention microbiome modulation strategies.

Table 1. Characteristics of the 12 included studies, including study design, sample characteristics, machine-learning algorithms, microbiome sequencing methods, clinical outcomes, and reported performance metrics.

Author (Year)	Design & Sample	ML Algorithm; Sequencing Method	Clinical Output / I-6 Basis	Performance Metrics
Wu 2025 [45]	Retrospective; n = 3,606 (1,563 healthy + 2,043 overweight); BMI ≥25 (GMrepo)	XGBoost, SVM, LR, DT; 16S rRNA (GMrepo DB)	Classification; SHAP taxa and probiotic validation (upstream dietary-stratification evidence)	AUC = 0.90; Acc = 0.814; Precision = 0.819; Recall = 0.848; F1 = 0.833
Liu W 2022 [46]	Cross-sectional; n = 2,262 adults	SVM-RBF, RF, others; WGS metagenomics	BMI classification/regression; microbiome-based obesity stratification (upstream evidence)	Acc = 0.716; R^2 = 0.485; AUC: NR
Liu Y 2022 [47]	Multi-cohort; n = 870; 5 countries	XGBoost + Deep RL; WGS shotgun	Classification plus DRL personalised modulation strategy (computational dietary/probiotic recommendation)	AUC = 0.85
Wang T 2025 [51]	Multi-study DL; 6 dietary intervention studies, including overweight/obesity trial	McMLP (DL); 16S rRNA + metabolomics	Metabolite response prediction after dietary interventions (direct response-prediction evidence; not obesity-only RCT)	Benchmarked against LASSO/Ridge/RF for metabolite-response prediction; no binary discrimination AUC because the primary outcome was metabolite response.
Yang 2025 [44]	Large-scale; n = 8,677; 65 cohorts	KPCA-Contrastive Learning (DL); WGS metagenomics	Gut-health classification linked to FGMDI dietary recommendations (computational recommendation evidence with obesity/dietary-stratification relevance)	AUC = 0.861; Acc = 0.791
Chanda & De 2024 [48]	Meta-analysis; n = 3,329; 17 countries	RF + MetAML; 16S rRNA + WGS	Obesity signatures/pathways; dietary-stratification relevance (upstream evidence)	CV AUC median = 0.67 (range 0.57–0.88); CSV median = 0.60; independent validation AUC = 0.85 (species), 0.80 (pathway)
Bischoff 2022 [53]	Longitudinal; n = 15; BMI = 43.1	LASSO, Ridge, Elastic Net; WGS metagenomics	Long-term weight-loss prediction after nonsurgical therapy (weight-management outcome)	7 baseline genus-level predictors and 8 species-level maintenance predictors identified; no standard discrimination/calibration metric reported.

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 1. (continued)

Author (Year)	Design & Sample	ML Algorithm; Sequencing Method	Clinical Output / I-6 Basis	Performance Metrics
Gong 2022 [49]	Meta-analysis; n = 4,282 (2,236 controls + 1,152 overweight + 894 obesity); 8 source datasets	Random Forest; 16S V3-V4 DADA2	Obesity biomarkers/pathways; diet-modulable signatures (upstream evidence)	AUC = 0.83
Tily 2022 [52]	Adult PPGR cohort; n = 550 adults (US; 66% female; not obesity-only)	Mixed-effects LR + ML ensemble; Metatranscriptomics (Viome)	PPGR prediction translated to dietary recommendations (direct response-output evidence)	Pearson R = 0.77 for PPGR prediction; no classification AUC because the primary outcome was continuous response.
Lin 2023 [50]	Meta-analysis; 18 studies	XGBoost, RF, Neural Network; 16S rRNA; 18 studies	Cross-study obesity signatures; classification benchmark (upstream evidence)	AUC median = 0.608 (10-fold CV across all 18 studies); 0.771 (8 obesity-specific subset)
Zeevi 2015 [4]	Cohort + small blinded dietary validation; n = 800 + 100 + 26	Gradient Boosting; WGS + CGM + diet	PPGR prediction and small blinded personalised-diet validation (direct response evidence)	R = 0.68; significant PPGR reduction in blinded dietary validation
Berry 2020 [5]	PREDICT 1; n = 1,002 adults/twins	ML ensemble (multi-model); WGS + CGM + SNPs	PPGR/TG prediction in prospective controlled-feeding cohort (intervention-adjacent evidence)	Variance explained = 46%; AUC = 0.73–0.77

**NR = Not Reported; NA = Not Applicable to the primary task; AUC = Area Under ROC Curve; Acc = Accuracy; PPGR = Postprandial Glycaemic Response; DL = Deep Learning; RL = Reinforcement Learning; CL = Contrastive Learning; KPCA = Kernel PCA; CV = Cross-Validation; CSV = Cross-Study Validation; WGS = Whole-Genome Shotgun metagenomics; CGM = Continuous Glucose Monitoring; TG = Triglyceridaemia; DT = Decision Tree; LR = Logistic Regression. To clarify cross-study metric availability and support the metric-reporting interpretation in Section III.B.4, AUC-ROC was reported in 7/12 studies [5], [44], [45], [47]-[50], Accuracy in 3/12 [44]-[46], and F1-Score, Precision, and Recall/Sensitivity in only one of seven classification-capable studies [45]. Specificity was not reported in any classification-capable study, while calibration plots or Brier scores were absent across all 12 studies. Studies reporting alternative metrics - Pearson R [4], [5], [52], R^2 [46], or qualitative model ranking [51] - are flagged in the Performance Metrics column. This explicit reporting matrix substantiates the metric-reporting deficit highlighted in the Results and Discussion sections.*

Random Forest/MetAML showed variable secondary performance across pooled and meta-analytic datasets. Gong et al. [49] reported an AUC of 0.83, while Chanda and De [48] reported a median within-dataset AUC of 0.67, falling to 0.60 under cross-study validation, with partial recovery to AUC 0.85 for species features under independent robust-signature validation. Lin et al. [50] reported a median AUC of 0.608 across 18 harmonised 16S studies, rising to 0.771 in the obesity-focused subset. In each of these multi-cohort studies, the cross-study or leave-one-dataset-out estimate was lower than the corresponding within-dataset estimate (the interpretation of this pattern for deployment is considered in Section IV).

2. Deep Learning Architectures

Deep learning was less frequent but useful for tasks beyond static classification. Wang et al. [51] developed McMLP for food-microbe-metabolite response prediction across six dietary intervention studies. Yang et al. [44] applied KPCA-contrastive learning to 8,677 metagenomic samples from 65 cohorts, reaching AUC 0.861 and Accuracy 0.791 for gut-health classification

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

linked to FGMDI dietary recommendations. Neither study reported calibration metrics or a prospective clinical-deployment test.

3. Classical ML: SVM, Regularised Regression, Mixed-Effects Models

Classical ML methods retained value where interpretability or small-sample modelling was prioritised. Liu W et al. [46] reported SVM-RBF accuracy of 0.716 and R^2 of 0.485 for BMI-related prediction. Bischoff et al. [53] used regularised regression in a 24-month cohort of 15 adults with severe obesity, identifying baseline microbial predictors of long-term weight-loss success. Tily et al. [52] reported Pearson R = 0.77 for PPGR prediction using metatranscriptomic data, linking functional microbiome activity to dietary ranking.

4. Cross-Cohort Generalisation Gap and Metric-Reporting Deficit

Across the 12-study evidence base, reported discrimination estimates ranged from peak internal AUC 0.90 in a single-cohort setting [45] to approximately 0.60 in cross-study/meta-analytic benchmarks [48], [50]. These values come from different studies and validation designs; no single model was observed to decline directly from AUC 0.90 to AUC 0.60. Metric reporting was incomplete: specificity was not reported in any classification-capable study, calibration/Brier score was absent in all 12 studies, and AUC-ROC was available in 7/12 studies. The five non-AUC studies reported non-AUC classification/regression metrics, response-prediction correlations, biomarker outputs, or qualitative model ranking rather than a shared AUC endpoint. F1, precision, and sensitivity were each reported in only one of seven classification-capable studies. These omissions prevented extraction of false-positive rates and calibrated operating thresholds from the published reports.

