Journal of Science Research and Reviews
PRINT ISSN: 1595-9074
E-ISSN: 1595-8329
DOI: https://doi.org/10.70882/josrar.2024.v1i2.40
Homepage: https://josrar.esrgngr.org
Original Research Article

Leveraging Machine Learning for Personalized Dietary Recommendations, Nutritional Patterns, and Health Outcome Predictions

[1]Olutunde, Timothy, *[1]Ani, Chukwuemeka L. and [2]Adesue, Godwin A.

[1]Department of Statistics, Faculty of Science, Air Force Institute of Technology, Mando, Kaduna, Nigeria.
[2]Department of HIV Prevention, AIDS Healthcare Foundation, Lafia, Nasarawa State, Nigeria.
*Corresponding Author's email: emekaani605@gmail.com ORCID iD: https://orcid.org/0000-0002-2546-3875

KEYWORDS
Machine learning,
Nutrition,
Food,
Health,
Dietary patterns,
Data integration.

ABSTRACT
Unhealthy dietary patterns are key contributors to chronic diseases such as obesity, diabetes, and cardiovascular conditions. This study employs machine learning (ML) techniques to analyze dietary intake, identify patterns, and assess their relationships with health outcomes. The aim is to provide personalized dietary recommendations and insights to promote healthier eating habits.
Data for this research were sourced from a Kaggle dataset on foods and nutrients and the National Health and Nutrition Examination Survey (NHANES) on health outcomes. Preprocessing steps included data cleaning, feature selection, and transformation using one-hot encoding and scaling techniques. Machine learning algorithms were applied to build a food recommendation system and a diet health check system. Visualizations such as correlation heatmaps, scatter plots, and dashboards further illustrated the relationships between demographic factors, nutrient intake, and health outcomes. The food recommendation system effectively identified foods with similar nutritional profiles to user preferences. For instance, it suggested nutrient-rich alternatives like rice noodles and kale, achieving similarity scores above 0.99 in multiple test cases. The diet health check system analyzed nutrient intake against predefined thresholds and provided tailored recommendations. Excessive carbohydrate, protein, fat, and cholesterol consumption were linked to conditions such as diabetes, coronary heart disease, and cancer, with specific dietary adjustments suggested for improvement. This study demonstrates the power of machine learning in personalizing dietary advice and enhancing health outcomes. By leveraging advanced algorithms and diverse datasets, the developed systems present a scalable solution for promoting balanced diets and mitigating chronic disease risks. Further refinement and broader implementation of these tools are recommended to maximize their impact on public health.

CITATION
Olutunde, T., Ani, C. L., & Adesue, G. A. (2024). Leveraging Machine Learning for Personalized Dietary Recommendations, Nutritional Patterns, and Health Outcome Predictions. *Journal of Science Research and Reviews*, *1*(2), 43-56. https://doi.org/10.70882/josrar.2024.v1i2.40

INTRODUCTION

Nutrition plays a pivotal role in human health, with dietary habits closely associated with various health outcomes, particularly chronic diseases such as obesity, diabetes, and cardiovascular conditions (Schulz et al., 2021; Hassan et al., 2021). The prevalence of nutrition-related chronic diseases is extensive, stemming from multifaceted origins that necessitate a comprehensive and nuanced approach to data analysis for effective intervention. Recent advancements in technology and computational methodologies have empowered researchers to utilize high-dimensional data techniques, facilitating deeper understanding of these diseases and other complex inquiries (Curion & Theis, 2024). The shift towards high-dimensional data analysis is motivated by several factors, as outlined by Schwedhelm et al. (2021). Primarily, the intricate nature of food composition—where each food item comprises various nutrients that interact in complex ways—renders it impractical to isolate and evaluate their individual impacts on health outcomes. Furthermore, typical dietary patterns are characterized by a combination of diverse foods exhibiting substitution effects; an increase in the consumption of certain foods often corresponds with a decrease in others. This interdependence complicates the analysis of individual food items within an analytical model, as multi-collinearity arises from the complex interactions among them, making it challenging to draw definitive conclusions regarding specific foods (Shang et al., 2023). In contrast, dietary patterns encapsulate the intricate relationships among different foods and nutrients as a cohesive whole. They reflect actual eating behaviors and provide more comprehensive insights into the associations between multiple nutrients and health outcomes. Notably, dietary patterns tend to exhibit greater stability over time and exert a more significant influence on health than isolated nutrients. Consequently, the analysis of dietary patterns is increasingly recognized as a valuable complement to traditional studies focused solely on individual nutrients or foods (Wang et al., 2020). Over recent decades, innovative statistical methods have emerged that leverage dietary information collected from diverse populations to construct dietary patterns. In nutritional epidemiology, regardless of the specific statistical techniques employed for dietary pattern analysis, the overarching aim remains the exploration of the connections between these patterns and health outcomes (Sakiyo et al., 2025; Wang et al., 2020). Moreover, numerous novel methodologies have been applied to dietary pattern analyses that have either received insufficient review or have necessitated data transformations to align with these methods (Shang et al., 2023; Zhao et al., 2021). The complexity inherent in understanding the relationships among foods, nutrients, and health outcomes is further compounded by various influencing factors, including cultural, social, economic, and environmental contexts (Shang et al., 2023).

Traditional research methodologies in nutrition—such as dietary surveys and food diaries—often present limitations that include potential biases and incomplete data capture (Fleischhacker et al., 2020). These conventional approaches typically operate under the assumption of no synergistic effects among dietary components, which can introduce biases if such interactions are not adequately addressed. In this context, machine learning algorithms emerge as a promising alternative to mitigate these limitations (National Academies of Sciences, Engineering, and Medicine et al., 2024). Machine Learning (ML), a subset of Artificial Intelligence (AI), employs algorithmic processes that have demonstrated potential in identifying causes and solutions for various nutrition-related non-communicable diseases characterized by complex and multifactorial origins (Kirk et al., 2023; Gupta et al., 2022; Ma et al., 2021). Within nutritional science, there is an escalating demand for sophisticated tools capable of generating and analyzing intricate datasets. While ML has the potential to enhance existing analytical techniques, its application must be judiciously executed to prevent biased models and unrealistic conclusions (An, 2023). This study aims to investigate the application of machine learning in analyzing dietary patterns alongside health outcomes to enhance our comprehension of the interplay between foods, nutrients, health outcomes, and dietary habits.

MATERIALS AND METHODS

Data Collection

The data employed in this study was sourced from a combination of dietary intake records, medical histories, and demographic information. Specifically, the dataset is a secondary source comprising 7,413 observations. The dataset is structured into several key variables, including Demographic Variables (age and sex), Covariates (weight, height, and body mass index), Food Type, Food Mix, Dietary Patterns (carbohydrate, cholesterol, fiber, protein, calories, fats, vitamins, minerals, water, and sugar), and Health Outcomes (heart disease, stroke, diabetes, cancer, chronic respiratory diseases such as asthma, and liver diseases). To facilitate the analysis, two distinct datasets were collected for this study: one focusing on food and nutrient intake and the other on health outcomes. The first dataset was obtained from Kaggle (https://www.kaggle.com/), specifically curated by Shruti Saxena (https://www.kaggle.com/datasets/shrutisaxena/food-nutrition-dataset). The second dataset was sourced from the National Health and Nutrition Examination Survey (NHANES), a comprehensive program designed to assess the health and nutritional status of adults and children in the United States (https://nces.ed.gov/fCSM/nhanes.asp).

