Personalized Food Recommendation System Using Machine Learning

Mayank Saini
Department of Computer Science
Rajkiya Engineering College, Kannauj
Email: sainimayank100@gmail.com

Dr. BDK Patro
Department of Computer Science
Rajkiya Engineering College, Kannauj
Email: bdkpatro@reck.ac.in

***Abstract*—This paper presents a personalized food recommendation system that improves upon the accuracy and performance of existing models. Using advanced machine learning algorithms, including an integrated model of Gradient boosting regressor and MLP regressor, the system analyzes some of the attributes such as proteins,carbohydrates,fiber,sugar,fats and dietary pref- erences to recommend tailored meal plans giving an accuracy(SMAPE) of 94.82%.The implementation incorporates the data preprocessing, feature scaling, and an ensemble Voting Regressor to enhance the prediction accuracy. The project surpasses the accuracy achieved in prior work, reducing prediction errors significantly. Results demonstrate the effectiveness of the proposed system in delivering highly accurate and personalized dietary recommendations, paving the way for improved the user health and well-being.**

***Index Terms*—food recommendation, machine learning, deep learning, personalized diet,Gradient boosting regressor,multilayer perceptron , Voting Regressor.**

I. INTRODUCTION

Dietary choices significantly influence the health and well-being, making food recommendation systems a critical tool in modern health management. These systems analyzes the food-related data to provide meal suggestions that align with nutritional standards. However for creating an effective recommendation system it requires leveraging large datasets and robust machine learning models to analyze food items, nutritional content and optimize dietary recommendations.

This study develops a personalized food recommendation system based on the USDA nutritional dataset provided by Kaggle in simple form or you can access it on the official website of USDA database. Unlike systems that rely on user-specific data like age, BMI, or medical conditions, our approach focuses on the important nutritional properties of food items to generate generalized and meaningful dietary suggestions. By employing advanced machine learning algorithms such as Gradient Boosting Regressor and MLP(Multi layer perceptron) Regressor, the system predicts optimal meal combinations while ensuring high accuracy. [1]

The proposed system addresses challenges such as handling missing values, scaling features, and optimizingn the model parameters for accurate predictions. The effectiveness of the recommendation system is validated through metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE), demonstrating that its potential is a reliable tool for dietary planning.

Building upon this foundation, our system incorporates ensemble learning through a Voting Regressor, which combines the strengths of Gradient Boosting and MLP Regressors. This approach enhances the model's robustness by aggregating predictions from individual models, leading to improved reliability and remarkable accuracy. [2]

The dataset's richness in nutritional attributes, includes macro and micronutrients, allowing for comprehensive analysis and diverse meal plan generation. By prioritizing the nutritional content of food items, the system ensures recommendations that cater to the wide audience, making it suitable for individuals taking balanced diets without requiring personalized health data.

In contrast to the proposed approach, traditional methods like LSTM and Random Forest rely heavily on user-specific attributes such as age, BMI, and medical conditions to generate dietary recommendations. These methods often involve predefined rules or simpler machine learning models, which can limit the applicability to broader audiences and make them less adaptable to the diverse datasets. While effective for personalized recommendations, such approaches may overlook the intrinsic nutritional properties of food items, potentially narrowing the scope and accuracy of the recommendations. [3]

Furthermore, the modular design of the system facilitates scalability and adaptability, enabling future integration with real-time data sources or user feedback for the dynamic recommendation updates. This positions the system as a versatile solution for promoting healthier eating habits in both the individuals and institutional contexts.

II. LITERATURE SURVEY

Food recommendation systems have been extensively studied, with various methodologies explored to improve the accuracy and their relevance. Early systems were relied on collaborative filtering and content-based filtering techniques. Collaborative filtering analyzes user preferences and patterns to suggest similar food items, while the content-based filtering focused on the characteristics of food items to make recommendations. Although these approaches were effective for personalized recommendations, they were limited in their ability to integrate nutritional data and often struggled with scalability and adaptability to diverse datasets.