C. Clinical Outcomes in Obesity

1. Obesity/BMI Classification and Obesity-Relevant Microbiome Stratification

Classification or biomarker-modelling with obesity or dietary-stratification relevance was the dominant evidence family (7/12 studies; Section III.B). Wu et al. [45] reported the highest obesity-classification AUC (AUC = 0.90, SHAP-based feature importance, in vivo + in vitro probiotic confirmation). Liu W et al. [46] reached an accuracy of 0.716 across 2,262 WGS-profiled Chinese adults. Liu Y et al. [47] achieved an AUC of 0.85 in cross-country validation across five nations, identifying 42 geographically conserved species signatures. Yang et al. [44] reached AUC = 0.861 in a large gut-health classification framework linked to dietary recommendations and was retained as computational recommendation evidence rather than as a direct obesity-treatment trial. Chanda and De [48] reported CV median AUC 0.67, CSV median AUC 0.60, and independent external validation AUC 0.85 for a pooled species-level robust signature.

2. Personalised Dietary Recommendations and Postprandial Response Prediction

Two studies reported prospective or intervention-adjacent dietary-response evidence. Zeevi et al. [4] measured 46,898 postprandial glycaemic responses (PPGRs) via 7-day continuous glucose monitoring (CGM) in 800 participants, using Gradient Boosting on WGS microbiota plus blood parameters, dietary logs, physical activity, and anthropometric measurements to predict personalised PPGRs at Pearson R = 0.68 (replicated at R = 0.68 in a 100-person validation cohort). A small blinded personalised-diet validation (n = 26) showed significantly lower PPGRs on algorithm-generated favourable diets than on algorithm-identified unfavourable diets [4].

Berry et al. [5] extended this in PREDICT 1 (n = 1,002 UK/US twins). Multi-omics modelling showed gut microbiome composition explained 7.1% of postprandial triglyceridaemia and 6.0% of glycaemic variance beyond meal macronutrients (3.6%), with the full multi-omics model explaining 46% of glycaemic variation [5].

3. Weight Loss Prediction and Biomarker Identification

Three types of weight-management-related evidence were observed across Bischoff et al. [53], Liu Y et al. [47], and Wu et al. [45]: baseline microbiota biomarkers associated with weight-loss success, computational strategies for microbiota modulation, and experimental validation of a candidate taxon.

Bischoff et al. [53] showed baseline microbiota profiles enriched in Akkermansia, Alistipes, Pseudoflavonifractor, and Symbiobacterium (four of seven weight-loss-predictive genera identified) distinguished participants achieving long-term weight-loss success (a priori defined as relative weight loss ≥10% at 24 months; the persistently successful subgroup achieved a mean relative weight loss of 18.2% versus 0.2% in the non-successful subgroup). Liu Y et al. [47] reported a DRL modulation framework identifying minimum dietary or probiotic interventions across five geographic backgrounds. Wu et al. [45] validated Bifidobacterium pseudocatenulatum through in vivo inhibition of lipid accumulation in C. elegans and in vitro suppression of C3H10T1/2 adipocyte differentiation.

D. Risk of Bias Assessment

Two reviewers independently appraised the 12 studies using design-appropriate instruments, with disagreements resolved by consensus and third-reviewer adjudication. PROBAST was applied to all 12 as prediction-model evidence; the Newcastle-Ottawa Scale to 4 observational studies [5], [46], [52], [53]; and Cochrane RoB 2.0 was applied to the Zeevi personalised-diet validation [4]. RoB 2.0 was not

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 2. Risk-of-bias and methodological-quality assessment of the 12 included studies using PROBAST, the Newcastle–Ottawa Scale, and Cochrane RoB 2.0, as applicable.

Study	PROBAST (Overall)	NOS (Score/9)	RoB 2.0 / source-trial considerations
Wu 2025 [45]	Low	NA	NA
Liu W 2022 [46]	Unclear	7/9	NA
Liu Y 2022 [47]	Low	NA	NA
Wang T 2025 [51]	High (Part.)	NA	RoB 2.0 not directly applicable; source-dataset heterogeneity noted narratively
Yang 2025 [44]	Low	NA	NA
Chanda & De 2024 [48]	Unclear	NA	NA
Bischoff 2022 [53]	High (Part.)	6/9	NA
Gong 2022 [49]	Unclear	NA	NA
Tily 2022 [52]	Unclear	7/9	NA
Lin 2023 [50]	Unclear	NA	NA
Zeevi 2015 [4]	Low	NA	Low
Berry 2020 [5]	Low	8/9	NA

**NA = Not Applicable (instrument not indicated for this study design). NOS maximum possible score = 9. PROBAST overall rating combines assessments across Participants, Predictors, Outcome, and Analysis domains. The RoB 2.0 rating for Zeevi et al. [4] refers to the small blinded personalised-diet validation arm within the PPGR cohort. RoB 2.0 was not applied directly to Wang et al. [51] because it is a prediction-model paper based on pooled dietary-intervention datasets; source-dataset heterogeneity was considered narratively.*

applied directly to Wang et al. [51] as a prediction-model paper; heterogeneity in its pooled dietary-intervention source datasets was noted narratively. Summary results are in Table 2.

PROBAST across its four domains showed 5 studies [4], [5], [44], [45], [47] at low overall risk (adequate sample size, pre-specified outcomes, appropriate cross-validation, transparent preprocessing reporting); 5 studies [46], [48], [49], [50], [52] at unclear/moderate risk driven by the Analysis domain (insufficient cross-validation partitioning reporting risking data leakage, absent calibration, incomplete class-imbalance handling); and 2 studies [51], [53] at high risk in Participants - Wang [51] because heterogeneity across six independently designed dietary intervention cohorts constrained generalisability, and Bischoff [53] because n = 15 limited model stability. Domain-level PROBAST ratings (Participants/Predictors/Outcome/Analysis) for each study are shown in Table 5. Newcastle-Ottawa scores for the 4 observational studies ranged 6-8/9, with adequate Selection and Outcome but variable Comparability. Cochrane RoB 2.0 was applied directly to Zeevi et al. [4], which was rated Low. Five of twelve studies [4], [5], [44], [45], [47] met combined low-risk criteria across applicable instruments.

Across the four PROBAST domains, Analysis was rated Unclear in 7/12 studies, the most frequent driver of moderate-risk classifications. Recurring weaknesses (cross-validation partitioning, calibration, pipeline isolation) are precisely what TRIPOD+AI is designed to mandate. Read directly from the standardised reporting matrix (Table 4), AUC-ROC was the only metric reported by a majority of studies (7/12), accuracy was reported by 3/12, and external validation on an independent cohort or via leave-one-cohort-out was reported by 5/12; among the seven classification-capable studies, sensitivity, F1, and precision were each reported by only one (1/7), specificity by none (0/7), and no study among all twelve reported a calibration curve or Brier score (0/12). This row-by-row pattern, rather than any single low score, is the quantitative basis for the reporting-deficit conclusion drawn throughout the Results and Discussion.

IV. Discussion

A. Interpretation of Results

This review indicates that XGBoost-based classifiers currently have the strongest support for static obesity/BMI classification within the included comparative evidence, with SHAP improving interpretability in the best-performing study. No model should be treated as universally preferred. The

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

strongest result (AUC 0.90) came from internal validation in a single large public-database setting [45], whereas cross-study benchmarks were substantially lower [48], [50]. The appropriate interpretation is therefore methodological: model choice must be paired with leakage-safe preprocessing, external validation, calibration, and transparent reporting. The consistent observation that cross-study and leave-one-dataset-out estimates fell below within-dataset estimates across the meta-analytic studies [48], [50] implies that, for deployment, cross-cohort study design and harmonised obesity case definitions matter more than an isolated peak AUC obtained in a single cohort. A model selected on internal performance alone is therefore likely to overstate its real-world discrimination.

