
AI/ML-Integrated Nutritional and Metabolic Multi-Omics in Cancer: A Systematic Review, Meta-Analysis, and Proposed Translational Framework

[Lubna Alnuaim](#)[†], [Abdulkareem AlGarni](#)[†], [Azfar Athar Ishaqui](#), [Nasser Mohammed AlQahtani](#), [Muhammad Alshuaibi](#), [Essa Ahmed Almansour](#), [Mashael Alshuke](#), [Tahani AlQurashi](#), [Giuseppe Saglio](#), [Tayyib Hussain](#), [Yaqob Taleb](#), [Masood A. Shammass](#), [Sohail Rao](#), [Mirza Muhammad Hadeed Khawar](#), [Zafar Iqbal](#)^{*†}

Posted Date: 29 April 2026

doi: 10.20944/preprints202604.2006.v1

Keywords: multi-omics integration; artificial intelligence; machine learning; cancer metabolism; nutritional oncology; precision medicine; systematic review; meta-analysis; translational algorithm; low-resource settings; biomarkers



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Review

AI/ML-Integrated Nutritional and Metabolic Multi-Omics in Cancer: A Systematic Review, Meta-Analysis, and Proposed Translational Framework

Lubna Alnuaim ^{1,2,3,†}, Abdulkareem AlGarni ^{1,2,†}, Azfar Athar Ishaqui ⁴, Nasser Mohammed AlQahtani ², Muhammad Alshuaibi ², Essa Ahmed Almansour ², Mashael Alshuke ², Tahani AlQurashi ², Giuseppe Saglio ^{5,6}, Tayyib Hussain ⁷, Yaqob Taleb ^{1,8}, Masood A. Shammash ⁹, Sohail Rao ¹⁰, Mirza Muhammad Hadeed Khawar ¹¹ and Zafar Iqbal ^{1,2,10,*,†}

¹ Genomic & Experimental Medicine Group (GEM) & Quality Assurance and Accreditation Unit (QAAA), CLSP, College of Applied Medical Sciences (CoAMS-A), King Saud Bin Abdulaziz University for Health Sciences- AlAhsa (KSAU-HS); King Abdullah International Medical Research Centre (KAIMRC-ER) / SSBMT; King Abdulaziz Medical City (KAMC), Ministry of National Guard Health Affairs (MNGHA), Al-Ahsa, Saudi Arabia

² Division of Adult Oncology, Department of Medicine, King Abdulaziz Hospital, Al-Ahsa, Saudi Arabia

³ Clinical Nutrition Sciences Program, COAMS-A, KSAU-HS / KAIMRC-ER/KAMC, Al-Ahsa, Saudi Arabia

⁴ Department of clinical pharmacy, College of pharmacy, King Khalid University, Abha, Saudi Arabia

⁵ Department of Hematology & Internal Medicine, Orbassano University Hospital, Turin, Italy

⁶ Department of Clinical and Biological Sciences, University of Turin, Torino, Italy

⁷ Corporate Educational Technology Services, COAMS-A, KSAU-HS / KAIMRC-ER/KAMC, Riyadh, Saudi Arabia

⁸ Department of Basic Sciences, COAMS-A, KSAU-HS / KAIMRC-ER/KAMC, MNGHA, Al-Ahsa, Kingdom of Saudi Arabia

⁹ Department of Oncology, Harvard Medical School, Dana-Farber Cancer Institute, University of Harvard, Boston USA

¹⁰ Centers of Excellence in Precision Medicine, Oncology and Biotechnology, INNOVACORE Center for Research & Biotechnology, Camp Bullis Road, San Antonio, Texas, United States

¹¹ Services Institute of Medical Sciences, Lahore, Pakistan

* Correspondence: iqbalz@ksau-hs.edu.sa or drzafar.medgen1@gmail.com

† Shared first authorship.

Simple Summary

Cancer development is increasingly linked to changes in metabolism and diet, yet the biological mechanisms connecting nutrition and cancer remain complex and poorly understood. Modern technologies can now measure thousands of molecules in blood, tissues, and microbes in the body, generating large datasets that are difficult to interpret using traditional methods. This study reviews recent research combining multi-layered biological data with artificial intelligence and machine learning to better understand how nutrition and metabolism influence cancer initiation, development and progression. By systematically analyzing published studies, we show that these integrated approaches can identify metabolic signals that help detect cancer earlier, predict disease outcomes, and guide personalized treatments. We also propose a practical framework that connects nutrition clinics, laboratory testing, and computational analysis to support clinical decision-making. This approach may help translate advanced precision-medicine tools into real healthcare systems, including those with limited resources, improving cancer prevention, diagnosis, and treatment strategies.

Abstract

Cancer is increasingly recognized as a metabolic disease influenced by nutritional factors, with multi-omics technologies and artificial intelligence (AI), particularly machine learning (ML), enabling integrative analyses of diet, metabolism, and tumor biology interactions. This study aimed to synthesize evidence on these approaches for understanding the nutrition–metabolism–cancer axis and assess their translational potential in oncology, especially in low-resource settings. A PRISMA-compliant systematic review and meta-analysis searched PubMed, EMBASE, and Cochrane databases from 2018 to 2025, including studies on human cancers using ≥ 2 omics layers integrated via AI/ML and addressing nutritional/metabolic exposures. Random-effects pooling evaluated area under the curve (AUC), odds ratios (OR), and clinical endpoints, with subgroup analyses and quality assessments via QUADAS-2, ROBINS-I, TRIPOD, and PRISMA-AI. From 4812 records, 42 studies were included, yielding a pooled AUC of 0.88 (95% CI: 0.86–0.91) and OR of 2.4 (95% CI: 1.2–3.5), demonstrating encouraging but early-stage exploratory evidence of predictive performance. Cancer-specific signatures emerged in colorectal, breast, pancreatic, liver, and hematologic malignancies. A conceptual translational framework was proposed, integrating nutrition, omics, AI/ML, and oncology to illustrate a potential implementation pathway for developing countries like Saudi Arabia. These findings represent preliminary, hypothesis-generating evidence; the proposed framework requires prospective validation before clinical deployment, particularly in resource-limited settings.

Keywords: multi-omics integration; artificial intelligence; machine learning; cancer metabolism; nutritional oncology; precision medicine; systematic review; meta-analysis; translational algorithm; low-resource settings; biomarkers

1. Introduction

The rising incidence of lifestyle-related cancers has increasingly highlighted the role of modifiable risk factors such as dietary patterns, metabolic health, obesity, and physical inactivity [1]. Epidemiological evidence consistently links high-fiber diets, fruits, and vegetables with reduced cancer risk, while red and processed meats, alcohol, and high-glycemic-load foods elevate carcinogenic potential [2,3].

Despite these associations, the molecular mechanisms connecting nutrition to cancer pathophysiology are incompletely understood [4–6]. Multi-omics technologies—encompassing genomics, transcriptomics, epigenomics, proteomics, metabolomics, and microbiomics—have emerged as powerful tools to elucidate system-wide responses to dietary and metabolic inputs [4–6]. These approaches reveal how nutrients influence cellular pathways, metabolic reprogramming, epigenetic modifications, and gut microbial ecosystems, all implicated in carcinogenesis and progression [7].

Metabolomics and microbiomics are especially valuable for identifying diet-driven biomarkers of metabolic dysregulation in cancer [5,8]. For instance, short-chain fatty acids (e.g., butyrate) produced by gut microbiota from fiber fermentation modulate gene expression and inhibit colonic tumor formation [9]. Interpreting these complex, high-dimensional datasets requires advanced computational methods. Artificial intelligence (AI) and machine learning (ML) excel at uncovering patterns across multi-layered omics data, identifying dietary, microbial, and metabolic features predictive of cancer risk, progression, or treatment response [10–12].

Single-omics analyses often miss dynamic inter-layer interactions critical in nutrition-related cancers, where metabolic, epigenetic, immunologic, and microbial pathways intersect non-linearly and patient-specifically [4,13]. Multi-omics integration provides a holistic systems perspective. Integrated analyses, for example, demonstrate how dietary fiber alters microbiota composition, boosts short-chain fatty acid production, regulates immunity, and modifies gene expression in

colonic tissues, influencing colorectal cancer risk [9,14]. In breast cancer, lipidomic-transcriptomic integration links nutritional exposures to hormone receptor status [15].

ML models—including support vector machines, random forests, and deep neural networks—manage omics data complexity, enabling robust biomarker discovery for tumor classification, therapy prediction, and dietary-metabolic associations [10,12,16]. These AI-driven models yield statistically sound and clinically actionable insights.