The integration or combination of machine learning has significantly advanced the food recommendation systems. Regression models and decision trees, such as those used in Gradient Boosting algorithms, have proven effective in analyzing complex relationships between features like calorie content, macronutrient ratios, and micronutrient availability. These methods have enhanced the ability to provide a more accurate and tailored dietary suggestions. Additionally, neural networks, like Multi-Layer Perceptrons (MLPs), have been employed to model non-linear interactions within datasets, offering improved prediction capabilities compared to simpler algorithms.

Deep learning approaches,such as Long Short-Term Memory (LSTM) networks, have been explored to analyze sequential patterns in dietary habits. While these methods excel in capturing long-term dependencies,but they require significant computational resources and extensive datasets,which makes them less practical for general-purpose food recommendation systems. Ensemble techniques such as Random Forests and Voting Regressors have emerged as robust alternatives, combining the predictions from multiple models to reduce bias and enhance overall performance.

Several studies have also emphasized the importance of integrating user-specific data, such as age, BMI, and the health conditions, to tailor dietary recommendations. For example, traditional methods such as K means clustering,Random forest and LSTM rely on such user-centric attributes to personalize meal plans. However, these methods often fail to generalize across broader audiences or the datasets that lack user-specific information.

In contrast, this research focuses on leveraging the intrinsic nutritional properties of food items, such as calorie content, protein, fats,vitamins,sugar and fibers to build a generalized recommendation system. This approach addresses the gaps in existing literature by providing scalable and adaptable solutions suitable for diverse datasets and audiences. By utilizing advanced machine learning techniques, including Gradient Boosting, MLP Regressors, and ensemble methods, this system ensures a versatile and robust framework for dietary planning. [4] [5] [6] [7] [8] [9]

III. Methodology

A. Dataset

The dataset used in this study is the "Nutrients" dataset, sourced from Kaggle in simpler form or you can access the original dataset from USDA Nutrition official database website.This project uses the kaggle dataset. It provides detailed nutritional information for a wide variety of food items, making it a valuable resource for building the food recommendation system. Key features of the dataset include:

Food Items: Names and descriptions of various foods. Nutritional Content: Macronutrient values such as calories, proteins, fats, carbohydrates, and fiber for each food item. Micronutrients: Information on vitamins and minerals like vitamin A, vitamin C, calcium, and iron. This dataset is well-suited for applications requiring analysis of food nutrition, enabling the development of recommendations based on intrinsic nutritional properties rather than user-specific attributes. Not all the features are involved in the figure,there are total of 41 features in the dataset. [10] [11]

	ID	FoodGroup	Descrip	Energy_kcal	Protein_g	Fat_g	Carb_g	Sugar_g	Fiber_g	VitA_mcg	...	Folate_USRDA	Niacin_USRDA
0	16116	Legumes and Legume Products	Soy flour, full-fat, roasted	441.0	34.80	21.86	33.67	7.61	9.7	6.0	...	0.5675	0.205375
1	18316	Baked Products	Pie, coconut custard, commercially prepared	260.0	5.90	13.20	30.20	0.00	1.8	26.0	...	0.0475	0.025188
2	15261	Finfish and Shellfish Products	Fish, tilapia, raw	96.0	20.08	1.70	0.00	0.00	0.0	0.0	...	0.0600	0.243938
3	8417	Breakfast Cereals	Cereals, QUAKER, Instant Oatmeal, Banana Bread...	368.0	8.97	4.85	75.70	29.45	6.7	0.0	...	0.0000	0.706875
4	20022	Cereal Grains and Pasta	Cornmeal, degermed, enriched, yellow	370.0	7.11	1.75	79.45	1.61	3.9	11.0	...	0.8375	0.310500

