Machine Learning-Based Nutritional Prediction of Recipes Using Structured Ingredient Data

Pratiksha Dabhade[1], Pratiksha Mulay[2], Dr. Kavita Dhakad[3]
[1,2]*M.Sc. CA, School of Information Technology, Indira University, Pune, India*
[3]*Asst. Prof., School of Information Technology, Indira University, Pune, India*
doi.org/10.64643/ijirt.208533-459

***Abstract*—As more recipe and nutrition data becomes available, machine learning can now automatically analyze the nutritional value of food recipes. Manually judging whether a recipe is healthy or nutrient-dense can be slow and tough especially when you're reviewing lots of recipes. This study presents a machine learning approach to multi-label classification, helping identify the nutritional profile of recipes. We worked with 725 recipes, cleaning the data, filling in missing info, prepping ingredients, picking key features, and building new ones to make them useful. The final dataset had 15 key attributes, and ingredient info was turned into numbers to ready it for training. The study focuses on six key nutrition categories: Healthy, Protein_Rich, Fibre_Rich, Calcium_Rich, Iron_Rich, four machine learning algorithms—Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost were trained and evaluated using a dataset split of 580 samples for training and 145 for testing. XGBoost came out on top with 75.68% accuracy, while SVM scored the best F1—0.5760. The results show machine learning can help automatically sort recipes by various nutritional labels—but how well it works varies by category.**

I. INTRODUCTION

Main Introduction

A healthy lifestyle is mostly dependent on food, and a meal's nutritional content can vary greatly based on its preparation and ingredients. The increasing quantity of recipes accessible via digital platforms makes it challenging and time-consuming to manually analyze and categorize each recipe based on its nutritional qualities. Additionally, a recipe may fall under more than one nutritional category; for instance, it may be high in both fiber and protein. Therefore, dietary analysis and nutrition-aware apps may benefit from an automated method that can recognize several nutritional features of a recipe.

Recipe ingredients and nutritional data can be analyzed using machine learning to find trends that might be difficult to spot with manual examination. In order to categorize dishes based on various nutritional attributes, a dataset comprising 725 recipes was examined. Healthy, Protein_Rich, Fibre_Rich, Calcium_Rich, Iron_Rich, and Vitamin_Rich are the six target categories that the study takes into account. Four machine learning algorithms—Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost—are used in the suggested method for data preprocessing, feature selection, ingredient-based feature engineering, and multi-label classification.

The study's objective is to compare these algorithms' performance using common evaluation metrics and assess how well they can categorize recipes into various nutritional categories.

Importance of Nutritional Classification

Nutritional classification makes it easier to understand nutritional facts and helps arrange recipes based on their nutritional qualities. Six target categories are taken into consideration in this study.

A recipe is considered Healthy if it meets the predetermined nutritional standards used in the study to assess overall health. Recipes that meet the stated protein threshold are identified by Protein_Rich, and recipes that meet the dietary-fibre criterion are identified by Fibre_Rich. Similarly, recipes that fulfill the given iron requirement are represented by Iron_Rich, and recipes that meet the defined calcium threshold are represented by Calcium_Rich. Vitamin_Rich finds recipes based on the study's established vitamin-related rule.

The precise cutoff points and guidelines applied to these classifications are:

Table No.1: Nutritional Classification

Target Label	The project's target label definition
Healthy	Healthy = 1 if Calories ≤ 200 kcal; otherwise, 0.
Protein_Rich	Protein_Rich = 1 if Protein ≥ 8.76 g; otherwise, 0.
Fibre_Rich	Fibre_Rich = 1 if Fibre ≥ 2.27 g; otherwise, 0.
Calcium_Rich	Calcium_Rich = 1 if Calcium ≥ 58.57 mg; otherwise, 0.
Iron_Rich	Iron_Rich = 1 if Iron ≥ 1.23 mg; otherwise, 0.
Vitamin_Rich	Vitamin_Rich = 1 if Vitamin C ≥ 24.8675 mg; otherwise, 0. Missing Vitamin C values are retained as missing (NA).

Motivation for Machine Learning

The more recipes and nutritional characteristics there are, the more challenging it is to manually classify them. A number of details, such as ingredients, calories, carbohydrates, protein, fats, fibre, minerals, and vitamins, are included in each recipe. It can take a lot of time and possibly result in inconsistent classification to evaluate each of these characteristics separately for hundreds or thousands of recipes.

By identifying connections between nutritional categories and recipe variables from available data, machine learning can assist in automating this procedure. In this work, pertinent nutritional characteristics and ingredient data are transformed into features that machine learning models can process.

This enables the models to forecast several categories for a new recipe and learn patterns linked to certain nutritional traits.

Thus, the study assesses the efficacy of four distinct algorithms for multi-label nutritional recipe classification: Decision Tree, Random Forest, SVM, and XGBoost.

Research Objective

Main Goal: This study's main goal is to create and assess a machine learning-based multi-label classification method for determining a recipe's nutritional qualities using ingredient and nutritional data.

Supporting Goals

1. To prepare and preprocess a nutritional dataset and recipe data for machine learning analysis.
2. To choose and design pertinent nutritional, ingredient, and recipe-related characteristics for categorization.
3. To use a multi-label technique to categorize dishes into Healthy, Protein_Rich, Fibre_Rich, Calcium_Rich, Iron_Rich, and Vitamin_Rich categories.
4. To evaluate the Accuracy, Precision, Recall, and F1-Score of Decision Tree, Random Forest, SVM, and XGBoost.

Paper Organization

The rest of the paper is structured as follows: Section II reviews the literature on food classification, nutritional analysis, machine learning, and multi-label classification; Section III discusses the research gap found in the existing literature; Section IV describes the dataset and data collection process; Section V presents the data preprocessing procedures; Section VI discusses feature engineering and selection; Section VII describes the proposed methodology and multi-label classification approach; Section VIII presents the machine learning algorithms used in the study; Section IX describes the experimental setup and evaluation metrics; Section X presents and analyzes the experimental results; Section XI discusses the findings and their implications; Section XII outlines the study's limitations; Section XIII wraps up the research; and Section XIV presents the future scope of the study..

II. LITERATURE REVIEW

1. Food and Recipe Classification

Authors/year: Bolaños, Ferrà and Radeva (2017)

Research focus/methodology: This paper defines the problem of ingredient recognition as a multi-label learning task. It modifies a CNN to predict multiple ingredients from a food image, including ingredients that were not present in the training set. Two datasets were created, and the paper provides its motivation based on the importance of automated food diaries and healthier eating.

Limitation or difference: The primary outcome of this research is ingredient recognition, not a nutritional or healthy/unhealthy label. In other words, it focuses on computer vision to identify what is in the food. The

recognition task is an upstream step to estimating calories, carbs, or an overall healthy/unhealthy classification; this paper tackles a different problem.

Relevance to the proposed research
This paper is related to the proposed project in that it uses ingredients to represent food. Notably, the ingredients are used as labels for the CNN. It will be relevant to the current project in that ingredient lists are likely to be used as secondary or auxiliary data. The current project's primary task, unlike this paper, is not about computer vision. Ingredients in the current project are more of an input, potentially a structured or semi-structured input.

