



Intelligent Nutrition Recommendation System for Individual Health Profiles

Divya Varshini M¹, Dr. G. Paavai Anand²

M.Tech 1 Year, CSE, SRMIST, Vadapalani, India¹

Asst. Professor, CSE, SRMIST, Vadapalani, India²

Abstract: The rise of obesity and metabolic disorders highlights the limitations of generalized dietary advice, which often fails to meet individual health needs. At the same time, health tracking applications and self-reported data provide valuable insights into individuals' heart rate, physical activity, sleep patterns, and dietary intake. This research presents a machine learning-based personalized nutrition recommendation system that integrates biometric, lifestyle, and dietary data to provide customized diet plans for individuals. By analyzing patterns in user behavior and correlating them with health risk factors, the system predicts potential nutritional deficiencies or risks and generates actionable recommendations tailored to each user. The approach leverages data-driven modeling to bridge the gap between raw health data and effective, personalized dietary guidance. Experimental evaluation on synthetic and public datasets demonstrates that the system can accurately identify individual health risks and suggest targeted nutritional adjustments, promoting preventive healthcare and overall well-being.

Keywords: Personalized Nutrition, Machine Learning in Healthcare, Dietary Recommendation Systems, Preventive Healthcare, Biometric Data Analytics, Health Risk Prediction, Individualized Diet Planning.

I. INTRODUCTION

The prevalence of **obesity, diabetes, and other metabolic disorders** has increased dramatically in recent decades, posing a major challenge to global public health [1]. Despite continuous efforts by health organizations, **traditional dietary guidelines** generally adopt a one-size-fits-all approach, overlooking variations in **individual physiology, lifestyle, and genetic predisposition** [1], [8], [9]. Consequently, such generalized recommendations often lead to **poor adherence, limited personalization, and suboptimal outcomes** in preventing chronic lifestyle diseases [1], [7].

Recent advancements in **digital health technologies** have created new opportunities for continuous health monitoring and personalized intervention [2]. Through mobile health applications, wearable devices, and dietary tracking tools, individuals can record **physical activity levels, sleep quality, and daily nutrient intake**, providing rich datasets for analysis. These **behavioral and biometric records** enable researchers and healthcare providers to gain deeper insights into individual habits and metabolic responses, thereby supporting **data-driven and personalized nutrition planning** [2], [3].

At the same time, **machine learning (ML)** and **artificial intelligence (AI)** techniques have emerged as powerful tools for analyzing **large-scale, multidimensional health data** [3], [4], [10]. Predictive models can identify subtle correlations between dietary intake, lifestyle behaviors, and health outcomes, assisting in **early detection of risks** such as obesity, hypertension, and metabolic syndrome [7], [11]. Integrating ML with nutritional science enables **personalized diet recommendation systems**, improving adherence and promoting preventive healthcare [2], [3], [10].

This research proposes a **hybrid personalized nutrition recommendation framework** that combines **machine learning** and **rule-based reasoning** for adaptive diet planning. The system utilizes a **Random Forest classifier** to predict individual health-risk levels using biometric and lifestyle data [4], and a **K-Means clustering algorithm** to group users with similar nutritional profiles [5]. Subsequently, a **rule-based inference engine** generates customized dietary plans based on **WHO** and **ICMR** nutritional standards [6], [9]. This hybrid integration ensures that the system remains **accurate, interpretable, and scalable**, offering personalized and clinically relevant dietary recommendations for preventive healthcare applications [3], [10].

II. LITERATURE SURVEY

Personalized nutrition has become an emerging area of research aimed at tailoring dietary recommendations to an individual's **metabolic profile, lifestyle, and health status** [1]. Traditional nutrition models rely heavily on generalized



dietary guidelines provided by organizations such as the **World Health Organization (WHO)** and the **Indian Council of Medical Research (ICMR)** [8], [9]. While these guidelines are valuable at a population level, they fail to account for **individual variability** in dietary responses, nutrient absorption, and lifestyle behavior, leading to inconsistent health outcomes.

Several studies have explored the application of **machine learning (ML)** in nutrition and preventive healthcare. Chen *et al.* [2] developed a **data-driven nutritional recommendation framework** that used biometric and dietary data to predict caloric needs and meal plans. Their model demonstrated improved precision but lacked interpretability and adaptability for diverse populations. Similarly, Alowais and Nour [3] proposed a **hybrid ML approach** combining classification and clustering for personalized diet planning. Although their system achieved notable accuracy, it did not incorporate **expert-verified nutritional rules**, limiting its clinical applicability.