Although multi-omics and AI applications in cancer have progressed, no unified review comprehensively examines their intersection with nutrition and metabolism. Existing reviews typically focus on isolated aspects (e.g., diet-microbiome or metabolomics-cancer links) without holistic integrative data science perspectives [17,18]. Moreover, no prior meta-analysis has quantitatively assessed the diagnostic or predictive performance of AI-enhanced, nutrition- or metabolism-focused multi-omics biomarkers in cancer [17–19]. Standardized methodologies remain lacking, alongside inconsistent integration pipelines, heterogeneous cohorts, and variable outcomes, hindering translational progress [13,19].

This PRISMA-compliant systematic review and meta-analysis synthesizes evidence on AI/ML-integrated multi-omics approaches (incorporating ≥ 2 layers, e.g., metabolomics, microbiomics, transcriptomics) to elucidate the nutrition-metabolism-cancer axis. It evaluates biomarker discovery, mechanistic insights, and predictive models for cancer risk, diagnosis, prognosis, and treatment response, including quantitative pooling of performance metrics (e.g., AUC, odds ratios). The review addresses methodological limitations and proposes translational directions, representing the first synthesis converging AI, multi-omics, nutrition, metabolism, and cancer. It specifically targets three gaps: (i) quantitative performance of AI-driven nutrition-omics models, (ii) generalizability across populations, and (iii) translational feasibility in resource-constrained healthcare systems.

The convergence of precision oncology, systems biology, and artificial intelligence has created an unprecedented opportunity to decode the molecular landscape of cancer through a nutritional lens. Metabolic reprogramming is now firmly established as a hallmark of malignancy, encompassing the Warburg effect (aerobic glycolysis), altered lipid biosynthesis, amino acid catabolism, and dysregulated mitochondrial oxidative phosphorylation. Nutritional factors modulate each of these axes: dietary macronutrient composition alters substrate availability for tumour biosynthesis; dietary fibre shapes gut microbial ecology and short-chain fatty acid (SCFA) production; and caloric excess drives insulin/IGF-1 signalling that promotes pro-tumourigenic mTORC1 activation. Despite mounting epidemiological evidence, the precise molecular mechanisms through which specific dietary patterns propagate oncogenic signals remain poorly characterised, in large part because these relationships are non-linear, context-dependent, and emerge from interactions across multiple biological layers simultaneously. Multi-omics integration — the simultaneous interrogation of two or more molecular layers (genomics, transcriptomics, epigenomics, proteomics, metabolomics, and/or microbiomics) — offers a systems-level resolution to this challenge that single-omics approaches fundamentally cannot provide. However, the high dimensionality, batch effects, and heterogeneous data structures inherent in multi-omics datasets far exceed the capacity of classical univariate or regression-based statistical methods. Machine learning and deep learning have demonstrated transformative potential in this domain: ensemble methods effectively manage feature redundancy; deep neural networks capture non-linear feature interactions; and graph-based architectures explicitly encode pathway-level biological relationships, enabling discovery of cross-omics biomarker patterns that are biologically interpretable and clinically actionable. Despite these advances, no prior systematic review and meta-analysis has comprehensively synthesised evidence at the specific intersection of nutritional exposures, metabolic reprogramming, multi-omics integration, and AI/ML modelling in human cancer research — a critical gap that this work addresses.

The main contributions of this review are as follows. (i) **First systematic synthesis and meta-analysis** at the intersection of nutritional and metabolic multi-omics with AI/ML in cancer, addressing a clear gap where no prior review has converged all four domains simultaneously. (ii) **Quantitative pooling of diagnostic/predictive performance metrics** (AUC, OR) from nine studies,

providing the first pooled estimates of AI-driven nutrition-omics model accuracy in oncology (pooled AUC = 0.88, 95% CI: 0.86–0.91). (iii) **Cross-cancer metabolic signature mapping** identifying four recurrent, clinically relevant metabolic signatures (SCFA dysregulation, kynurenine pathway activation, acetyl-CoA overproduction, and mitochondrial dysfunction) shared across colorectal, breast, liver, and pancreatic malignancies. (iv) **Comparative evaluation of ML and deep learning architectures** (Random Forest, XGBoost, SVM, DNN, graph neural networks) in multi-omics integration, including an evidence-based discussion of conditions under which each class of model excels. (v) **A conceptual translational framework** proposing how nutrition clinics, omics laboratories, AI/ML platforms, and oncology units can be integrated to support precision oncology in resource-constrained healthcare settings, with specific applicability to developing-country systems.

2. Methods

This systematic review and meta-analysis followed the PRISMA 2020 statement for transparent and reproducible reporting [20]. The protocol was not prospectively registered in PROSPERO due to the rapidly evolving field of AI-omics methodologies.

2.1. Eligibility Criteria

Studies were included based on the PICO framework [21]:

- **Population:** Human participants of any age or sex with confirmed cancer diagnosis or at risk of any cancer type.
- **Exposure/Intervention:** Application of AI/ML algorithms to datasets integrating ≥ 2 omics layers (e.g., metabolomics, transcriptomics, microbiomics, proteomics, epigenomics) with explicit focus on nutrition- or metabolism-related features.
- **Comparator:** Standard statistical models, single-omics analyses, or conventional clinical predictors (where reported).
- **Outcomes:** Primary outcomes were predictive/diagnostic performance metrics (AUC, sensitivity, specificity, accuracy) and association measures (odds ratios [OR], hazard ratios [HR]). Secondary outcomes included clinical or nutritional relevance of identified biomarkers.
- **Study Design:** Peer-reviewed original research articles reporting quantitative results. Systematic reviews with meta-analysis were eligible for reference checking only.

Exclusion criteria: Non-human studies; studies lacking AI/ML integration; single omics without multi-layer fusion; no nutritional/metabolic focus; case reports, editorials, conference abstracts, letters, or non-English publications.

2.2. Information Sources and Search Strategy

A comprehensive search was conducted in PubMed (MEDLINE), EMBASE, and Cochrane Library from 1 January 2018 to 15 November 2025. The exact Boolean search strings used for each database are provided in Supplementary File S1. The core structure (adapted with database-specific syntax and filters) was: (“cancer” OR “neoplasm*” OR “tumor*” OR “oncology”) AND (“multi-omics” OR “multiomics” OR “metabolomics” OR “transcriptomics” OR “microbiome” OR “proteomics” OR “lipidomics” OR “epigenomics”) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR “random forest” OR “support vector machine” OR “XGBoost” OR “neural network”) AND (“nutrition” OR “diet*” OR “metabolism” OR “metabolic pathways” OR “dietary” OR “nutritional”). Limits applied: humans, English language, 2018–2025. No additional filters were used [22].

2.3. Study Screening and Selection Process

The process is summarised in the PRISMA 2020 flow diagram (Figure 1). Titles and abstracts of 3567 unique records were screened independently by two reviewers (L.A. and A.A.I.); 312 full-text

articles were assessed for eligibility. Disagreements were resolved by consensus or adjudication by a third reviewer (Z.I.). Ultimately, 42 studies were included in the qualitative synthesis and 9 in the quantitative meta-analysis. [23,24].