Data Preprocessing

Before conducting our analysis, it is essential to preprocess the collected data. Data preprocessing is a fundamental step in any machine-learning project as it ensures that the data is clean, consistent, and suitable for analysis. The process our study followed involves the following critical steps:

1. Data Cleaning: This step involves cleaning the raw data to remove duplicate entries and handle missing values. Missing values are either imputed or removed based on their significance to the analysis. Additionally, outliers are identified and addressed to ensure that they do not distort the results.
2. Feature Engineering: Once the data is cleaned, feature engineering is performed to transform it into a format that enhances its utility for machine learning models. This includes creating new variables or modifying existing ones to better capture the relationships between dietary patterns and health outcomes. Feature selection is also conducted during this stage to identify the most relevant variables for analysis. For instance, features such as nutrient intake (e.g., carbohydrates or vitamins), caloric consumption, and food group preferences were selected due to their potential influence on health outcomes.
3. One-Hot Encoding: To handle categorical variables like "food type," one-hot encoding was applied. This technique converts categorical variables into binary variables that can be easily interpreted by machine learning algorithms. For example, each category within "food type" is represented as a separate binary variable.
4. Data Scaling: After encoding categorical variables into binary formats, scaling was performed to ensure consistency across all features. Scaling involves transforming numerical variables so that they have a uniform scale or range. This step is critical for improving model performance since many machine-learning algorithms are sensitive to differences in feature magnitudes.

Finally, we adopted these preprocessing steps—data cleaning, feature engineering (including feature selection and one-hot encoding), and scaling—the datasets, so as to ensure that the employed data is not only accurate but also optimized for uncovering meaningful relationships between dietary patterns and health outcomes.

Diet Health Check System

The diet health check system employed a rule-based approach to assess users' dietary patterns and provide personalized recommendations (Varshney et al., 2023). The system compared users' nutrient intake to recommended intake levels and generated recommendations for balancing nutrient intake and promoting a healthier diet.

Software and Tools

The data preprocessing, model-training and analysis are conducted using Python programming language. Analysis for Food and Nutrient are conducted using Power Bi. Libraries such as pandas, NumPy, Scikit-learn, and Tensor-Flow are utilized for data manipulation, algorithm implementation, and evaluation.

RESULTS AND DISCUSSION

Descriptive and Summary Statistics for Food & Nutrient and Demography & Health Outcomes

This sub-section presents the descriptive and summary statistics for both the food & nutrient and demography and health outcomes variables in our dataset. The measures of central tendency used are mean and median, while, the measures of dispersion is standard deviation as shown in the Table 1 and also, Figure 1 shows the graphical representation of the distribution of some food and nutrient variables.

Table 1: Descriptive Statistics for Food and Nutrient

Measures	Carbohydrate	Protein	Fats	Vitamin	Mineral	Water
Count	7413	7413	7413	7413	7413	7413
Mean	21.785381	10.809883	19.196398	887.814803	862.982380	55.168469
Std. Dev.	27.123491	10.483772	31.706753	4483.305844	1241.360189	30.906416
Min	0	0	0	0	0	0
Median	9.290000	7.270000	7.660000	51.330000	647.169000	64.280000
Max	100.000000	88.320000	198.780000	130000.00000	38791.460000	100.000000