Critical observation
This paper's most critical part relates to the approach of multi-label learning. Specifically, the ingredients were labels, which is logical, but they were learned as multi-component entities. This is informative to the current project in that it suggests that ingredients might need cleaning and encoding with that in mind.

2. A Survey on Food Computing
Authors/year: Min, Jiang, Liu, Rui and Jain (2019)

Research focus and methodology:
This survey classifies works on food computing in the areas of perception, recognition, retrieval, recommendation and monitoring. The article highlights the abundance of food-related data in various forms, including images, recipes, videos, and food diaries, which can be processed, analyzed, and utilized in healthy eating promotion, food service personalization, and other domains.

Limitation or distinction:
While this survey covers a wide range of research topics, it does not specifically address the problem of structured recipe nutrition prediction or healthy/unhealthy recipe classification.

Relevance to the proposed research: This survey provides foundational context for this study, highlighting the diversity of food-related problems addressed by researchers. It helps emphasize the distinction between different tasks, such as recipe-level nutrition prediction and the analysis of food diary data.

Critical observation: For this project, this survey serves to set the boundaries of the problem: while recipe nutrition is a part of food computing, it is a specific task that is best approached by considering recipe-specific considerations.

3. Food prediction based on recipe using Machine Learning Algorithms
Authors/date: Kunta Neha, Pallerla Sanjan, Hariharan, Seelam Namitha et al. (2023)

Focus of research/methodology:
The conference paper explores recipe-driven food prediction modeling using machine learning algorithms. In particular, the authors develop Parallel-CNN, Naive Bayes, fuzzy rule-based system, and Artificial Neural Network. Recipe information was used, as well as users' dietary preferences.

Limitation/difference:
The work is more extensive, covering both recommendation and prediction tasks, while I focus on nutrition-related binary classification. Also, the public version of the article only provides bibliographic and abstract data; thus, I cannot comment on specific experiments.

Relation to my project:
The article suggests using recipe text as input for machine learning algorithms. Also, the authors compare multiple algorithms, while I intend to use conventional tabular methods.
The most interesting part: The paper highlights the importance of recipe-driven food prediction while proposing various methods, including modern deep learning approaches.

4. Development of a Machine Learning Model for Classifying Cooking Recipes According to Dietary Styles
Authors/year: Yamaguchi, Araki, Hamada, Nojiri and Nishi (2024)

Research focus and methodology:
The article uses 8,183 instances characterized by 604 features and constructs six predictive models for classifying recipes into Japanese, Chinese and Western styles. The data is split into training, validation and testing sets, associations are made

between features and classes, and performance metrics are reported for all models, most of which are over 0.8.

Limitation or distinction:
The focus of the research is on dietary style, not nutritional healthiness, and the data is taken from a particular source and culture.

Relevance to the proposed research:
This article directly demonstrates that recipe features can be used for supervised learning and that feature-class associations can be made at a large scale across different classifiers.
Critical observation: The paper provides an important methodological justification for the proposed comparative classification experiment.

5. Recipe1M+: A dataset for learning cross-modal embeddings for cooking recipes and food images

Authors/year: Marin, Biswas, Ofli, Hynes, Salvador, Aytar, Weber and Torralba (2019)

Research focus/methodology:
This paper presents Recipe1M+ – a large-scale recipe dataset with 1M recipes and 13M food images. The authors train cross-modal neural networks that can perform recipe-image retrieval.

Limitation or distinction:
The authors focus on cross-modal representation learning rather than on nutritional or healthiness estimation.

Relevance to the proposed research:
Recipe1M+ demonstrates the potential of computational analysis of recipes.
Critical observation: The current project builds upon such analysis by attempting to operationalize recipe texts in terms of nutritional healthiness.

Nutritional Analysis
These studies focus on nutritional attributes, ingredient-to-nutrient relationships, calorie estimation, micronutrient prediction, and nutritional information associated with recipes or food products.

1. Multi-Task Learning for Calorie Prediction on a Novel Large-Scale Recipe Dataset Enriched with Nutritional Information

Authors/date: Ruede, Heusser, Frank, Roitberg, Haurilet and Stiefelhagen (2020/2021) – note: 2020 – publication, 2021 – version with updates.
Focus of research/methodology: The authors introduce the pic2kcal benchmark, which includes 308,000 images of over 70,000 recipes, providing photographs of food, ingredients, and instructions. They predict calories while simultaneously performing tasks such as protein, carbohydrates, and fat estimation as well as ingredient detection. Multi-task learning is employed to facilitate calorie estimation.

Limitation/difference:
The main goal of the work is to find associations between nutrition data and images; thus, the paper focuses on visual calorie estimation. The task is regression, not classification.

Relation to my project: The work demonstrates the importance of a large-scale recipe dataset while also implementing multi-task learning.
The most interesting part: This article provides an excellent comparison of different deep learning methods for calorie estimation. In contrast, I work with traditional machine learning algorithms.

2. A Multimodal Deep Learning Framework for Nutritional Estimation and Health-Oriented Recipe Analysis

Authors/date: Morales-Garzón, Quiñones-Muñoz, Gutiérrez-Batista and Martin-Bautista (2026) – note: 2026 is a preliminary date; possibly, the paper will appear earlier.

Focus of research/methodology:
The framework combines vision and language to estimate the caloric content of recipes using CNN image embeddings while also utilizing text sentence-transformer representations.
The authors evaluate the healthiness of diets using nutritional information, expert evaluation, and user feedback.

Limitation/difference:
The framework employs a hybrid modality and is computationally heavier than the traditional approach of working with tabular data. In addition, this is a recent work, and I have only accessed the abstract, not the full paper.

Relation to my project:
This article underlines the importance of analyzing the healthiness of recipes, which is the central point of my project as well. Moreover, the authors also take into account multiple sources of data, including user ratings.

The most interesting part:
The paper shows that multimodal methods have better performance in healthiness estimation. However, I intend to implement a more classic machine learning approach.

3. Composing Recipes Based on Nutrients in Food in a Machine Learning Context

Authors/year: 2020

Research focus and methodology:
The paper uses a set of over 220,000 recipes and their nutritional content for research. Kernel CCA analysis is conducted to examine associations between nutrients and recipes, and autoencoders, non-negative matrix factorization and regularized least squares are applied.

Limitation or distinction:
The goal of the paper is to compose recipes, not classify them as healthy or unhealthy, and the methods are more complex than necessary for the current task.

Relevance to the proposed research:
The paper demonstrates that nutritional characteristics can be used for computational analysis at a recipe level.

Critical observation:
The data-driven approach described in the paper supports the proposed methodology.

4. Machine Learning Models to Predict Micronutrient Profile in Food After Processing

Authors/year: Naravane and Tagkopoulos (2023)
Research focus/methodology: The work predicts micronutrient contents of cooked food from raw food compositions. The authors trained their models using 820 single-ingredient foods and focused on 14 micronutrients, including seven vitamins and seven minerals. The R2 values vary significantly across different nutrients-processing methods combinations, as the results demonstrate.