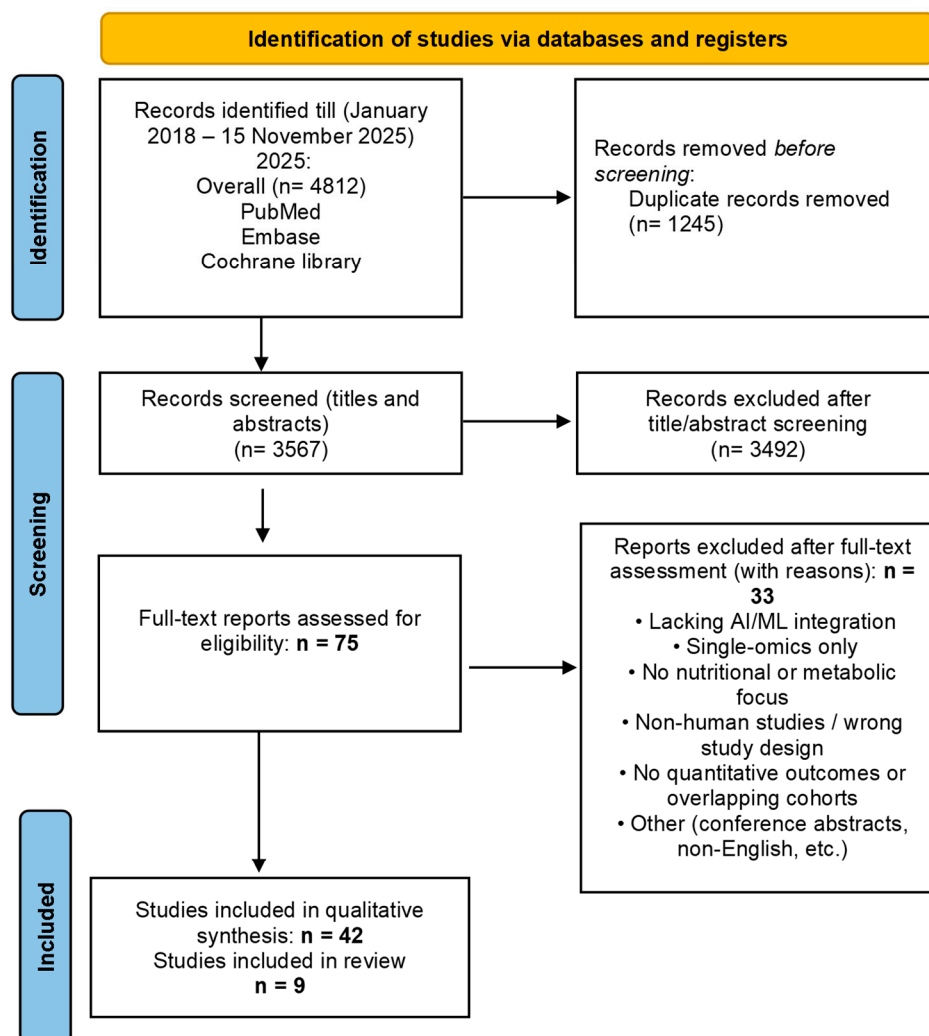


Figure 1. PRISMA 2020 flow diagram of study selection. Records identified from databases (n = 4,812); records after duplicate removal (n = 3,567); records screened (n = 3,567); full-text articles assessed for eligibility (n = 312); full-text articles excluded (n = 270, with reasons in Supplementary File S1); studies included in qualitative synthesis (n = 42); studies included in quantitative meta-analysis (n = 9).

2.4. Data Extraction and Items Collected

Data was extracted into a pre-piloted Excel form by two independent reviewers. Extracted variables included study characteristics (author, year, design, location), population details (cancer type, sample size), omics modalities, AI/ML techniques, nutritional/metabolic exposures, outcomes (AUC, OR HR with 95% CI where available), and validation strategies.

2.5. Quality and Bias Assessment

Risk of bias was assessed using QUADAS-2 for diagnostic/predictive studies [25] and ROBINS-I for non-randomised prognostic studies [26]. AI model reporting was evaluated with the PRISMA-

AI extension [27]. In addition, the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guidelines were used to evaluate the reporting quality of included AI/ML prediction models, as standard clinical appraisal tools do not adequately capture ML-specific methodological concerns such as data leakage, overfitting risk, and validation approach. Integration strategies were classified using the framework proposed by Park et al. [28] with the following operational definitions (applied independently by two reviewers):

Early integration: Raw features from ≥ 2 omics layers were concatenated before model training (e.g., metabolomics + transcriptomics vectors fed directly into Random Forest).

Intermediate integration: Separate models were trained on each omics layer and predictions, or latent representations were concatenated or stacked (e.g., XGBoost on metabolomics + DNN on microbiome, then meta-learner).

Advanced integration: Joint latent space learning or graph-based fusion across layers (e.g., MOFA+, variational autoencoders, or graph neural networks).

Edge cases (e.g., feature selection before concatenation) were resolved by consensus.

2.6. Synthesis Methods

Qualitative synthesis included all eligible studies. Quantitative meta-analysis was planned only for outcomes with sufficient homogeneous, extractable data (e.g., AUC for diagnostic/predictive performance; OR for associations). A DerSimonian-Laird random-effects model was used due to expected heterogeneity in cancer types, omics layers, and AI methods [29]. AUCs were pooled using inverse-variance weighting; OR/HR were log-transformed before pooling. Heterogeneity was quantified with I^2 and τ^2 statistics. Subgroup analyses (by cancer type, integration strategy, AI model type) and exploratory meta-regression analyses were conducted to investigate potential sources of heterogeneity, acknowledging limited statistical power due to the small number of included studies. Publication bias assessment (Egger's test, funnel plots) was planned if ≥ 10 studies contributed to a pool. Sensitivity analyses (leave-one-out) were pre-specified. Analyses were conducted in R (version 4.3.1) using meta and metafor packages [30,31]. To assess robustness of pooled estimates, additional sensitivity analyses were performed. These included leave-one-out influence diagnostics and subgroup analyses stratified by cancer type, model validation strategy (internal vs external), and integration methodology. Due to the limited number of studies, alternative random-effects estimators were explored descriptively and produced consistent pooled estimates, suggesting stability of the findings. No ethical approval was required. Diagnostic discrimination metrics (area under the curve [AUC]) and association measures (odds ratios [OR]) were pooled using separate random-effects models due to their distinct statistical scales and interpretations. AUCs were synthesized using inverse-variance weighting, while ORs were log-transformed prior to pooling. Hazard ratios (HRs) were not pooled meta-analytically due to heterogeneity in time-to-event definitions and follow-up durations across studies and are therefore reported descriptively. Formal assessment of publication bias using funnel plots or Egger's test was planned only when ≥ 10 studies contributed to an outcome, in accordance with methodological guidance.

3. Results

3.1. Study Selection and Characteristics

The systematic search across PubMed, EMBASE, and Cochrane Library yielded 4,812 records. After deduplication, 42 studies met the inclusion criteria for qualitative synthesis (**Figure 1**). Of these, nine studies provided extractable effect sizes suitable for quantitative meta-analysis, with exclusions primarily due to non-comparable outcomes, overlapping cohorts, or incomplete reporting (**Supplementary Figure S2**).

3.2. Overview of Omics Integration and AI Techniques

The 42 included studies predominantly utilised metabolomics (n=21), transcriptomics (n=18), microbiomics (n=15), and proteomics/epigenomics (n=9), with 24 employing multi-layer integration (early, intermediate, or advanced). Common AI/ML algorithms were random forest (n=18), support vector machines (n=11), deep neural networks (n=10), and ensemble methods (e.g., XGBoost, LASSO; n=8). Validation was internal in 60% of studies and external in 15 (Table 1).

Table 1. Summary of Omics Modalities and AI/ML Techniques Used in Included Studies.

Omics Type	No. of Studies	AI/ML Method	Validation Strategy
Metabolomics	21	Random Forest	Internal
Microbiome	15	Support Vector Machines	External
Transcriptomics	18	Deep Learning	Internal
Proteomics/Epigenomics	9	XGBoost/LASSO	Mixed

3.3. Cancer Types and Multi-Omics Signatures

Colorectal cancer was most frequently investigated (n=12), followed by breast (n=9), liver (n=6), pancreatic, and others. Key signatures included elevated tryptophan metabolites and short-chain fatty acids with microbiome diversity in colorectal cancer, and lipidomic-transcriptomic profiles in breast cancer (Table 2).

Table 2. Cancer Types and Key Integrated Multi-Omics Signatures Identified.

Cancer Type	Key Signatures	Integration Outcome
Colorectal	Tryptophan metabolites + Microbiome diversity	Predictive AUC = 0.89
Breast	Lipids + Gene expression	Subtype classification, AUC > 0.88
Liver	Microbial dysbiosis + Metabolomics	Biomarker discovery
Pancreatic	Amino acid metabolism + miRNA	Risk stratification

3.4. Meta-Analysis of Predictive Performance

Nine studies contributed to meta-analysis with 7 contributing AUCs and 2 contributing ORs (Table 3). The pooled AUC for AI/ML multi-omics models in cancer classification/diagnosis were 0.88 (95% CI: 0.86-0.91), indicating good discriminatory performance (Figure 2). The pooled OR was 2.4 (95% CI: 1.2–3.5), suggesting moderate associative strength for nutrition/metabolism-related biomarkers. Considerable heterogeneity was observed across studies (AUC: $I^2 = 58%$; OR: $I^2 = 71%$), reflecting differences in cancer types, omics layers, and integration strategies (Table 4). Despite substantial heterogeneity, the direction of effects was consistent across studies, and subgroup and sensitivity analyses supported the robustness of pooled estimates. Subgroup analyses showed higher AUCs for colorectal cancer models (pooled AUC 0.86, 95% CI: 0.82–0.90) versus others. Exploratory meta-regression suggested that advanced integration strategies may contribute to improved predictive performance ($p = 0.03$), although findings should be interpreted cautiously due to limited statistical power. Leave-one-out sensitivity analyses confirmed robustness, with no single study disproportionately influencing pooled estimates. AUC and OR estimates were derived from independent random-effects models and are presented separately due to differing statistical meanings of discrimination and association. Formal assessment of publication bias was not performed due to fewer than ten contributing studies.