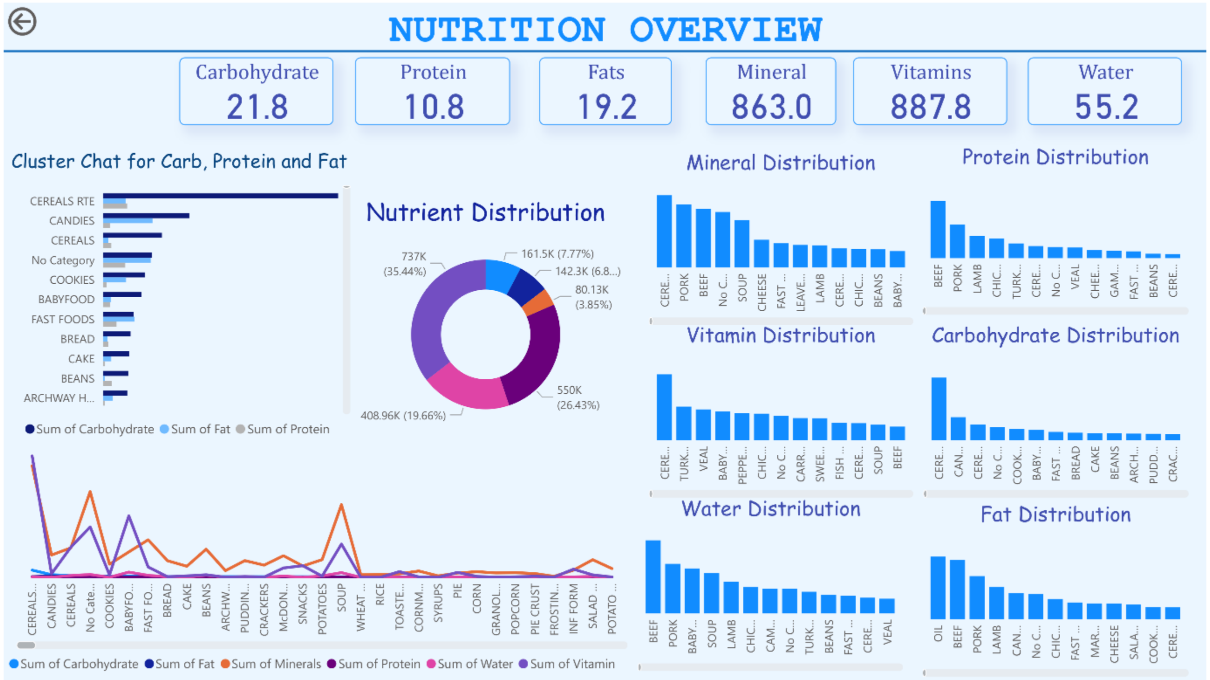


Figure 1: Nutrition overview

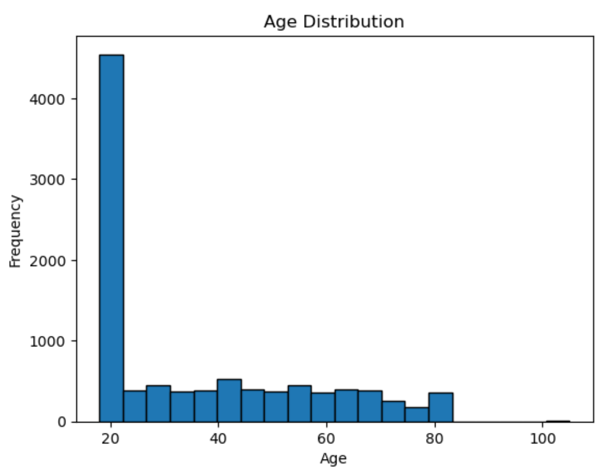
From Table 1, the mean carbohydrate intake is approximately 21.79 grams. This value represents the average daily consumption of carbohydrates; the standard deviation of approximately 27.12 indicates substantial variability in carbohydrate intake meaning some individuals in the dataset have significantly higher or lower carbohydrate consumption than the mean. Carbohydrate intake ranges from 0 grams (indicating some participants have very low or no carbohydrate consumption) to 100 grams (suggesting that some participants have high carbohydrate intake). The mean protein intake is about 10.81 grams, reflecting the average daily protein consumption from the dataset, similar to carbohydrates, protein intake exhibits variability with a standard deviation of approximately 10.48, this implies that protein consumption varies widely among participants, also protein intake ranges from 0 grams to 88.32 grams. The mean fat intake is approximately 19.20 grams, the standard deviation for fat intake is relatively high at around 31.71, and fat intake varies from 0 grams to 198.78 grams. Similar to carbohydrates, fat intake also shows a wide range of values, with a notable standard deviation. This suggests significant variability in dietary fat consumption among the individuals in the dataset. The mean vitamin intake is notably high at 887.81, suggesting that, on average, participants in the dataset consumes a substantial amount of vitamins daily, the standard deviation for vitamin intake is quite large (approximately 4483.31), Vitamin intake spans from 0 to high maximum value of 130,000 (suggesting extreme outliers in vitamin consumption). The mean mineral intake is 862.98, the standard deviation for mineral intake is approximately 1241.36, and mineral intake ranges from 0 to a maximum of 38,791.46, which may indicate extreme outliers. The mean water intake is 55.17%, indicating that, on average, water constitutes about 55.17% of participants' daily nutrient intake. The standard deviation for water intake is relatively low at approximately 30.91, Water intake varies from 0% to a maximum of 100%, which might indicate that some people rely solely on water for their nutrient intake. Zhao et al. (2021) suggest that statistical transformations can reduce extreme outliers in dietary data. However, this study reports very high standard deviations for vitamin (4483.31) and mineral intake (1241.36), indicating extreme outliers. Possible reasons for disagreement may possibly be that, the dataset includes highly variable food intake records, possibly influenced by incomplete tracking or extreme dietary habits. And also, the study did not explicitly apply statistical transformations such as log-scaling or robust standardization, which Zhao et al. (2021) recommend.

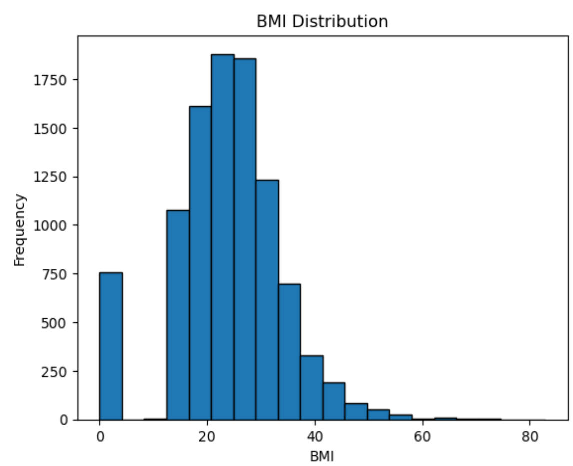
Similarly, the descriptive statistics for demography and health outcomes are shown in Table 2. While, the histogram chart is used to display the distribution of the variables (which includes age, BMI, weight, and some selected health outcomes) are shown in Figure 2.

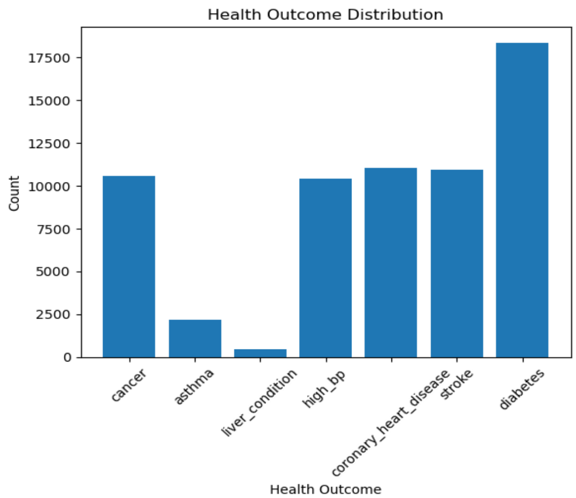
Table 2: Descriptive Statistics for Health Outcomes

Measures	Age	BMI	Cancer	Asthma	Liver Condition	High BP	Coronary Heart Disease	Stroke	Diabetes
Mean	35.657495	23.694742	1.079996	0.046571	1.059615	1.123408	1.116478	1.54561	1.869561
Std. Dev.	20.316483	10.483772	0.969690	0.633311	0.447364	0.902849	1.038082	0.996215	0.510293
Min	18	0	0	0	0	0	0	0	0
Median	26.000000	23.900000	2.000000	0	0	1.0	2.0	2.0	2.0
Max	105.00000(	82.900000	2.000000	9.000000	9.000000	9.000000	9.000000	9.000000	9.000000

The mean age in the dataset is 23.69 years, with a standard deviation of 20.31 years, the minimum age in the dataset is 18 years, and the maximum age is 105 years. The mean BMI in the dataset is 21.99, with a standard deviation of 1.079996. The minimum BMI in the dataset is 0, and the maximum BMI is 35.9. The frequency of cancer in the dataset is 2.19912%, for asthma is 0.219912%, for liver disease is 1.059615%, for high blood pressure is 1.123408%, high blood pressure is a major risk factor for heart disease and stroke, the rate of coronary heart disease is 1.116478%, coronary heart disease is the leading cause of death in the United States. The presence of stroke is 1.869561%, which is slightly higher than the rate of coronary heart disease, but it is still relatively low. Stroke is the fifth leading cause of death in the United States. Diabetes has an occurrence rate of 1.869561% and diabetes is a chronic disease that can lead to serious complications such as heart disease, stroke, blindness, and kidney failure. An (2023) warns that ML-based health predictions can sometimes overestimate disease risks if models are not carefully validated. This study's Diet Health Check System assigns high-risk warnings for relatively moderate nutrient excesses. For instance, a user with 400g of carbohydrates is classified at risk for diabetes, high BP, and coronary heart disease. However, clinical guidelines suggest that carbohydrate intake alone may not be the sole determinant of diabetes risk. Now, possible reasons for disagreement may include, the model might be over-fitting to extreme cases in the dataset. Likewise, lack of clinical validation—the study does not test predictions against real-world patient outcomes. Furthermore, Shang et al. (2023) argue that dietary habits are strongly influenced by cultural and socioeconomic factors, which must be accounted for in ML models and not captured in our study. However, this study relies on NHANES (U.S.-based) and Kaggle datasets, which may not fully represent global dietary patterns. Potential issue include, the recommendation system may not work well for populations with distinct eating habits (e.g., African, Asian, or Mediterranean diets).