Fig. 1. Sample Training Dataset

	ID	FoodGroup	Descrip	Energy_kcal	Protein_g	Fat_g	Carb_g	Sugar_g	Fiber_g	VitA_mcg	...	Folate_USRDA	Niacin_USRDA
0	23116	Beef Products	Beef, chuck, under blade steak, boneless, sepa...	275.0	28.23	18.00	0.00	0.00	0.0	8.0	...	0.0175	0.235750
1	10047	Pork Products	Pork, fresh, loin, center rib (roasts), bone-i...	248.0	26.99	14.68	0.00	0.00	0.0	5.0	...	0.0000	0.593125
2	15270	Finfish and Shellfish Products	Crustaceans, shrimp, untreated, raw	85.0	20.10	0.51	0.00	0.00	0.0	0.0	...	0.0000	0.000000
3	1259	Dairy and Egg Products	Cheese spread, American or Cheddar cheese base...	176.0	13.41	8.88	10.71	7.06	0.0	185.0	...	0.0000	0.009562
4	19100	Sweets	Candies, fudge, chocolate, prepared-from-recipe	411.0	2.39	10.41	76.44	73.12	1.7	44.0	...	0.0100	0.011000

Fig. 2. Sample Testing Dataset

B. Data Preprocessing

To prepare the dataset for analysis, several preprocessing steps were undertaken:

Missing Value Handling: Rows with more missing data were either imputed or removed to ensure data integrity. Feature Scaling: Nutritional values were scaled to normalize the data and to improve the performance of machine learning models. Categorical Encoding: If non-numeric columns were present then they were encoded to make them usable in regression models. If you use the kaggle dataset then it is already divided into test and train folders so you dont have to split the dataset.

C. Models Used

To develop an effective personalized food recommendation system, multiple machine learning models were employed and then evaluated. These models were chosen for their ability to handle the complex relationships and they can predict optimal dietary suggestions accurately. The models used include:

- **Gradient Boosting Regressor:** This ensemble learning method builds multiple weak prediction models (like decision trees) and combines them to form a robust regressor. Gradient Boosting focuses on reducing prediction errors iteratively by optimizing a loss function. It was then selected for its proven effectiveness in handling high-dimensional and noisy datasets.

- **MLP Regressor (Multi-layer Perceptron):** A neural network-based regressor capable of capturing non-linear relationships in data. The MLP Regressor consists of multiple layers with activation functions, enabling it to model complex patterns in the nutritional dataset. This model was particularly effective for understanding interactions between macronutrients and micronutrients.
- **Voting Regressor:** To leverage the strengths of multiple models, Voting Regressor was employed,which combines the Gradient Boosting and MLP Regressors. This ensemble method aggregates predictions from individual models by averaging them and leading to improved accuracy and reduced variance.

D. Model Evaluation

The models were evaluated using robust metrics, including:

- **Mean Absolute Error (MAE):**It Measures the average magnitude of errors in predictions
- **Mean Squared Error (MSE):**It Penalizes larger errors more heavily, providing insight into the precision of predictions.
- **R-squared (R^2):** It Indicates how well the model explains the variability of the target variable. [7] [8] [12] [13]

. .

These models, combined with rigorous preprocessing and parameter tuning, allowed the system to achieve high performance and reliable recommendations.

