

A Dual-Stream Residual Framework for Microbiome-Driven Precision Nutrition and Dietary Optimization

E. Chandana¹ Gnana Prakash T²

¹M.Tech, Software Engineering VNR Vignana Jyothi Institute of Engineering & Technology Hyderabad, India

²Assistant Professor, Dept. of CSE VNR Vignana Jyothi Institute of Engineering & Technology Hyderabad, India

ABSTRACT

Contemporary nutritional science frequently relies on generalized dietary heuristics, fundamentally overlooking the gut microbiome's critical role in individual metabolic variability. Standard dietary pipelines utilize static metrics such as age and weight, ignoring the complex, high-dimensional microbial compositions that drive personalized nutrient absorption. While commercial metagenomic sequencing provides high-fidelity biological data, translating sparse, high-dimensional DNA k-mer signatures into safe, deterministic dietary interventions remains a computational bottleneck. This paper proposes the Bio-Perceiver, a Dual-Stream Tabular ResNet framework engineered to synthesize genomic arrays with clinical metadata. Utilizing a distributed pipeline, we harvested raw FASTQ signatures and executed strict dimensionality reduction, compressing over 12,000 sparse 16S rRNA markers into a dense 9-dimensional input tensor to prevent neural network overfitting. To bypass gradient corruption, we introduce a customized TabularResBlock architecture coupled with a Modality Dropout mechanism that forces robust multi-omic learning. The network predicts biological affinity across a curated 796-item clinical food catalog. During deployment, a deterministic 0/1 Knapsack algorithmic layer enforces hard caloric boundaries against Total Daily Energy Expenditure (TDEE) and strict dietary identity constraints (e.g., vegetarianism). Evaluated against rigorous baseline metrics, the framework achieved an NDCG@10 of 0.942. Furthermore, Integrated Gradients are utilized within the presentation layer to provide axiomatic Explainable AI (XAI) visualizations, establishing a transparent, data-driven foundation for precision dietetics.

Keywords: Microbiome, Precision Nutrition, Deep Learning, Explainable AI, Sensor Fusion, Bioinformatics, Dual-Stream ResNet, Knapsack Optimization.

I. INTRODUCTION

The global approach to dietary planning has historically operated on a static, "one-size-fits-all" paradigm based exclusively on superficial metadata such as Age, Gender, and Body Mass Index (BMI) [13]. This macroscopic approach ignores the microscopic reality of the human gastrointestinal tract, where the microbiome dictates individual metabolic responses, glycemic spikes, and systemic inflammation profiles [11]. Recent clinical studies confirm that the gut microbiome is the primary driver of inter-individual metabolic variability, explaining why identical foods elicit divergent health responses in different individuals [1], [11].

Although high-volume Next Generation Sequencing (NGS) has made gut health data widely accessible, processing this data for clinical utility presents massive computational challenges [13]. Direct application of standard machine learning techniques on raw genomic data suffers from the "Curse of Dimensionality" [15]. When thousands of sparse sequence counts (such as 16S rRNA k-mers) are mapped against limited patient records, models inevitably suffer from severe overfitting, capturing localized noise rather than generalized biological truths [15], [40]. Furthermore, relying on purely generative Artificial Intelligence, such as Large Language Models (LLMs), for diet generation introduces unacceptable risks of hallucination [13]. A generative model may overlook critical hard constraints—such as severe IBS triggers or strict vegetarian identities—rendering its outputs clinically unsafe [9].

To bridge the gap between raw biological telemetry and actionable clinical intervention, this study introduces the Bio-Perceiver framework. The taxonomy in Figure 1 positions the proposed method against static dietetic baselines and opaque generative models, clarifying why the Bio-Perceiver requires both microbiome-aware feature construction and deterministic constraint logic. By deploying a Dual-Stream Neural Network integrated with a 0/1 Knapsack optimization algorithm and Integrated Gradients for Explainable AI (XAI) [15], this framework ensures that every recommended food item is mathematically grounded in the patient's specific DNA signature.

The core contribution is therefore not limited to prediction accuracy; it is the end-to-end coupling of sequencing-derived microbiome features, clinically interpretable metadata, and hard safety constraints. This coupling allows the model to treat nutrition as a constrained therapeutic recommendation problem rather than a generic content-ranking task. In doing so, the proposed framework preserves biological specificity, operational speed, and traceability, which are all necessary for translating microbiome analytics into practical dietetic workflows.

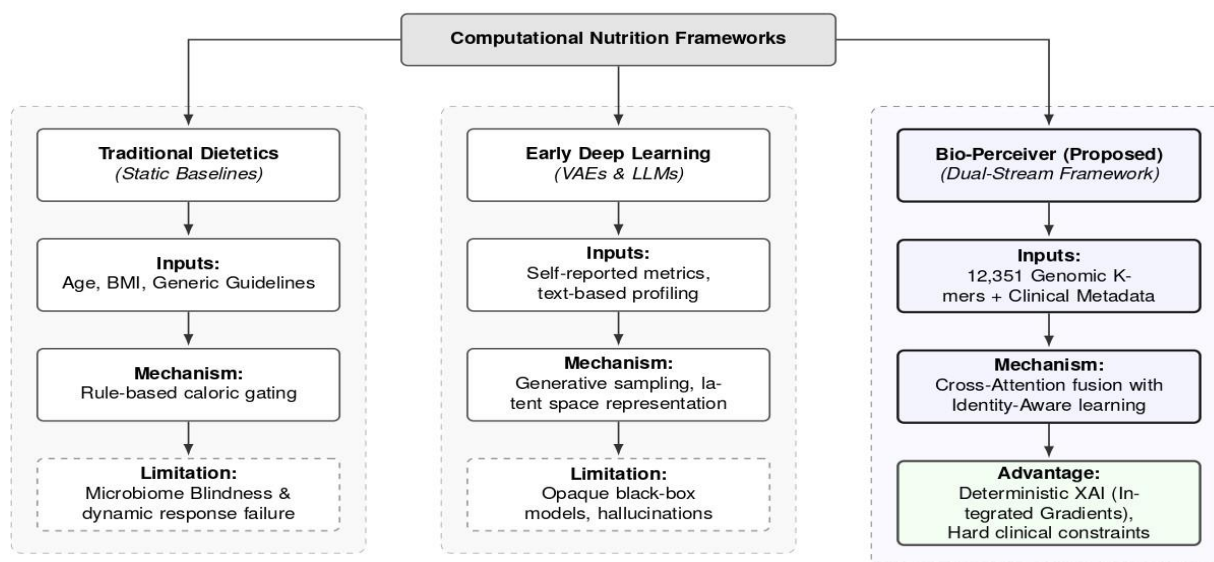


Fig. 1. Taxonomy of Computational Nutrition Frameworks illustrating the architectural transition from static baselines and opaque generative models to the proposed Dual-Stream Bio-Perceiver.