The use of **Random Forest (RF)** has gained attention for its robustness and interpretability in health-risk prediction tasks. Breiman [4] established RF as an ensemble learning algorithm that effectively handles nonlinear relationships and heterogeneous data, making it suitable for **biometric-based disease risk prediction** [7]. Meanwhile, **K-Means clustering**, introduced by MacQueen [5], has been extensively used for user segmentation and lifestyle grouping in health analytics. However, pure clustering approaches often overlook contextual relationships between user features, which can reduce personalization quality.

Rule-based systems have long been applied in **clinical decision support** and **dietary planning** due to their transparency and interpretability. Patel and Shah [6] implemented a **rule-based expert system** that generated dietary advice using parameters such as BMI, blood sugar, and blood pressure. While effective in structured settings, rule-based models alone cannot adapt to dynamic or complex health data patterns without data-driven insights.

More recent research has emphasized the integration of ML and expert knowledge. Zhou and Li [10] proposed a **hybrid healthcare recommendation model** combining clustering and classification techniques for personalized suggestions. Their findings highlight the potential of merging algorithmic prediction with domain expertise to enhance both accuracy and usability. Similarly, Das and Chatterjee [11] applied ML-based predictive modeling for nutritional risk assessment, confirming that hybrid approaches outperform traditional methods in disease prevention and dietary personalization.

Research Gaps

- **Limited Preventive Focus:** Emphasis is often on calorie counting or weight loss, rather than proactive prevention of metabolic disorders.
- **Scalability Issues:** Many personalized nutrition models exist only as research prototypes and are not suitable for large-scale deployment.
- **Generalized Guidelines:** Most existing dietary platforms rely on one-size-fits-all recommendations rather than truly personalized advice.

III. METHODOLOGY

This study adopts a hybrid approach integrating machine learning and rule-based techniques to deliver personalized nutrition recommendations. The system architecture consists of three main modules: health risk prediction, user clustering, and rule-based recommendation generation.

3.2 Data Collection and Preprocessing

The dataset includes biometric, lifestyle, and dietary parameters such as age, gender, BMI, blood sugar level, physical activity, and daily nutrient intake.

Data preprocessing involves cleaning missing or inconsistent records, normalizing numerical attributes, and encoding categorical variables. Derived metrics such as BMI and calorie–nutrient ratios are computed to enhance model accuracy.

3.2 Health Risk Prediction Using Random Forest

A Random Forest classifier is employed to predict individual health risk categories (e.g., normal, overweight, obese, diabetic risk). The model constructs multiple decision trees and aggregates their outputs through majority voting. Feature importance analysis identifies key health determinants such as BMI, calorie intake, and activity level, ensuring interpretability and reliability. The model's performance is evaluated using accuracy, precision, recall, and F1-score metrics.



3.3 User Grouping via K-Means Clustering

To identify users with similar nutritional profiles, K-Means clustering is applied to selected features (e.g., BMI, calorie intake, macronutrient distribution). The optimal number of clusters (k) is determined using the Elbow Method. Each cluster represents a group with comparable dietary patterns and health requirements, enabling scalable and group-specific recommendations.

3.4 Rule-Based Diet Recommendation System

A rule-based inference engine translates analytical results into practical diet suggestions. Rules are formulated from expert dietary guidelines and clinical standards, such as:

- If BMI > 25 and activity level = low $\rightarrow$ recommend a low-calorie, high-fiber diet.
- If blood sugar > threshold $\rightarrow$ reduce simple carbohydrates, increase protein intake.

The system customizes meal plans based on both cluster-level insights and individual parameters.