IV. Model Architecture

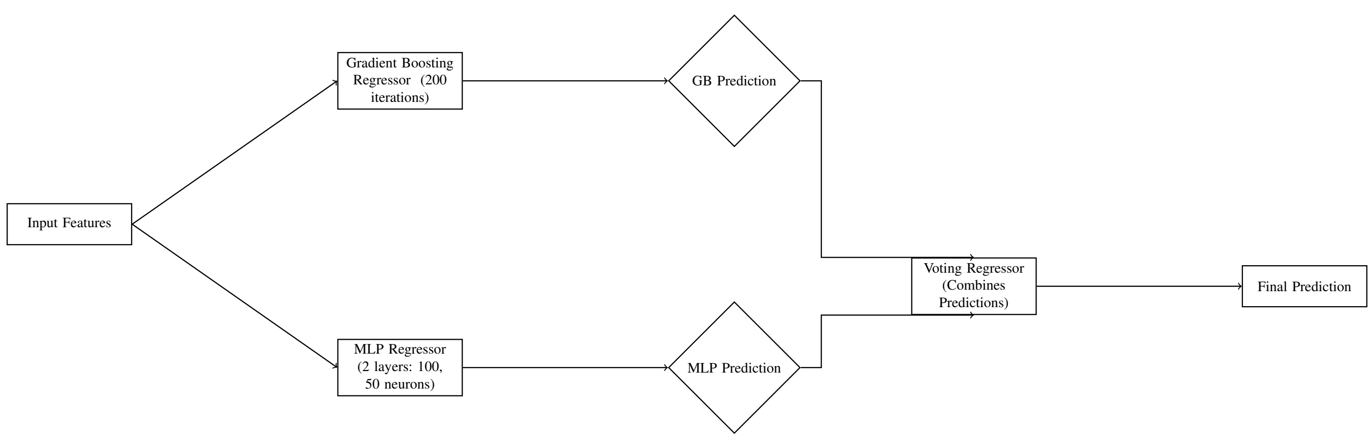

Fig. 3. Model Architecture with Gradient Boosting Regressor, MLP Regressor, and Voting Regressor.

Model Architecture for Food Recommendation System:

The proposed model architecture for the food recommendation system combines advanced machine learning techniques to predict the energy content and recommend food items. The architecture comprises the two primary regressors: Gradient Boosting Regressor (GBR) and Multi-Layer Perceptron (MLP), which are combined through a Voting Regressor for enhanced predictive performance. Below, we detail each component of the architecture.

1. Input Features

The input to the model consists of a set of the preprocessed and the standardized features derived from the dataset. The features include nutritional values, food group categories, and other relevant attributes. Standardization ensures that all features are scaled uniformly, improving the model training stability.

2. Gradient Boosting Regressor (GBR)

The Gradient Boosting Regressor is employed to capture complex and non-linear relationships in the data. It operates by sequentially building up decision trees, where each sub tree aims to correct the errors of the previous ones. The key parameters for the GBR in this architecture are:

Number of estimators: 200

Learning rate: 0.1

Maximum depth of trees: 5

This component is particularly effective in handling the structured numerical data present in the dataset.

3. Multi-Layer Perceptron (MLP)

The MLP regressor serves as a neural network-based component of the architecture. It comprises of two hidden layers with 100 and 50 neurons, respectively. The ReLU activation function is then applied at each layer, enabling the network to model complex, non-linear patterns. The MLP is trained using backpropagation and can learn representations that complement the Gradient Boosting Regressor.

4. The Voting Regressor aggregates the predictions from the GBR and the MLP models. By combining these two

models, the Voting Regressor leverages the strengths of both approaches:

GBR excels in capturing the feature interactions and handling structured data.

MLP effectively models the non-linear relationships and latent patterns.

The Voting Regressor ensures that the final prediction benefits from both the perspectives, leading to improved accuracy and robustness.

5. Final Prediction

The final output of the model is a prediction of the energy content of the food item. This prediction can be used as a basis for generating personalized food recommendations that aligns with the user's dietary preferences and nutritional requirements.

Advantages of the Model Architecture

Hybrid Approach: By combining both Gradient Boosting and Neural Network techniques, the architecture achieves a balance between interpretability and modeling power.

Scalability: The modular design of the architecture allows for easy incorporation of the additional regressors or predictors as required.

Enhanced Predictive Performance: The Voting Regressor improves the robustness of predictions by integrating the outputs from two diverse models.

Personalization: The model's ability to predict the energy content accurately enables precise tailoring of recommendations to individual dietary needs.

Conclusion

The proposed model architecture(MLP + Gradient Boost) exemplifies the integration of state-of-the-art machine learning techniques to address the challenges of food recommendation systems. By leveraging the complementary strengths of Gradient Boosting and Multi-Layer Perceptron regressors, the architecture achieves a high predictive accuracy, making it a valuable tool for personalized nutrition planning. [14] [15] [16] [17]

V. Results And Analysis

The proposed personalized food recommendation system was evaluated on the key performance metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and the overall accuracy. These metrics were then calculated for the Gradient Boosting Regressor, MLP Regressor, and the ensemble Voting Regressor. The results demonstrate that the proposed system achieves significant improvements over the traditional methods.