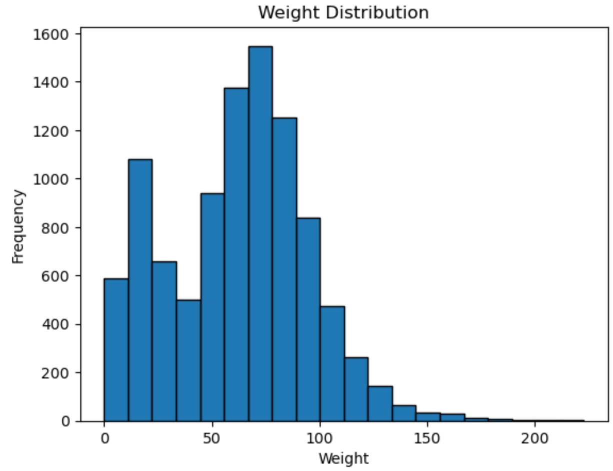


Figure 2: Histograms showing distributions for age, BMI, health outcomes, and weight

Food Recommendation System

A food recommendation system was built on Python that recommends food with similar nutrients to the user's search. The system works by first identifying the key nutrients in the user's search query. Then, the system uses a machine-learning algorithm to recommend a list of foods that are high in those nutrients. The food recommendation system has a number of benefits. First, it can help users find new foods to try that are nutritionally similar to the foods they already enjoy. This can help users to expand their diets and ensure that they are getting all of the nutrients they need. Second, the system can help users to discover new and exciting foods that they may not have otherwise known about. Finally, the system can help users to save time by eliminating the need to manually search for new foods to try. Table 3 shows an example of the output of our food recommendation system. The user's search queries are 'Rice, Beans, Butter, Spinach, Water' and the system recommends five foods that are similar in nutritional content to the user's search query.

Table 3: Result for food Recommendation for User Search Term

	Food Name	**Similarity Score**
Rice	*rice noodles*	0.993764
	carob-flavor beverages mix	0.990725
	jams preserves	0.989143
	marmalade	0.989116
	cocktail mix	.988610
Beans	lima beans	0.990349
	chickpeas	0.932445
	Pigeon peas	0.922726
	cowpeas	0.907938
	cowpeas	0.897683
Butter	butter oil	0.995961
	margarine	0.995880
	butter-margarine blend	0.995410
	margarine-butter blend	0.994934
	oil	0.985539
Juice	cranberry juice	0.999310
	hominy	0.999110
	Pineapple & grapefruit juice drink	0.998983
	Orange & apricot juice drink	0.998912
	pomegranate juice	0.998852

	kale	0.998379
	mustard greens	0.998225
Spinach	dandelion greens	0.997108
	collards	0.996990
	lambs quarters	0.996735
	carb beverage	0.999989
	Quaker oats	0.999651
Water	energy drink	0.999514
	wine	0.999224
	breakfast type drink	0.998795

The similarity score indicates how similar the food is to the user's search term. A higher similarity score indicates that the food is more similar to the user's search term. For instance in Table 3, the food recommendation system recommends rice noodles as the most similar food to rice, followed by carob-flavor beverage mix, jams and preserves, marmalade, and cocktail mix. These results are reasonable, as all of the recommended foods are made from grains or fruits, which are similar to rice in terms of nutritional content. The National Academies (2024) emphasize that nutrition guidelines should be dynamic and personalized rather than generic. This study aligns with this perspective by developing a personalized food recommendation system that suggests alternatives based on users' dietary intake. Example from this study, the system suggests nutrient-rich alternatives like kale for spinach or rice noodles for rice with high similarity scores (>0.99).

Diet-Health Check System

In this section, we look into the Diet Health Check System. The goal is to assess users' consumption of various nutrients and provide insights into potential health outcomes associated with imbalances or excessive intake. To evaluate nutrient intake, we established thresholds for various nutrients, such as carbohydrates, proteins, fats, vitamins, and minerals. These thresholds serve as benchmarks to categorize users' nutrient consumption as either within a healthy range or potentially excessive. If a user's nutrient intake exceeds the defined threshold, the system displays specific health outcomes associated with excessive nutrient consumption. This includes conditions such as diabetes, high blood pressure, and coronary heart disease. If a user's nutrient intake does not exceed the system displays that the user is in a healthy range. Table 4 shows the results we obtained while testing this system.

Table 4: Diet Health Check Result

Nutrients	Intake in grams	Health Outcomes	Recommendations
Carbohydrate	400	diabetes, high_bp, coronary_heart_disease	Please try and reduce your carbohydrate intake to 100g-350g daily to avoid the above health outcomes in the nearest future
Protein	300	cancer, liver_condition, coronary_heart_disease, stroke	Please try and reduce your protein intake to 40g-60g daily to avoid the above health outcomes in the nearest future
Fats	300	cancer, diabetes, coronary_heart_disease	Please try and reduce your fat intake to 10g-17g daily to avoid the above health outcomes in the nearest future
Cholesterol	400	coronary_heart_disease	Please try and reduce your cholesterol intake to 200-239g daily to avoid the above health outcomes in the nearest future
Mineral	1500	cancer, diabetes, coronary_heart_disease, high_bp, stroke	Please try and reduce your Mineral intake to 1000g daily to avoid the above health outcomes in the nearest future

Fleischhacker et al. (2020) criticize traditional food diaries and dietary surveys for being prone to biases and incomplete data capture. This study supports their claim by demonstrating how ML-based tools can minimize human error and provide more reliable dietary assessments. For example, the Diet Health Check System objectively analyzes macronutrient intake and provides personalized feedback, unlike traditional self-reported dietary assessments. To encourage healthier dietary habits, the system recommends adjusting nutrient intake to maintain a balanced diet. Users with carbohydrate intake exceeding the recommended threshold of 350 grams displayed health outcomes associated with excessive carbohydrate consumption such as diabetes, high blood pressure, and coronary heart disease The system's recommendation emphasized the importance of reducing carbohydrate intake to maintain a healthier dietary balance. Excessive protein intake, surpassing the suggested range of 40g60g daily, led to health outcomes including cancer, liver conditions, and coronary heart disease. Recommendations highlighted the importance of moderating protein consumption to reduce potential health risks. Fat intake beyond the recommended 10g-17g daily range faced health outcomes such as cancer, diabetes, and coronary heart disease. The system pointed out the significance of adjusting fat intake for overall well-being. Cholesterol intake exceeding the advised 200-239g daily range was linked to coronary heart disease. Recommendations emphasized the need to reduce cholesterol intake to maintain cardiovascular health. Users with mineral intake surpassing 1000g daily are at risk of facing health outcomes like cancer, diabetes, high blood pressure, and stroke. The system encouraged a reduction in mineral intake to promote a balanced and healthier dietary pattern. The Diet Health Check System has proven to be a valuable tool in evaluating users' dietary patterns and offering actionable points to promote healthier living which is also re-emphasize by both Schulz et al. (2021) and Wang et al. (2020) that also emphasize that dietary patterns, rather than individual nutrients, have a greater impact on long-term health. This study aligns with their findings, as machine learning models were used to analyze dietary patterns rather than focusing solely on isolated nutrients. The results obtained serve as a foundation for ongoing improvements and a commitment to supporting users on their journey towards good well-being.