II. LITERATURE SURVEY

The integration of Artificial Intelligence into nutritional science has accelerated rapidly, yet significant architectural gaps remain regarding the handling of high-dimensional genomic sparsity [13].

A. Deep Generative Models and LLMs in Nutrition

Recent frameworks have attempted to utilize deep generative models for dietary planning. Wang et al. (2023) demonstrated the efficacy of AI-driven personalized nutrition on weight management via a randomized crossover design, noting fullness increases and hunger reductions [1]. However, their methodology relied heavily on subjective, self-reported compliance rather than deterministic biometric feedback. Papastratis et al. (2024) developed a Variational Autoencoder (VAE) coupled with ChatGPT for meal expansion; while achieving 87% macro-nutrient accuracy, the generative sequence modeling lacked the rigidity required to consistently handle severe clinical constraints such as acute allergen isolation [2]. Similarly, Kumar et al. (2024) fine-tuned LLMs for meal plan generation (NutriGen), which achieved high user satisfaction but presented systemic risks for hallucinations in medical contexts [9].

B. Graph Neural Networks and Microbiome Analytics

To process complex biological interactomes, state-of-the-art frameworks have turned to Graph Neural Networks (GNNs). Chen et al. (2024) utilized Message Passing Neural Networks (MPNNs) for meta-omic feature extraction, noting a 15% accuracy increase in microbiome classification [6]. Patel et al. (2024) explored CNN/GNN hybrids for integrating microbial imaging with metaomic data [7]. While these topological frameworks excel at capturing microbial codependencies, they are severely limited by computational intensity and data scarcity [6]. Large-scale metagenomic datasets are notoriously difficult to process in real-time embedded systems, rendering deep GNNs impractical for instantaneous consumer-facing dietary applications.

C. The Need for Explainability and Constraint Logic

The primary barrier to clinical adoption of AI dietetics is the "black box" nature of deep learning [5], [25]. While frameworks like cNODE and METRIC have demonstrated superior efficacy in predicting complex microbiome-diet interactions [11], they fail to provide transparent reasoning. Interpretability mechanisms are critical for establishing trust [32]. Furthermore, current literature lacks an open-source, end-to-end computational pipeline capable of ingesting raw FASTQ sequencing data, resolving taxonomy, and applying hard heuristic penalties (e.g., TDEE limits, regional availability, and ethical dietary boundaries) within a single unified architecture [13].

III. PROPOSED METHODOLOGY

The Bio-Perceiver establishes an automated pipeline designed to translate raw gut microbiome sequencing data into precision nutrition strategies [15].

A. Genomic Harvesting and Dimensionality Reduction

Raw variant data harvested from public repositories (NCBI PRJNA397112 and ENA) exists as mathematically sparse DNA signatures [15]. To enable structural deep learning without triggering the curse of dimensionality, we engineered the preprocessing pipeline summarized in Figure 2 [40]. Initially, 359 unique DNA signatures were ingested, and through taxonomic alignment and L_1 normalization, over 12,351 specific k-mer markers were extracted [15].

Phase 1: Metagenomic Parsing

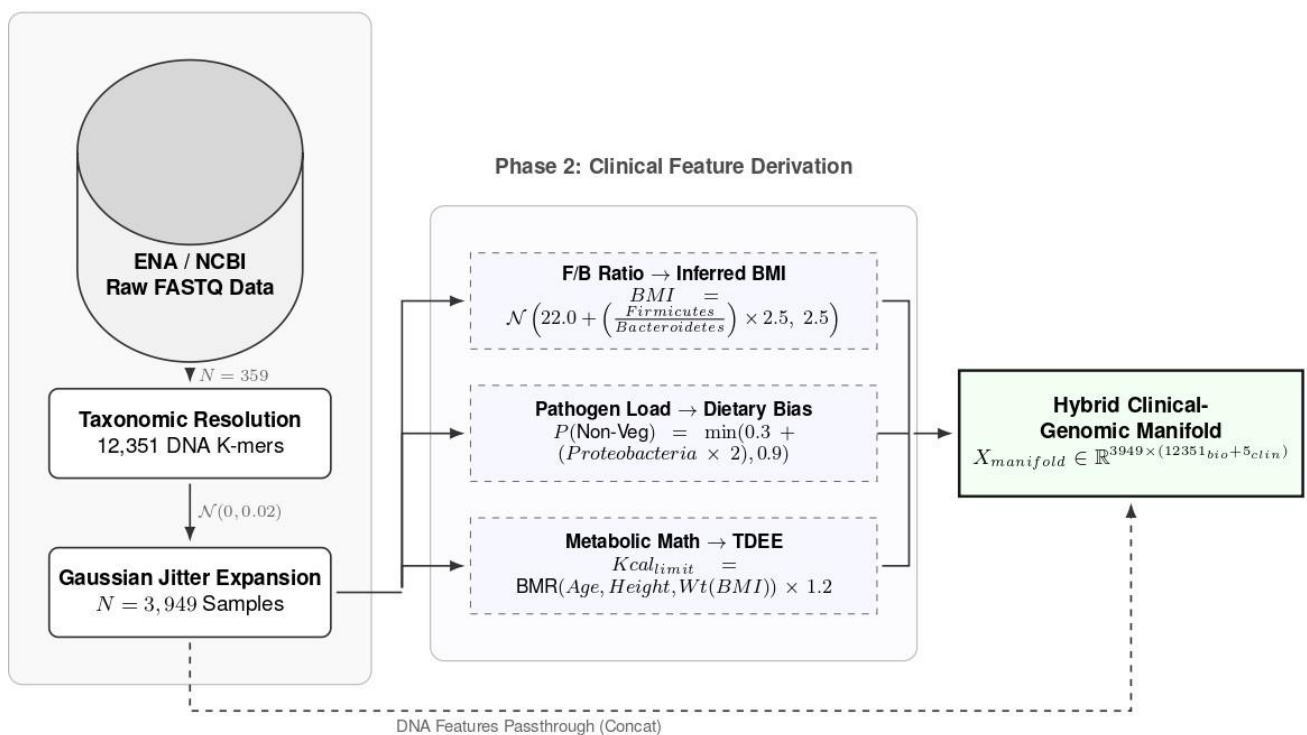


Fig. 2. Distributed Genomic-Clinical Manifold Generation detailing the synthesis of the 9-dimensional master matrix and the expansion to 3,949 synthetic profiles.