Limitation or difference: The data used come from single-ingredient foods and involve transformations during cooking rather than recipes. The aim was nutrient regression, rather than healthiness classification.

Relation to my proposed research: This work demonstrates that ML can learn the relationship between food components and nutrients, which contributes to my research question. Additionally, it touches upon issues of data scaling, which is also relevant to my study, and serves as an important comparison.

Criticality: I find this paper critical to my research because it relates to the micronutrient level analysis, an essential part of my proposed study. Additionally, it discusses the challenges of working with real-life data, which will be relevant to my study.

5. Application of Machine Learning for Estimating Label Nutrients Using USDA Global Branded Food Products Database (BFPD)

Authors/year: Ma et al. (2021)

Research focus/methodology: The article evaluates five ML algorithms' ability to predict nutrient contents of branded food products based on their ingredient statements using USDA Global BFPD database. The authors develop machine-readable representations of both ingredients and target nutrients. They report that best performance was demonstrated by MLP for carbohydrate prediction.

Limitation/difference: The work focuses on branded products, not recipes, using databases, not real-life data, and nutrient regression, rather than healthiness classification.

Relation to my proposed research: This paper relates to my research question, since it also focuses on finding connections between ingredients and nutrients using ML.
Additionally, it emphasizes the importance of representation of ingredients, which is vital in feature engineering.
Criticality: I believe this article is critical to my research due to its focus on prediction of nutrients from food items based on their description. It directly

addresses the feasibility of my approach and provides support for it.

6. Deep Learning Accurately Predicts Food Categories and Nutrients Based on Ingredient Statements

Authors/year: Ma et al. (2022)

Research focus/methodology: The 134k BFPD dataset is used for food classification and nutrient estimation with an MLP-TF-SE model. The reported classification accuracy is up to 99%, while the estimation accuracy (R^2) reaches 0.93–0.97 for most of the nutrients and 0.98 for calcium.

Limitation or distinction: This is dietary-level prediction of nutrients, rather than recipe-level healthiness estimation.

Relevance to the proposed research: This study demonstrates that the ingredient texts can be used for predicting nutritional properties with high accuracy.

Critical observation: The current proposal builds upon this study by aiming to estimate recipe healthiness based on the nutritional facts.

7. Nutrition5k: Towards automatic nutritional understanding of generic food

Authors/year: Thames, Karpur, Norris, Xia, Panait, Weyand and Sim (2021)

Research focus/methodology: The paper introduces Nutrition5k – a database of 5,000 generic dishes with video, depth, weight per component, and high-quality nutritional annotations. The authors train a computer-vision model to estimate calories and macronutrients.

Limitation or distinction: This study focuses on video analysis of natural food items, rather than recipe texts.

Relevance to the proposed research: This paper provides domain-specific background about accurate nutritional estimation and highlights the importance of high-quality annotations.

Critical observation: The current project focuses on text-based recipe analysis, aiming to avoid the technical complexity of vision-based approaches.

8. Classifying Food Ingredients Using Machine Learning on Nutritional and Biochemical Data

Authors/year: Rai, Chandan, Marangappanavar, Indira et al. (2025)

Research focus/methodology: Classifies ingredients into healthy and unhealthy using nutritional and biochemical properties with Decision Tree, Random Forest, SVM, Logistic Regression, KNN and XGBoost algorithms. XGBoost reaches around 94% average accuracy, with Random Forest and KNN around 92%.

Limitation or distinguishing feature: Ingredient-level analysis and smaller dataset; does not attempt to model a whole recipe with its nutrient profile.

Relevance: Critical for the current project as a recent study that informs the healthiness classification.

Criticality: Underlines that the current project deals with a recipe-level estimation of nutritional data rather than claiming to be the first healthy/unhealthy food classifier.

Machine Learning for Food/Nutrition

These studies demonstrate the application and comparison of machine learning, ensemble learning, deep learning, and other computational approaches to food and nutritional problems.

1. Development of a Machine Learning Model for Classifying Cooking Recipes According to Dietary Styles

Authors/year: Yamaguchi, Araki, Hamada, Nojiri and Nishi (2024)

The article uses 8,183 instances characterized by 604 features and constructs six predictive models for classifying recipes into Japanese, Chinese and Western styles. The data is split into training, validation and testing sets, associations are made between features and classes, and performance metrics are reported for all models, most of which are over 0.8. The paper provides an important methodological justification for the proposed comparative classification experiment.

2. Diet and Recipe Recommendation System Using Machine Learning and Deep Learning Models

Authors/year: Lakshmanarao, Obulreddy, Priya, Ganesh and Jayaram (2026)

Research focus: The paper designs a recommendation system that applies traditional machine learning,

ensemble methods and deep learning to diet and recipe data. Features include both nutritional contents and user-specific properties, and the authors evaluate Decision Tree, KNN and Random Forest algorithms.
Limitation or distinction: The paper focuses on personalization, which requires user input not available in the current data frame.
Relevance to the proposed research: The paper supports the general idea that machine learning can be applied to nutritional data, and it provides recent evidence in favor of traditional methods' continued relevance.
Critical observation: The current project will focus on recipes as objects of analysis, not user preferences.

3. A Machine Learning-Based Food Recommendation System with Nutrition Estimation

Authors/year: Nandeppanavar, Kudari, Bammigatti and Vakkund (2024)
Research focus/methodology: The article trains a model that can estimate the number of calories a person should consume per day, based on age, weight, gender, height and activity level. These features are used as input, KNN, Decision Tree and Random Forest are compared, and the accuracy is evaluated at 96.4%, 97.1% and 96.8% respectively. In addition, the paper describes a diet recommendation system.

Limitation or distinction: The focus of the paper is on calorie estimation rather than recipe classification, and the input data includes personal features.
Relevance to the proposed research: This paper provides proof of concept that traditional machine learning methods are effective for nutritional data.
Critical observation: The methods described in the paper will be applied in the comparative classification experiment, but the system described in the article is too different to consider direct applicability.

4. Multi-Task Learning for Calorie Prediction on a Novel Large-Scale Recipe Dataset Enriched with Nutritional Information

Authors/date: Ruede, Heusser, Frank, Roitberg, Haurilet and Stiefelhagen (2020/2021)
The authors use multi-task learning for calorie, protein, carbohydrate and fat estimation as well as ingredient detection. The work demonstrates the application of machine learning/deep learning to multiple food-related prediction tasks.

5. Food Ingredients Recognition through Multi-Label Learning

Authors/year: Bolaños, Ferrà and Radeva (2017)
The paper is particularly relevant to the multi-label aspect because multiple ingredients can be predicted simultaneously. It demonstrates the use of machine learning to represent food as a combination of multiple components.

Multi-Label Classification

These studies are particularly relevant to the concept that a single food or recipe can belong to multiple categories simultaneously.