Correlation Analysis for Food and Nutrient, Dietary Patterns & Health Outcomes and Visualization

In this section, we plotted a correlation heat map for demographic data (age and BMI) and (food & nutrients and health outcomes) as shown in Figure 3, 4 and 5, respectively. Correlation analysis for dietary patterns and food nutrients was taken as shown in Table 5. In addition, we created scatter plots between nutrients intake and each health outcomes as shown in Figure 6 below. Lastly, pair-plot for food and nutrient were plotted as shown in Figure 7. The correlation analysis in this study as presented in Table 5 below supports Schulz et al. (2021)'s assertion that food components interact synergistically, making pattern-based approaches more informative.

Table 5: Correlation Analysis for Food Nutrient and Health Outcomes

	protein	carb	cholesterol	fat	mineral
protein	1.000000	0.552661	0.706274	0.716635	0.873234
carb	0.552661	1.000000	0.330044	0.681633	0.732269
cholesterol	0.706274	0.330044	1.000000	0.593432	0.586158
fat	0.716635	0.681633	0.593432	1.000000	0.790497
mineral	0.873234	0.732269	0.586158	0.790497	1.000000
vitamin	0.362927	0.338416	0.297095	0.333151	0.605029
water	0.171592	0.043855	0.115299	0.112427	0.203712
fiber	0.514365	0.59545	0.218713	0.509666	0.67631
cancer	0.013381	0.015417	-0.004665	0.007086	-0.00092
asthma	0.041404	0.008367	0.021618	-0.015781	0.027829
liver_condition	0.004208	-0.070072	-0.024485	-0.010785	0.023518
high_bp	-0.011091	0.001313	-0.011575	0.005189	-0.00293
coronary_heart_disease	-0.005065	0.006228	-0.01127	-0.002519	0.012612
stroke	-0.001452	-0.008958	-0.005448	-0.007293	-0.0225
diabetes	-0.003539	0.004485	-0.010084	0.0013	0.008266

	vitamin	**water**	**fiber**	**cancer**	**asthma**
protein	0.362927	0.171592	0.514365	0.013381	0.041404
carb	0.338416	0.043855	0.59545	0.015417	0.008367
cholesterol	0.29795	0.115299	0.218713	-0.004665	0.021618
fat	0.333151	0.112427	0.509666	0.007086	-0.01578
mineral	0.60529	0.203712	0.67631	-0.000921	0.027829
vitamin	1.000000	0.114138	0.419634	-0.016751	0.029866
water	0.114138	1.000000	0.194792	0.004439	0.015398
fiber	0.419634	0.194792	1.000000	0.007194	0.106338
cancer	-0.016751	0.004439	0.007194	1.000000	0.057039
asthma	0.029866	0.015398	0.106338	0.057039	1.000000
liver_condition	0.05553	-0.010167	-0.125743	0.090174	-0.03111
high_bp	-0.020595	0.024142	-0.010652	0.133212	0.091652
coronary_heart_disease	0.010579	0.008191	0.010754	0.032827	-0.00562
stroke	-0.016514	-0.010025	-0.007757	0.02423	0.068614
diabetes	0.000547	0.002683	-0.005481	0.050042	0.046243

	liver_condition	**high_bp**	**Coronary_heart_disease**
protein	0.004208	-0.011091	-0.005065
carb	-0.070072	0.001313	0.006228
cholesterol	-0.024485	-0.011575	-0.011277
fat	-0.010785	0.005189	-0.002519
mineral	0.023518	-0.002927	0.012612
vitamin	0.055053	-0.020595	0.010579
water	-0.010167	0.024142	0.008191
fiber	-0.125743	-0.010652	0.010754
cancer	0.090174	0.133212	0.032827
asthma	-0.031113	0.091652	0.005619
liver_condition	1.000000	0.01649	0.004642
high_bp	0.01649	1.000000	0.028671
coronary_heart_disease	0.004642	0.028671	1.000000
stroke	-0.002482	0.098409	0.062991
diabetes	0.091016	0.147528	0.001933

	stroke	**diabetes**
protein	-0.001452	-0.003539
carb	-0.008958	0.004485
cholesterol	-0.005448	-0.010084
fat	-0.007293	0.00130
mineral	-0.022501	0.008266
vitamin	-0.016514	0.000547
water	-0.0125	0.002683
fiber	-0.007757	-0.005481
cancer	0.024230	0.050042
asthma	0.068614	0.046243
liver_condition	-0.002482	0.091016
high_bp	0.098409	0.147528
coronary_heart_disease	0.062991	0.001933
stroke	1.000000	0.061954
diabetes	0.061954	1.000000

Schulz et al. (2021) suggest that certain individual nutrients (e.g., cholesterol, fiber) have strong direct correlations with specific diseases. However, this study finds relatively weak direct correlations between individual nutrients and diseases. For example, cholesterol intake showed only a weak correlation (0.33) with heart disease, contrary to expectations based on previous studies. The Possible reasons for disagreement maybe due to the fact that, this study analyzes a mixed dataset (NHANES + Kaggle), which may introduce greater variability than controlled clinical studies. Also, the interaction of multiple nutrients makes it difficult to isolate single-nutrient effects.

Correlation heat map for demographic data (age and BMI) and (food nutrients and health outcomes) as shown in Figure 3, 4 and 5, respectively.

