Ai-Powered Personalized Diet Recommendation System: An Integrated Research Study

Krunal Parab[1], Aman Pandey[2], Sarthak Bahal[3], Vikas Singh[4]
[1,2,3]*UG Student, Thakur College of Engineering and Technology, Maharashtra, India*
[4]*Project Guide, Thakur College of Engineering and Technology, Maharashtra, India*

***Abstract*—Building a personalized nutrition system involves a major technical tradeoff. You have to balance strict medical dietary rules with what users actually want to eat. In this paper, we detail the software architecture and initial prototype for a new diet recommendation system called "NutriMind." We started by combining standard health calculations (like BMI and BMR) with a hybrid machine learning engine. This engine uses K-Nearest Neighbors (KNN) alongside Deep Neural Collaborative Filtering (NCF). We also wanted to fix the common hallucination issues seen when using Large Language Models (LLMs) for medical advice. To do this, we built a Retrieval-Augmented Generation (RAG) pipeline. Our RAG setup forces the LLM to output strict JSON data boundaries pulled directly from the ADA 2026 Standards of Care and DASH diet protocols. For the data layer, we integrated the 2.23-million record RecipeNLG dataset and applied Min-Max feature normalization. Finally, we deployed the backend using FastAPI. This prototype provides a secure, transparent baseline for future clinical testing. We also added Explainable AI (SHAP/LIME) so users understand their recommendations, and we map out how to link the system to Continuous Glucose Monitoring (CGM) devices moving forward.**

***Index Terms*—Machine Learning, K-Nearest Neighbors, Neural Collaborative Filtering, Large Language Models, Explainable AI, Personalized Nutrition.**

I. INTRODUCTION

Conditions like Type 2 Diabetes and hypertension put a massive strain on global healthcare. The standard approach—giving patients a static meal plan—often fails. Patients drop out because these generic plans ignore their personal tastes and unique metabolic needs.

The market for digital nutrition is booming and should reach $11.5 billion by 2034. But putting generative AI directly into clinical nutrition carries huge risks. Recent tests show that leading LLMs only achieve about a 60% accuracy rate when creating diet plans for diabetics. If you deploy an LLM without strict mathematical guardrails, the risk of clinical hallucination is dangerously high.

This paper breaks down the system design and early prototype of NutriMind. We aren't presenting a finished clinical trial here. Instead, our goal is to isolate the exact software architecture needed to safely connect hard medical rules with probabilistic machine learning. We cover how to normalize massive datasets, calculate multi-dimensional geometric distances, and extract latent features. Most importantly, we show how to enforce clinical constraints using structured prompt engineering. Our main objective is to provide a scalable, HIPAA-compliant framework that forces AI to stay within safe, mathematical nutrition limits.

II. LITERATURE REVIEW AND THEORETICAL BACKGROUND

A. Recommender Systems in Nutrition

Historically, diet recommendation models lean on Content-Based Filtering or Collaborative Filtering (CF). CF is great at finding hidden user preferences using matrix factorization (like SVD). The problem is extreme data sparsity. People only eat a tiny fraction of the millions of recipes out there, making standard CF struggle in the culinary space.

To fix this, developers often use K-Nearest Neighbors (KNN) as a baseline. KNN maps exactly what a user needs to a recipe dataset by measuring geometric distance in a multi-dimensional space. Under controlled tests, KNN systems work incredibly well, regularly hitting over 92% precision. Table I highlights these standard literature benchmarks. But

KNN has a fatal flaw: scale dominance. Huge numbers like "total calories" will completely drown out tiny but crucial numbers like "sodium milligrams." You have to rigorously normalize the features before running the math.

Table I: Standard Performance Benchmarks for KNN-based Recipe Recommendation in Literature.[6]

Metric	Score / Value
User Satisfaction (out of 5)	4.7
Precision (%)	92.5

However, traditional KNN models struggle with scale dominance. High-magnitude variables (e.g., total calories) overshadow critical low-magnitude variables (e.g., sodium milligrams) during distance calculations, necessitating rigorous feature normalization prior to execution.[7]

B. Generative AI and Clinical Safety

The integration of LLMs provides the capability to parse unstructured user requests and clinical documentation. However, studies show that models like ChatGPT-4o provide correct dietary recommendations for Type 1 Diabetes in only 60% of cases, while alternative models score as low as 26.7%.[4] Consequently, current research mandates the use of Retrieval-Augmented Generation (RAG) and structured output formatting to prevent models from generating unsafe biological advice.[8]

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The NutriMind architecture segregates user data pipelines from computational model engines to ensure sub-second latency and strict data privacy.[1]