1. Food Ingredients Recognition through Multi-Label Learning

Authors/year: Bolaños, Ferrà and Radeva (2017)
This paper defines ingredient recognition as a multi-label learning task. It modifies a CNN to predict multiple ingredients from a food image, including ingredients that were not present in the training set.
The most critical part relates to the approach of multi-label learning. Specifically, the ingredients were labels, which is logical, but they were learned as multi-component entities. This is informative to the current project in that it suggests that ingredients might need cleaning and encoding with that in mind.
Relation to the current project: Unlike the paper, where ingredients themselves are the labels, the current project uses recipe information to generate multiple nutritional target labels.

2. Multi-Task Learning for Calorie Prediction on a Novel Large-Scale Recipe Dataset Enriched with Nutritional Information

Authors/date: Ruede et al. (2020/2021)
The study simultaneously performs calorie, protein, carbohydrate, fat and ingredient-related prediction tasks. This demonstrates the usefulness of considering multiple food-related properties rather than restricting analysis to a single nutritional outcome.

3. Deep Learning Accurately Predicts Food Categories and Nutrients Based on Ingredient Statements

Authors/year: Ma et al. (2022)
The study performs food classification and nutrient estimation using ingredient statements. Its relevance to the current project comes from the use of ingredient information to estimate nutritional properties.

Relation to the Six Target Labels

The current project extends the multi-label concept to recipe-level nutritional classification. A single recipe can simultaneously be classified as:

- Healthy
- Protein_Rich
- Fibre_Rich
- Calcium_Rich
- Iron_Rich
- Vitamin_Rich

Therefore, these labels are not mutually exclusive. One recipe may receive multiple positive labels depending on its nutritional characteristics.

Comparison of Existing Studies

Across 30 works, the researchers explored six themes related to AI and food: ingredient recognition, recipe representation, nutrient prediction, recipe-level classification, nutrition-aware recommendation, and personalized or multimodal systems. Each paper focuses on several inputs or outputs of an imaginary unified system in the domain. A proper review of the literature should therefore prioritize comparing the findings related to these research areas.

Ingredient recognition studies demonstrate how recipes can be represented by their ingredients. Prediction studies therefore show that ingredient statements or vectors of food items can be converted into numbers representing nutritional parameters of meals. Recipe-level classification works prove that features of ingredients can be used to predict dietary styles. Healthiness-classification papers illustrate how nutritional and biochemical parameters can be used for binary classification of food items. Recommendation research goes as far as selecting or modifying items according to constraints or preferences.

An important observation relates to the distinction between prediction and recommendation. The former methods predict an attribute of an item, while the latter recommend an item or modify it to meet specific requirements. The current project will adhere to the tradition of prediction/classification research, as it will not attempt to create new meal plans or tailor individual medical advices currently.

A significant finding relates to the difference between the types of input data used in the studies. While image-based methods like Nutrition5k or ingredient recognition rely on computer vision, multimodal approaches combine multiple input types. Recipe-level recognition can therefore be based on structured data in the form of tables with nutrients and ingredients. The current project will use the last-mentioned framework since it will be sufficient to implement a comprehensive comparison of ML methods within the capacity of a single student.

The literature review also prioritized methods which balanced the complexity of the model and the characteristics of the input data. Thus, deep neural networks which require large amounts of data for training were employed in works with Nutrition5k or images, while Random Forest, SVM, Logistic Regression, Decision Tree, KNN, or XGBoost were used for structured data tables. A comparison of these methods will therefore be an integral point of the current research study even with the limited sample size.

The most significant distinction across the works in the chosen area of research were the targets of classification which ranged from calorie prediction to ingredient recognition or dietary style detection. Some works attempted to predict macronutrients like proteins, carbohydrates, or fats, while others operationalized dietary style as a target variable.

One 2025 paper classified ingredients into healthy and unhealthy categories at the ingredient level. The current project will therefore prioritize defining a specific target, namely, recipe-level healthy/unhealthy classification, which will be informed by existing nutritional guidelines.

Datasets, Features, Algorithms and Evaluation Practices

The size of the dataset varied significantly from hundreds of rows of data to over a million recipes. However, a large sample size is not necessarily better, as it depends on the richness of the information contained. Recipe1M+ is therefore suitable for multimodal studies, while BFPD contains more relevant features for ingredient-to-nutrient prediction. The current project will adhere to the principles of research design and select the database based on the presence of features which will be relevant for analysis.

The input data used in the studies ranged from raw text of ingredients, their quantities, images, depth, or food composition to nutritional values, users' characteristics, or dietary constraints. The ingredients text, quantities, and nutritional items therefore have

the best chance for success in the current project, as they represent the inputs for the system. Other variables like user characteristics were used in recommendation studies which go beyond the scope of the current research. Ingredient texts can be used as they are or processed with NLP techniques.

The preprocessing steps had a significant effect on the performance of the models. This conclusion follows from the detailed description of steps taken in the micronutrient prediction article. The same applies to recipe systems, where ingredients must be normalized, and multimodal approaches, where images and text must be aligned. The current project will therefore describe each preprocessing step and attempt to standardize instructions for all models.

The choice of algorithm followed from the choice of input data, as can be seen from the array of methods used in the prediction studies. Logistic Regression, Decision Tree, or Random Forest can be applied to any structured data with numerical or categorical variables. SVM, KNN, or XGBoost will be used for the same dataset to compare results, while Logistic Regression will be used for interpretability. Algorithms can be implemented in R for simplicity and to evaluate performance in terms of underfitting and overfitting.

Evaluation metrics should go beyond accuracy to include precision, recall, F1-score, confusion matrix, and, for binary classification, ROC-AUC. Cross-validation will be used for model selection, and a large test set will be held for final evaluation. All duplicates or near-duplicates must be identified and removed in order to prevent leakage.

Consolidated Limitations and Research Challenges

Healthiness is not a universal ground-truth variable. Some literature explores nutrition targets, while others attempt to use expert judgment or established rules to classify meals as healthy/unhealthy. Thus, a nutrition-based Healthy/Unhealthy label must be explicitly defined as an operational research variable.

Recipes require measuring ingredients to estimate their impact on health. NutritionRec shows that ingredient amounts affect a recipe's healthiness. Multimodal studies note that the volume estimation for a given serving is a recurring challenge. If the selected secondary dataset contains nutritional information about the recipe, it must be kept as is unless otherwise stated.

Hotness is a function of culture and is a critical contributor to model generalizability. A model may conflate cultural elements unique to certain cuisines with nutritional information, particularly when cultural elements are present as explicit features. Thus, EDA should look into the distribution of cuisines and the frequency of ingredients if these features are present.

Class imbalance within the selected label is a significant concern. The paper should report class distribution statistics and employ methods to address the imbalance, such as stratified sampling or weighing of classes, if relevant to the research questions or objectives.

Feature leakage is a recurring challenge when the label is determined by nutritional facts. It is critical to report on which features were employed to define the label and then train a model that attempts to recreate these facts from features in an effort to detect and mitigate feature leakage. A sensitivity analysis may be performed if multiple thresholds are available for defining healthy/unhealthy.