The combination of tree-based classification and SHAP explanations is valuable because it links patient-level predictions to taxa-level contributions and enables mechanistic follow-up [45], [77], [79]. However, post-hoc interpretability does not establish causality or clinical safety. Feature attribution should therefore not be reported in isolation. It should be accompanied by calibration, external validation, and decision-curve analysis, and, where possible, by experimental or prospective dietary validation [33], [78]-[81]. Deep learning appears most useful for multi-omics, metabolite-response, and temporal or nonlinear

Table 3. Domain-level PROBAST risk-of-bias assessment for the 12 included studies across Participants, Predictors, Outcome, and Analysis domains.

Study	Participants	Predictors	Outcome	Analysis	Notes on weakest domain(s)
Wu 2025 [45]	L	L	L	L	All domains are low risk; the strongest classification metric reporting in pool (5/6 metrics + SHAP); specificity remains unreported
Liu W 2022 [46]	L	L	L	U	Analysis: limited CV-fold preprocessing isolation reporting
Liu Y 2022 [47]	L	L	L	L	Multi-cohort design with external validation
Wang T 2025 [51]	H	L	L	U	Participants: 6 independent dietary intervention cohorts with non-uniform inclusion; Analysis: heterogeneity-pooling justification incomplete
Yang 2025 [44]	L	L	L	L	Largest pooled cohort; appropriate CV strategy
Chanda & De 2024 [48]	L	U	L	U	Predictors: heterogeneous reference-database versions across pooled cohorts; Analysis: cross-cohort harmonisation not pre-specified
Bischoff 2022 [53]	H	L	L	U	Participants: n = 15 limits model stability; Analysis: small-sample CV reporting incomplete
Gong 2022 [49]	L	U	L	U	Predictors: 16S V-region heterogeneity across pooled datasets; Analysis: absent calibration
Tily 2022 [52]	L	U	L	U	Predictors: proprietary metatranscriptomic platform limits independent reproducibility; Analysis: no calibration / decision-curve analysis
Lin 2023 [50]	L	U	L	U	Predictors: 18-study heterogeneity in sequencing and processing; Analysis: CV partitioning not fully documented
Zeevi 2015 [4]	L	L	L	L	Small blinded dietary validation arm; reports Pearson R and prospective dietary response rather than standard classification metrics
Berry 2020 [5]	L	L	L	L	Twin design controls genetic confounding; large cohort

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 4. Standardised performance-metric reporting matrix across the 12 included studies.

Study	AUC	Acc	Sens	Spec	F1	Prec	Ext. Val
Wu 2025 [45]	✓	✓	✓	NR	✓	✓	NR
Liu W 2022 [46]	NR	✓	NR	NR	NR	NR	NR
Liu Y 2022 [47]	✓	NR	NR	NR	NR	NR	✓
Wang T 2025 [51]	NA	NA	NA	NA	NA	NA	NR
Yang 2025 [44]	✓	✓	NR	NR	NR	NR	NR
Chanda & De 2024 [48]	✓	NR	NR	NR	NR	NR	✓
Bischoff 2022 [53]	NA	NA	NA	NA	NA	NA	NR
Gong 2022 [49]	✓	NR	NR	NR	NR	NR	NR
Tily 2022 [52]	NA	NA	NA	NA	NA	NA	NR
Lin 2023 [50]	✓	NR	NR	NR	NR	NR	✓
Zeevi 2015 [4]	NA	NA	NA	NA	NA	NA	✓
Berry 2020 [5]	✓	NR	NA	NA	NA	NA	✓
Total reported (n/12)	7/12	3/12	1/7 appl.	0/7 appl.	1/7 appl.	1/7 appl.	5/12

**Note. The matrix is built from explicit metric values disclosed in each primary publication. NR = not reported; NA = not applicable to the primary task because the outcome was continuous response prediction, metabolite-response modelling, or longitudinal biomarker discovery rather than threshold-based classification. Specificity was not reported for any classification-capable study (0 of 7), and calibration/Brier score was reported in 0/12 studies. AUC-ROC is available in 7/12 studies; among classification-capable studies, F1, Precision, and Sensitivity are each reported in only one study; and external validation on an independent cohort or leave-one-cohort-out scheme is reported in 5/12. These deficits support the metric-reporting interpretation in Section III.B.4.*

phenotype modelling, rather than as a default replacement for tree-based classifiers [29]-[32], [51], [82], [98]-[100]. Its role in microbiome-guided nutrition should be evaluated against simpler baselines under nested cross-validation and external test cohorts, because sparse compositional datasets remain vulnerable to overfitting.

B. Comparison with Previous Studies

This synthesis differs from prior precision-nutrition and microbiome-ML reviews by restricting the included evidence to adult obesity/BMI-stratified or obesity-relevant dietary-response microbiome ML studies and by explicitly separating classification benchmarks from dietary-response and intervention evidence [30], [31], [36]-[42]. The low cross-study AUCs observed here are consistent with broader microbiome-ML benchmarks, where obesity is among the more difficult phenotypes to classify reproducibly across cohorts [17], [18]. A direct comparison of the five most informative classification studies clarifies why peak accuracy and deployable accuracy diverge. Wu et al. [45] achieved the highest internal discrimination (AUC 0.90; accuracy 0.814) but used a single retrospective public-database cohort (n = 3,606) with an internal split rather than an external test set, so the value is an upper bound under favourable conditions. Yang et al. [44] reported a slightly lower but more robust AUC of 0.861 (accuracy 0.791) using by far the largest pool (n = 8,677 across 65 cohorts) and contrastive deep learning, indicating that scale and cohort diversity cost little discrimination. Liu Y et al. [47] reported AUC 0.85 with explicit five-country external validation (n = 870), the most credible cross-population estimate in the pool because it was not derived from a single-cohort split. By contrast, the two meta-analytic studies that enforced cross-study validation reported substantially lower values: Chanda and De [48] fell from a within-dataset median AUC of 0.67 to 0.60 under cross-study validation (n = 3,329; 17 countries), recovering to 0.85 only for a curated species-level robust signature, and Lin et al. [50] reported a 10-fold cross-validation median AUC of 0.608 across 18 studies, rising to 0.771 in the obesity-specific subset. Read together, validation design not algorithm family or headline AUC is the dominant determinant of reported performance: single-cohort internal splits cluster near 0.85–0.90, whereas leave-one-dataset-out or pooled cross-study designs cluster near 0.60–0.77 for the same task.

The strongest clinical translation evidence remains the PPGR work of Zeevi et al. [4] and Berry et al. [5]. Zeevi et al. combined WGS microbiome, CGM, diet, activity, and clinical data and tested algorithm-derived

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

diets in a small blinded personalised-diet validation. Berry et al. extended this to a larger multi-omics cohort and showed that microbiome features add predictive information beyond macronutrients. These studies support early efficacy signals but are insufficient to establish broad obesity-treatment CDSS effectiveness. The generalisation gap also parallels evidence from broader microbiome disease prediction and multi-cohort harmonisation studies [17], [18], [57], [58], [60]. None of the included studies fully resolved batch effects, compositionality, and database-version heterogeneity prospectively, making cross-cohort validation a minimum requirement for future microbiome-ML obesity studies.

Table 5. Recurrent bioinformatics pipeline challenges in microbiome-based machine-learning studies and recommended mitigation strategies.