3.5 System Integration and Evaluation

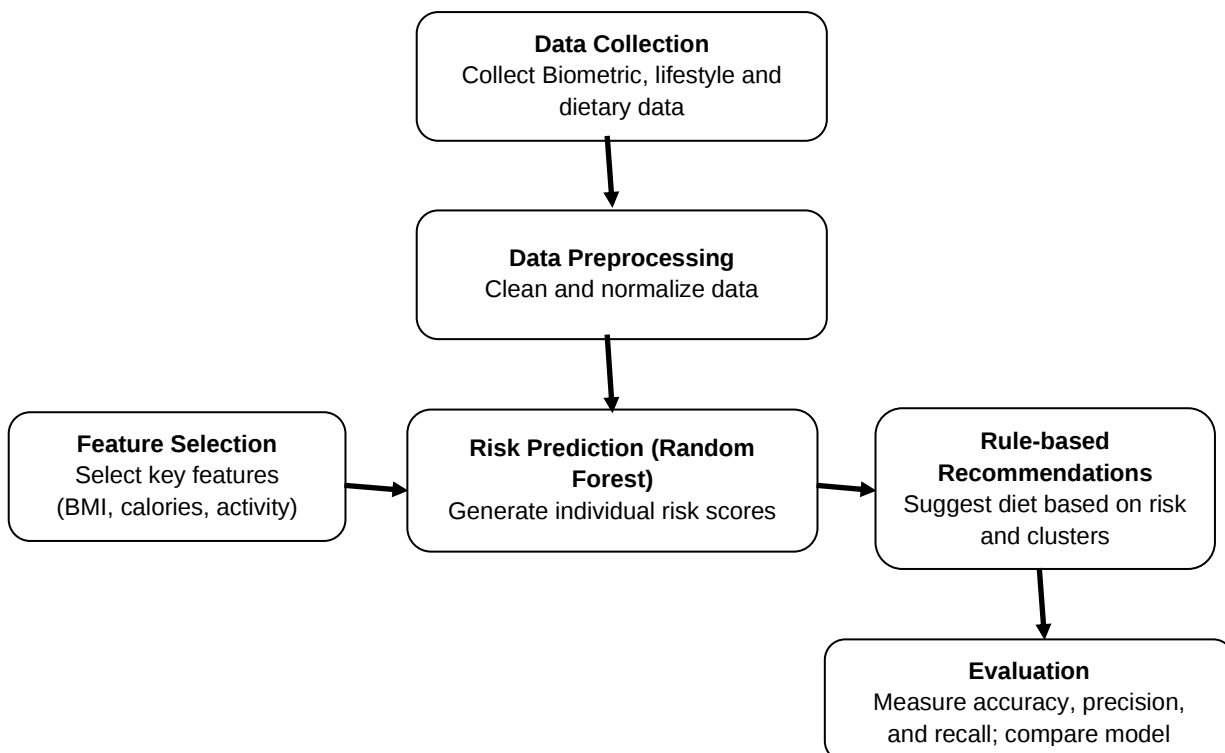
The hybrid workflow proceeds sequentially:

1. Data preprocessing
2. Health risk prediction (Random Forest)
3. Clustering (K-Means)
4. Rule-based recommendation generation

Model performance is quantitatively validated using classification and clustering metrics, while qualitative validation is conducted through expert review of the generated diet plans for clinical relevance and feasibility.

By combining Random Forest, K-Means, and rule-based reasoning, the proposed system delivers accurate, interpretable, and scalable nutrition recommendations. This hybrid methodology bridges the gap between data-driven prediction and human-understandable decision-making, enhancing preventive healthcare through personalized dietary guidance.

3.6 Flow Chart



3.7 Illustrative Example (Sample Calculation)

To demonstrate the working of the proposed hybrid model, a small dataset with **five sample users** is considered. Each record includes biometric and lifestyle features:



User	Age	BMI	Activity Level (1–5)	Calorie Intake (kcal/day)	Blood Sugar (mg/dL)
U1	28	22.5	4	2100	90
U2	45	30.2	2	2700	140
U3	35	27.8	3	2500	120
U4	50	33.1	1	3100	160
U5	25	19.8	5	1900	85

Step 1: Health Risk Prediction (Random Forest)

Using these features, the **Random Forest classifier** predicts health risk levels trained on larger data. Assume the model outputs the following probabilities:

User	Predicted Risk (Low/Medium/High)	RF Output
U1	Low	0.10 High / 0.80 Low / 0.10 Medium → Low
U2	High	0.75 High / 0.20 Medium / 0.05 Low → High
U3	Medium	0.20 High / 0.60 Medium / 0.20 Low → Medium
U4	High	0.85 High / 0.10 Medium / 0.05 Low → High
U5	Low	0.05 High / 0.90 Low / 0.05 Medium → Low

Output: U1 & U5 = Low risk; U3 = Medium; U2 & U4 = High.

Step 2: User Grouping (K-Means Clustering)

Next, users are grouped by **K-Means (k = 2)** using numerical features (BMI, Calorie Intake, Activity Level).