Interpretability is a significant concern for the proposed research since nutritional recommendations often take this form. The study should consider employing models that offer built-in interpretability, such as linear regression, or report feature importance statistics, such as coefficients for linear models or permutation importance for non-linear models, based on the selected algorithm.

Generalizability should always be viewed as a challenge, particularly when the sample size is small or large enough to include all potential categories. The paper should discuss the possibility of overfitting, especially in the context of model performance comparisons.

Research Gap Derived from the 30 Papers

The 30 papers demonstrate that ingredient-to-nutrient prediction, recipe representation learning, diet-style categorization, nutrition-aware recommendation, and healthiness categorization have all been extensively explored. Thus, the current research proposal does not support the assertion that no prior work has attempted to classify food items as healthy or unhealthy. A more suitable research gap is the intersection between recipe-level structured data, multi-nutrient

consideration, operational healthiness label definition, and classical ML evaluation comparative analysis.

While most works leverage individual nutritional facts, image representations or operate at the ingredient level, the current proposal focuses on recipe-level tabular data analysis and multi-nutrient considerations.

Many works address dietary recommendations, kiosks, personalization, or ingredient-level healthiness. The current research question seeks to evaluate the possibility of a recipe-level classification with clearly defined nutritional facts while performing a classical ML evaluation comparative analysis across multiple algorithms. Thus, the proposed research aims to (1) establish whether recipe-level nutritional facts can be employed to predict a healthiness status and (2) determine the most accurate and interpretable ML model for this task. The research gap is evident in the selection of methods, specifically the choice to operate with recipe-level data and utilize nutritional facts as predictive features while focusing on healthiness as a target variable. The gap is both in the theoretical (documenting a process of secondary data preparation, nutritional facts analysis, target definition, and classical ML model comparison) and practical sense (developing an ML-driven solution that can be deployed within a recipe recommendation system).

The research gap is well positioned for master's-level research since it offers opportunities to contribute to both the theory and practice of ML. The study can be completed within the constraints of a single researcher, who would be able to prepare the secondary data, describe the nutritional facts analysis process, determine the healthiness label, and compare classical ML approaches without the need to gather primary data.

III. RESEARCH GAP

Gap Identification

The review of 30 papers shows that ingredient-to-nutrient prediction, recipe representation learning, dietary-style classification, nutrition-aware recommendation, and healthiness categorization have already been explored. Therefore, this study does not claim that healthy/unhealthy food classification has never been attempted. The identified research gap lies in the intersection of recipe-level structured data, multiple nutritional attributes, an explicitly defined healthiness label, and comparative evaluation of classical machine learning algorithms. While many existing studies focus on individual nutritional properties, food images, ingredient-level classification, dietary recommendation, or personalized nutrition, comparatively less attention is given to recipe-level tabular analysis considering multiple nutritional attributes together.

The study therefore focuses on determining whether recipe-level nutritional information can be used to classify recipe healthiness and on comparing the performance of multiple classical machine learning algorithms using the same recipe-level data.

How This Research Addresses the Gap

This research addresses the identified gap by developing a structured recipe-level machine learning framework that combines recipe name/ingredient information with multiple nutritional attributes to predict recipe healthiness.

The study uses secondary recipe data and an explicitly defined nutrition-related prediction task. It applies and compares four classical machine learning algorithms—Decision Tree, Random Forest, SVM, and XGBoost—using a consistent evaluation approach. The study also emphasizes systematic nutritional attribute analysis, target-label definition, model performance comparison, and examination of overfitting and feature importance.

Thus, the contribution of the study is not simply the development of an ML model. It provides a documented process for secondary data preparation, nutritional analysis, healthiness-label definition, recipe-level classification, and comparative evaluation of classical ML algorithms. The resulting framework can also serve as a foundation for future development of ML-driven recipe recommendation systems.

IV. DATASET AND DATA COLLECTION

A. Dataset Description

725 recipe records with 26 original attributes make up the dataset used in this study. A recipe's name, ingredients, cuisine, preparation time, and nutritional makeup are all included in each record.

The dataset was cleaned and reduced to the characteristics necessary for machine learning and nutritional categorization for the study. 15 attributes were kept in the final working dataset following

preprocessing. The recipe name, preparation time, cuisine, cleaned ingredients, and nutritional measures (calories, carbs, protein, fats, free sugar, fiber, salt, calcium, iron, vitamin C, and folate) are some of these characteristics. There are various kinds of data in the dataset. While preparation time and nutritional measurements are **numerical data*, recipe titles, cuisine details, and ingredient descriptions are mostly *text/categorical data. Before being utilized as an input feature for the machine-learning models, the ingredient data was cleaned and altered.

The final dataset's columns were:
final_food_name, TotalTimeInMins, Cuisine, Cleaned-Ingredients, Calories (kcal), Carbs (g), Protein (g), Fats (g), Free Sugar (g), Fiber (g), Sodium (mg), Calcium (mg), Iron (mg), Vitamin C (mg), and Folate (µg).
As a result, the study's final dataset included *725 recipes and 15 pertinent attributes*.

B. Data Collection
Recipe-level data, including *recipe names, ingredient lists, cuisine details, and nutritional information*, was used to create the dataset. Every recipe was handled as a separate document, and the nutritional values and related items were recorded as characteristics.
In order to minimize needless differences from the original ingredient descriptions, the ingredient data was cleansed during preprocessing. Machine learning analysis and feature extraction were then performed on the cleaned ingredient column.
The nutritional factors were chosen because they were pertinent to the study's goals. For creating the nutrient-related target labels, in particular, *protein, fiber, calcium, iron, vitamin C, and folate* were deemed crucial. To give a more comprehensive picture of each recipe's nutritional makeup, calories, carbs, fats, free sugar, and sodium were also kept.
During preprocessing, the records were examined for missing or inappropriate values. *95 recipes with missing vitamin C or folate levels were included in the dataset. While superfluous attributes from the original dataset were eliminated, the pertinent attributes were kept for the study.
Thus, *725 viable recipe records* made up the final dataset, which was then utilized for feature processing, target-label generation, feature selection, and machine-learning model building.

C. Dataset Demography
Recipes from many cuisines are included in the dataset. *Indian, North Indian Recipes, Continental, and South Indian Recipes are the biggest categories. The types of recipes represented in the dataset vary somewhat as a result.

Table No.2: Data Collection

Cuisine Category	Number of Recipes	Percentage
Indian	144	19.86%
North Indian Recipes	135	18.62%
Continental	102	14.07%
South Indian Recipes	83	11.45%
Other Cuisine Categories	261	35.99%
Total	725	100%

Prior to using the machine-learning methods, the target labels were also analyzed to determine their distribution. There are 326 healthy recipes and 399 unhealthy recipes in the *Healthy* category. Less positive cases are found in the nutrient-specific categories, where about one-fourth of the recipes are categorized as rich in the corresponding nutrient.