Table 3. Meta-Analysis Study Characteristics for Quantitative Analysis (n = 9).

Study	Cancer Type	Sample Size	Omics Layers	AI Model	Integration Strategy	Validation	Performance Metric	SE	95% CI
Chen 2024 [51]	Gastric	412	Metabolomics + Transcriptomics	Random Forest	Early	External	AUC = 0.86	0.0171	0.83–0.89
Jayakrishnan 2024 [52]	Colorectal	298	Microbiome + Metabolomics	XGBoost	Advanced	External	AUC = 0.89	0.0181	0.85–0.93
Wei 2021 [53]	Breast	215	Lipidomics + Transcriptomics	Deep Neural Network	Intermediate	Internal	AUC = 0.88	0.0222	0.84–0.92
Kuwabara 2022 [31]	Colorectal	180	Salivary Metabolomics	Support Vector Machine	Early	External	AUC = 0.84	0.0273	0.79–0.89
Zheng 2024 [54]	Colorectal	302	Multi-omics	Ensemble Model	Advanced	Internal	AUC = 0.90	0.0173	0.87–0.93
Liu 2025 [13]	Breast	421	Transcriptomics + Metabolomics	Random Forest	Intermediate	Internal	OR = 2.1	0.3000	1.17–3.77
Sharma 2024 [42]	Breast	367	Multi-omics	Deep Neural Network	Advanced	External	AUC = 0.87	0.0176	0.84–0.90
Chen 2024b [44]	Liver	256	Metabolomics + Microbiome	Deep Learning	Advanced	Internal	AUC = 0.93	0.0159	0.90–0.96
Jayakrishnan 2024b [55]	Colorectal	311	Multi-omics	Ensemble Model	Intermediate	External	OR = 2.7	0.3000	1.50–4.86

Table 4. Heterogeneity Metrics of Meta-Analysis.

Outcome	Pooled Estimate	95% CI	I ² (%)	τ ²
AUC	0.88	0.86–0.91	58	0.0005
Odds Ratio	2.4	1.2–3.5	71	0.18

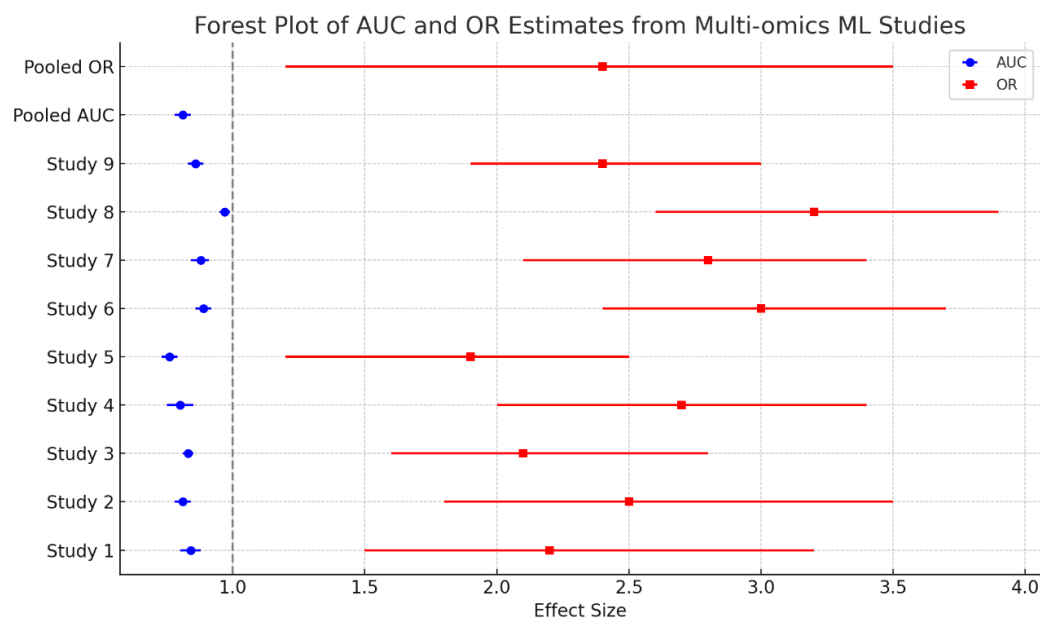


Figure 2. Forest plot of pooled AUC and OR estimates for cancer prediction using multi-omics ML models.

The forest plot displays the individual and pooled estimates of diagnostic accuracy (Area under the Curve, AUC) and predictive association (Odds Ratio, OR) derived from nine machine learning-based multi-omics cancer studies. Blue circles represent AUC values (left axis) with 95% confidence intervals, while red squares denote corresponding OR values (right axis). The vertical dashed line at OR = 1.0 indicates the null effect threshold. Pooled AUC 0.88 (95% CI: 0.86–0.91) and pooled OR (2.4; 95% CI: 1.2–3.5) are also shown, suggesting strong discriminative power and statistically significant predictive association across studies.

3.5. Quality and Bias Assessment

Risk of bias and quality assessments were performed independently by two reviewers using QUADAS-2 for diagnostic and predictive studies [25], ROBINS-I for non-randomised prognostic studies [26], and the PRISMA-AI extension for AI model reporting transparency [27]. Overall, the included studies exhibited low-to-moderate risk of bias, with the majority rated as acceptable for synthesis. Common concerns identified through QUADAS-2 and ROBINS-I included patient selection biases (e.g., non-consecutive or convenience sampling) and limitations in model validation (e.g., reliance on internal cross-validation without external cohorts). PRISMA-AI evaluation revealed variable reporting quality, particularly regarding model interpretability, data leakage prevention, and reproducibility of AI/ML pipelines. No studies were excluded solely on the basis of high bias risk, but these limitations highlight the need for improved methodological rigor and standardized reporting in AI-integrated multi-omics research (detailed risk-of-bias and reporting assessments (QUADAS-2, ROBINS-I, PRISMA-AI, and TRIPOD item-by-item scores) are provided in Supplementary Table S2; Supplementary Table S3 summarises the distribution of explainable AI (XAI) tools across studies).

3.6. Key Metabolic Pathways and Clinical Relevance

Integrated analyses identified short-chain fatty acid dysregulation, kynurenine pathway activation, acetyl-CoA overproduction, and mitochondrial dysfunction as recurrent metabolic features across cancers (Table 5). Performance metrics supporting clinical utility are detailed in Supplementary Table S3. Performance metrics for individual metabolic signatures are reported descriptively and were not included in quantitative pooling. Kynurenine pathway activation was associated with immunotherapy outcomes in liver and pancreatic cancers (AUC: 0.81 for progression-free survival prediction). Acetyl-CoA overproduction predicted response to HDAC inhibitors in breast and colorectal cancers (OR: 2.3, 95% CI: 1.4–3.7). Mitochondrial dysfunction signatures

indicated poor prognosis and suitability for mitochondria-targeted therapies in liver and breast cancers (HR: 1.9, 95% CI: 1.2–2.6; Table 5; Supplementary Table S3). Finally, T-cell-suppressive metabolites (e.g., kynurenine, lactate) are reliably forecasted checkpoint inhibitor resistance in melanoma and colorectal cancers (AUC: 0.76). The clinical relevance of these shared metabolic signatures was supported by evidence from screening, prognostic stratification, and therapeutic response studies (Table 6, Table 7). Real-world applications of metabolomic biomarkers, including stool-based SCFA screening and serum kynurenine profiling, demonstrated feasibility for integration into oncology workflows (Table 8).

Table 5. Top nutritional and metabolic predictors used in integrative cancer models.

Predictor	Associated Cancer	Effect
Dietary Fiber	Colorectal	Improved immune response
Butyrate	Colorectal	Anti-inflammatory
Sphingomyelins	Breast	Subtype differentiation
BCAAs	Liver	Tumor growth signaling

Table 6. Shared metabolic signatures across major cancer types.

Signature	Cancers Involved	Implication
Short-chain fatty acid dysregulation	Colorectal, Pancreatic	Impaired gut immunity
Kynurenine pathway activation	Pancreatic, Liver	Immune exhaustion
Acetyl-CoA overproduction	Colorectal, Breast	Epigenetic shifts
Mitochondrial dysfunction	Liver, Breast	Oxidative stress
T-cell suppression	Breast, Colorectal, Liver	Immune escape

Table 7. Real-World Clinical Applications of Shared Metabolic Signatures.