Challenge	Mitigation Strategy	Key References
1. Compositionality bias of OTU/ASV relative abundances	Apply centred log-ratio (CLR) or additive log-ratio transformation; consider arcsine-contrast methods under high zero-inflation	[53], [60], [62], [14], [15], [20]
2. Cross-study batch effects	Apply ComBat or MMUPHin harmonisation; report Leave-One-Dataset-Out (LODO) validation as the minimum cross-cohort benchmark	[17], [48], [50]
3. Reference database inconsistency (SILVA, RDP, Greengenes, GTDB)	Adopt species-level genome bin (SGB)–based profilers such as MetaPhlAn 4 and version-pin reference databases per study	[57], [58], [64], [65], [66], [67]
4. Data leakage during preprocessing	Fit normalisation, feature selection, and dimensionality reduction inside cross-validation folds; pre-register the pipeline	[33]
5. Class imbalance (obese vs healthy-weight)	Apply stratified cross-validation, SMOTE oversampling, or class-weight adjustment; report minority-class Recall and F1	[45], [33], [68]
6. Small per-site sample sizes	Adopt federated learning consortia for multi-site training without centralising patient data	[69], [70], [71], [72], [73], [74], [75], [76]
7. Absent calibration and decision-curve analyses	Report calibration plots, Brier scores, and decision-curve analysis per TRIPOD+AI guidance	[33]
8. Cross-sectional, single-time-point designs	Adopt longitudinal sampling and temporal-aware deep-learning architectures (CNN-LSTM, Transformer-based sequence models)	[53]
9. Inadequate performance metric reporting (Sensitivity, Specificity, AUC)	Mandate TRIPOD+AI checklist at journal submission; report full confusion matrix and discrimination metrics	[43], [33]
10. Opaque feature attribution (black-box ML)	Integrate SHAP, LIME, and counterfactual explanation frameworks; pair ML predictions with mechanistic experimental validation	[77], [45], [47], [78], [79]

C. Limitations and Weaknesses

The limitations of this work fall into four groups: review-level, included-study, methodological, and clinical-translation limitations.

1. Review-level limitations.

A formal meta-analysis was not appropriate for the five reasons specified in Section II.E: incompatible outcomes and metrics; confidence intervals absent in 9/12 studies; non-harmonisable classification thresholds; heterogeneous validation schemes; and outcomes ranging from binary obesity status to PPGR and long-term weight loss. The review was not prospectively registered, which limits protocol transparency despite pre-specified PICO and extraction procedures. The search did not include Embase, Cochrane Library, trial registries, preprint servers, or grey-literature sources. Formal inter-rater agreement statistics were not calculated, although disagreements were resolved by consensus and third-reviewer adjudication.

2. Included-study limitations

The included studies used heterogeneous sequencing platforms, bioinformatics pipelines, and reference

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

databases, including 16S rRNA, WGS, metatranscriptomics, DADA2, QIIME2, MetaPhlAn, and bioBakery workflows [57], [58], [64]-[67]. In addition, 11/12 studies used single-time-point profiling, limiting inference about dietary dynamics.

3. Methodological limitations

Compositional-data handling was inadequately reported. CLR transformation could be unambiguously confirmed in only 1/12 studies [53], and CLR alone is not sufficient unless zero replacement, sequencing-depth normalisation, sparsity thresholds, and sensitivity analyses are pre-specified [14], [15], [20], [60], [62], [63]. Data leakage remains a central reliability risk. Global normalisation, prevalence filtering, feature selection, dimensionality reduction, batch-effect harmonisation, class balancing, or hyperparameter tuning performed before splitting can inflate internal performance. Future studies should fit every data-dependent preprocessing step inside each training fold, report fold-level pipeline isolation explicitly, and include at least one external test cohort never used in preprocessing or model selection [33], [68].

4. Clinical-translation limitations.

Finally, the body of evidence is not yet sufficient for clinical deployment. Only one small blinded dietary-validation arm (n = 26 [4]) and one prospective controlled-feeding cohort [5] tested algorithm-derived dietary guidance prospectively; no included study reported specificity or calibration, no study performed decision-curve analysis, and training cohorts overrepresented high-income populations, so subgroup performance, false-positive dietary-recommendation rates, and operating-threshold safety remain unquantified. These constraints mean the present synthesis can establish methodological readiness but not clinical-deployment readiness. Synthesising these limitations, ten recurrent bioinformatics pipeline challenges emerge as priority targets for next-generation microbiome-ML obesity research, with evidence-based mitigation strategies summarised in Table 5.

D. Implications for Clinical Practice and CDSS

The clinical evidence supports cautious translation rather than immediate deployment. Zeevi et al. [4] provide a small blinded personalised-diet validation within a PPGR cohort, while Berry et al. [5] provide strong prospective controlled-feeding evidence. These findings support an early signal of efficacy for microbiome-informed dietary ranking, but safety, calibration, reproducibility across populations, and prospective workflow integration remain unproven.

CDSS architecture recommendations should therefore be treated as implementation considerations rather than empirical findings from the 12 included studies. Future systems will require standardised BIOM-like input schemas, version-controlled preprocessing, locked or regulator-approved adaptive models, interpretable output, uncertainty communication, and post-deployment monitoring [33]-[35], [61], [83], [84]. Federated learning is relevant because microbiome cohorts are small and geographically biased, but privacy, fairness, governance, and secure computation remain unresolved implementation issues [69]-[76], [85], [86]. Regulatory-facing translation should define minimum external-validation cohorts, sensitivity/specificity floors, calibration targets, and algorithm-change protocols before microbiome-guided dietary CDSS can be evaluated as software as a medical device. Current included studies do not meet this evidentiary threshold, so the appropriate conclusion is methodological readiness, not clinical deployment readiness. Equity is also central. Gut microbiota varies with geography, diet, socioeconomic context, medication exposure, and environment, while current training cohorts overrepresent high-income settings [87]-[90]. Globally representative federated consortia and harmonised metadata standards are therefore needed to avoid dietary recommendations that generalise poorly to underrepresented populations.

E. Future Research Directions

Five priorities follow directly from the evidence gaps: (1) longitudinal multisite WGS cohorts with CGM and standardised diet records; (2) compositional-data-aware preprocessing with zero-handling and sensitivity analyses; (3) multi-omics modelling that integrates metagenomics, metabolomics, host genetics, and digital phenotyping; (4) TRIPOD+AI-aligned reporting of discrimination, calibration, confusion matrices, and external validation; and (5) privacy-preserving international consortia for representative training and prospective testing [5], [33], [43], [57]-[76], [91]-[100].

F. Minimum Reporting Standards for Microbiome-ML Obesity Studies

Future primary studies and systematic review updates should make each candidate microbiome-ML study traceable to an evidence level: classification-only, response prediction, intervention-adjacent, or randomised dietary intervention evidence. This distinction is essential because obesity classification models can support stratification but do not by themselves prove that an ML-derived diet improves clinical outcomes. For reproducibility, full database strings, screening counts, deduplication rules, full-text exclusion reasons, extraction sheets, and risk-of-bias judgments should be supplied as supplementary files. In addition, the main manuscript should report included studies separately from background, methodological, CDSS, and regulatory references [4], [5], [33], [43]-[56].

For primary studies in this field, the minimum reporting panel should include AUC-ROC,

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

sensitivity/recall, specificity, precision, F1-score, confusion matrix, calibration plot or Brier score, decision-curve analysis, external validation, preprocessing order, and whether every data-dependent step was refitted inside each cross-validation fold. Microbiome-specific reporting should further specify three things. First, the data should be characterised by zero-handling, sequencing-depth distributions, and rare-feature filtering. Second, the compositional treatment should be stated, namely the log-ratio or alternative transformation and the reference-database version. Third, a sensitivity analysis should be reported across at least two preprocessing pipelines [14], [15], [20], [33], [57]-[68]. These requirements are not cosmetic; they determine whether high internal AUC reflects biological signal or preventable leakage and batch effects. For CDSS translation, the evidence threshold should be higher than publication-level prediction performance. A clinically interpretable microbiome-nutrition CDSS should report the operating threshold chosen for recommendation, the false-positive and false-negative consequences of that threshold, calibration drift across populations, subgroup performance across geography and diet pattern, and model-version governance. Federated learning, privacy-preserving computation, and post-deployment monitoring may address multi-site data limitations, but they cannot replace prospective dietary trials or representative validation cohorts [69]-[86]. Accordingly, this review treats current models as promising methodological prototypes rather than clinically deployable systems.