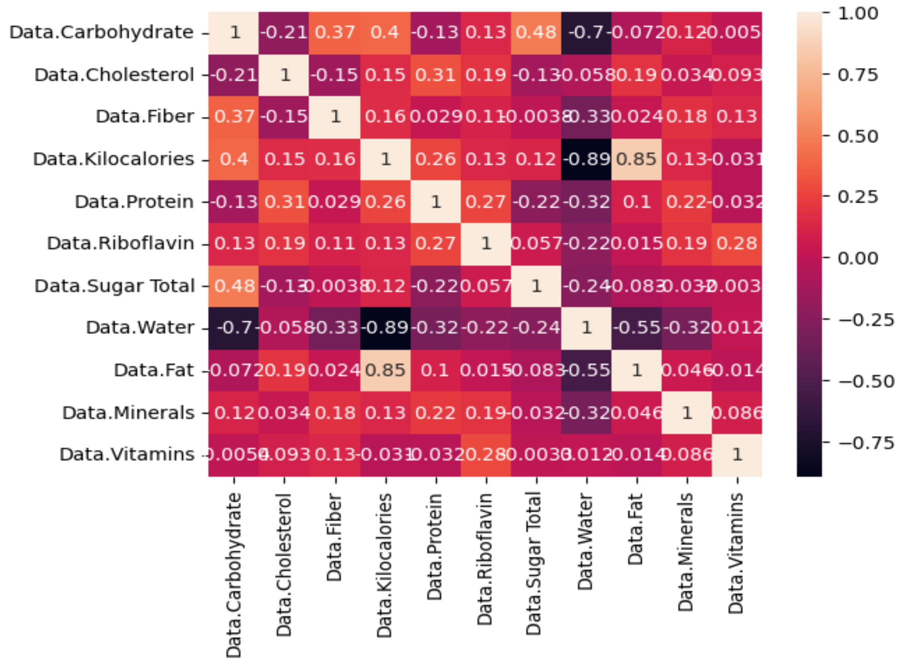


Figure 3: Correlation heat map for food nutrients

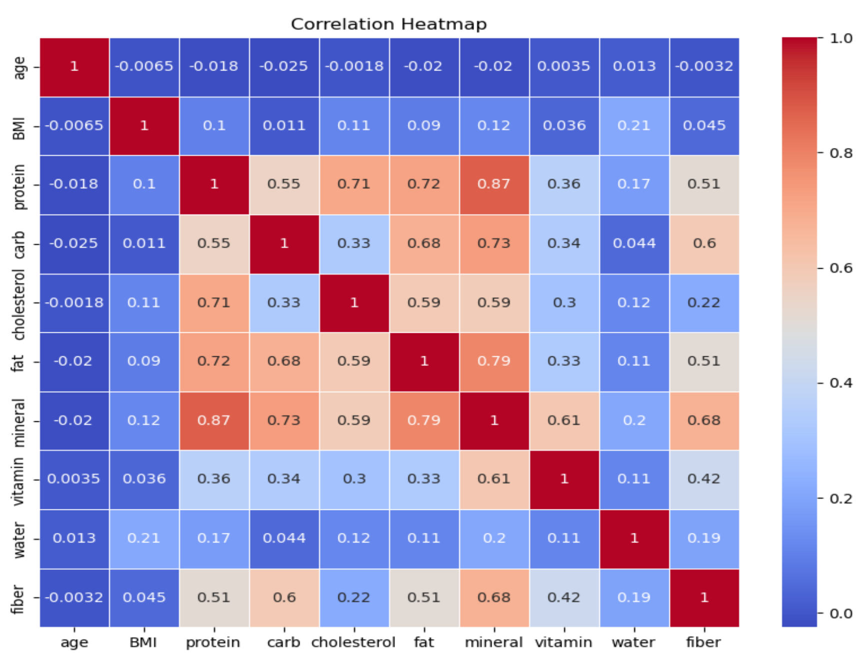


Figure 4: Correlation heat map for demographic data and food nutrients

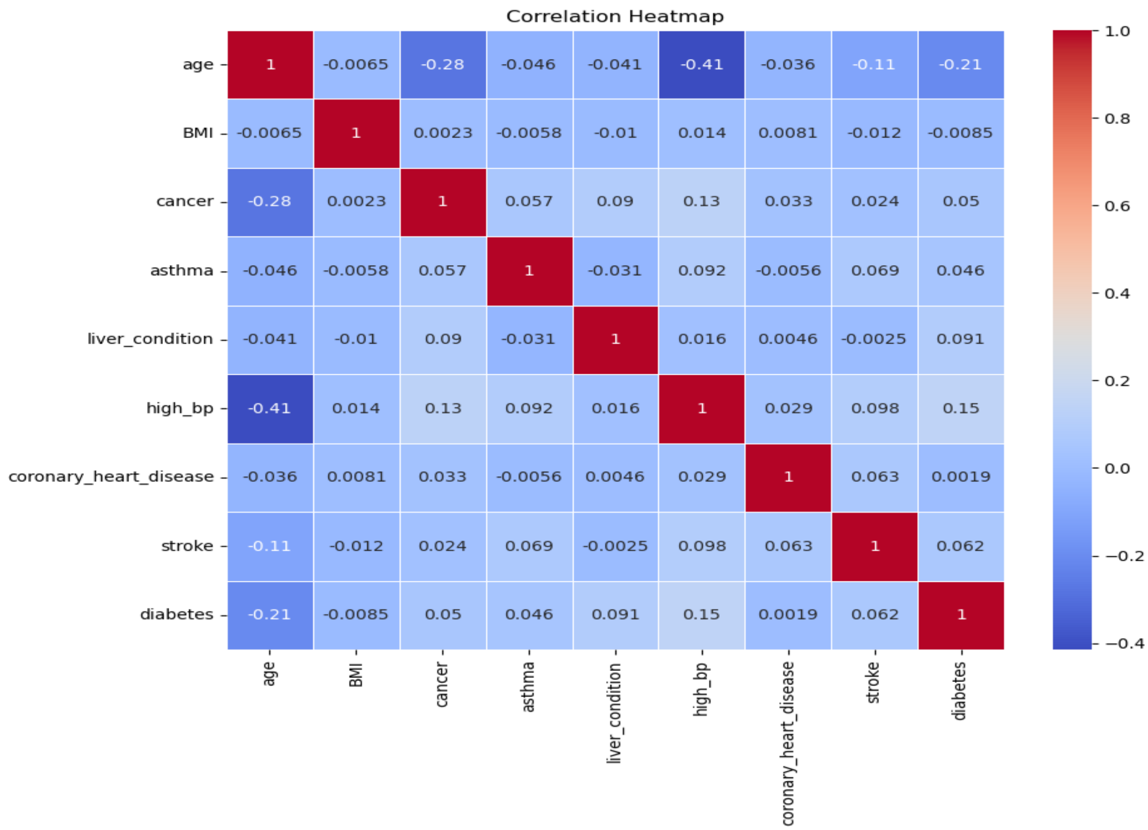
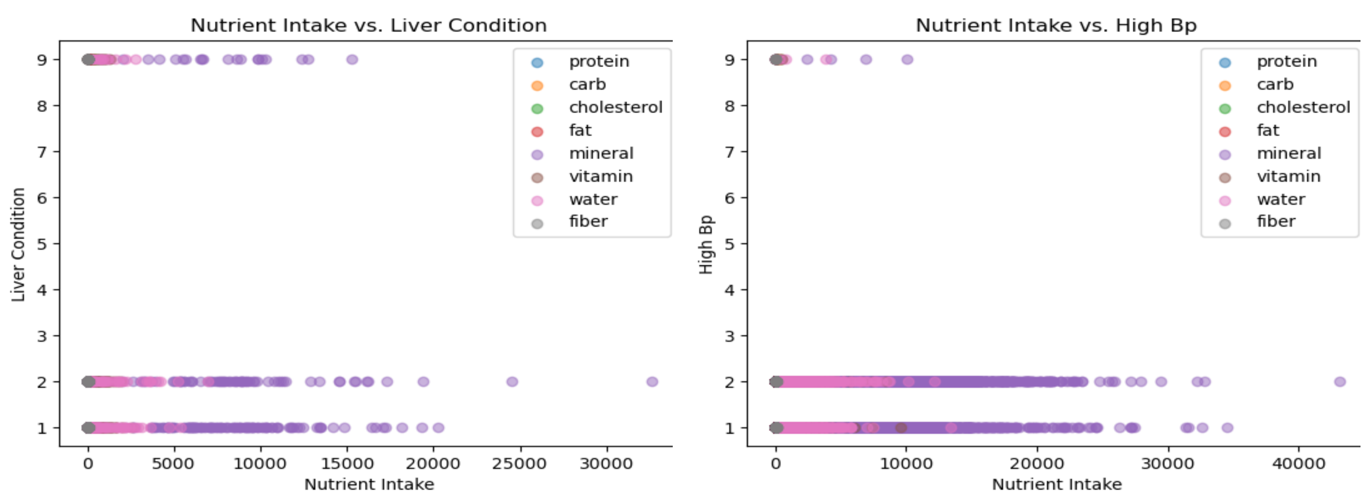