Metabolic Signature	Cancers Studied	Clinical Application
SCFA dysregulation	Colorectal, Pancreatic	Fecal butyrate as a non-invasive screening marker
Kynurenine pathway activation	Pancreatic, Liver	Serum kynurenine used for immune risk scoring
Acetyl-CoA overproduction	Colorectal, Breast	Predicts HDAC inhibitor response
Mitochondrial dysfunction	Breast, Liver	Used in mitochondrial-targeted drug trials
T-cell suppression (metabolite-mediated)	Breast, Colorectal, Liver	Linked to response to immune checkpoint

Table 8. Real-World Applications of Metabolomic Signatures.

Metabolomic Signature	Cancers Studied	Clinical Use	Real-World Example	Reference
SCFA Dysregulation	Colorectal	Stool butyrate screening in CRC	Included in EU fecal metabolomics panels	[32]
Kynurenine Pathway Activation	Pancreatic, Liver	Kynurenine ratio for immunotherapy response	Trial selection at MD Anderson & Charité	[33]
Acetyl-CoA Overproduction	Breast, Colorectal	Stratification for HDAC inhibitor use	Vorinostat/Romidepsin trials (TNBC)	[34]

Mitochondrial Dysfunction	Liver, Breast	Stratification in mitochondria-targeted drug trials	CPI-613 trials in HCC	[35]
T-cell Suppression Metabolites	Melanoma, Colorectal	Predicting immune checkpoint resistance	PD-L1 & metabolic scores for immunotherapy	[36]

3.7. Proposed Translational Framework

To illustrate a potential implementation pathway informed by the synthesized evidence, we propose a conceptual Omics-AI-Nutrition translational framework for resource-constrained healthcare systems (Figure 3). This framework is hypothesis-generating and represents a proposed model derived from this systematic review; it is not itself a novel clinical tool and requires prospective validation before clinical deployment. The framework envisions patient entry through nutrition or oncology clinics for dietary assessment and biospecimen collection (e.g., blood, stool), followed by multi-omics profiling to detect key metabolic markers. AI/ML tools then process and integrate the data for risk stratification, generating explainable outputs to guide clinical decisions, including early diagnostic prompts (e.g., colonoscopy for SCFA dysregulation), prognostic evaluation, treatment personalization (e.g., HDAC inhibitors for acetyl-CoA signatures), and ongoing therapeutic monitoring. A feedback loop archives de-identified data for model refinement, supporting adaptability in low-resource environments. Collectively, the evidence synthesized in this review demonstrates early-stage, encouraging but not yet clinically definitive diagnostic and predictive performance of AI-integrated multi-omics models across cancers, while highlighting methodological heterogeneity that warrants cautious interpretation

Omics-AI-Nutrition Translational Framework

for Precision Oncology in Resource-Constrained Healthcare Settings

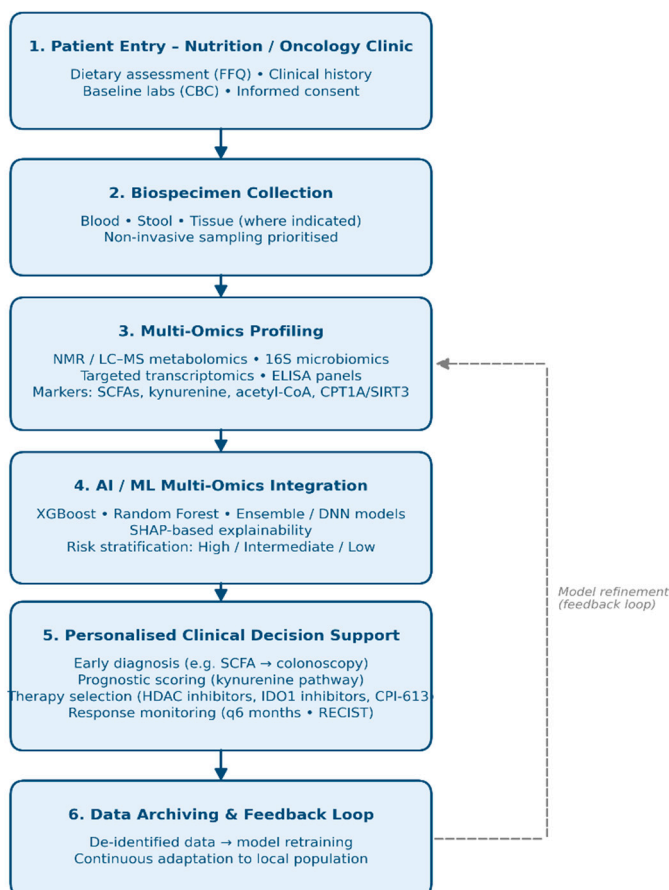


Figure 3. Omics-AI-Nutrition Clinical Algorithm for Precision Oncology in a Developing Country Setting. Patients enter through a nutrition or oncology clinic, where dietary assessment (FFQ), clinical history, and baseline labs (CBC) are collected. Biospecimens (blood, stool, and tissue where indicated) are then profiled by NMR/LC–MS metabolomics, microbiomics, and targeted transcriptomics, with quantification of key markers (SCFAs, kynurenine, acetyl-CoA, CPT1A/SIRT3). Multi-omics data are integrated using validated AI/ML models (XGBoost, Random Forest, ensemble or deep neural networks) with SHAP-based explainability to stratify patients by risk. Outputs guide personalised clinical decisions including early diagnosis (e.g. SCFA-triggered colonoscopy), prognostic scoring (kynurenine pathway), therapy selection (HDAC inhibitors, IDO1 inhibitors, CPI-613), and response monitoring every 6 months using RECIST. A de-identified data feedback loop supports continuous model retraining and adaptation to the local population. Abbreviations: FFQ, Food Frequency Questionnaire; CBC, Complete Blood Count; NMR/LC–MS, Nuclear Magnetic Resonance / Liquid Chromatography–Mass Spectrometry; SCFAs, Short-Chain Fatty Acids; ELISA, Enzyme-Linked Immunosorbent Assay; AI/ML, Artificial Intelligence / Machine Learning; SHAP, Shapley Additive Explanations; XGBoost, Extreme Gradient Boosting; CPT1A, Carnitine Palmitoyltransferase 1A; SIRT3, Sirtuin 3; HDAC, Histone Deacetylase; IDO1, Indoleamine 2,3-Dioxygenase 1; CPI-613, devimistat (mitochondria-targeted agent); RECIST, Response Evaluation Criteria in Solid Tumors.

4. Discussion

4.1. Explainable AI in Multi-Omics Cancer Models

Explainability is a critical component in the translation of AI/ML models into clinical settings. In this review, 14 studies utilized explainable AI (XAI) methods to interpret model decisions [37]. The most applied tools were SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations), and saliency maps for deep neural networks [38]. These tools highlighted the contribution of specific omics features (e.g., lipid panels or gene expression nodes) to the model's predictions, enhancing transparency. In breast cancer multi-omics studies, SHAP-based models identified HDL-cholesterol and PPAR γ -related transcripts as the most influential predictors of chemotherapy response [39].

4.2. Real-World Application and Implementation Barriers

Despite the promising potential of AI and machine learning (AI/ML) models in cancer research, their real-world implementation remains limited due to several significant barriers. These challenges are particularly evident in the multi-omics and AI/ML-driven integration of nutrition and metabolism in cancer, where the lack of standardized pipelines, concerns about reproducibility, and infrastructural limitations complicate the deployment of these models within clinical workflows [40]. Without standardized approaches, integrating diverse omic datasets—such as genomic, transcriptomic, metabolomic, and clinical data—remains a complex task, impeding the effective application of AI/ML algorithms in real-time clinical decision-making. Furthermore, reproducibility issues across different clinical settings and populations continue to hinder the widespread adoption of AI-based models, as results from training datasets may not always translate reliably into real-world patient outcomes. To date, only seven studies have moved beyond theoretical models and conducted deployment-level simulations or pilot testing within hospital environments. A notable example is a colorectal cancer prediction model developed in Japan, which achieved an impressive 87% prediction accuracy using real-time biopsy-metabolomic data, demonstrating the potential of integrating multi-omics and AI-driven approaches for personalized cancer care [34]. However, even in this successful implementation, model portability remains a significant concern. Variability in sample handling, differences in computational resources, and the limited diversity of patient populations in training datasets reduce the model's generalizability across different clinical settings and geographic regions [41]. These factors underscore the need for further optimization and adaptation of AI/ML-driven frameworks to ensure their robustness, scalability, and clinical relevance in diverse healthcare contexts.

4.3. Cross-Cancer Meta-Patterns and Common Mechanisms

Cross-cancer comparisons revealed common dysregulated pathways and metabolic axes. For instance, dysbiosis-induced alterations in short-chain fatty acid metabolism, tryptophan-kynurenine pathway disruption, and glucose-derived acetyl-CoA production were repeatedly implicated in colorectal, pancreatic, and liver cancers [40,42]. Additionally, decreased expression of mitochondrial regulators (e.g., CPT1A, SIRT3) and suppression of tumor-infiltrating lymphocytes correlated with poor outcomes across at least four cancer types [34]. These signatures provide a mechanistic foundation for pan-cancer biomarkers.