V. Conclusion

This PRISMA 2020-guided systematic review of 12 studies supports three conclusions. The review set out to evaluate machine-learning models that use gut microbiome data for obesity-relevant classification, dietary-response prediction, and personalised nutrition, and to judge their readiness for clinical decision support. From 1,240 records, 12 studies met the criteria and were grouped into seven classification or biomarker-modelling studies, two dietary or postprandial-response prediction studies, and three intervention or intervention-adjacent studies. First, XGBoost-based classifiers currently have the strongest support for static gut microbiome-based obesity/BMI classification within heterogeneous comparative evidence, with SHAP improving interpretability in the best-performing study; Random Forest/MetAML and selected deep-learning models show value for secondary classification, metabolite-response, and large-scale phenotype tasks. Second, the main barrier to CDSS translation is not algorithm availability but reproducibility: published estimates range from a peak internal AUC of 0.90 in a single-cohort setting down to cross-study/meta-analytic benchmarks near 0.60, with specificity reported in none of the seven classification-capable studies (0/7) and calibration absent from all 12 included studies. Third, future microbiome-guided personalised-nutrition studies should require leakage-safe preprocessing, compositional-data sensitivity analysis, external validation, full metric reporting, prospective dietary testing, and representative multisite training before clinical deployment is considered. Until prospective dietary trials with complete discrimination and calibration reporting are available, microbiome-ML models for obesity should be regarded as promising methodological prototypes rather than clinically deployable decision-support tools.

Acknowledgments

The authors acknowledge Universitas Karya Husada, Semarang, for providing access to research databases and computational resources used during this systematic review.

Funding

This work was supported by a Publication Grant from LLDIKTI VI, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, Contract No. 031/LL6/AL.04/BP/2025. The funding body had no role in the study design, literature search, screening, data extraction, interpretation of results, manuscript preparation, or decision to submit the article.

Data Availability

No primary datasets were generated or analysed in this systematic review. Search strategies, screening records, full-text exclusion reasons, eligibility assessments, risk-of-bias evaluations, and metric-reporting summaries are available in the Supplementary Materials. Risk-of-bias details are reported in Tables 2 and 3, while metric-reporting completeness is summarised in Table 4. The PRISMA 2020 checklist is provided as a separate supplementary file. All included evidence derives from previously published peer-reviewed studies.

Author Contributions

Fitra Adi Prayogo: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Supervision. Isy Royhanaty: Methodology, Investigation, Data Curation, Validation, Writing – Review & Editing. Poppy Fransisca Amelia: Investigation, Validation, Data Curation, Writing – Review & Editing. Heni Wijayanti: Investigation, Data Curation, Validation, Writing – Review & Editing. Egy Sunanda Putra: Investigation, Validation, Interpretation of Nutritional Outcomes, Writing – Review & Editing. All authors approved the final manuscript.

Declarations

Ethics Approval and Consent to Participate: Not applicable. This systematic review used data

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

exclusively from previously published studies and did not involve recruitment of human participants, collection of identifiable personal information, or access to individual-level patient data.

Consent for Publication: Not applicable. The manuscript does not contain identifiable individual participant data, images, or personal information.

Competing Interests: The authors declare that they have no financial or non-financial competing interests that could have influenced the conduct, interpretation, or reporting of this study.

Protocol Registration: The systematic review protocol was not prospectively registered in PROSPERO, OSF, or another registry. The eligibility criteria, PICO framework, extraction schema, and risk-of-bias assessment procedures were defined before screening.

Use of Generative AI and AI-Assisted Technologies: Generative AI-assisted tools were used only for English-language editing, structural refinement, and consistency checking. They were not used for study selection, eligibility assessment, data extraction, statistical analysis, reference fabrication, or final scientific interpretation. All AI-assisted content was reviewed and approved by the authors, who take full responsibility for the manuscript.

References

[1] World Health Organization (WHO), "Obesity and overweight," Fact Sheet, Geneva, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [Accessed: May 2026].

[2] Y. C. Chooi, C. Ding, and F. Magkos, "The epidemiology of obesity," Metabolism, vol. 92, pp. 6–10, Mar. 2019, doi: 10.1016/j.metabol.2018.09.005.

[3] K. D. Hall and S. Kahan, "Maintenance of lost weight and long-term management of obesity," Med. Clin. North Am., vol. 102, no. 1, pp. 183–197, Jan. 2018, doi: 10.1016/j.mcna.2017.08.012.

[4] 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.

[5] S. E. Berry et al., "Human postprandial responses to food and potential for precision nutrition," Nat. Med., vol. 26, no. 6, pp. 964–973, Jun. 2020, doi: 10.1038/s41591-020-0934-0.

[6] P. J. Turnbaugh et al., "An obesity-associated gut microbiome with increased capacity for energy harvest," Nature, vol. 444, no. 7122, pp. 1027–1031, Dec. 2006, doi: 10.1038/nature05414.

[7] R. E. Ley, P. J. Turnbaugh, S. Klein, and J. I. Gordon, "Microbial ecology: human gut microbes associated with obesity," Nature, vol. 444, no. 7122, pp. 1022–1023, Dec. 2006, doi: 10.1038/4441022a.

[8] P. D. Cani and C. Knauf, "How gut microbes talk to organs: the role of endocrine and nervous routes," Mol. Metab., vol. 5, no. 9, pp. 743–752, Sep. 2016, doi: 10.1016/j.molmet.2016.05.011.

[9] A. Koh, F. De Vadder, P. Kovatcheva-Datchary, and F. Bäckhed, "From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites," Cell, vol. 165, no. 6, pp. 1332–1345, Jun. 2016, doi: 10.1016/j.cell.2016.05.041.

[10] F. Asnicar et al., "Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals," Nat. Med., vol. 27, no. 2, pp. 321–332, Feb. 2021, doi: 10.1038/s41591-020-01183-8.

[11] J. L. Sonnenburg and F. Bäckhed, "Diet-microbiota interactions as moderators of human metabolism," Nature, vol. 535, no. 7610, pp. 56–64, Jul. 2016, doi: 10.1038/nature18846.

[12] F. Magne, M. Gotteland, L. Gauthier, A. Zazueta, S. Pesoa, P. Navarrete, and R. Balamurugan, "The Firmicutes/Bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients?" Nutrients, vol. 12, no. 5, art. no. 1474, May 2020, doi: 10.3390/nu12051474.

[13] A. A. Kolodziejczyk, D. Zheng, and E. Elinav, "Diet-microbiota interactions and personalized nutrition," Nat. Rev. Microbiol., vol. 17, no. 12, pp. 742–753, Dec. 2019, doi: 10.1038/s41579-019-0256-8.

[14] J. Aitchison, "The statistical analysis of compositional data," J. R. Stat. Soc. B, vol. 44, no. 2, pp. 139–177, 1982, doi: 10.1111/j.2517-6161.1982.tb01195.x.

[15] G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, and J. J. Egozcue, "Microbiome datasets are compositional: and this is not optional," Front. Microbiol., vol. 8, art. no. 2224, Nov. 2017, doi: 10.3389/fmicb.2017.02224.

[16] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[17] M. Li et al., "Performance of gut microbiome as an independent diagnostic tool for 20 diseases: cross-cohort validation of machine-learning classifiers," Gut Microbes, vol. 15, no. 1, art. no. 2205386, 2023, doi: 10.1080/19490976.2023.2205386.

[18] E. Pasolli, D. T. Truong, F. Malik, L. Waldron, and N. Segata, "Machine learning meta-analysis of large metagenomic datasets: tools and biological

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

insights," PLoS Comput. Biol., vol. 12, no. 7, art. no. e1004977, Jul. 2016, doi: 10.1371/journal.pcbi.1004977.