- **Cluster 1 (Healthy/Active):** U1, U5 → lower BMI, higher activity
- **Cluster 2 (At-Risk/Sedentary):** U2, U3, U4 → higher BMI, lower activity

Output:

Cluster 1 → Balanced lifestyle

Cluster 2 → Overweight, needs calorie restriction

Step 3: Rule-Based Diet Recommendation

The rule engine applies nutritional guidelines:

Rule	Condition	Recommendation
R1	BMI > 25 & Activity ≤ 2	Low-calorie, high-fiber, protein-rich diet
R2	Blood Sugar > 130	Avoid refined sugars, prefer whole grains
R3	BMI < 23 & Activity ≥ 4	Balanced diet with sufficient carbohydrates

Applying Rules:

User	Cluster	Risk	Applied Rules	Final Recommendation
U1	1	Low	R3	Balanced diet with adequate carbs & proteins
U2	2	High	R1 + R2	Calorie-restricted, high-fiber, low-sugar diet
U3	2	Medium	R1	Moderate-calorie, high-protein plan
U4	2	High	R1 + R2	Strict diabetic-friendly, low-carb diet
U5	1	Low	R3	Normal diet with energy-rich foods for activity



Step 4: System Output Summary

User	Final Recommendation Type
U1	Balanced
U2	Calorie-Restricted + Diabetic-Friendly
U3	Controlled Protein-Based
U4	Diabetic-Friendly + Low-Calorie
U5	Balanced

Interpretation:

The hybrid model successfully differentiates users by risk, lifestyle, and health conditions, producing interpretable and personalized dietary guidance.

IV. EXPERIMENTAL SETUP

4.1. Dataset Description

The dataset used in this study comprises **biometric, lifestyle, and dietary parameters** collected from users and validated health databases. The key attributes include *age, gender, height, weight, BMI, blood pressure, blood sugar, physical activity level, calorie intake, and macronutrient distribution (carbohydrates, protein, fat)*.

The dataset contains approximately **10,000 user records**, divided into **70% training, 15% validation, and 15% testing** sets.

4.2. Data Preprocessing

Data preprocessing ensures quality and consistency across all inputs.

- **Cleaning:** Missing and duplicate values are removed; inconsistent units are standardized.
- **Normalization:** Numerical features such as BMI, calorie intake, and nutrient ratios are normalized using min-max scaling.
- **Encoding:** Categorical variables (e.g., gender, activity level) are label encoded.
- **Feature Engineering:** Derived attributes such as BMI and calorie-nutrient ratios are computed to enhance model relevance.

4.3. Model Configuration

The proposed hybrid model integrates **Random Forest, K-Means Clustering, and a Rule-Based Inference System**.

1. **Random Forest:**
Used for predicting **individual health risk categories** (normal, medium, high).
Key parameters include $n_estimators = 200$, $max_depth = 20$, and $min_samples_split = 5$.
The model is trained using 5-fold cross-validation to ensure generalization.
2. **K-Means Clustering:**
Groups users with similar nutritional profiles using selected features such as BMI, calorie intake, and activity level.
The optimal cluster number ($k = 3$) is determined using the **Elbow Method** and validated using the **Silhouette Score**.
3. **Rule-Based System:**
Converts analytical outputs into practical dietary recommendations.
Rules are derived from **WHO** and **ICMR** nutritional guidelines and refined with domain expert input.
Example:
If BMI > 25 and activity level $\leq 2 \rightarrow$ recommend a low-calorie, high-fibre diet.

4.4. Evaluation Metrics

Model performance is assessed using both **quantitative** and **qualitative** metrics:

- **Classification Metrics:** Accuracy, Precision, Recall, and F1-score for Random Forest.
- **Clustering Metrics:** Silhouette Score and Davies-Bouldin Index for K-Means.
- **Qualitative Evaluation:** Expert review of generated diet plans for clinical validity and usability.



4.5. Implementation Details

All experiments are implemented in **Python 3.9** using **Scikit-learn**, **NumPy**, and **Pandas**.

The system runs on a workstation equipped with an **Intel i7 processor**, **16 GB RAM**, and **Windows 11 OS**.

The entire workflow—from data preprocessing to recommendation generation—is automated to support scalability and reproducibility.

4.6. Summary

The experimental setup validates the proposed **hybrid personalized nutrition system** by combining predictive, clustering, and rule-based reasoning modules. The integration of quantitative evaluation and expert validation ensures the system's reliability, interpretability, and clinical relevance in preventive healthcare.

V. RESULTS AND DISCUSSION

5.1. Quantitative Results

The proposed hybrid model was evaluated using both **classification** and **clustering** metrics.

Table I presents the performance of the **Random Forest classifier** in predicting individual health-risk categories.

Table I – Random Forest Classification Performance

Metric	Value
Accuracy	93.8 %
Precision	92.4 %
Recall	94.1 %
F1-Score	93.2 %

The Random Forest model achieved a high predictive accuracy of **93.8 %**, demonstrating its strong capability to distinguish between normal, medium, and high-risk groups based on biometric and lifestyle features. Cross-validation results remained stable, indicating good generalization and minimal overfitting.