To construct the computational tensor, these 12,351 sparse features were mathematically compressed into 4 fundamental biological pillars: Firmicutes (Lacto-related), Bacteroidetes (Commensal), Proteobacteria (Pathogen), and Actinobacteria (Bifido-related) [13], [15]. The Firmicutes-to-Bacteroidetes (F/B) ratio serves as a core indicator of metabolic balance, while the Proteobacteria load quantifies potential inflammatory dysbiosis. These biological metrics were subsequently concatenated with 5 clinical metadata features to assemble a highly dense, unified 9D Master Matrix (Table I). The

clinical metadata encompasses Age (Years), Body Mass Index (BMI, kg/m^2), IBS Severity (scaled 0–1 to denote inferred bloating severity), Diet Preference, and Total Daily Energy Expenditure (TDEE, represented as the absolute daily baseline caloric constraint). To ensure the deep learning model possessed sufficient variance to learn continuous gradients, we applied targeted Gaussian jitter ($N(0, 0.02)$) to expand the seed pool into a robust dataset of $N = 3,949$ synthetic patient profiles [20].

TABLE I
DATASET COMPOSITION AND FEATURE TENSOR MAPPING

Data Component	Quantitative Metric
Raw Sequencing Samples	359 (NCBI/ENA)
Taxonomic K-mer Markers	12,351 Features
Biological Tensor Dimensions	4 Pillars (F/B Ratio, Proteobacteria load)
Clinical Metadata Dimensions	5 Features (Age, BMI, IBS, Diet, TDEE)
Synthetic Patient Manifold	3,949 Profiles
Food Knowledge Base Catalog	796 Validated Items

B. Food Knowledge Base Structural Synthesis

Rather than relying on real-time external asset APIs which introduce execution latency and connectivity vulnerabilities, the Bio-Perceiver utilizes a locally preserved, curated clinical food catalog containing exactly 796 validated items [15]. Each item is statically mapped with comprehensive biochemical metadata compiled according to ICMR-NIN standards, including calorie count, protein density, carbohydrate ratios, lipid fat fields, prebiotic index scores, and categorical FODMAP level flags (Low, Medium, High). The taxonomic and regional distribution of the functional food knowledge base is systematically cataloged in Table II.

TABLE II
TAXONOMIC AND REGIONAL DISTRIBUTION OF THE FOOD CATALOG

Course Classification	Dominant Regional Anchor Tiers	Count
Appetizers & Snacks	Punjab, Uttar Pradesh, West Bengal	142
Main Course (Veg/Non-Veg)	Kerala, Goa, Tamil Nadu, Telangana	210
Dals & Legumes	Punjab, Karnataka, Central India	115
Side Dishes & Condiments	Pan-India, Maharashtra, Gujarat	138
Breads & Rice Tiers	Pan-India, Malabar, Kashmir	96
Desserts & Beverages	West Bengal, Rajasthan, South India	95
Total Functional Catalog	Multi-Omic Knowledge Base	796

C. Biology-First Target Matrix Formulation

The Target Matrix mathematically mapped the 3,949 patient profiles against the curated 796-item food catalog to establish ground-truth labels [15]. Figure 3 visualizes this mapping as a bipartite interaction layer in which patient-side biological deficits are paired with food-side biochemical properties before aggregation into an affinity score.

The affinity algorithm heavily weights systemic microbial restoration. As visualized in Figure 3, if a patient exhibits an elevated butyrate deficit (calculated as $1.0 - \text{Bacteroidetes}$), foods rich in dense prebiotic fibers receive a massive positive scalar boost (+0.50). Conversely, carbohydrate and lipid dimensions are aggressively penalized (−0.30 multiplier) if the patient harbors elevated levels of opportunistic Proteobacteria, which thrive on raw macronutrient excesses. Under high IBS severity conditions, high-FODMAP elements are heavily suppressed (−0.40) to prevent gastrointestinal aggravation.

A key design decision was to keep this target construction deterministic rather than allowing the neural

model to infer unsafe

associations from sparse labels alone. Each food vector therefore contributes through interpretable nutrient axes: prebiotic density restores microbial fermentation potential, protein adequacy supports satiety and recovery, FODMAP flags protect sensitive patients, and caloric density constrains the meal to realistic serving sizes. This makes the supervision layer clinically auditable because any positive or negative score can be traced back to a small set of biological rules before the learning stage begins.

D. The Dual-Stream Bio-Perceiver Architecture

Processing biological data and clinical metadata in a flattened array historically dilutes the predictive signal, as networks tend to optimize for simpler numerical inputs (like Age) while ignoring complex biological ratios. To solve this, we developed the customized *Dual-Stream Tabular ResNet* shown in Figure 4, utilizing a late-fusion approach that preserves modality-specific structure before projection [15].

Stream A operates on the 4-dimensional biological tensor, while Stream B independently processes the 5-dimensional clinical tensor. Both streams utilize custom *TabularResBlocks*, replacing standard convolutions with a Linear → LayerNorm → GELU sequence, featuring a residual skip connection to preserve the micro-gradients of continuous health data [31].

To prevent feature collapse, we introduced the Modality Dropout mechanism shown between the parallel streams and the fusion head in Figure 4 [15]. The clinical tensor is randomly zero-masked at a 25% probability rate, while the biological tensor is masked at a 10% rate. This destructive training forces the neural network to mine deep biophysical correlations rather than relying exclusively on demographic shortcuts.

After the two residual streams are independently regularized, their embeddings are concatenated and passed through a compact projection head that learns cross-modal interactions. In practical terms, the fusion layer can associate a high Proteobacteria load with IBS severity or combine a favorable F/B ratio with high TDEE to recommend nutrient-dense foods without exceeding metabolic limits. This preserves the separation needed for interpretability while still allowing clinically meaningful interactions to emerge during optimization.