[19] N. T. Baxter, A. W. Schmidt, A. Venkataraman, K. S. Kim, C. Waldron, and T. M. Schmidt, "Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers," mBio, vol. 10, no. 1, art. no. e02566-18, Jan. 2019, doi: 10.1128/mBio.02566-18.

[20] T. P. Quinn, I. Erb, M. F. Richardson, and T. M. Crowley, "Understanding sequencing data as compositions: an outlook and review," Bioinformatics, vol. 34, no. 16, pp. 2870–2878, Aug. 2018, doi: 10.1093/bioinformatics/bty175.

[21] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Ann. Stat., vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[22] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Disc. Data Min., 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[23] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[24] R. Tibshirani, "Regression shrinkage and selection via the Lasso," J. R. Stat. Soc. B, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[25] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," J. R. Stat. Soc. B, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 5998–6008, arXiv:1706.03762, doi: 10.48550/arXiv.1706.03762.

[28] R. B. Ghannam and S. M. Techtmann, "Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring," Comput. Struct. Biotechnol. J., vol. 19, pp. 1092–1107, Jan. 2021, doi: 10.1016/j.csbj.2021.01.028.

[29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[30] L. J. Marcos-Zambrano, K. Karaduzovic-Hadziabdic, T. Loncar Turukalo, P. Przymus, V. Trajkovik, O. Aasmets, M. Berland, A. Gruca, J. Hasic, K. Hron, et al., "Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment," Front. Microbiol., vol. 12, art. no. 634511, Feb. 2021, doi: 10.3389/fmicb.2021.634511.

[31] R. Hernández Medina, S. Kutuzova, K. N. Nielsen, J. Johansen, L. H. Hansen, M. Nielsen, and S. Rasmussen, "Machine learning and deep learning applications in microbiome research," ISME Commun., vol. 2, art. no. 98, Oct. 2022, doi: 10.1038/s43705-022-00182-9.

[32] G. Cammarota, G. Ianiro, A. Ahern, C. Carbone, A. Temko, M. J. Claesson, A. Gasbarrini, and G. Tortora, "Gut microbiome, big data and machine learning to promote precision medicine for cancer," Nat. Rev. Gastroenterol. Hepatol., vol. 17, no. 10, pp. 635–648, Oct. 2020, doi: 10.1038/s41575-020-0327-3.

[33] G. S. Collins, K. G. M. Moons, P. Dhiman, R. D. Riley, A. L. Beam, B. Van Calster, et al., "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," BMJ, vol. 385, art. no. e078378, Apr. 2024, doi: 10.1136/bmj-2023-078378.

[34] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat. Med., vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7.

[35] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," N. Engl. J. Med., vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/NEJMra1814259.

[36] A. M. Valdes, J. Walter, E. Segal, and T. D. Spector, "Role of the gut microbiota in nutrition and health," BMJ, vol. 361, art. no. k2179, Jun. 2018, doi: 10.1136/bmj.k2179.

[37] P. D. Cani, "Human gut microbiome: hopes, threats and promises," Gut, vol. 67, no. 9, pp. 1716–1725, Sep. 2018, doi: 10.1136/gutjnl-2018-316723.

[38] R. Miotto et al., "Deep learning for healthcare: review, opportunities and challenges," Brief. Bioinform., vol. 19, no. 6, pp. 1236–1246, Nov. 2018, doi: 10.1093/bib/bbx044.

[39] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

[40] X. Wu et al., "A scoping review of artificial intelligence for precision nutrition," Adv. Nutr., vol. 16, no. 4, art. no. 100398, Apr. 2025, doi: 10.1016/j.advnut.2025.100398.

[41] A. Muijsenberg, E. E. Canfora, and E. E. Blaak, "Metabolic phenotypes, genotypes, and gut microbiome signatures in obesity: implications for precision nutrition strategies in type 2 diabetes prevention," Nutr. Rev., vol. 84, no. 1, pp. 158–188, Jan. 2026, doi: 10.1093/nutrit/nuaf088.

[42] S. Mansour, S. M. I. Alkhaaldi, A. F. Sammanasunathan, S. Ibrahim, J. Farhat, and B. Al-Omari, "Precision nutrition unveiled: gene–nutrient interactions, microbiota dynamics, and lifestyle factors in obesity management," Nutrients, vol. 16, no. 5, art. no. 581, Feb. 2024, doi: 10.3390/nu16050581.

[43] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," BMJ, vol. 372, art. no. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[44] Z. Yang, X. Gao, Y. Wu, H. Xie, H. Wang, W. Lu, F. Tian, J. Zhao, H. Zhang, W. Chen, and J. Zhu, "Nonlinear machine learning-based gut microbiota modelling for health prediction and dietary recommendation," Food Biosci., vol. 74, art. no. 107962, Dec. 2025, doi: 10.1016/j.fbio.2025.107962.

[45] H. Wu et al., "Machine learning prediction of obesity-associated gut microbiota: identifying Bifidobacterium pseudocatenulatum as a potential therapeutic target," Front. Microbiol., vol. 15, p. 1488656, Feb. 2025, doi: 10.3389/fmicb.2024.1488656.

[46] W. Liu, X. Fang, Y. Zhou, L. Dou, and T. Dou, "Machine learning-based investigation of the relationship between gut microbiome and obesity status," Microbes Infect., vol. 24, no. 2, art. no. 104892, Mar. 2022, doi: 10.1016/j.micinf.2021.104892.

[47] Y. Liu, J. Zhu, H. Wang, W. Lu, Y. K. Lee, J. Zhao, and H. Zhang, "Machine learning framework for gut microbiome biomarkers discovery and modulation analysis in large-scale obese population," BMC Genomics, vol. 23, art. no. 850, Dec. 2022, doi: 10.1186/s12864-022-09087-2.

[48] D. Chanda and D. De, "Meta-analysis reveals obesity associated gut microbial alteration patterns and reproducible contributors of functional shift," Gut Microbes, vol. 16, no. 1, p. 2304900, Jan. 2024, doi: 10.1080/19490976.2024.2304900.

[49] J. Gong, Y. Shen, H. Zhang, M. Cao, M. Guo, J. He, B. Zhang, and C. Xiao, "Gut microbiota characteristics of people with obesity by meta-analysis of existing datasets," Nutrients, vol. 14, no. 14, art. no. 2993, Jul. 2022, doi: 10.3390/nu14142993.

[50] Y. Lin, Z. Xu, Y. K. Yeoh, H. M. Tun, W. Huang, W. Jiang, F. K. L. Chan, and S. C. Ng, "Combining fecal microbial community data to identify consistent obesity-specific microbial signatures and shared metabolic pathways," iScience, vol. 26, no. 4, art. no. 106476, Apr. 2023, doi: 10.1016/j.isci.2023.106476.

[51] T. Wang, H. D. Holscher, S. Maslov, F. B. Hu, S. T. Weiss, and Y.-Y. Liu, "Predicting metabolite response to dietary intervention using deep learning," Nat. Commun., vol. 16, p. 815, Jan. 2025, doi: 10.1038/s41467-025-56165-6.

[52] H. Tily, E. Patridge, Y. Cai, V. Gopu, S. Gline, M. Genkin, H. Lindau, A. Sjue, I. Slavov, A. Perlina, N. Klitgord, H. Messier, M. Vuyisich, and G. Banavar, "Gut microbiome activity contributes to prediction of individual variation in glycemic response in adults," Diabetes Ther., vol. 13, no. 1, pp. 89–111, Jan. 2022, doi: 10.1007/s13300-021-01174-z.

[53] S. C. Bischoff et al., "Gut microbiota patterns predicting long-term weight loss success in individuals with obesity undergoing nonsurgical therapy," Nutrients, vol. 14, no. 15, p. 3182, Aug. 2022, doi: 10.3390/nu14153182.