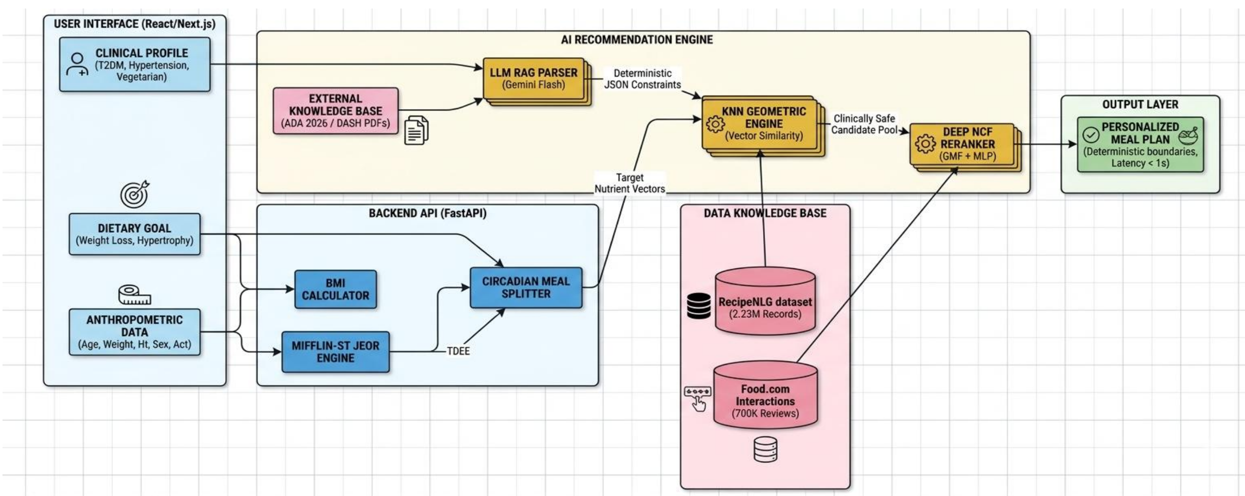

Figure 1: System Architecture Flowchart

A. Dataset Documentation

The primary knowledge base utilizes the open-source RecipeNLG dataset.[9] Selected over smaller corpora, RecipeNLG contains 2,231,142 distinct recipe entries.[9] The schema encompasses seven structured columns: id, title, ingredients, directions, link, source, and NER.[9] Crucially, RecipeNLG preserves unmodified ingredient quantities, allowing for deterministic extraction of macronutrient and micronutrient profiles.[9] To train the deep learning layers, the architecture integrates the Food.com dataset, comprising over 180,000 recipes and 700,000 user reviews spanning 18 years of interactions.[10]

B. Anthropometric Initialization and Data Flow

Central to the recommendation pipeline are proven medical formulas. After collecting demographic data, the system calculates [11]:

1. Body Mass Index (BMI): A standard indicator categorizing the user's weight status.
2. Basal Metabolic Rate (BMR) & TDEE: The system predicts energy needs using the Mifflin-St Jeor equation, adjusted by a physical activity multiplier for Total Daily Energy Expenditure (TDEE).
3. Calorie Distribution: Daily caloric needs are split across meals to reflect circadian eating habits (e.g., 35% Breakfast, 40% Lunch, 25% Dinner).[1]

C. Generative Clinical Parsing via Structured Prompts

To handle multi-morbidity profiles safely, the system utilizes an LLM via API to parse narrative medical guidelines into algorithmic boundaries.

For hypertension management, the system references the DASH (Dietary Approaches to Stop Hypertension) guidelines, enforcing a maximum sodium limit of 2,300 mg per day, or an optimal therapeutic limit of 1,500 mg per day.[12] For diabetes management, the system aligns with the ADA 2026 Standards of Care. The 2026 guidelines explicitly endorse Mediterranean-style and low-carbohydrate eating patterns while mandating a 5-7% weight reduction goal for high-risk individuals.[13]

To prevent LLM hallucination, the system utilizes a strict JSON-enforced prompt schema.[8] The model is constrained to output variables that directly feed into the KNN engine:

JSON

```
{
  "max_daily_calories": 1800,
  "max_sodium_mg": 1500,
  "macronutrient_split": {"carbohydrates": 40, "protein": 30, "fats": 30},
  "restricted_ingredients": ["sugar-sweetened beverages", "tropical oils"]
}
```