Table No.3: Rich in Corresponding Nutrient

Target Label	Positive (1)	Negative (0)	Missing	Positive %*
Healthy	536	189	0	73.93%
Protein_Rich	184	541	0	25.38%
Fibre_Rich	190	535	0	26.21%
Calcium_Rich	183	542	0	25.24%
Iron_Rich	186	539	0	25.66%
Vitamin_Rich	158	472	95	21.79%

Because the positive class accounts for about one-fourth of the recipes, the distribution indicates that the *Healthy* label is comparatively more balanced, while the nutrient-rich labels are more imbalanced. When assessing the machine-learning models' performance, this distribution was taken into account.

The dataset is appropriate for the study's goal of categorizing recipes based on their general health status and nutritional qualities since it offers a combination of *recipe, ingredient, cuisine, and nutritional information.

V. DATA PREPROCESSING

A. Data Cleaning

There were 26 attributes and 725 recipes in the original recipe dataset. Before using the machine-learning models, the dataset was analyzed to find duplicate records, unnecessary characteristics, inconsistent values, and formatting problems.

During preprocessing, redundant and superfluous properties were eliminated but the pertinent recipe, ingredient, cuisine, and nutritional data were kept. The values of numerical nutritional parameters, such as kcal, g, mg, and µg, were examined to make sure they were consistently represented.

To create a more uniform depiction of recipe ingredients, the ingredient data was also cleansed. The dataset included 15 pertinent attributes for additional analysis following the cleaning and feature-selection procedures.

Descriptive statistics and visualization were used to examine the numerical attributes in order to find anomalous values and possible outliers. Prior to model training, the analysis was utilized to comprehend the distribution of nutritional variables.

B. Missing Value Handling

The processed dataset was subjected to missing-value analysis prior to model building. The Vitamin C and Folate characteristics were linked to the majority of missing results.

There were 95 recipes with missing vitamin C or folate levels. Before the final modeling dataset was created, these missing variables were taken into account during the preprocessing phase.

The preprocessing code used in the final experiment should be precisely followed when reporting the missing-value treatment. Unless it was actually used, no unsupported imputation method should be mentioned in the paper. The investigation should specifically differentiate between data that were imputed and records that were eliminated.

Therefore, the missing-value process should be explained as follows for the final paper:

During exploratory data analysis, missing values were found, and the main characteristics impacted were vitamin C and folate. These properties have missing values in a total of 95 recipes. In order to prevent incomplete nutritional data from negatively impacting the analysis that followed, the missing-value treatment was used during preprocessing prior to model building.

C. Ingredient/Text Preprocessing

To create a uniform text representation appropriate for machine-learning analysis, the recipe ingredient data was processed. The main ingredient representation was the Cleaned-Ingredients property.

During feature engineering, the constituent text was transformed into a numerical representation that could be read by machines. The constituent data was preprocessed and then converted into a feature matrix with 1,882 numerical characteristics.

After that, the dataset was split into 145 testing records and 580 training records.

D. Target Label Preparation

Nutritional recipe classification is formulated as a multi-label classification issue in the study. There are six target categories taken into account:

Protein-rich, fiber-rich, calcium-rich, iron-rich, vitamin-rich

A binary value is used to represent each target, where:

0 indicates that the recipe does not meet the relevant nutritional requirement.

1 = The recipe meets the relevant nutritional requirement.

The multi-label method permits a recipe to have multiple nutritional labels, in contrast to single-label classification.

For instance, if a recipe meets the requirements, it can be categorized as Protein_Rich, Fibre_Rich, and Iron_Rich at the same time.

Target-label rules

The final paper should use the exact threshold/rule implemented in the project code for each target Table:

VI. FEATURE SELECTION AND FEATURE ENGINEERING

A. Feature Selection

Following preprocessing and data cleaning, the features were chosen according to their significance for nutritional classification at the recipe level. The major input representation for the study is ingredient/text data, and the target nutritional categories are defined and analyzed using the nutritional attributes.

1. Text/Ingredient Features

Cleaned-Ingredients: Because ingredients reveal details about each recipe's composition, they serve as the primary textual predictor.

TF-IDF vectorization was used to transform the ingredient text into numerical features.

2. Recipe Details

During dataset analysis, final_food_name was kept as recipe identification/text information; however, the cleaned ingredient information served as the basis for the model's primary text representation.

Unless specifically added in a subsequent experiment, TotalTimeInMins and Cuisine were kept in the cleaned dataset for analysis but excluded from the final TF-IDF feature matrix used for the reported model training.

3. Nutritional Features

The cleaned dataset contained:

Calories (kcal)
Carbohydrates (g)
Protein (g)
Fats (g)
Free Sugar (g)
Fibre (g)
Sodium (mg)
Calcium (mg)
Iron (mg)
Vitamin C (mg)
Folate (µg)

These nutritional attributes were important for nutritional analysis and target-label construction. The six target labels were:

Healthy
Protein_Rich
Fibre_Rich
Calcium_Rich
Iron_Rich
Vitamin_Rich

B. Text Feature Representation

TF-IDF (Term Frequency–Inverse Document Frequency) vectorization was used to transform the Cleaned-Ingredients column into numerical features.

Because ingredient information is textual and includes numerous unique ingredient phrases, TF-IDF was used. It lessens the impact of terms that appear often in many recipes while giving greater weight to terms that are instructive for specific dishes.

The feature matrix that was produced was:

Table No.4: Feature Matrix

Dataset	Samples	TF-IDF Features
Training Set	580	1,882
Testing Set	145	1,882
Total	725	1,882

C. Numerical Feature Processing

During preprocessing, the nutritional columns were transformed into numerical format for use in target-label generation, statistical analysis, and visualization.

The following numerical characteristics were taken into account:

Calories
Carbohydrates
Protein
Fats
Free Sugar
Fibre
Sodium
Calcium
Iron
Vitamin C
Folate

The TF-IDF feature matrix utilized in the presented experiment was not subjected to an extra normalization or standardization step by the project.

The TF-IDF format was utilized directly in the model pipeline for algorithms like SVM. As a result, throughout the comparative trial, the preprocessing was constant.

D. Final Feature Matrix

There were 725 recipes in the final dataset. It was separated into:

580 recipes, or 80% of the training data

145 recipes, or 20% of the testing data

For both datasets, the TF-IDF transformation yielded 1,882 features in the same feature space.

Goal Labels

Final Feature Matrix

Table No.5: Final Feature Matrix

Feature Group	Representation	Dimension
Cleaned Ingredients	TF-IDF	1,882
Training Samples	TF-IDF Matrix	580 × 1,882
Testing Samples	TF-IDF Matrix	145 × 1,882
Target Labels	Binary Multi-Label	6
Total Recipes	—	725

A single recipe may concurrently fall under more than one nutritional category because the classification problem is multi-label.

One recipe, for instance, might have:

Protein-rich, fiber-rich, calcium-rich, iron-rich, vitamin-rich 1 1 0 1 0

As a result, there are multiple mutually exclusive classes in the model. Rather, it forecasts six separate binary (0/1) nutritional labels for every recipe.