[54] R. F. Wolff, K. G. M. Moons, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, and S. Mallett (PROBAST Group), "PROBAST: a tool to assess the risk of bias and applicability of prediction model studies," Ann. Intern. Med., vol. 170, no. 1, pp. 51–58, Jan. 2019, doi: 10.7326/M18-1376.

[55] G. A. Wells et al., "The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomized studies in meta-analyses," Ottawa Hospital Research Institute, Ottawa, 2011. [Online]. Available: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [Accessed: May 2026]

[56] J. A. C. Sterne, J. Savović, M. J. Page, R. G. Elbers, N. S. Blencowe, I. Boutron, et al., "RoB 2: a revised tool for assessing risk of bias in randomised trials," BMJ, vol. 366, art. no. l4898, Aug. 2019, doi: 10.1136/bmj.l4898.

[57] B. J. Callahan et al., "DADA2: high-resolution sample inference from Illumina amplicon data," Nat. Methods, vol. 13, no. 7, pp. 581–583, Jul. 2016, doi: 10.1038/nmeth.3869.

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

[58] A. Blanco-Míguez et al., "Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4," Nat. Biotechnol., vol. 41, no. 11, pp. 1633–1644, Nov. 2023, doi: 10.1038/s41587-023-01688-w.

[59] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/ [Accessed: May 2026]

[60] A. M. Thomas, P. Manghi, F. Asnicar, et al., "Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation," Nat. Med., vol. 25, no. 4, pp. 667–678, Apr. 2019, doi: 10.1038/s41591-019-0405-7.

[61] M. Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar, "Foundation models for generalist medical artificial intelligence," Nature, vol. 616, no. 7956, pp. 259–265, Apr. 2023, doi: 10.1038/s41586-023-05881-4.

[62] Y. Zhang, J. Schluter, L. Zhang, X. Cao, R. R. Jenq, H. Feng, J. Haines, and L. Zhang, "Review and revamp of compositional data transformation: a new framework combining proportion conversion and contrast transformation," Comput. Struct. Biotechnol. J., vol. 23, pp. 4088–4107, Nov. 2024, doi: 10.1016/j.csbj.2024.11.003.

[63] S. Weiss, Z. Z. Xu, S. Peddada, A. Amir, K. Bittinger, A. Gonzalez, C. Lozupone, J. R. Zaneveld, Y. Vázquez-Baeza, A. Birmingham, et al., "Normalization and microbial differential abundance strategies depend upon data characteristics," Microbiome, vol. 5, art. no. 27, Mar. 2017, doi: 10.1186/s40168-017-0237-y.

[64] C. Quince, A. W. Walker, J. T. Simpson, N. J. Loman, and N. Segata, "Shotgun metagenomics, from sampling to analysis," Nat. Biotechnol., vol. 35, no. 9, pp. 833–844, Sep. 2017, doi: 10.1038/nbt.3935.

[65] E. Bolyen, J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar, et al., "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2," Nat. Biotechnol., vol. 37, no. 8, pp. 852–857, Aug. 2019, doi: 10.1038/s41587-019-0209-9.

[66] F. Beghini, L. J. McIver, A. Blanco-Míguez, L. Dubois, F. Asnicar, S. Maharjan, A. Mailyan, P. Manghi, M. Scholz, A. M. Thomas, et al., "Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3," eLife, vol. 10, art. no. e65088, May 2021, doi: 10.7554/eLife.65088.

[67] R. Knight, A. Vrbanac, B. C. Taylor, A. Aksenov, C. Callewaert, J. Debelius, A. Gonzalez, T. Kosciolek, L.-I. McCall, D. McDonald, et al., "Best practices for analysing microbiomes," Nat. Rev. Microbiol., vol. 16, no. 7, pp. 410–422, Jul. 2018, doi: 10.1038/s41579-018-0029-9.

[68] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[69] S. R. Abbas, Z. Abbas, A. Zahir, and S. W. Lee, "Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration," Healthcare, vol. 12, no. 24, art. no. 2587, Dec. 2024, doi: 10.3390/healthcare12242587.

[70] B. A. Mir, S. R. Abbas, and S. W. Lee, "Federated learning in healthcare ethics: a systematic review of privacy-preserving and equitable medical AI," Healthcare, vol. 14, no. 3, art. no. 306, Jan. 2026, doi: 10.3390/healthcare14030306.

[71] P. Kairouz et al., "Advances and open problems in federated learning," Found. Trends Mach. Learn., vol. 14, nos. 1–2, pp. 1–210, 2021, doi: 10.1561/2200000083.

[72] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Stat. (AISTATS), PMLR, vol. 54, 2017, pp. 1273–1282, arXiv:1602.05629, doi: 10.48550/arXiv.1602.05629.

[73] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: concept and applications," ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, art. no. 12, pp. 1–19, Mar. 2019, doi: 10.1145/3298981.

[74] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, et al., "The future of digital health with federated learning," NPJ Digit. Med., vol. 3, art. no. 119, Sep. 2020, doi: 10.1038/s41746-020-00323-1.

[75] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," Sci. Rep., vol. 10, art. no. 12598, Jul. 2020, doi: 10.1038/s41598-020-69250-1.

[76] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

federated machine learning in medical imaging," Nat. Mach. Intell., vol. 2, no. 6, pp. 305–311, Jun. 2020, doi: 10.1038/s42256-020-0186-1.

[77] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 4766–4777, arXiv:1705.07874, doi: 10.48550/arXiv.1705.07874.

[78] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should I trust you?": explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD), 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[79] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell., vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[80] Z. C. Lipton, "The mythos of model interpretability," Commun. ACM, vol. 61, no. 10, pp. 36–43, Oct. 2018, doi: 10.1145/3233231.

[81] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for HealthCare: predicting pneumonia risk and hospital 30-day readmission," in Proc. 21st ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD), 2015, pp. 1721–1730, doi: 10.1145/2783258.2788613.

[82] T. Rozera, E. Pasolli, N. Segata, and G. Ianiro, "Machine learning and artificial intelligence in the multi-omics approach to gut microbiota," Gastroenterology, vol. 169, no. 3, pp. 487–501, Aug. 2025, doi: 10.1053/j.gastro.2025.02.035.

[83] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," JAMA, vol. 319, no. 13, pp. 1317–1318, Apr. 2018, doi: 10.1001/jama.2017.18391.

[84] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," Nat. Med., vol. 25, no. 1, pp. 24–29, Jan. 2019, doi: 10.1038/s41591-018-0316-z.

[85] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health careaddressing ethical challenges," N. Engl. J. Med., vol. 378, no. 11, pp. 981–983, Mar. 2018, doi: 10.1056/NEJMp1714229.

[86] E. Vayena, A. Blasimme, and I. G. Cohen, "Machine learning in medicine: Addressing ethical challenges," PLoS Med., vol. 15, no. 11, art. no. e1002689, Nov. 2018, doi: 10.1371/journal.pmed.1002689.

[87] T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, et al., "Human gut microbiome viewed across age and geography," Nature, vol. 486, no. 7402, pp. 222–227, Jun. 2012, doi: 10.1038/nature11053.

[88] G. Falony, M. Joossens, S. Vieira-Silva, J. Wang, Y. Darzi, K. Faust, A. Kurilshikov, M. J. Bonder, M. Valles-Colomer, D. Vandeputte, et al., "Population-level analysis of gut microbiome variation," Science, vol. 352, no. 6285, pp. 560–564, Apr. 2016, doi: 10.1126/science.aad3503.

[89] A. Zhernakova, A. Kurilshikov, M. J. Bonder, E. F. Tigchelaar, M. Schirmer, T. Vatanen, Z. Mujagic, A. V. Vila, G. Falony, S. Vieira-Silva, et al., "Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity," Science, vol. 352, no. 6285, pp. 565–569, Apr. 2016, doi: 10.1126/science.aad3369.