Figure 5: Correlation heat map for demographic data and health outcomes

Scattered plots illustrating nutrient and health relationships is shown in figure 6 below



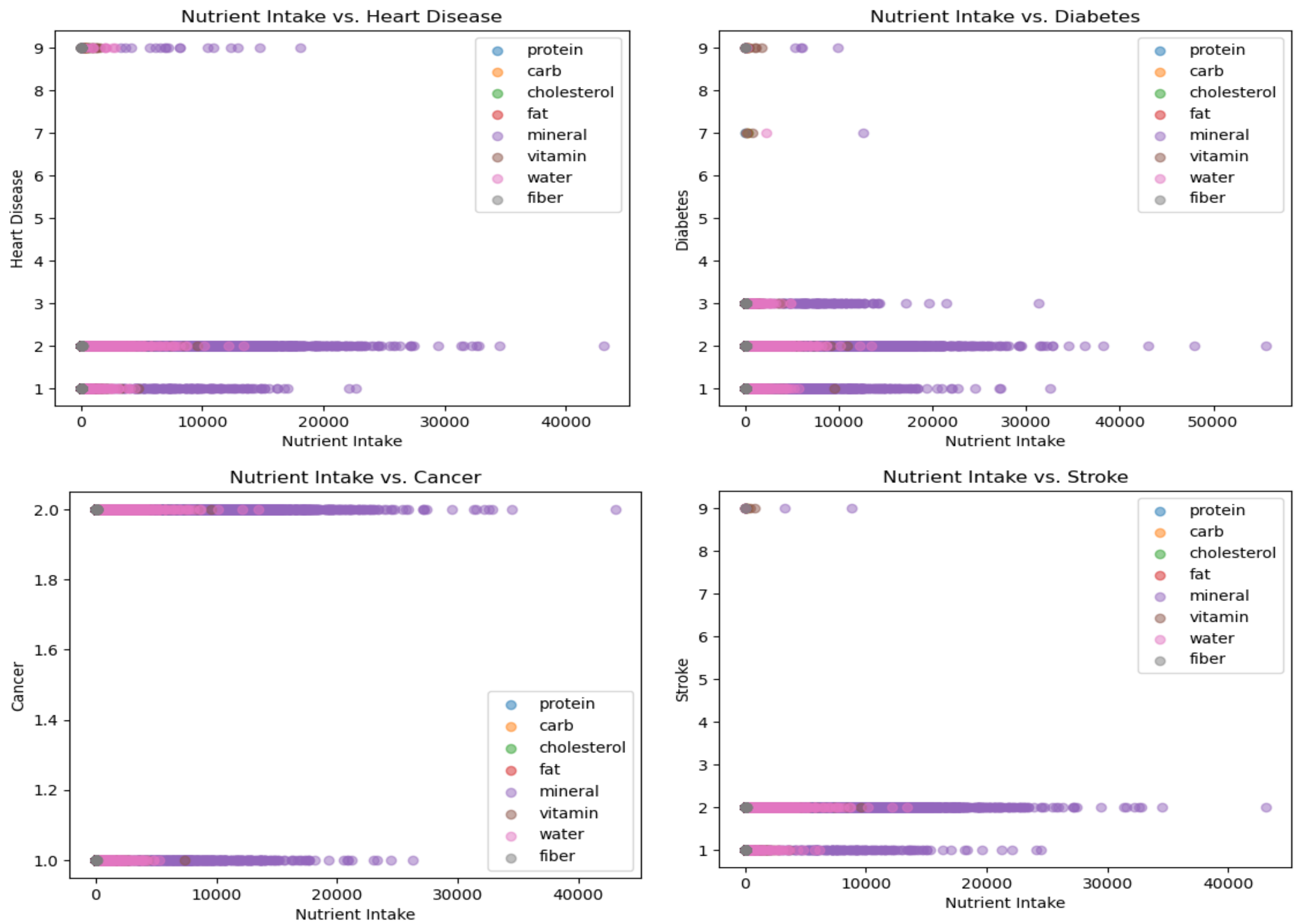


Figure 6: Scattered plots for food nutrients and health outcomes

Shang et al. (2023) in their study analyzed how unhealthy dietary patterns contribute to chronic diseases such as diabetes, heart disease, and cancer. Thus, the Diet Health Check System in this study provides similar evidence, showing that: Excessive carbohydrate intake (>400g) is linked to diabetes, high blood pressure, and heart disease and also, high fat and cholesterol intake increase the risk of coronary heart disease and cancer. Hence, the consistency in results between this study and Shang et al. (2023) strengthens the validity of ML-based nutritional analysis. Pair plot for food and nutrient is shown in figure 7.