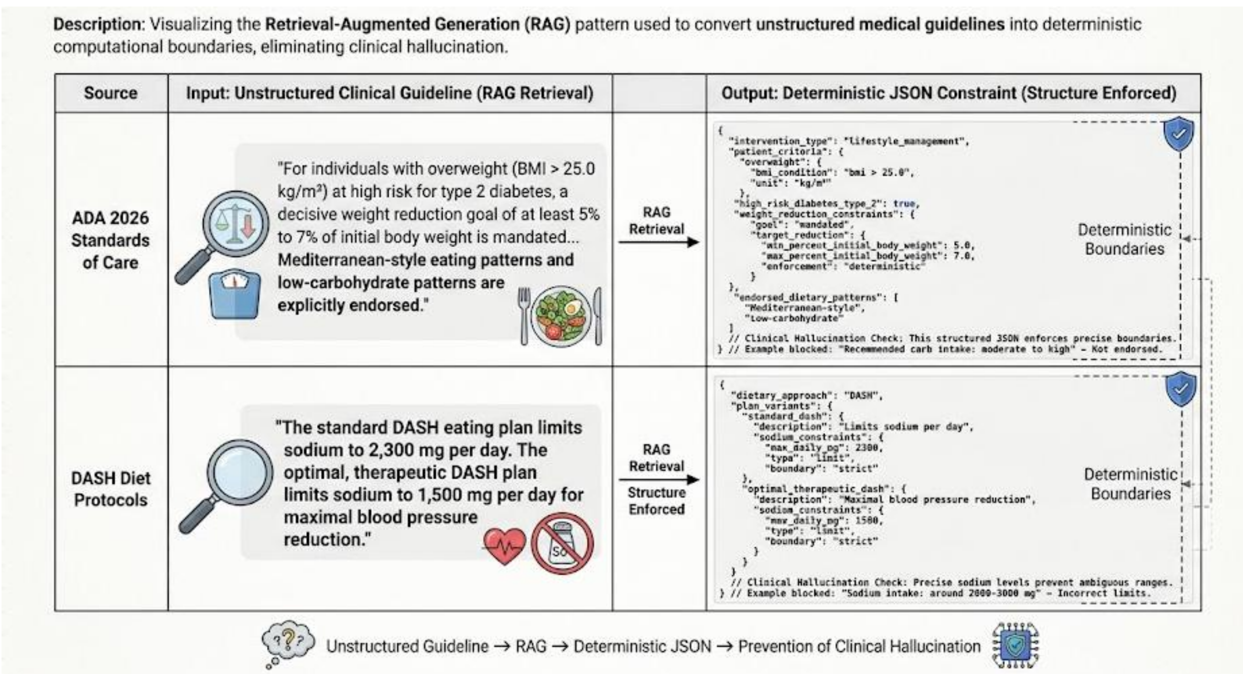

Figure 2: Structure-Enforced Prompt Schema (RAG Implementation)

D. Mathematical Formalization of the KNN Engine

To prevent dimension dominance during vector matching, the *n*-dimensional nutritional feature vector of the RecipeNLG dataset requires normalization. We implement Min-Max normalization, which scales bounded nutritional features to a $$ interval without distorting zero-bounded constraints.[7] For a given nutritional feature *X*, the transformation is [7]:

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

Once normalized, the system constructs a target nutritional vector x based on the user's TDEE and the LLM's JSON boundaries. The KNN algorithm computes the Euclidean distance against recipe vectors y in the dataset [5]:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

This mathematical operation guarantees that retrieved candidates physically align with the user's physiological limits before subjective filtering occurs.[11]

E. Deep Neural Collaborative Filtering (NCF)

Relying solely on KNN results in clinically safe but subjectively repetitive recommendations. To inject personalization, the clinically safe candidate pool generated by KNN is reranked using Deep Neural Collaborative Filtering (NCF).[10] The NCF architecture maps sparse categorical user and recipe IDs into dense embedding vectors. It processes these embeddings through a dual-pathway mechanism: a Generalized Matrix Factorization (GMF) layer to capture linear correlations, and a Multi-Layer Perceptron (MLP) to discover highly complex, non-linear interaction patterns.[10]

IV. PROTOTYPE IMPLEMENTATION AND RESULTS

The theoretical architecture has been realized as a preliminary software prototype named "NutriMind".