[90] D. Rothschild, O. Weissbrod, E. Barkan, A. Kurilshikov, T. Korem, D. Zeevi, P. I. Costea, A. Godneva, I. N. Kalka, N. Bar, et al., "Environment dominates over host genetics in shaping human gut microbiota," Nature, vol. 555, no. 7695, pp. 210–215, Mar. 2018, doi: 10.1038/nature25973.

[91] A. Cotillard, S. P. Kennedy, L. C. Kong, E. Prifti, N. Pons, E. Le Chatelier, M. Almeida, B. Quinquis, F. Levenez, N. Galleron, et al., "Dietary intervention impact on gut microbial gene richness," Nature, vol. 500, no. 7464, pp. 585–588, Aug. 2013, doi: 10.1038/nature12480.

[92] E. Le Chatelier, T. Nielsen, J. Qin, E. Prifti, F. Hildebrand, G. Falony, M. Almeida, M. Arumugam, J.-M. Batto, S. Kennedy, et al., "Richness of human gut microbiome correlates with metabolic markers," Nature, vol. 500, no. 7464, pp. 541–546, Aug. 2013, doi: 10.1038/nature12506.

[93] L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, and P. J. Turnbaugh, "Diet rapidly and reproducibly alters the human gut microbiome," Nature, vol. 505, no. 7484, pp. 559–563, Jan. 2014, doi: 10.1038/nature12820.

[94] O. Manor, C. L. Dai, S. A. Kornilov, B. Smith, N. D. Price, J. C. Lovejoy, S. M. Gibbons, and A. T. Magis, "Health and disease markers correlate with gut microbiome composition across thousands of people," Nat. Commun., vol. 11, art. no. 5206, Oct. 2020, doi: 10.1038/s41467-020-18871-1.

[95] H. K. Pedersen, V. Gudmundsdottir, H. B. Nielsen,

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

T. Hyotylainen, T. Nielsen, B. A. H. Jensen, K. Forslund, F. Hildebrand, E. Prifti, G. Falony, et al., "Human gut microbes impact host serum metabolome and insulin sensitivity," Nature, vol. 535, no. 7612, pp. 376–381, Jul. 2016, doi: 10.1038/nature18646.

[96] J. Lloyd-Price, C. Arze, A. N. Ananthakrishnan, M. Schirmer, J. Avila-Pacheco, T. W. Poon, E. Andrews, N. J. Ajami, K. S. Bonham, C. J. Brislawn, et al., "Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases," Nature, vol. 569, no. 7758, pp. 655–662, May 2019, doi: 10.1038/s41586-019-1237-9.

[97] K. Forslund, F. Hildebrand, T. Nielsen, G. Falony, E. Le Chatelier, S. Sunagawa, E. Prifti, S. Vieira-Silva, V. Gudmundsdottir, H. Krogh Pedersen, et al., "Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota," Nature, vol. 528, no. 7581, pp. 262–266, Dec. 2015, doi: 10.1038/nature15766.

[98] D. Reiman, A. A. Metwally, J. Sun, and Y. Dai, "PopPhy-CNN: a phylogenetic tree embedded architecture for convolutional neural networks to predict host phenotype from metagenomic data," IEEE J. Biomed. Health Inform., vol. 24, no. 10, pp. 2993–3001, Oct. 2020, doi: 10.1109/JBHI.2020.2993761.

[99] M. Oh and L. Zhang, "DeepMicro: deep representation learning for disease prediction based on microbiome data," Sci. Rep., vol. 10, art. no. 6026, Apr. 2020, doi: 10.1038/s41598-020-63159-5.

[100] N. LaPierre, C. J.-T. Ju, G. Zhou, and W. Wang, "MetaPheno: a critical evaluation of deep learning and machine learning in metagenome-based disease prediction," Methods, vol. 166, pp. 74–82, Aug. 2019, doi: 10.1016/j.ymeth.2019.03.003.

Author Biography



Poppy Fransisca Amelia is a lecturer in the Study Program of Biomedical Sciences, Universitas Karya Husada, Semarang, Indonesia. She holds a degree in Biomedical Sciences and has built her academic career around the intersection of laboratory science and clinical application. Her primary research interests span biomedical science, microbiology, biochemistry, research methodology, and health sciences education. She is actively involved in teaching undergraduate courses in microbiology and biochemistry, where she integrates evidence-based pedagogy with laboratory practicals. Her scholarly work focuses on connecting biomedical laboratory findings with clinical and public health questions, with particular attention to translational research that bridges in vitro observations and community health outcomes. She has contributed to collaborative research on infectious disease biomarkers and health promotion strategies in Central Java.



Fitra Adi Prayogo received his Bachelor's and Master's degrees in Biology from Diponegoro University, Indonesia. He is a lecturer in Biomedical Sciences at Universitas Karya Husada, Semarang, Indonesia, and the corresponding author of this manuscript. His primary research focus lies at the intersection of microbiome profiling, biochemistry, bioinformatics, and computational biology applied to health sciences. He has developed expertise in systematic review methodology, machine-learning-based biomarker discovery, and evidence synthesis for clinical and nutritional decision support. His current research programme investigates how gut microbiome data can be leveraged through supervised learning algorithms to inform personalised dietary and metabolic interventions. He is actively involved in mentoring undergraduate research in biomedical data analysis and has contributed to several peer-reviewed publications in biomedical and health informatics journals.



Isy Royhanaty received her Master's degree in Biomedical Science from Diponegoro University, Indonesia. She is a lecturer in the Biomedical Sciences Study Program at Universitas Karya Husada, Semarang, Indonesia. Her academic interests encompass biomedical science, health research methodology, clinical laboratory interpretation, and evidence synthesis for biomedical and public health applications. She has particular expertise in systematic review methodology and quality-of-evidence appraisal, contributing to research that critically evaluates diagnostic and prognostic tools in the biomedical literature. Her teaching responsibilities include research methodology and biostatistics courses for undergraduate health science students. She is actively involved in collaborative research projects examining the application of laboratory-based biomarkers to community health screening, and she has co-authored peer-reviewed articles on evidence-based health interventions in the Indonesian public health context.



Heni Wijayanti is a lecturer in the Study Program of Biomedical Sciences, Universitas Karya Husada, Semarang, Indonesia. She holds a background in biomedical sciences with a focus on human anatomy and physiology. Her academic interests span biomedical science, anatomy, physiology, general medicine, and

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

the integration of basic biomedical concepts with applied health research. She has extensive teaching experience in anatomy, physiology, and pathophysiology courses for health science undergraduates, where she emphasises the clinical relevance of fundamental biological mechanisms. Her research contributions include collaborative studies on metabolic health indicators and nutritional status assessment in community populations. She is committed to interdisciplinary approaches that link anatomical and physiological understanding with emerging computational and data-driven methods in health sciences, and she actively participates in faculty research groups addressing non-communicable disease prevention in Central Java.



Egy Sunanda Putra received his degree in Nutrition Science from Universitas Sebelas Maret, Indonesia. He is a lecturer in Health Promotion at Poltekkes Kemenkes Jambi and is affiliated with the Centre of Excellence for Science and Technology, Poltekkes Kemenkes Jambi, Indonesia. His primary research interests include nutrition science, adolescent health, obesity prevention, health promotion, clinical data interpretation, and community-based intervention research. He has developed particular expertise in dietary assessment methodologies and nutritional epidemiology applied to adolescent and adult populations in Indonesia. His current research focuses on the design and evaluation of community-level nutritional interventions aimed at reducing obesity prevalence and improving metabolic health outcomes. He actively contributes to public health policy discussions on nutrition education and has co-authored studies examining the relationship between dietary patterns, gut health indicators, and body composition in Southeast Asian cohorts.

Manuscript received 30 April 2026; Revised 31 May 2026; Accepted 2 September 2026; Available online 25 October 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1722
Copyright © 2026 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).