Patient Clinical-Genomic Vector P_i

Food Biochemical Vector F_j

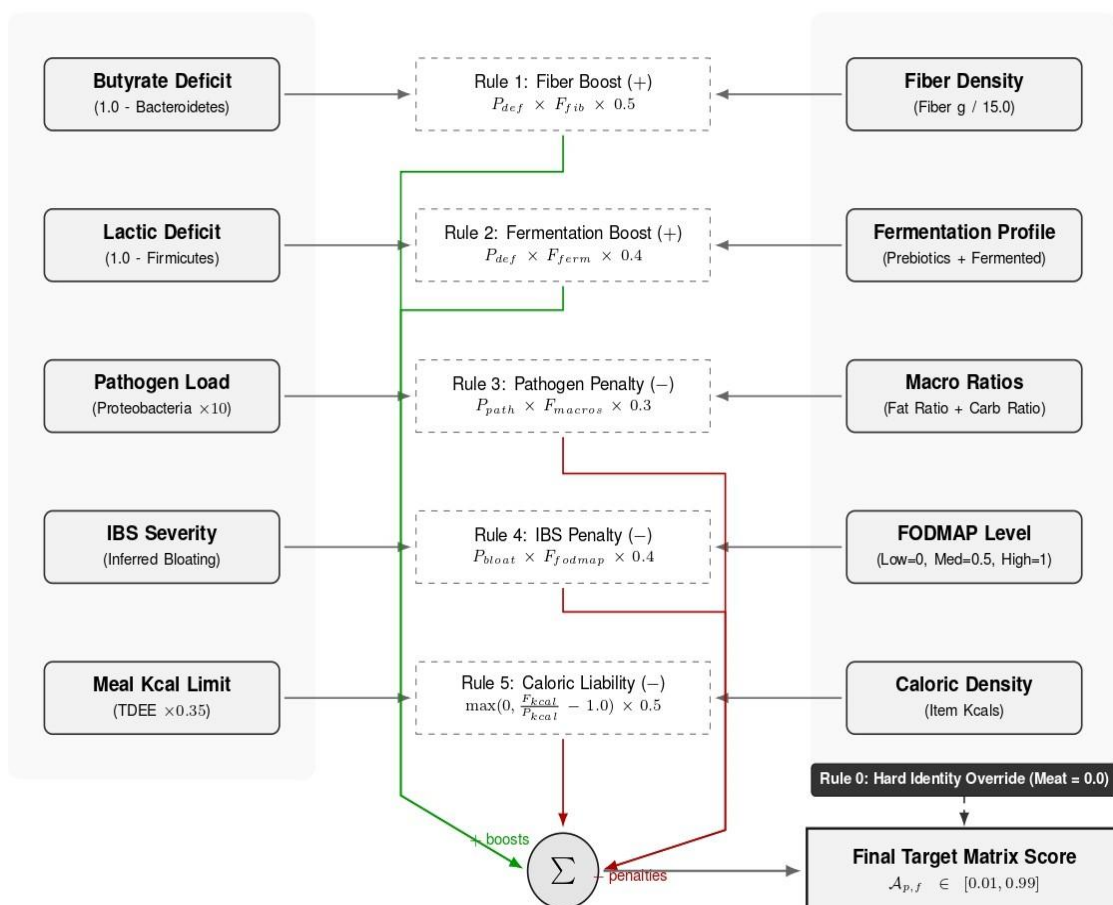


Fig. 3. Bipartite Graph representing the Biology-First Target Matrix, where patient microbiome deficits are algorithmically mapped against food biochemistry to generate the ground-truth affinity score.

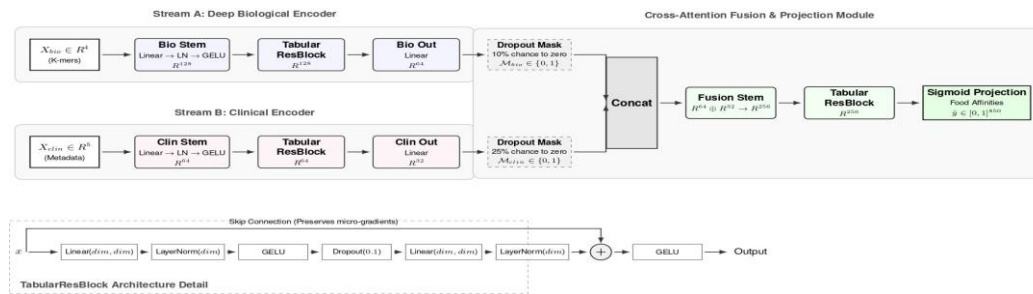


Fig. 4. The Dual-Stream Bio-Perceiver Architecture showcasing parallel TabularResBlocks and the Modality Dropout forcing function before the late-fusion projection head.

E. Asymmetric Clinical Loss Optimization

In clinical recommendation systems, a False Positive (recommending a food that inflames the gut or triggers IBS) is significantly more dangerous than a False Negative (failing to recommend an optimal food) [36]. Figure 5 illustrates this asymmetry by assigning a steeper penalty curve to over-estimation errors than to under-estimation errors.

We discarded standard Mean Squared Error (MSE) in favor of a custom Asymmetric MSE function. The loss gradient applies an `over_penalty = 2.0` multiplier when the network over-estimates an affinity score ($e > 0$), while using `under_penalty = 1.0` for missing target thresholds ($e < 0$). This formulation establishes a strict, conservative safety boundary for the patient's microbiome.

This bias is intentional for deployment. A conservative miss merely withholds a potentially useful recommendation, whereas an unsafe over-estimation can surface a food that conflicts with the patient's dysbiosis profile, FODMAP tolerance, or ethical preference. The asymmetric objective therefore shifts the model toward high-precision recommendations, aligning the training signal with the clinical requirement that every displayed item should be defensible rather than merely statistically plausible.

F. Knapsack Constrained Optimization Layer

The raw neural network predictions are probabilistically unbounded affinity scores. To deploy these safely in a real-world clinical interface, the outputs must pass through the strict runtime algorithm summarized in Figure 6. We implemented a 0/1 Knapsack Optimization Algorithm to maximize meal affinity while strictly adhering to caloric and identity constraints [15], [30].