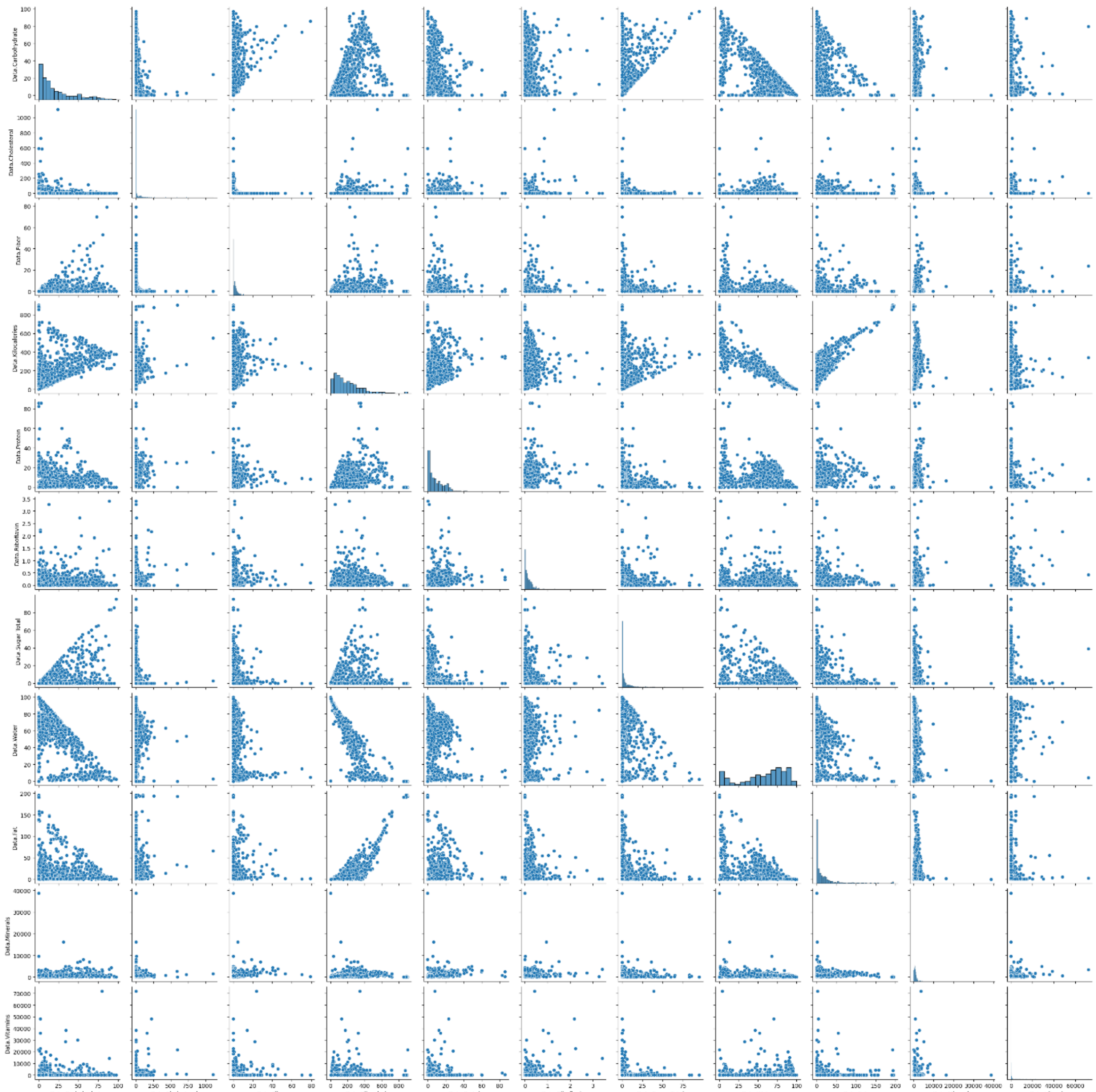


Figure 7: Pair-plot for food and nutrients

CONCLUSION

This study has successfully demonstrated the application of machine learning techniques in analyzing dietary patterns and their relationships with health outcomes. The findings reveal significant variability in nutrient intake among participants, emphasizing the importance of personalized dietary recommendations tailored to individual needs rather than relying on generalized guidelines. The development of a food recommendation system showcased how machine learning can effectively identify nutrient-rich alternatives based on user preferences, achieving similarity scores above 0.99 in multiple test cases. This indicates a robust capacity for providing practical dietary advice that aligns with individual tastes while promoting healthier eating habits. Furthermore, the diet health check system's ability to assess nutrient intake against established thresholds offers a valuable tool for evaluating dietary habits and suggesting necessary adjustments to mitigate risks associated with chronic diseases such as obesity, diabetes, and cardiovascular conditions. Despite these promising results, there are challenges ahead in ensuring that machine learning models are adequately refined and validated across diverse populations. Future research should focus on expanding datasets to include a broader range of demographic variables and cultural contexts that influence dietary habits. Additionally, ongoing evaluation of these systems' effectiveness in real-world settings will be crucial for understanding their impact on public health. In summary, this study highlights the transformative potential of machine learning in personalizing dietary recommendations and enhancing health outcomes. By integrating advanced algorithms with nutritional science, we can deepen our understanding of complex dietary patterns and their implications for health. This work sets a foundation for further exploration into innovative approaches within public health initiatives aimed at reducing chronic disease risks across various populations.

REFERENCES

AnQ, R. S. (2023). A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. *Sensors*, 23(9). https://doi.org/10.3390/s23094178.4178.

Curion, F., & Theis, F. J. (2024). Machine learning integrative approaches to advance computational immunology. Genome Medicine, 16(1), 80. https://doi.org/10.1186/s13073-024-01350-3.

Fleischhacker, S. E., Woteki, C. E., Coates, P. M., Hubbard, V. S., Flaherty, G. E., Glickman, D. R., Harkin, T. R., Kessler, D., Li, W. W., Loscalzo, J., Parekh, A., Rowe, S., Stover, P. J., Tagtow, A., Yun, A. J., & Mozaffarian, D. (2020). Strengthening national nutrition research: Rationale and options for a new coordinated federal research effort and authority. *American Journal of Clinical Nutrition*, 112(3), 721-769. https://doi.org/10.1093/ajcn/nqaa179.

Gupta, V., Mishra, V. K., Singhal, P., & Kumar, A. (2022). An overview of supervised machine learning algorithm [Paper presentation]. *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)*, Moradabad, India, 87-92. https://doi.org/10.1109/SMART55829.2022.10047618.

Hassan, U. M., Ani, C. L., Ndaware, M. M. & Adesue, G. A. (2021). Modelling the Trend and Determinants of Stunted Children Age 0-59 Months in Nigeria. *Science World Journal*, 16(1): 22-28. https://www.scienceworldjournal.org/article/view/21496.

Kirk, D., Kok, E., Tufano, M., Tekinerdogan, B., Feskens, E. J. M., & Camps, G. (2023). Machine learning in nutrition research. *Advances in Nutrition*, 14(1), nmac103. https://doi.org/10.1093/advances/nmac103.

Ma, P., Li, A., Yu, N., Li, Y., Bahadur, R., Wang, Q., & Ahuja, J. K. (2021). Application of machine learning for estimating label nutrients using USDA Global Branded Food Products Database (BFPD). *Journal of Food Composition and Analysis*, 102, 103857. https://doi.org/10.1016/j.jfca.2021.103857.

National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Food and Nutrition Board, & Food Forum. (2024). *Dietary patterns to prevent and manage diet-related disease across the lifespan: Proceedings of a workshop* (M. Snair, Ed.). National Academies Press. https://doi.org/10.17226/38648303.

Sakiyo, D. C., Yusuf, F. A., Bristone, B., & Okuh-Ikeme, P. C. (2025). Comparative Effect of Single and Mixed Organic Manure on Cucumber Growth Parameters in Girei Local Government Area of Adamawa State. Journal of Science Research and Reviews, 2(1), 45-52. https://doi.org/10.70882/josrar.2025.v2i1.24.

Schulz, C. A., Oluwagbemigun, K., & Nöthlings, U. (2021). Advances in dietary pattern analysis in nutritional epidemiology. *European Journal of Nutrition*, 60(8), 4115–4130. https://doi.org/10.1007/s00394-021-02545-9.

Schwedhelm, C., Lipsky, L. M., Shearrer, G. E., & et al. (2021). Using food network analysis to understand meal patterns in pregnant women with high and low diet quality. *International Journal of Behavioral Nutrition and Physical Activity*, 18, 101. https://doi.org/10.1186/s12966-021-01172-1.

Shang, X., Liu, J., Zhu, Z., Zhang, X., Huang, Y., Liu, S., Wang, W., Zhang, X., Tang, S., Hu, Y., Yu, H., Ge, Z., & He, M. (2023). Healthy dietary patterns and the risk of individual chronic diseases in community-dwelling adults. *Nature Communications*, 14(1), 6704. https://doi.org/10.1038/s41467-023-42523-9.

Varshney, N., Jadhav, N., Gupta, K., Mate, N. R., Rose, A., & Kumar, P. (2023). *Personalized dietary recommendations using machine learning: A comprehensive review [Paper presentation]*. 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India. https://doi.org/10.1109/ICAIIHI57871.2023.10489126.

Wang, W., Liu, Y., Li, Y., Luo, B., Lin, Z., Chen, K., & Liu, Y. (2020). Dietary patterns and cardiometabolic health: Clinical evidence and mechanism. *MedComm*, 4(1), e212. https://doi.org/10.1002/mco2.212.

Zhao, J. L. & Zheng, S. (2021). A review of statistical methods for dietary pattern analysis. *Nutr J*, 20: 37. https://doi.org/10.1186/s12937-021-00692-7.