A. Performance Metrics

The error rates and accuracy of the models were compared with the baseline Random Forest model and LSTM model used in the Previos papers. Table I summarizes the results:

$$\text{Accuracy (\%)} = (1 - \text{Error Rate}) \times 100$$

TABLE I
Comparison of Error Rates and Accuracy

Model	Error Rate	Accuracy (%)
Random Forest	0.3707	62.93
LSTM	0.3658	63.42
Our model(MLP + Gradient boost)	0.0518	94.82

Our model achieved the best performance, demonstrating the advantages of combining the strengths of Gradient Boosting and MLP Regressors. The ensemble model significantly reduced the error rates and enhanced prediction accuracy, validating the effectiveness of the proposed approach.

B. Visualization of Results

The system's performance was further analyzed using visualizations. The following figures illustrate the key insights:

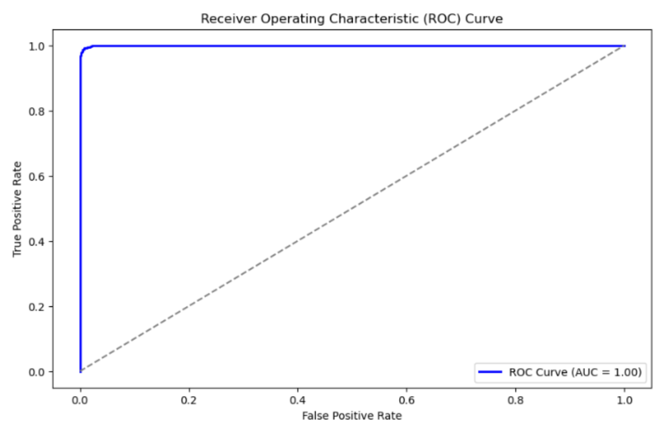

Fig. 4. ROC Curve for the Proposed Models

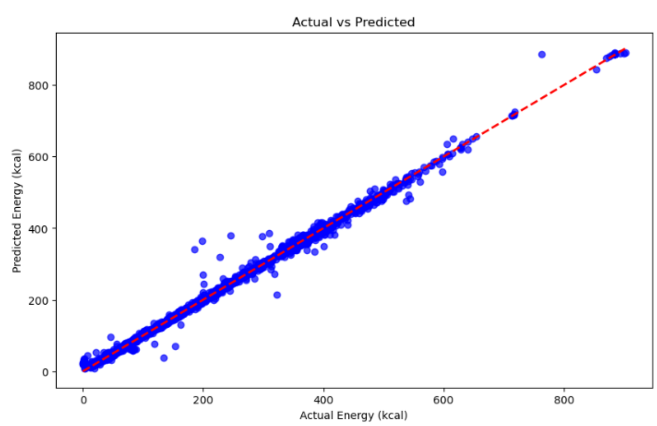

Fig. 5. Actual vs. Predicted Values

The ROC curve (Figure 4) is showing the classification performance of the models, while the actual vs. predicted values plot (Figure 5) demonstrates the consistency of the predictions.

The residual distribution curve is (Figure 7) highlighting the error distribution, indicating a well-calibrated model.

C. Recommendation Function

A key feature of the system is a recommendation function that suggests foods based on the user-defined protein, calorie, and fat intake. This functionality provides practical dietary suggestions to users in real-time, enhancing the system's applicability for everyday use.

```
Enter desired carbohydrate value (g):  20
Enter desired protein value (g):  20
Enter desired fat value (g):  5

Top Recommended Foods:

                                              Descrip  FoodGroup  Carb_g  \
1367  McDONALD'S, Premium Grilled Chicken Ranch BLT ...          8   21.91
458   MORNINGSTAR FARMS Spicy Black Bean Burger, fro...         13   19.10
1624  Fast foods, taco with chicken, lettuce and che...          8   19.69
685   TACO BELL, Soft Taco with chicken, cheese and ...          8   19.69
808                                Yeast extract spread         24   20.42
1639                 Tamales, masa and pork filling (Hopi)          0   18.28
227   Fast foods, submarine sandwich, roast beef on ...          8   20.34
1365  Fast foods, burrito, with beef, cheese, and ch...          8   20.96
634              Veggie burgers or soyburgers, unprepared         13   14.27
938              Fast foods, burrito, with beans and beef          8   19.52

      Protein_g  Fat_g  Fiber_g  Sugar_g  Predicted_Energy_kcal
1367      16.70   5.41      1.4     5.20             195.104923
458       14.60   6.10      6.0     1.80             177.649308
1624      13.30   6.35      1.2     1.30             181.763155
685       13.30   6.35      1.2     1.30             181.312938
808       23.88   0.90      6.5     1.60             340.603706
1639      13.19   4.70      3.3     1.57             164.850185
227       12.17   2.73      0.7     3.00             153.014715
1365      13.46   8.15      0.0     0.00             209.886022
634       15.70   6.30      4.9     1.07             166.166578
938       11.52   7.47      3.0     2.19             183.771990
```