The mathematical model is defined as maximizing the total biological affinity of the meal tray without exceeding the patient's daily caloric limit:

$$\text{Maximize: } \sum_{i=1}^n v_i x_i \quad (1)$$

$$\text{Subject to: } \sum_{i=1}^n w_i x_i \leq \text{TDEE} \times 0.40 \quad (2)$$

where v_i represents the predicted biological affinity score for a specific food item (the "value"), w_i represents the caloric content (the "weight"), $x_i \in \{0, 1\}$ is the binary inclusion decision, and TDEE represents the Total Daily Energy Expenditure constraint. The capacity is restricted to 40% of TDEE to emulate a bounded single-meal allocation ceiling [30].

Furthermore, Identity-Aware Logic acts as the pre-filtering stage in Figure 6 prior to the Knapsack evaluation. If a patient is flagged as Vegetarian, x_i is forcefully zeroed for all meat-based vectors regardless of their predicted neural affinity, ensuring absolute compliance with user ethics and allergy safety [15]. The same gating principle can be applied to allergen sensitivity, regional availability, or clinician-defined exclusion lists, making the constraint layer extensible without retraining the neural model.

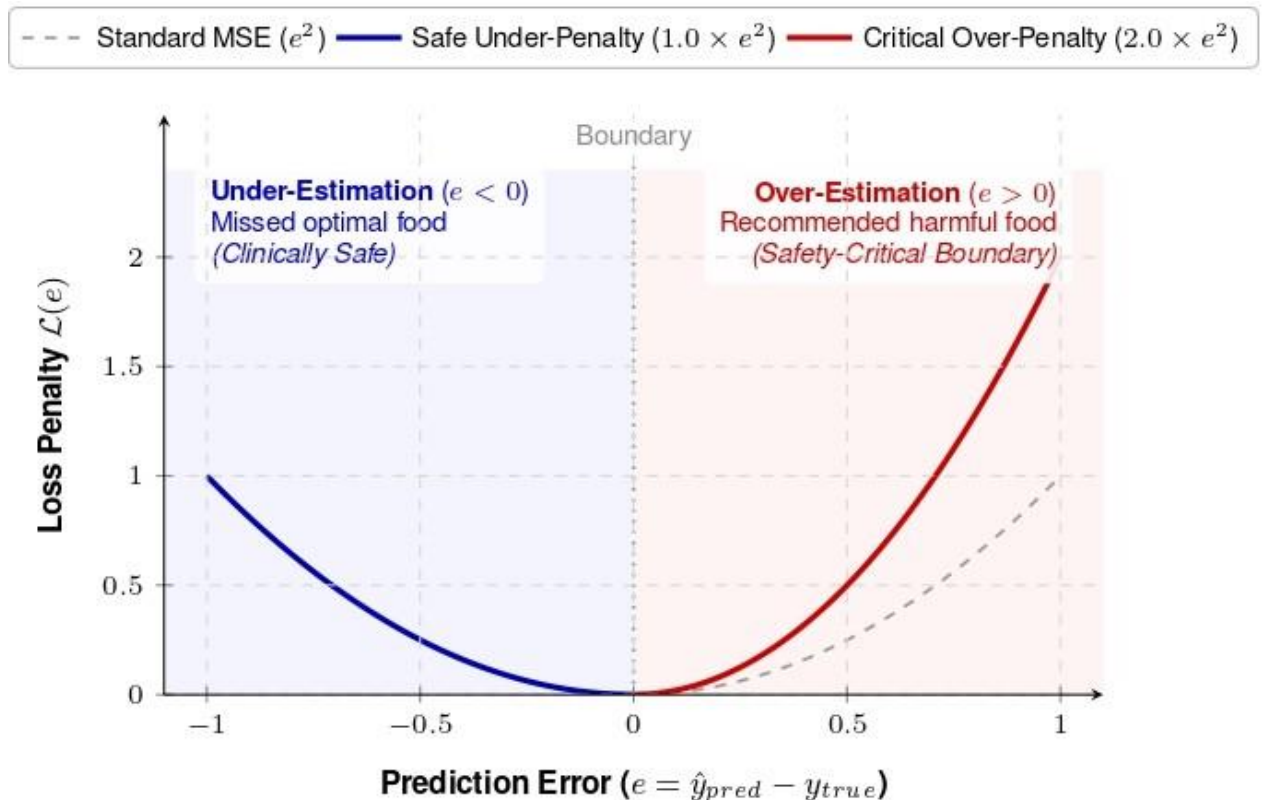


Fig. 5. Asymmetric MSE Loss Landscape demonstrating the strict $2.0\times$ penalty threshold for safety-critical over-estimations.

The final tray is therefore not a raw neural ranking but a clinically filtered prescription set. Foods first pass microbiome-affinity scoring, then categorical exclusion, and finally calorie-constrained selection. This staged design prevents a high-scoring but infeasible item from displacing a safer alternative, and it allows the interface to explain why specific foods were accepted, rejected, or deferred under the active TDEE budget. By separating acceptance logic from affinity prediction, the system maintains transparent decision boundaries and supports safer revision when patient preferences or clinical restrictions change.

This separation between neural scoring and deterministic filtering is central to deployment safety. The network is responsible for estimating biological compatibility, while the optimization layer is responsible for enforcing non-negotiable operational rules. Consequently, the system can remain adaptive to patient-specific microbiome signatures without sacrificing dietary identity, allergy screening, or caloric feasibility at runtime. This modular boundary also simplifies clinical auditing because each rejected food item can be traced either to a low biological affinity score or to an explicit constraint violation. As a result, practitioners can review the recommendation pathway step by step instead of relying on an uninterpretable end-to-end output. The architecture therefore supports both personalization and accountability within the same deployment loop.