VII. PROPOSED METHODOLOGY

A. Overall Methodology

Data preparation, feature extraction, target-label development, model training, prediction, and evaluation are the steps that make up the suggested methodology. The recipe dataset is the starting point of the entire procedure, which then moves on to data cleansing and missing-value analysis. After cleaning and standardizing the ingredient data, TF-IDF is used to transform it into numerical characteristics.

The six target labels—Healthy, Protein_Rich, Fiber_Rich, Calcium_Rich, Iron_Rich, and Vitamin_Rich—are created using the nutritional characteristics. To prevent data leaking, the machine-learning models do not get the nutritional values needed to generate these labels as input characteristics. The resulting TF-IDF feature matrix is divided into training and testing sets. The training data is used to develop the selected classification algorithms, while the testing data is kept for evaluating their performance. The trained models then predict the nutritional labels for unseen recipes, and the predictions are evaluated using measures such as accuracy, precision, recall, and F1-score.

Workflow:

Dataset → Data Cleaning → Missing-Value Handling → Ingredient/Text Preprocessing → TF-IDF Feature Engineering → Target Label Generation → 80:20 Train/Test Split → Multi-Label Model Training → Prediction → Evaluation

The final feature-processing stage produced 1,882 TF-IDF features, with 580 training samples and 145 testing samples.

B. Train-Test Split

There are 725 recipes in the entire dataset. There were 580 training recipes and 145 testing recipes after it was split into an 80:20 training-to-testing ratio.

While the testing set was kept apart and used to gage how well the taught models performed on untested recipes, the training set was used to fit the machine-learning models.

Neither a random_state value nor a specific multi-label stratification procedure are specified in the project documentation that is currently available. Therefore, unless they are verified by the final notebook or code, these should not be mentioned in the paper. Reporting an assumed value is not as good as this.

Proportion of Dataset Samples

Training Set: 580–80%

Testing Set 145 20%

Total: 725 100%

C. Multi-Label Classification

Instead of treating nutritional categorization as a single-class classification problem, the suggested solution tackles it as a multi-label classification problem.

This is due to the fact that a dish can simultaneously meet many nutritional requirements. For instance, the same meal may have a Healthy label and be categorized as both Protein_Rich and Fibre_Rich. This multi-label character is confirmed by the project documentation there were six binary target variables

produced: Protein-rich, fiber-rich, calcium-rich, iron-rich, vitamin-rich

Every target denotes a distinct nutritional characteristic, with 1 denoting a recipe that falls into the category and 0 denoting a food that does not.

As the primary predictive input, the ingredient data was converted into TF-IDF numerical characteristics. To avoid data leaking, the nutritional values themselves were not included in the direct model input and were simply utilized to create the target labels.

The study assessed Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost as classification models. The same prepared multi-label dataset was used to compare these models in order to assess how well they predicted the nutritional categories. These four model comparison procedures are identified in the project documentation.

VIII. MACHINE LEARNING ALGORITHMS

A. Decision Tree:

A supervised learning approach called a decision tree divides the dataset into smaller groups recursively according to feature values. The method chooses a feature and split that optimally divides the classes at each node. Until appropriate halting circumstances are met, this process keeps going.

Decision trees were employed in this work as an interpretable baseline model. Understanding how features contribute to classification is rather simple due to its tree structure.

Implementation: The 580 × 1,882 TF-IDF training matrix was used to train the model, and the 145-test-sample matrix was used to evaluate it. A typical classification Decision Tree was utilized in the implementation.

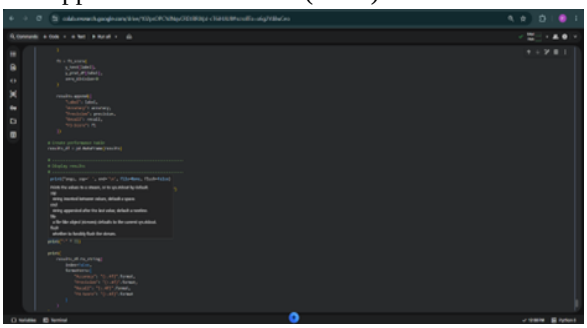
B. Random Forest

An ensemble learning system called Random Forest combines several Decision Trees. The final prediction is generated by combining the predictions of each tree, which is trained using various subsets of the data and characteristics.

In general, using many trees improves generalization and lowers the variation associated with a single Decision Tree.

The TF-IDF feature matrix and 80:20 train-test split utilized for the other algorithms in this study were also employed to train Random Forest. A typical Random Forest classifier was employed to create the model, and its tree-based ensemble structure was utilized to increase resilience when compared to a single Decision Tree.

C. Support Vector Machine (SVM)



The Support Vector Machine (SVM) looks for the best hyperplane to divide classes. Kernel functions can convert data that cannot be separated linearly into a higher-dimensional space where separation might be achievable.

For high-dimensional sparse text representations like TF-IDF, where the number of features might be significantly greater than the number of observations, SVM is especially well suited.

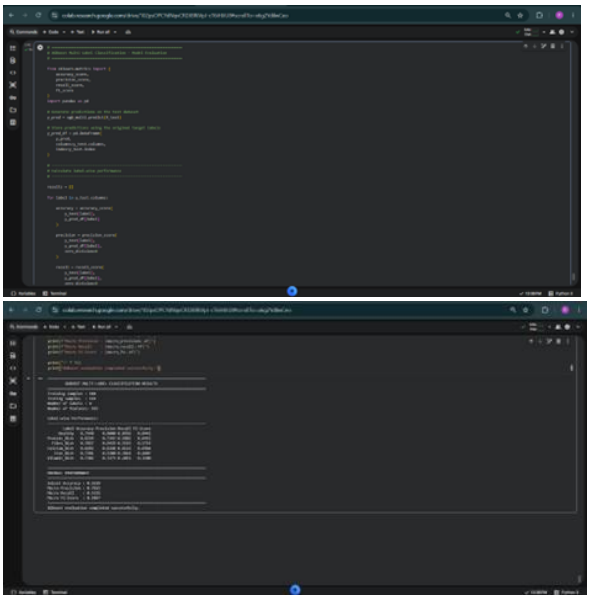
The 1,882 TF-IDF features and the identical 80:20 train-test split were used to train the SVM in this investigation. Since linear SVM is computationally suitable for this kind of data and TF-IDF generates a high-dimensional sparse feature space, a linear kernel was employed.

D. XGBoost

An ensemble technique based on gradient-boosted decision trees is called XGBoost (Extreme Gradient Boosting). XGBoost creates trees sequentially rather than independently, with each subsequent tree aiming to fix mistakes produced by the preceding trees.

This incorporates regularization to help control overfitting while enabling the model to gradually improve its predictions.

The same 580 training samples, 1,882 TF-IDF features, and 145 testing samples used for the other models were employed in this work to train XGBoost. The XGBoost classifier's boosting-based tree-learning methodology was used to create the model.



E. Model Configuration

The TF-IDF representation of the cleaned ingredient data was used to train the machine-learning models. The identical train-test configuration was used to compare the four chosen methods. These variables are neither assumed nor manually provided because the published research documentation does not give the whole hyperparameter configuration or random_state values for every model.