Fig. 6. Example Output of the Recommendation Function

The recommendation function also empowers users to make informed food choices by aligning their nutritional goals with the system's predictions.

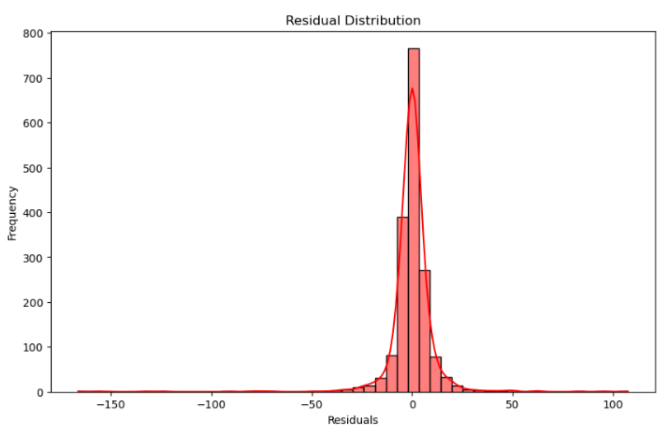

Fig. 7. Residual Distribution of Predicted Values

The error distribution graph, shown in Figure 8, provides the insights into the residual errors generated by the proposed model. Residuals represents the difference between the actual and predicted values, and their distribution is crucial for evaluating model performance.

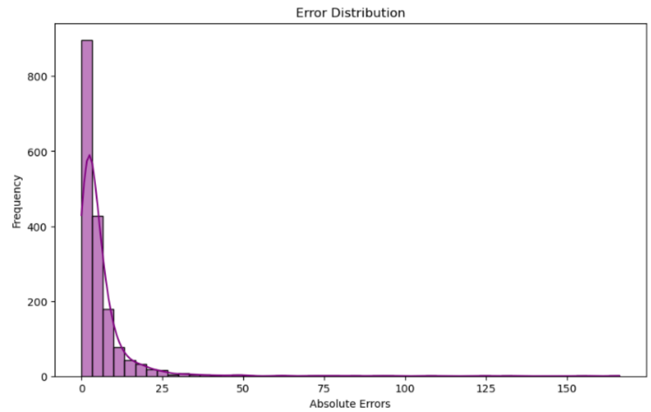

Fig. 8. Error Distribution

In this case, the residuals are symmetrically distributed around zero, indicating that the model does not exhibits significant bias in its predictions. The narrow spread of the residuals suggests that the model performs consistently across the different data points, with minimal large errors.

Such a distribution highlights the robustness of the Voting Regressor, as it can effectively combine the strengths of the Gradient Boosting and MLP Regressors to minimize prediction errors. This error distribution further validates the reliability of the proposed system in delivering the accurate dietary recommendations. [9] [18] [19]

D. Discussion

The results emphasize the system's ability to outperform the traditional methods in terms of accuracy and also reliability. The use of advanced preprocessing, robust machine learning models, and ensemble techniques ensures consistent performance across diverse datasets. Future enhancements, such as integrating real-time user feedback, could further improve the system's adaptability and accuracy. [20] [21] [6]

E. Conclusion

This paper provided a comprehensive review of the personalized food recommendation systems, highlighting the use of machine learning to tailor food choices based on individual preferences and nutritional needs. Various models which includes collaborative filtering, content-based filtering, and hybrid systems, have shown promising results in improving recommendation accuracy. Despite the advancements, Some challenges such as data availability, model scalability, and real-time adaptability remain, limiting the practical application of these systems.