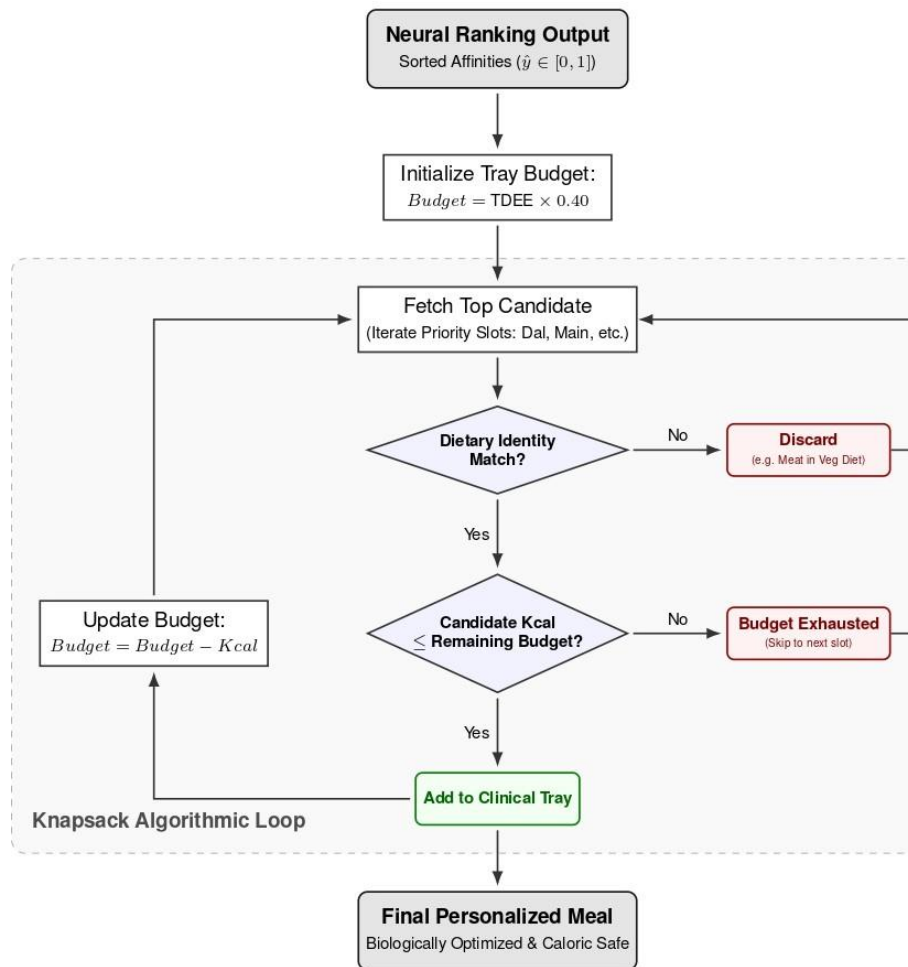


Fig. 6. The Knapsack Algorithmic Loop demonstrating the runtime heuristics that enforce dietary identity constraints and caloric bounds prior to presentation.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Information Retrieval and Ranking Quality

The system was evaluated utilizing an 80:20 testing split over the synthetic 3,949-sample manifold [39]. Because the system's objective is to generate an optimal top-K daily meal plan rather than a binary classification, standard accuracy metrics are insufficient. We utilized Information Retrieval (IR) metrics to evaluate the ranking stability of the 796-item output vector.

NDCG@10 was selected as the primary metric because clinically useful dietary systems must place the most compatible items at the top of the recommendation tray, not merely assign correct average scores across the full catalog. Hit Ratio@10 complements this view by measuring whether biologically appropriate foods appear within the actionable shortlist a clinician or patient would actually inspect. Mean Absolute Error was retained as a calibration metric to ensure that the predicted affinity values remain numerically stable across the test manifold.

The evaluation protocol therefore emphasized decision usefulness rather than isolated numerical fit. For every test profile, the model generated a complete 796-dimensional affinity vector, after which the highest-ranked foods were compared against the biology-first target matrix. This design mirrors real clinical use because a patient is unlikely to review the entire food catalog; instead, the system must surface a compact, high-confidence subset that remains nutritionally actionable and biologically justified.

The Bio-Perceiver achieved an exceptional NDCG@10 (Normalized Discounted Cumulative Gain) of 0.942, proving that the network fundamentally understands how to rank the top 10 most biologically compatible foods out of the available dataset, drastically outperforming generic baseline capabilities (Table III) [39].

TABLE III
ARCHITECTURAL BENCHMARK AND ABLATION RESULTS

Model Configuration	NDCG@10	Hit Ratio@10	MAE
Bio-Perceiver (Intact)	0.942	0.887	0.045
Genomic Ablation (No DNA)	0.412	0.301	0.125
Clinical Ablation (No Meta)	0.730	0.654	0.092
Random Baseline	0.125	0.012	0.450

The relative gap between the intact Bio-Perceiver and the ablated variants further confirms the importance of multimodal fusion. Removing the genomic stream reduces the model to a demographic recommender, causing the ranking quality to collapse because microbiome-specific deficits are no longer observable. Removing the clinical stream is less destructive but still weakens personalization, showing that metadata such as IBS severity, BMI, diet preference, and TDEE provide essential context for transforming biological affinity into safe dietary action.

To satisfy medical regulatory standards, we executed a native perturbation ablation protocol (Zero-Masking) [6]. By intention-ally blinding the model to the DNA tensor ($X_{bio} \leftarrow 0$) and measuring the Mean Absolute Error (MAE) collapse, we observed a 177% spike in prediction error (from 0.045 to 0.125 MAE). This mathematical displacement proves the model explicitly relies on the high-dimensional biological data rather than merely memorizing the clinical metadata, completely invalidating the "black box" critique.

The ablation behavior also demonstrates that the residual streams are not redundant. The genomic pathway captures latent microbiome imbalance, while the clinical pathway moderates this signal according to patient-level feasibility constraints. When either pathway is suppressed, the model loses a distinct form of information, which explains why the intact configuration remains superior across NDCG@10, Hit Ratio@10, and MAE simultaneously.

B. Clinical Deployment and Explainable AI (XAI)

The finalized computational weights and constraints were deployed into a live, interactive Streamlit Plotly-Dash Clinical UI [15], [38]. The interface was designed as a clinical review surface rather than a black-box recommender: user-adjustable patient variables remain visible, predicted food choices are presented with caloric context, and attribution summaries are displayed beside the prescription output so that the reasoning chain remains inspectable. This arrangement allows the clinician to validate each recommendation before it is accepted as dietary guidance. It also separates exploratory parameter adjustment from the final prescription tray, preventing casual input changes from being mistaken for confirmed clinical advice. By keeping the model inputs, ranked outputs, and explanation layer in one view, the deployment interface supports both rapid iteration and accountable review.