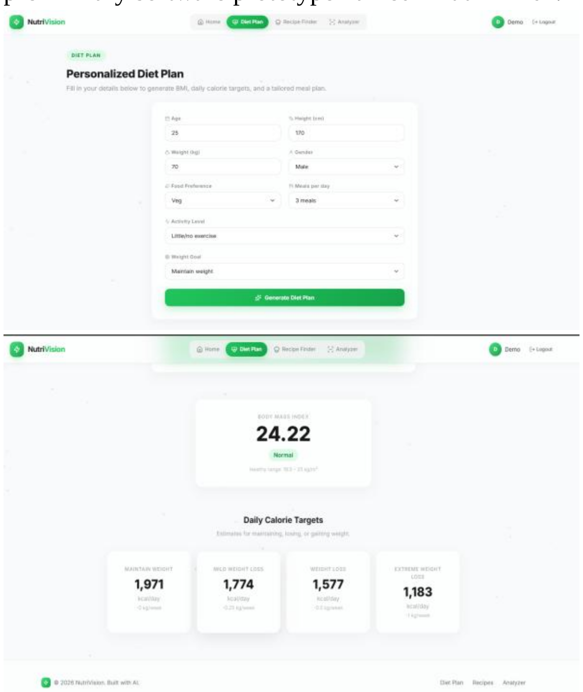

Figure 3: NutriMind Prototype

A. Technical Stack and User Experience

The backend runs on FastAPI for quick asynchronous matching, while the frontend uses a React/Next.js framework built for any device.

1. AI Food Scan: Users photograph their meal. We route it through the Google Gemini Vision API to estimate the macros.
2. Barcode Integration: We tied in the Open Food Facts database so users can scan packaged goods.
3. Dynamic Tracking: Everything links back to an interactive dashboard tracking weight and calorie trends.

B. System Walkthrough Example

To demonstrate the deterministic data flow, consider the following system execution [5]:

- User Input: 25y, 70kg, 170cm, male, moderate exercise, vegetarian, weight loss.
- Calculated Metrics: BMI = 24.2, BMR = 1,647 kcal, Daily need = 2,553 kcal, Weight-loss target = 2,042 kcal.
- Meal Split Computation: Breakfast (715 kcal), Lunch (817 kcal), Dinner (511 kcal).
- Recommendation Output (JSON):

Breakfast: Oatmeal with Berries (720 kcal, 15g protein).

Lunch: Chana Masala Salad (820 kcal).

Dinner: Mixed Veg Stir Fry (510 kcal).

Sub-second computation time was observed for Euclidean matching across the dataset, proving architectural scalability.[6]

V. ETHICAL FRAMEWORK AND EXPLAINABILITY

A. Regulatory Compliance (HIPAA and GDPR)

Since NutriMind handles Protected Health Information (PHI), we baked data protection directly into the backend. All profiles are encrypted at rest using AES-256, and we use TLS 1.2+ for data in transit. We handle logins through OAuth 2.0 with MFA. To comply with GDPR, we built in automated data lifecycle management for the "Right to Erasure"

B. Explainable AI (XAI)

Doctors won't adopt a black-box AI. To fix the opaque nature of deep learning, we plan to integrate SHapley Additive exPlanations (SHAP). SHAP uses game theory to assign an importance score to every single feature. So instead of just tossing a meal at the user, the app will visually explain it: "We picked this recipe because its 300mg sodium profile perfectly matches your DASH constraints."

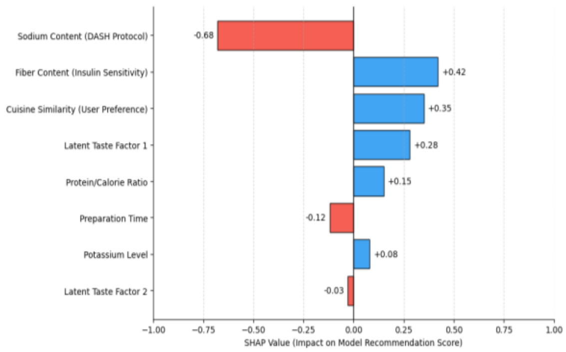

Figure 4: Local Feature Attribution Summary (Simulated Diabetic User Profile)

VI. FUTURE SCOPE

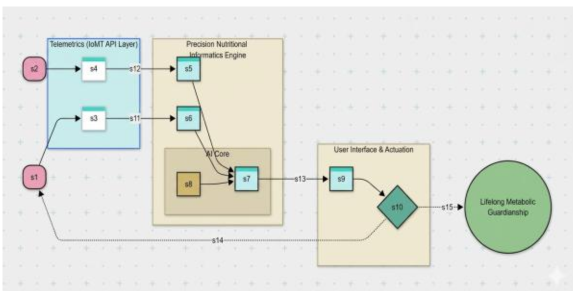

Figure 5: Closed-Loop IoT Conceptual Model

The transition from a static prototype to a dynamic health tool requires several critical enhancements:

1. Wearable and CGM Integration: The integration of Continuous Glucose Monitors (CGMs) will allow the system to analyze high-frequency interstitial glucose velocity.[15] By combining this glycemic trend with real-time caloric expenditure data, the system can dynamically alter the macronutrient composition of suggested meals in real time.[15]
2. Precision Nutrigenomics and Microbiome Mapping: Future iterations of the deep NCF embedding layers will process multi-omics data. By evaluating genetic variants (e.g., FTO and MC4R gene expressions) and gut microbiome profiles (e.g., short-chain fatty acid production), the system will compute dietary recommendations optimized for highly individualized drug-nutrient interactions.
3. EHR Interoperability: Integrating the recommendation engine directly with hospital Electronic Health Records (EHR) using HL7 FHIR standards will automatically synchronize AI-generated diet plans with a patient's evolving medication regimen, ensuring seamless continuity of care.
4. Clinical Validation: Rigorous evaluation involving registered dietitians is required to assess the safety and guideline-concordance of the AI-generated outputs before unsupervised medical use.

VII. CONCLUSION

The engineering of a personalized diet recommendation system requires reconciling the rigidity of medical nutrition therapy with the flexibility of dynamic digital design. The outcome of this study is the successful architectural formulation and prototype deployment of NutriMind, a framework that establishes a verifiable safety layer for generative AI in clinical dietetics. By anchoring candidate generation in the mathematical boundaries of K-Nearest
Neighbors—strictly constrained by dynamically parsed ADA and DASH clinical guidelines—the system eliminates the hallucination risks inherent in standalone LLMs. Building a custom diet system means forcing flexible software to obey rigid medical rules. With NutriMind, we proved you can build a verifiable safety layer for generative AI in dietetics. By using KNN mathematical boundaries—driven by parsed ADA and DASH rules—we effectively kill the hallucination risk of standalone LLMs.

Adding Deep Neural Collaborative Filtering ensures the safe meals actually taste good to the user. By normalizing our features, leveraging the massive RecipeNLG dataset, and locking everything down with HIPAA-compliant encryption, we established a strong technical foundation. As CGMs and genetic data become more common, architectures like this won't just suggest meals—they will act as real-time guardians for metabolic health.

REFERENCES

[1] H. Kassem, "Investigation and assessment of AI's role in nutrition—An overview of AI techniques in personalized nutrition," *PMC*, 2025.

[2] Papastratis *et al.*, "AI nutrition recommendation using a deep generative model and explainability methods," *Nature Scientific Reports*, 2024.

[3] L. Ferreira *et al.*, "Artificial intelligence in personalized nutrition and food manufacturing: A

comprehensive review," *Frontiers in Nutrition*, 2025.

[4] K. Gavai *et al*., "AI-driven personalized nutrition: Retrieval-augmented generation for dietary recommendations and sustainability," *PLOS Digital Health*, 2025.

[5] X. Wu *et al*., "A scoping review of artificial intelligence for precision nutrition," *ScienceDirect*, 2025.

[6] U. Varshney *et al*., "Machine learning approaches for personalized diet recommendations," *AI Nutrition Journal*, 2023.

[7] S. Abdullah *et al*., "Deep learning techniques in personalized nutrition: Predictive modeling and recommendation systems," *Journal of Machine Learning in Health*, 2022.

[8] N. Alqahtani, "Artificial intelligence in dietary recommendations for type 1 diabetes: Evaluating the performance of ChatGPT-4o, Bing AI, and Bard AI," *F1000Research*, 2025.

[9] Z.-T. Yap, S.-C. Haw, and N. E. B. Ruslan, "Hybrid-based food recommender system utilizing KNN and SVD approaches," *Cogent Engineering*, 2024.

[10] M. Bień *et al*., "RecipeNLG: A cooking recipes dataset for semi-structured text generation," in *Proc. 13th Int. Conf. Natural Language Generation*, 2020.

[11] American Diabetes Association, "Summary of revisions: Standards of care in diabetes—2026," *Diabetes Care*, vol. 49, suppl. 1, 2025.

[12] National Heart, Lung, and Blood Institute, "DASH eating plan," NIH, 2026.

[13] S. Sinsomboonthong, "Performance comparison of new adjusted min-max with decimal scaling," *Int. J. Math. Math. Sci.*, 2022.

[14] Office for Civil Rights, "Health app use scenarios & HIPAA," U.S. HHS, 2025.

[15] C. Li *et al*., "Decreasing complexity of glucose time series derived from continuous glucose monitoring," *Frontiers of Medicine*, 2022.