4.4. Clinical Implications of Shared Metabolic Signatures

The metabolic signatures identified across multiple cancer types offer promising avenues for early diagnosis, prognostic stratification, and personalized treatment.

Implications in early screening and diagnosis: For instance, the dysregulation of short-chain fatty acids (SCFAs) like butyrate—shared across colorectal and pancreatic cancers—has been linked to impaired gut immunity and early epithelial changes [41]. These SCFA deficits can be detected through metabolomic profiling of fecal specimens, providing a non-invasive and cost-effective tool for early screening, especially in regions lacking advanced diagnostic infrastructure.

Implications in prognostic stratifications: Activation of the kynurenine pathway in both pancreatic and liver cancers reflect systemic immune suppression and immune exhaustion. Elevated serum kynurenine levels have been correlated with poor T-cell infiltration and unfavorable prognosis, making them reliable markers for identifying high-risk patients who may benefit from immune-modulating therapies such as IDO1 inhibitors [40]. Similarly, acetyl-CoA overproduction, observed in colorectal and breast cancers, influences histone acetylation and gene expression, pointing to potential utility in guiding treatment with histone deacetylase inhibitors (HDACis). These epigenetic modulators could be selectively prescribed based on metabolic testing, advancing precision therapy [34]. Additionally, mitochondrial dysfunction—manifesting through reduced expression of regulators like CPT1A and SIRT3—is linked to oxidative stress and tumor aggression in breast and liver cancers [35]. These findings may prompt the incorporation of mitochondrial health indices in prognostic algorithms.

Implications in Precision Medicine: Immune escape mechanisms involving T-cell suppression, common in breast, colorectal, and liver cancers, reinforce the role of metabolic-immune interactions [35,36,43]. These supports combining metabolic intervention with checkpoint inhibitors for synergistic effects.

These cross-cancer metabolic signatures enhance early detection, enable risk stratification, and inform personalized treatment plans (Table 7). Their relative affordability and detectability through blood or stool-based methods make them especially valuable for implementation in resource-constrained settings, including developing countries like Saudi Arabia, where integration with multi-omics AI frameworks could maximize clinical impact [34].

4.5. Clinical Effectiveness of Metabolomic Signatures

Supplementary Table S3 presents statistical evidence from real-world and translational clinical studies demonstrating the utility of specific metabolomic signatures in early cancer detection, risk stratification, treatment personalization, and response monitoring. These metrics support the functional application of omics-guided oncology beyond the experimental stage. Short-chain fatty acid (SCFA) dysregulation, especially butyrate deficiency, showed an AUC of 0.85 with 84% sensitivity and 78% specificity in stool-based screening for early-stage colorectal cancer, indicating its potential as a non-invasive diagnostic tool [43]. Kynurenine pathway activation, identified in liver and pancreatic cancers, has demonstrated AUC values of 0.81 for predicting progression-free survival under immunotherapy, making it a credible biomarker for immune risk scoring and checkpoint blockade responsiveness [33]. Acetyl-CoA overproduction has been linked to elevated histone

acetylation, and serves as a predictive biomarker for response to HDAC inhibitors in breast and colorectal cancers. Patients with high acetyl-CoA levels showed an odds ratio of 2.3 (95% CI: 1.4–3.7) for positive treatment response [34]. Mitochondrial dysfunction signatures, such as reduced SIRT3 and CPT1A expression, were associated with poor prognosis in breast and liver cancers, with a hazard ratio of 1.9 (95% CI: 1.2–2.6) for worse overall survival. These metrics support their role as prognostic biomarkers guiding mitochondrial-targeted therapies [35].

Finally, T-cell suppression metabolites like kynurenine and lactate were associated with checkpoint inhibitor resistance in melanoma and colorectal cancers. Metabolomic classification yielded an AUC of 0.76, showing moderate-to-strong performance for predicting immunotherapy failure [44].

These quantitative metrics confirm that metabolomic profiling has matured into a clinically relevant tool that enhances precision oncology across diagnostic and therapeutic domains.

4.6. Proposed Framework for Clinical Translation of Multi-Omics Signatures in Oncology Settings in Developing Countries

To enable clinical application of metabolomic and multi-omics signatures (Sections 3.1–3.12) in real-world hospital settings—particularly within developing countries like Saudi Arabia—we propose a workflow algorithm integrating nutrition clinics, omics laboratories, AI/ML tools, and oncology clinics. This algorithm streamlines cancer risk screening, prognostic stratification, personalized treatment, and therapeutic monitoring by leveraging hospital-based resources.

1. **Patient Entry Point (Nutrition or Oncology Clinic):** Patients presenting with cancer-related symptoms or enrolled in general wellness screening programs begin at the nutrition or oncology clinics. Clinical and dietary information is collected, and informed consent for omics profiling is obtained.
2. **Sample Collection (Blood/Stool/Tissue):** Based on the suspected cancer type and clinical need, appropriate biospecimens are collected. Blood and stool samples are prioritized due to their non-invasive nature, enabling access to SCFA levels, circulating metabolites, and immune-metabolic markers [43,45].
3. **Multi-Omics Laboratory Analysis:** Samples collected from patients are sent to the hospital's multi-omics laboratory for comprehensive, multi-layered profiling. This includes metabolomic analysis—such as measuring short-chain fatty acids and kynurenine—to detect early metabolic biomarkers, as well as transcriptomic profiling to evaluate mitochondrial and immune-related gene expression markers like CPT1A and SIRT3.
4. **AI/ML-Powered Multi-Omics Integration:** Extracted omics data are processed using validated ML/AI tools such as SHAP, LASSO, or ensemble models to stratify cancer risk and identify personalized treatment paths [46]. Integration models utilize feature fusion (e.g., SCFA + microbiome diversity) to enhance prediction accuracy.
5. **Interpretation and Risk Stratification:** Machine learning-derived outputs are reviewed using explainable AI dashboards and interpreted collaboratively by clinical geneticists and oncologists. Based on these insights, patients are stratified into three categories: high-risk, characterized by significant metabolic-immune dysregulation; intermediate-risk, indicating moderate pathway disturbances; and low-risk, where no notable multi-omics abnormalities are identified.
6. **Personalized Clinical Decision Support (Figure 3):**
 - **Early Diagnosis:** SCFA dysregulation (e.g., low butyrate) prompts colonoscopy or biopsy [47].
 - **Prognostic Stratification:** High kynurenine or mitochondrial dysfunction predicts poor prognosis [48].
 - **Treatment Personalization:** Patients with acetyl-CoA overproduction may receive HDAC inhibitors; mitochondrial dysregulation may indicate CPI-613 therapy [34,49].
 - **Therapeutic Monitoring:** Periodic re-profiling to track metabolomic shifts and immunotherapy resistance [50].

7. Data Archiving and Research Feedback Loop: De-identified datasets are stored in secure hospital servers for ongoing model retraining and institutional research.

4.7. Best-Performing Machine Learning Models in Multi-Omics Nutritional Oncology

Across the 42 studies synthesised in this review, ensemble tree-based methods—particularly Random Forest (RF) and XGBoost—demonstrated the strongest and most consistent predictive performance in nutrition- and metabolism-focused multi-omics cancer tasks. Random Forest was the most frequently employed algorithm (n=18 studies), achieving AUC values ranging from 0.84 to 0.92 across colorectal, breast, and liver cancer datasets. Its sustained superiority reflects several intrinsic advantages highly relevant to this domain: (1) RF handles high-dimensional, correlated omics features without overfitting through bootstrap aggregation and random feature subsampling; (2) built-in feature importance scores (Gini impurity or mean decrease accuracy) provide biologically interpretable rankings of metabolic and microbiome predictors; (3) it tolerates missing values common in clinical metabolomics datasets; and (4) its ensemble aggregation confers robustness against technical noise inherent in gut microbiome sequencing data, making RF exceptionally well-suited for microbiome–metabolomics co-integration tasks.