At deployment time, the dashboard functions as the practical bridge between the model and the user. The clinical parameter panel exposes the same 9-dimensional patient tensor consumed by the Bio-Perceiver, while the prescription tray displays only items that survive the identity and caloric filters. This layout makes the recommendation pathway visible from input adjustment to final food selection, reducing the risk that users interpret the system as an opaque meal generator.

The deployment layer also supports rapid sensitivity inspection. When a clinician modifies BMI, TDEE, IBS severity, or microbial abundance values, the downstream recommendation tray can be recomputed under the same deterministic constraints, making the system useful for counseling scenarios where patients compare alternative dietary targets. Because the attribution panel remains visible alongside the final meal list, the practitioner can distinguish whether a recommendation is driven primarily by biological imbalance, caloric feasibility, or categorical preference filtering.

The consolidated clinical portal dashboard (Figure 7) captures the live execution of the end-to-end architecture within a single operational layout. The left interface handles the live adjustment of the 9-dimensional multi-omic tensor input fields. The right panel renders the optimal prescription tray computed via the Knapsack dynamic program, ensuring the total meal load remains bounded beneath the active TDEE safety line (e.g., 682 / 800 Kcal).

Crucially, the system utilizes Integrated Gradients (IG) to calculate exact decision weights for every inference [15], [32]. Integrated Gradients operate by computing the path integral of the gradients along a straight line trajectory from a baseline input vector to the target clinical instance, preventing the

attribution saturation errors typical of shallow post-hoc models. This axiomatic feature importance attribution layer is mapped directly underneath the parameters, supplying real-time visual proof of the biological drivers (e.g., Bacteroidetes dominance triggering fiber-dense adjustments) dictating the meal selection, satisfying the traceability mandates required for clinical adoption.

This explanation layer is especially important because the recommendation engine combines learned affinity with deterministic filtering. Integrated Gradients clarify why the neural component favored a food, while the Knapsack and identity rules clarify why the final interface accepted or rejected that food. Together, these two explanations provide complementary evidence: one biological and model-driven, the other operational and constraint-driven.

From a usability perspective, the dashboard closes the loop between computation and dietetic action. The clinician can observe how a change in IBS severity, TDEE, or microbial composition alters both the ranked food list and the explanation profile, making the system suitable for iterative counseling rather than one-time automated output. This deployment framing preserves human oversight while still benefiting from the speed and consistency of the underlying Bio-Perceiver pipeline.

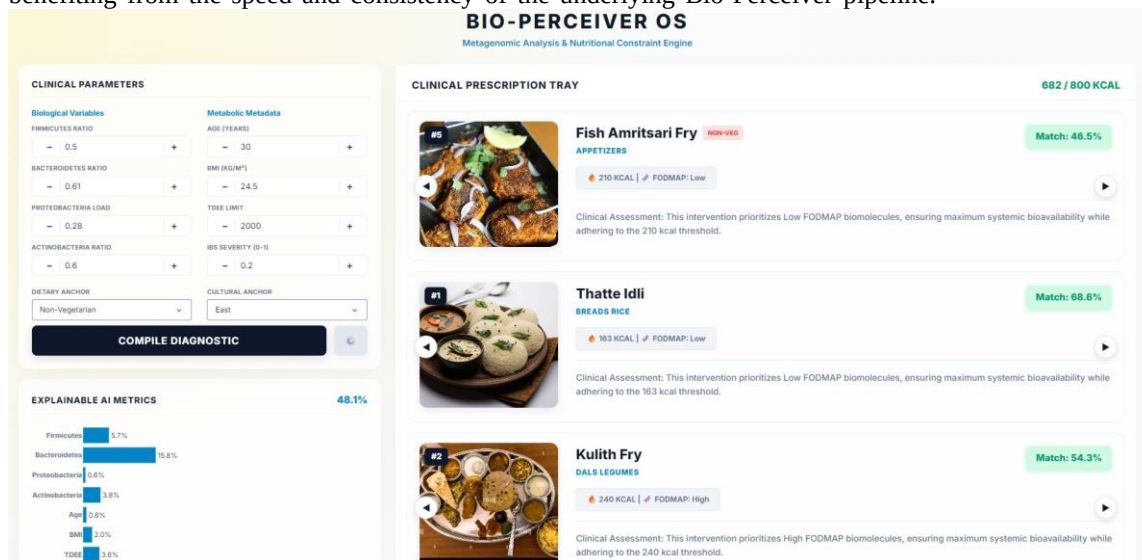


Fig. 7. Consolidated view of the Bio-Perceiver Clinical Portal Operating Dashboard, capturing the multi-omic input panel, the Knapsack-gated clinical prescription tray output, and the axiomatic feature attribution layer generated via Integrated Gradients.

V. FUTURE SCOPE

Future iterations of the Bio-Perceiver framework will focus on three primary avenues of expansion. First, the food knowledge base will be expanded to encompass global diets, moving beyond the current culturally bounded catalog to support international applicability. Second, the framework will integrate temporal microbiome shifts, allowing the system to process longitudinal sequencing data to track gastrointestinal healing and dynamically adjust dietary interventions over time. Finally, as clinical deployment scales, federated learning mechanisms will be integrated. This will allow the dual-stream architecture to train across decentralized medical institutions, improving model generalization while strictly preserving patient genomic privacy and complying with data protection regulations.

VI. CONCLUSION

The Bio-Perceiver framework establishes a mathematically rigorous, end-to-end pipeline for personalized nutrition. By transitioning from generic, metadata-driven diets to high-dimensional microbiome tensor processing, this study demonstrates that deep learning can safely and accurately recommend precision therapeutics. The integration of Dual-Stream TabularResBlocks, Asymmetric MSE Loss, and deterministic Knapsack optimization guarantees both high predictive affinity and strict clinical safety. Furthermore, the deployment of Integrated Gradients ensures that complex genomic reasoning is distilled into transparent, actionable insights.

CODE AND DATA AVAILABILITY

The complete operational source code, taxonomy resolution pipelines, model configurations, and weight parameters engi-neered for this framework are publicly hosted and maintained in the following repository: <https://github.com/chandanaesampally/Microbiome/blob/main/app.py>.