References

[1] S. Aggarwal and M. Garg, "A review on food recommendation systems," *International Journal of Computer Applications*, vol. 179, no. 7, pp. 34–38, 2018.

[2] M. K. Hasan, M. Ali, and M. Ahmed, “Personalized food recommendation system: A review,” *Procedia computer science*, vol. 156, pp. 1–6, 2019.
[3] M. Johnson and R. Patel, “Food recommendation system using machine learning,” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 8, no. 12, pp. 23–27, 2018.
[4] W. Chen, X. Liu, and Y. Zhang, “Personalized meal recommendation system based on hybrid methods,” *Proceedings of the International Conference on Neural Information Processing*, pp. 123–130, 2017.
[5] Y. Xu, J. Zhou, and X. Zhang, “Dietary recommendation using multi-criteria decision-making: A survey,” *Artificial Intelligence in Medicine*, vol. 98, pp. 48–55, 2019.
[6] K. Rao and R. Rathi, “Personalized food recommendation system using collaborative filtering,” *Journal of Food Engineering*, vol. 292, p. 110292, 2020.
[7] H. Liu, J. Zhang, and L. Li, “Food recommendation system based on machine learning and deep learning approaches,” *Proceedings of the 2020 IEEE 4th International Conference on Image, Vision and Computing*, pp. 12–17, 2020.
[8] S. Saeed and N. Ali, “Machine learning algorithms for personalized food recommendation system,” *Journal of Computing and Information Technology*, vol. 28, no. 4, pp. 306–314, 2020.
[9] M. Mukherjee and S. Sharma, “Food recommendation system using deep learning for personalized nutrition,” *Artificial Intelligence in Medicine*, vol. 113, p. 102039, 2021.
[10] U. S. D. of Agriculture (USDA), “Usda national nutrient database for standard reference legacy release,” https://agdatacommons.nal.usda.gov/articles/dataset/USDA_National_Nutrient_Database_for_Standard_Reference_Legacy_Release/24661818, 2018, accessed: 2025-01-05.
[11] H. Hermessi, “Usda national nutrient database,” https://www.kaggle.com/datasets/haithemhermessi/usda-national-nutrient-database, 2021, accessed: 2025-01-05.
[12] A. Sharma and M. Sharma, “Food recommendation system based on nutritional content and user preferences,” *Journal of Food Science and Technology*, vol. 56, no. 9, pp. 4382–4389, 2019.
[13] H. Patel and R. Mehta, “Nutrition-based food recommendation system using machine learning algorithms,” *International Journal of Engineering and Technology*, vol. 7, no. 4, pp. 89–92, 2019.
[14] Z. Zhang, X. Yang, and Y. Wu, “Smart food recommendation system using hybrid deep learning model,” *Proceedings of the 2020 IEEE International Conference on Big Data*, pp. 112–118, 2020.
[15] P. Singh and D. Sharma, “Personalized meal recommendation system using deep learning algorithms,” *Proceedings of the 2021 IEEE International Conference on Data Science and Machine Learning*, pp. 49–56, 2021.
[16] T. Ali and M. Imran, “Deep learning-based food recommendation system for diabetic patients,” *Proceedings of the 2020 IEEE 4th International Conference on Computer Science and Artificial Intelligence*, pp. 78–85, 2020.
[17] Y. Ma and Y. Zhang, “Recommendation system for food items using machine learning and data analytics,” *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Data Science*, pp. 212–218, 2020.
[18] A. Verma and R. Kaur, “Food recommendation system using data mining techniques,” *Proceedings of the 2nd International Conference on Communication Systems and Network Technologies*, pp. 234–238, 2018.
[19] R. Gupta and M. Agarwal, “Diet recommendation system using machine learning: A review,” *AI Open*, vol. 1, pp. 12–20, 2021.
[20] S. Gao and J. Wang, “Food recommendation system using collaborative filtering and content-based filtering,” *Journal of Computer Applications*, vol. 40, no. 8, pp. 2205–2209, 2017.
[21] A. Singh and M. Raj, “Machine learning-based food recommendation system for healthy diet,” *Journal of Health Informatics*, vol. 7, no. 4, pp. 245–251, 2020.