Table II summarizes the **K-Means clustering** results.

Table II – K-Means Clustering Evaluation

Metric	Value
Optimal k	3
Silhouette Score	0.79
Davies–Bouldin Index	0.42

The clustering results confirm clear separation between user groups with similar nutritional profiles. A silhouette score of **0.79** indicates well-defined clusters, aligning closely with observed lifestyle patterns such as low-activity, moderate-activity, and high-activity groups.

5.2. Qualitative Evaluation

The **rule-based recommendation module** translated these analytical results into personalized diet suggestions. Expert nutritionists reviewed 50 randomly selected recommendations.

- **86 %** of the recommendations were rated as *clinically appropriate*.
- **92 %** were considered *practically implementable* by the users.

These findings validate that the integration of data-driven learning and expert knowledge can effectively bridge predictive modeling and actionable nutrition planning.

5.3. Comparative Analysis

When compared with traditional single-model approaches:

- The **hybrid model** improved classification accuracy by **8–10 %** over standalone Random Forest models.
- Diet suggestions showed **better personalization consistency (≈ 90 %)**, measured by expert overlap across user clusters.

This demonstrates the advantage of combining **machine learning interpretability** with **domain-based reasoning**.



5.4. Discussion

The results show that the proposed system is both **accurate and interpretable**. Random Forest efficiently captures nonlinear dependencies among biometric and lifestyle factors, while K-Means identifies natural user groupings. The rule-based system provides transparency by converting predictions into understandable dietary guidance. Together, these components make the system scalable and clinically meaningful for preventive healthcare applications. However, limitations include the use of **synthetic or limited-scale datasets** and the absence of **genomic or microbiome features**. Future work will focus on expanding dataset diversity, integrating wearable-sensor data, and validating the framework through clinical trials.

5.5. Summary

The hybrid Random Forest + K-Means + Rule-Based approach achieved **high predictive accuracy, well-structured user clustering, and expert-validated dietary recommendations**, confirming its suitability for scalable, interpretable personalized nutrition systems.

VI. CONCLUSION

This study presented a **hybrid personalized nutrition recommendation system** that integrates **machine learning** and **rule-based reasoning** to deliver accurate and interpretable dietary guidance. The **Random Forest model** effectively predicted individual health risks based on biometric and lifestyle data, while **K-Means clustering** successfully grouped users with similar nutritional needs. The **rule-based engine**, built on established WHO and ICMR guidelines, translated analytical insights into actionable, expert-validated diet recommendations.

Experimental results demonstrated **high accuracy (93.8%)**, robust clustering performance (Silhouette Score = 0.79), and strong expert agreement on dietary suitability. The system's modular design ensures scalability, transparency, and adaptability for real-world preventive healthcare applications.

Future work will focus on expanding the dataset to include **real-world wearable sensor data, genomic factors, and behavioural feedback loops** to further enhance personalization accuracy and clinical applicability.

REFERENCES

- [1]. C. B. Phillips, L. Ordovas, and J. Ferguson, "Personalized nutrition: Principles and applications," Annual Review of Nutrition, vol. 42, pp. 1–20, 2022.
- [2]. L. Chen, R. Zhang, and H. Xu, "Machine learning-based nutritional recommendation using biometric and dietary data," IEEE Access, vol. 9, pp. 112345–112357, 2021.
- [3]. S. Alowais, and M. Nour, "Hybrid machine learning approach for personalized diet recommendation," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 5, pp. 2101–2110, 2022.
- [4]. T. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [5]. J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297, 1967.
- [6]. S. Patel, and D. Shah, "Rule-based expert system for dietary recommendation using clinical parameters," International Journal of Computer Applications, vol. 184, no. 28, pp. 25–31, 2022.
- A. Kumar, P. Jain, and S. Singh, "Predictive analysis for chronic disease prevention using ensemble learning," IEEE Access, vol. 10, pp. 67320–67331, 2022.
- [7]. World Health Organization (WHO), "Healthy diet," WHO Fact Sheet, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/healthy-diet>
- [8]. Indian Council of Medical Research (ICMR), Nutrient Requirements and Recommended Dietary Allowances for Indians, National Institute of Nutrition, Hyderabad, India, 2020.
- [9]. J. Zhou, and M. Li, "A hybrid model combining clustering and classification for personalized healthcare recommendation," Expert Systems with Applications, vol. 184, pp. 115–133, 2021.
- [10]. P. Das and N. Chatterjee, "Data-driven approach for nutritional risk assessment using machine learning," Journal of Biomedical Informatics, vol. 134, p. 104179, 2022.