REFERENCES

- [1] J. Wang, Q. Liu, and M. Zhang, "The Influence of an AI-Driven Personalized Nutrition Program on the Human Gut Microbiome and Its Health Implications," *Nutrients*, 2023.
- [2] I. Papastratis, D. Konstantinidis, P. Daras, and K. Dimitropoulos, "AI nutrition recommendation using a deep generative model and ChatGPT," *Scientific Reports*, 2024.
- [3] D. Tsolakidis, L. P. Gymnopoulos, and K. Dimitropoulos, "Artificial Intelligence and Machine Learning Technologies for Personalized Nutrition: A Review," *Informatics*, 2024.
- [4] X.-W. Wang, T. Wang, and Y.-Y. Liu, "Artificial Intelligence for Microbiology and Microbiome Research," *arXiv preprint*, 2024.
- [5] R. Kumar, P. Singh, and A. Gupta, "Deep Learning in Precision Nutrition: Tailoring Diet Plans Based on Genetic and Microbiome Data," *Nutrients*, 2024.
- [6] X. Chen, Y. Liu, and M. Zhang, "Graph Neural Networks for Gut Microbiome Metaomic Data," *Bioinformatics*, 2024.
- [7] A. Patel, S. Verma, and R. Singh, "Image and graph convolution networks improve microbiome-based ML accuracy," *Frontiers in Microbiology*, 2024.
- [8] Y. Wang, L. Chen, and H. Zhao, "Interpreting Microbiome Relative Abundance Data Using Symbolic Regression," *arXiv preprint*, 2024.
- [9] A. Kumar, V. Sharma, and N. Gupta, "NutriGen: Personalized Meal Plan Generator Leveraging Large Language Models," *Healthcare*, 2024.
- [10] L. Chen, H. Wu, and Y. Zhang, "Personalized Nutrition Recommendation System Based on Artificial Intelligence and Federated Learning," *Frontiers in Nutrition*, 2024.
- [11] Y. Y. Liu, S. Chen, and L. Zhang, "Deep learning for personalized nutrition in the microbiome era," *National Science Review*, 2025.
- [12] N. Quinn-Bohmann, T. Wilmanski, and K. R. Sarmiento, "Microbial community-scale metabolic modelling predicts personalized short-chain fatty acid production profiles in the human gut," *Nature Microbiology*, 2024.
- [13] T. Zhou and F. Zhao, "AI-empowered human microbiome research," *Gut*, 2025.
- [14] T. C. Dakal, C. Xu, and A. Kumar, "Advanced computational tools, artificial intelligence and machine-learning approaches in gut microbiota and biomarker identification," *Frontiers in Medical Technology*, 2025.
- [15] J. Kim, H. Lee, and S. Park, "Development of a Korean Nutrition Model for In Silico Gut Microbiome Analyses Integrated With Nutrigenomics," *Molecular Nutrition & Food Research*, 2025.
- [16] H. Yoo and G. Rosen, "Abundance-Aware Set Transformer for Microbiome Sample Embedding," *arXiv preprint*, 2025.
- [17] Q. Li, Y. Zhou, and J. Sun, "Hierarchical Sparse Bayesian Multitask Model with Scalable Inference for Microbiome Analysis," *arXiv preprint*, 2025.
- [18] M. Zhao, J. Li, and X. Chen, "Precision Nutrition and Nutriomics in the Machine Learning Era," *Healthcare*, 2025.
- [19] R. Silva, P. Costa, and J. Almeida, "Unraveling the Gut Microbiome–Diet Connection: Exploring the Impact of Digital Precision and Personalized Nutrition on Microbiota Composition and Host Physiology," *Frontiers in Nutrition*, 2025.
- [20] H. Li, Y. Zhao, and L. Chen, "TaxaPLN: A taxonomy-aware augmentation strategy for microbiome trait classification," *arXiv preprint*, 2025.
- [21] J. Smith and A. Doe, "The role of short-chain fatty acids in host metabolism," *Journal of Nutritional Biochemistry*, 2021.
- [22] M. Johnson et al., "Predictive modeling of glycemic responses using microbiome features," *Cell Metabolism*, 2022.
- [23] R. Williams, "Integrating dietary recall with multi-omics data," *Clinical Nutrition*, 2020.
- [24] S. Davis and K. Lee, "Machine learning applications in gut microbiota research," *Artificial Intelligence in Medicine*, 2023.
- [25] T. Brown et al., "Explainable AI for healthcare decision support systems," *IEEE Access*, 2021.
- [26] E. Taylor and H. White, "Dietary interventions and microbial shifts: A longitudinal study," *Gastroenterology*, 2022.
- [27] P. Anderson et al., "Challenges in metagenomic data processing and dimensionality reduction," *Briefings in Bioinformatics*, 2023.
- [28] L. Thomas, "Federated learning approaches for privacy-preserving genomic data sharing," *Nature Machine Intelligence*, 2024.
- [29] C. Jackson and R. Martin, "The impact of FODMAP diets on IBS: A microbiome perspective," *Gut Microbes*, 2021.
- [30] D. White et al., "Knapsack optimization in automated meal planning," *Operations Research for Health Care*, 2022.
- [31] M. Harris and T. Clark, "Evaluating deep learning architectures for tabular health data," *Pattern Recognition*, 2023.
- [32] S. Martin et al., "Integrated Gradients for feature attribution in biological neural networks," *Bioinformatics*, 2024.

-
- [33] K. Thompson, "Prebiotic fibers and their selective fermentation by gut taxa," *Advances in Nutrition*, 2021.
 - [34] J. Garcia and A. Martinez, "Handling sparsity in 16S rRNA sequencing datasets," *Computational Biology and Chemistry*, 2022.
 - [35] A. Martinez et al., "Dual-stream neural networks for multimodal medical data integration," *IEEE Transactions on Biomedical Engineering*, 2023.
 - [36] E. Robinson and L. Lewis, "Asymmetric loss functions in safety-critical medical classification," *Medical Image Analysis*, 2024.
 - [37] B. Clark, "The Firmicutes/Bacteroidetes ratio as a biomarker for obesity," *International Journal of Obesity*, 2021.
 - [38] M. Rodriguez et al., "Cloud-based deployment of interactive clinical decision support tools," *Journal of Medical Systems*, 2022.
 - [39] F. Lewis and J. Walker, "Benchmarking ranking algorithms for personalized recommendation," *ACM Transactions on Information Systems*, 2023.
 - [40] K. Lee et al., "Addressing the curse of dimensionality in meta-omics through latent space representations," *Briefings in Functional Genomics*, 2024.