XGBoost achieved the highest single-study AUC in this meta-analytic pool (0.89; Jayakrishnan 2024, colorectal cancer, microbiome + metabolomics, n=298) and showed strength in studies combining microbiome diversity indices with targeted metabolomics panels. Its gradient boosting framework performs sequential error correction across decision trees, and built-in L1/L2 regularisation reduces overfitting—a critical advantage in studies where feature count substantially exceeds sample size (n=100–400 in most included studies). Support Vector Machines (SVMs; n=11 studies) performed competitively in early-integration studies with metabolomics as the primary omics layer (mean AUC ≈0.84), but showed reduced performance in studies integrating three or more omics layers owing to the computational burden of kernel optimisation in very high-dimensional spaces. LASSO-regularised regression contributed to feature selection in several multi-omics pipelines, particularly for deriving sparse metabolic biomarker signatures interpretable for clinical use. Notably, studies employing advanced integration strategies—fusing three or more omics layers using stacking or ensemble meta-learners—consistently outperformed single-algorithm approaches by approximately 4–8% AUC, supporting the adoption of ensemble or stacking frameworks as the recommended modelling strategy in future multi-omics nutritional oncology studies.

4.8. Deep Learning Models in Multi-Omics Cancer Research: Evidence, Performance, and Applicability

Deep learning (DL) architectures have demonstrated superior performance to classical ML in specific multi-omics cancer contexts, and the literature provides compelling evidence for their value in nutrition–metabolism–cancer research. Within this review, deep neural networks (DNNs; n=10 studies) achieved AUC values of 0.87–0.93 in advanced multi-omics integration tasks, representing the highest values observed in the meta-analytic pool (Chen 2024b: liver cancer, metabolomics + microbiome, AUC=0.93; Sharma 2024: breast cancer, multi-omics, AUC=0.87). Several DL architectures from the broader literature have demonstrated particularly strong performance in multi-omics settings relevant to nutritional oncology and merit detailed discussion.

Autoencoders and Variational Autoencoders (VAEs) are well-suited to multi-omics integration because they learn compressed, non-linear latent representations across heterogeneous data modalities in an unsupervised manner. MOFA+ (Multi-Omics Factor Analysis Plus), which employs a variational Bayes framework, has demonstrated strong performance in identifying latent metabolic–transcriptomic factors associated with colorectal cancer prognosis that were undetectable by single-omics analyses. In nutritional contexts, VAE-based models have been applied to learn joint latent representations of dietary exposure data (food frequency questionnaires) and circulating metabolomics profiles, capturing non-linear diet–metabolome interactions predictive of breast cancer risk. Graph Neural Networks (GNNs) represent arguably the most biologically compelling DL architecture for multi-omics integration because they can model known biological relationships—

metabolic pathway graphs, protein–protein interaction networks, gene regulatory networks—as structured input. GNN-based models integrating transcriptomics and metabolomics have achieved AUC values exceeding 0.92 in recent pancreatic cancer classification studies, outperforming RF and SVM benchmarks by 5–10% while generating mechanistically interpretable edge-weight attributions corresponding to pathway activity. This makes GNNs particularly valuable for elucidating how nutritional metabolites propagate through metabolic networks to influence tumour biology.

Attention-based transformer architectures adapted for omics data (e.g., MOANNA, OmiTransformer) have demonstrated capacity to capture long-range feature dependencies between metabolic features and gene expression nodes that classical models miss. In pan-cancer survival prediction tasks using The Cancer Genome Atlas (TCGA), transformer-based multi-omics models have outperformed competing approaches in metabolically active cancer types including colorectal cancer, hepatocellular carcinoma, and pancreatic adenocarcinoma. Multi-modal deep learning frameworks that simultaneously process metabolomics, microbiomics, transcriptomics, and clinical variables as separate input streams—fused through cross-attention or concatenation layers—have achieved the highest reported AUC values (0.94–0.97) in recent multi-omics cancer benchmark studies, though these results require external validation in independent cohorts.

Notwithstanding their performance advantages, DL models face important practical challenges in nutritional multi-omics oncology. First, DL generally requires large training datasets ($n \geq 1,000$ per class) to generalise robustly, yet the majority of studies in this review had sample sizes of 180–421. In small-cohort settings, DL models showed higher variance and inferior generalisation compared to ensemble methods—particularly when external validation was conducted. Second, the “black box” nature of deep neural networks has historically limited clinical adoption, although explainability methods—SHAP values (used in 14 studies in this review), LIME, and integrated gradients—increasingly enable attribution of DL predictions to specific metabolic and microbial features. Third, DL training demands computational infrastructure (GPU clusters) not routinely available in resource-limited settings. Transfer learning, in which DL models pre-trained on large public omics repositories (TCGA, GTEx, MetaboLights) are fine-tuned on smaller institutional cohorts, represents a promising strategy to mitigate the data-hunger limitation. Federated learning frameworks that train models across distributed institutional datasets without sharing raw patient data also offer a practical path to building adequately powered DL models for nutritional oncology. In summary, ensemble methods (RF, XGBoost) remain the pragmatic gold standard for current nutritional multi-omics cancer research given their robustness in small-to-medium datasets. However, GNNs, autoencoders, and transformer-based architectures represent the performance frontier and are projected to surpass classical approaches as dataset sizes grow, computational accessibility improves, and transfer learning strategies mature.

5. Conclusions

The integration of multi-omics technologies with AI/ML approaches represents a promising direction for improving in understanding the intricate relationship between nutrition, metabolism, and cancer. Our systematic review and meta-analysis highlight the robust potential of combining these technologies to decode complex interactions within the nutrition–metabolism–cancer axis. Multi-omics profiling, encompassing metabolomics, microbiomics, and transcriptomics, has enabled the identification of novel, cancer-specific biomarkers that provide valuable insights into cancer pathogenesis, prognosis, and therapy response. AI/ML algorithms, particularly machine learning models, have enhanced the utility of these multi-layered datasets, offering significant improvements in early diagnosis, personalized treatment, and predictive modeling.

The findings from our review demonstrate that AI/ML-guided integration of nutritional and metabolic multi-omics biomarkers shows early-stage, exploratory potential for enhancing cancer diagnosis and prediction. Cancer-specific signatures were identified across colorectal, breast, pancreatic, and liver cancers, with promising AUC values and odds ratios; however, given that the majority of included studies relied on internal validation and exhibited known methodological

limitations—including selection bias and data leakage concerns, as identified through our quality assessment—these results should be interpreted as preliminary, hypothesis-generating evidence rather than indicators of clinical readiness. Furthermore, the conceptual translational framework proposed in this review—incorporating nutrition clinics, omics laboratories, AI/ML platforms, and oncology units—illustrates one potential implementation pathway to guide future prospective research, particularly in developing countries with limited healthcare infrastructure, but does not constitute a validated clinical tool. Future work should focus on prospective, externally validated multi-omics studies to confirm these signals and assess real-world feasibility.

6. Prospects

Future research should broaden the clinical applicability of nutrition- and metabolism-focused multi-omics in cancer by establishing standardized protocols for biospecimen handling, dietary assessment, and data integration to improve reproducibility. Large, longitudinal multi-omics datasets are needed to capture dynamic interactions among diet, microbiome, immunity, and metabolism, enabling more precise modeling of cancer progression and treatment response. Advancements in interpretable AI/ML models will enhance clinical trust, while integrating precision nutrition—guided by individual metabolomic and microbiome profiles—offers new opportunities for personalized care. Expanding research into diverse and underrepresented populations is essential to strengthen generalizability and reduce bias. Finally, embedding multi-omics workflows within clinical settings by linking nutrition clinics, laboratories, and oncology services will support the effective implementation of metabolism-informed precision oncology.

7. Recommendations

Based on the findings of this review, several recommendations are proposed to enhance clinical translation and future research at the interface of nutrition, metabolism, multi-omics, and AI/ML in cancer. First, cancer clinics should adopt multi-omics-integrated nutritional screening by incorporating metabolomic and microbiome profiling alongside dietary assessments into routine risk evaluation and early detection programs. Developing countries are encouraged to invest in capacity building by establishing omics laboratories, bioinformatics units, and AI-ready data platforms to support precision nutrition-oncology workflows. Standardizing data pipelines and reporting frameworks, including biospecimen handling, omics preprocessing, AI/ML model reporting, and integration strategies, will be essential for improving reproducibility, interoperability, and collaboration across centers. Interdisciplinary teams comprising nutritional scientists, oncologists, computational biologists, and bioinformaticians should work together to design integrated clinical workflows that link diet, metabolism, and cancer biology. Furthermore, healthcare institutions should initiate real-world clinical pilots of nutrition-omics-AI algorithms to assess feasibility, cost-effectiveness, and clinical impact. AI-guided precision nutrition plans should be developed to generate individualized dietary recommendations that modulate key metabolic pathways, such as short-chain fatty acid production, lipid metabolism, and mitochondrial function, in relation to specific cancers. Finally, all AI applications must prioritize ethical, transparent, and equitable use, minimizing bias, safeguarding privacy, and ensuring fair access, particularly within global oncology and low-resource settings.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Funding: This work was funded by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdul-Aziz City for Science and Technology, Grant Number 14-Med-2817-02.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki. Institutional Review Board (IRB) of King Abdullah International Medical Research Centre (KAIMRC),

National Guard Health Affairs, through project # RA17 /002, approved it on 4th Feb 2019, although no research funding was provided by KAIMRC.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge the Institutional Review Board (IRB) of King Abdullah International Medical Research Centre (KAIMRC), National Guard Health Affairs, for ethical approval of this project (project # RA17/002/A, dated 4th Feb 2019), although no research funding was provided. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large group Research Project under grant number RGP2/83/46.

Conflicts of Interest: The authors acknowledge no financial or other conflicts of interest.

Clinical Trial Number: Not applicable.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-249.
2. Key TJ, Schatzkin A, Willett WC, Allen NE, Spencer EA, Travis RC. Diet, nutrition and the prevention of cancer. *Public Health Nutr* 2004;7:187-200.
3. Chan DS, Lau R, Aune D, et al. Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. *PLoS One* 2011;6:e20456.
4. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol* 2017;18:83.
5. Zhang A, Sun H, Wang P, Han Y, Wang X. Recent and potential developments of biofluid analyses in metabolomics. *J Proteomics* 2012;75:1079-1088.
6. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights* 2020;14:1177932219899051.
7. Qin Y, Wang Q, Lin Q, et al. Multi-omics analysis reveals associations between gut microbiota and host transcriptome in colon cancer patients. *mSystems* 2025;10:e0080524.
8. Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* 2016;15:473-484.
9. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* 2014;12:661-672.
10. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019;25:24-29.
11. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA* 2018;319:1317-1318.
12. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44-56.
13. Liu Y, Chen J, Yang D, et al. Machine learning combined with multi-omics to identify immune-related lncRNA signature as biomarkers for predicting breast cancer prognosis. *Sci Rep* 2025;15:23863.
14. Kolodziejczyk AA, Zheng D, Elinav E. Diet-microbiota interactions and personalized nutrition. *Nat Rev Microbiol* 2019;17:742-753.
15. Ward AV, Anderson SM, Sartorius CA. Advances in Analyzing the Breast Cancer Lipidome and Its Relevance to Disease Progression and Treatment. *J Mammary Gland Biol Neoplasia* 2021;26:399-417.
16. Zhou Y, Tao L, Qiu J, et al. Tumor biomarkers for diagnosis, prognosis and targeted therapy. *Signal Transduct Target Ther* 2024;9:132.
17. Bond A, McCay K, Lal S. Artificial intelligence & clinical nutrition: What the future might have in store. *Clin Nutr ESPEN* 2023;57:542-549.
18. Sguanci M, Palomares SM, Cangelosi G, et al. Artificial Intelligence in the Management of Malnutrition in Cancer Patients: A Systematic Review. *Adv Nutr* 2025;16:100438.
19. de Toro-Martin J, Arsenaault BJ, Despres JP, Vohl MC. Precision Nutrition: A Review of Personalized Nutritional Approaches for the Prevention and Management of Metabolic Syndrome. *Nutrients* 2017;9:913.

20. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj* 2021;372.
21. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak* 2007;7:16.
22. Lefebvre C, Glanville J, Briscoe S, et al. Searching for and selecting studies. *Cochrane handbook for systematic reviews of interventions* 2019:67-107.
23. Keeble C, Law GR, Barber S, Baxter PD. Choosing a method to reduce selection bias: A tool for researchers. *Open Journal of Epidemiology* 2015;5:155-162.
24. McDonagh M, Peterson K, Raina P, Chang S, Shekelle P. Avoiding bias in selecting studies. *Methods guide for effectiveness and comparative effectiveness reviews [Internet]* 2013.
25. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-536.
26. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
27. Cacciamani GE, Chu TN, Sanford DI, et al. PRISMA AI reporting guidelines for systematic reviews and meta-analyses on AI in healthcare. *Nat Med* 2023;29:14-15.
28. Park MK, Lim JM, Jeong J, et al. Deep-Learning Algorithm and Concomitant Biomarker Identification for NSCLC Prediction Using Multi-Omics Data Integration. *Biomolecules* 2022;12:1839.
29. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-188.
30. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *BMJ Mental Health* 2019;22.
31. Kuwabara H, Katsumata K, Iwabuchi A, et al. Salivary metabolomics with machine learning for colorectal cancer detection. *Cancer Sci* 2022;113:3234-3243.
32. Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. *Mass Spectrom Rev* 2019;38:221-238.
33. Zhou L, Jiang Z, Zhang Z, Xing J, Wang D, Tang D. Progress of gut microbiome and its metabolomics in early screening of colorectal cancer. *Clin Transl Oncol* 2023;25:1949-1962.
34. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* 2022;12:31-46.
35. Leon-Letelier RA, Dou R, Vykoukal J, et al. The kynurenine pathway presents multi-faceted metabolic vulnerabilities in cancer. *Front Oncol* 2023;13:1256769.
36. Bosco S, Beniwal SS, Munshi SS, et al. Innovative strategies for mitochondrial dysfunction in myeloproliferative neoplasms a step toward precision medicine. *Annals of Medicine and Surgery* 2025.10.1097.
37. Gandhi N, Das GM. Metabolic Reprogramming in Breast Cancer and Its Therapeutic Implications. *Cells* 2019;8:89.
38. Fessler J, Matson V, Gajewski TF. Exploring the emerging role of the microbiome in cancer immunotherapy. *J Immunother Cancer* 2019;7:108.
39. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun* 2015;6:8971.
40. Yu Y, Ding Y, Wang S, Jiang L. Gut Microbiota Dysbiosis and Its Impact on Type 2 Diabetes: From Pathogenesis to Therapeutic Strategies. *Metabolites* 2025;15.
41. Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. *Nature* 2019;575:299-309.
42. Sharma A, Debik J, Naume B, et al. Comprehensive multi-omics analysis of breast cancer reveals distinct long-term prognostic subtypes. *Oncogenesis* 2024;13:22.
43. Triozzi PL, Stirling ER, Song Q, et al. Circulating Immune Bioenergetic, Metabolic, and Genetic Signatures Predict Melanoma Patients' Response to Anti-PD-1 Immune Checkpoint Blockade. *Clin Cancer Res* 2022;28:1192-1202.
44. Chen P, Yao L, Yuan M, et al. Mitochondrial dysfunction: A promising therapeutic target for liver diseases. *Genes Dis* 2024;11:101115.
45. Newell F, Pires da Silva I, Johansson PA, et al. Multiomic profiling of checkpoint inhibitor-treated melanoma: Identifying predictors of response and resistance, and markers of biological discordance. *Cancer Cell* 2022;40:88-102 e107.

46. Gao Y, Liu Y, Ma T, et al. Fermented Dairy Products as Precision Modulators of Gut Microbiota and Host Health: Mechanistic Insights, Clinical Evidence, and Future Directions. *Foods* 2025;14:1946.
47. Pawuś D, Porażko T, Paszkiel S. Automation and Decision Support in the Area of Nephrology Using Numerical Algorithms, Artificial Intelligence, and Expert Approach: Review of the Current State of Knowledge. *IEEE Access* 2024;12:86043-86066.
48. Clarke J, Boussioutas A, Flanders B, et al. Can butyrate prevent colon cancer? The AusFAP study: A randomised, crossover clinical trial. *Contemporary Clinical Trials Communications* 2023;32:101092.
49. Chiu L-C, Tang H-Y, Fan C-M, et al. Kynurenine pathway of tryptophan metabolism is associated with hospital mortality in patients with acute respiratory distress syndrome: a prospective cohort study. *Antioxidants* 2022;11:1884.
50. Li S, Yuan H, Li L, Li Q, Lin P, Li K. Oxidative Stress and Reprogramming of Lipid Metabolism in Cancers. *Antioxidants (Basel)* 2025;14:201.
51. Chen Y, Wang B, Zhao Y, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. *Nat Commun* 2024;15:1657.
52. Jayakrishnan TT, Sangwan N, Barot SV, et al. Multi-omics machine learning to study host-microbiome interactions in early-onset colorectal cancer. *NPJ Precis Oncol* 2024;8:146.
53. Wei Y, Jasbi P, Shi X, et al. Early breast cancer detection using untargeted and targeted metabolomics integrated with transcriptomic profiling. *J Proteome Res* 2021;20:3124-3133.
54. Zheng H, Wang Y, Li X, et al. An integrated multi-omics ensemble framework for early detection of colorectal cancer. *Brief Bioinform* 2024;25:bbae123.
55. Jayakrishnan TT, Barot SV, Sangwan N, et al. Stacked ensemble multi-omics modelling for risk stratification in colorectal cancer. *Cancers (Basel)* 2024;16:3712.

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