Table No.6: Model Configuration

Algorithm	Key Hyperparameters	Preprocessing Requirements	Random State	Class / Multi-Label Handling
Decision Tree	Not specified in documented results	TF-IDF feature representation	Not specified	Six binary target labels
Random Forest	Not specified in documented results	TF-IDF feature representation	Not specified	Six binary target labels
SVM	Not specified in documented results	TF-IDF feature representation	Not applicable / not specified	Six binary target labels
XGBoost	Model hyperparameters used during training; exact values to be taken from final code	TF-IDF feature representation	Not specified in documented results	Six binary target labels

IX. EXPERIMENTAL SETUP AND EVALUATION METRICS

A. Experimental Environment

Python was used to carry out the suggested machine-learning experiments. Model training, assessment, feature engineering, and data preprocessing were all done with Python.

Table No.7: Working Environment

Component	Specification
Programming Language	Python
Operating System	Windows 11
Development Environment	Jupyter Notebook / Python environment
Data Processing	Pandas, NumPy
Machine Learning	Scikit-learn
Boosting Algorithm	XGBoost
Text Feature Extraction	TF-IDF Vectorizer from Scikit-learn
Visualization	Matplotlib
Dataset Size	725 recipes
Training Samples	580
Testing Samples	145
Feature Dimension	1,882 TF-IDF features
Target Labels	6 binary labels

To guarantee a fair comparison, all four algorithms employed the identical preprocessing pipeline, feature representation, training set, and testing set.

B. Evaluation Metrics

Each recipe can concurrently fall into more than one nutritional category because the suggested system uses

multi-label classification. Accuracy, Precision, Recall, and F1-Score were used to assess the models' performance.
The weighted averaging method, which takes into consideration the amount of samples in each class, was used to produce the metrics for the stated overall model comparison.

1. Correctness
The percentage of accurately classified forecasts among all predictions is known as accuracy.
Accuracy = TP+TN / TP+TN+FP+FN
Where:
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative
Higher accuracy indicates better overall classification performance

2. Accuracy
Precision quantifies the proportion of samples that were correctly expected to be positive.
Precision = TP / TP+FP
A high precision means that fewer false-positive predictions are generated by the model.

3. Remember
Recall quantifies the proportion of real positive samples that the model accurately detected.
Recall = TP / TP+FN
The model's ability to detect positive nutritional categories is demonstrated by a high recall.

4. The F1-Score
The harmonic mean of precision and recall is known as the F1-score.
F1 = 2 × Precision + Recall / Precision × Recall
Because the F1-score takes into account both false positives and false negatives, it is especially helpful when there is a class imbalance.

C. Evaluation Strategy
All four machine-learning algorithms were trained and evaluated using the same training and testing datasets to ensure a fair comparison. The same TF-IDF feature representation and preprocessing pipeline were used for each model.
The models were evaluated using the same metric definitions: Accuracy, Precision, Recall, and F1-Score. The overall model comparison used weighted averaging to account for the class distribution of the target labels.
The dataset consisted of 580 training samples and 145 testing samples, with 1,882 TF-IDF features. No cross-validation was used in the reported experiments. The final performance was therefore measured on the common 20% testing set.
This consistent evaluation setup allowed the performance of Decision Tree, Random Forest, SVM, and XGBoost to be compared under the same experimental conditions.

X. RESULTS AND ANALYSIS

A. Overall Model Performance
The same training and testing data were used to assess the four machine-learning algorithms that were chosen. Accuracy, Precision, Recall, and F1-Score are used to display the overall performance.

Table No.8: Model Performances

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	66.69%	48.18%	56.48%	51.73%
Random Forest	75.58%	83.38%	32.27%	43.62%
SVM	73.29%	57.76%	58.05%	57.60%
XGBoost	75.68%	69.19%	45.36%	53.60%

The main metric is accuracy.
With an accuracy score of 75.68%, XGBoost outperformed Random Forest, which came in at 75.58%.

B. Label-Wise Performance
To determine how the models fared for each nutritional category, label-wise analysis was carried out.

Random Forest:

Table No.9: Random Forest Performance

Target Label	F1-Score
Healthy	66.10%
Protein_Rich	56.67%
Fibre_Rich	30.77%

Calcium_Rich	38.30%
Iron_Rich	27.45%
Vitamin_Rich	42.42%

XGBoost:

Table No.10: XGBoost Performance

Target Label	Accuracy	F1-Score
Healthy	69.66%	65.08%
Protein_Rich	82.76%	65.75%
Fibre_Rich	75.86%	49.28%
Calcium_Rich	81.38%	54.24%
Iron_Rich	75.86%	40.68%
Vitamin_Rich	68.55%	46.58%

Protein_Rich has the highest label-wise F1-score for XGBoost (65.75%), followed by Calcium_Rich (54.24%) and Healthy (65.08%). Iron_Rich's somewhat lower F1-score (40.68%) suggests that the model had a harder time classifying this category.

C. Model Comparison

The outcomes show that each of the four algorithms has unique advantages.

With a total accuracy of 75.68%, XGBoost outperformed Random Forest by just 0.10 percentage points at 75.58%. Additionally, XGBoost's F1-score was 53.60%, whereas Decision Tree's was 51.73% and Random Forest's was 43.62%.

SVM, on the other hand, had the highest recall of 58.05% and the highest overall F1-score of 57.60%. Random Forest had the highest precision (83.38%), but its recall (32.27%) was far lower.

Therefore, XGBoost was selected as the top-performing model based on the chosen primary metric, accuracy, with a score of 75.68%. However, SVM performed better when F1-Score and Recall were considered.

D. Visualization of Results

The experimental results should be presented using the following figures:

Figure 1: Overall Model Performance

A grouped bar chart comparing Accuracy, Precision, Recall, and F1-Score for:

- Decision Tree
- Random Forest
- SVM
- XGBoost
- Caption:

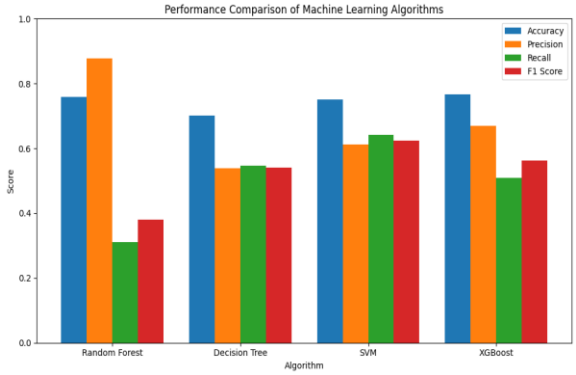


Fig. 1. Comparison of overall performance of the four machine-learning algorithms.

Figure 2: Label-Wise Performance of XGBoost

A grouped bar chart comparing Accuracy and F1-Score for:

- Healthy
- Protein_Rich
- Fibre_Rich
- Calcium_Rich
- Iron_Rich
- Vitamin_Rich
- Caption

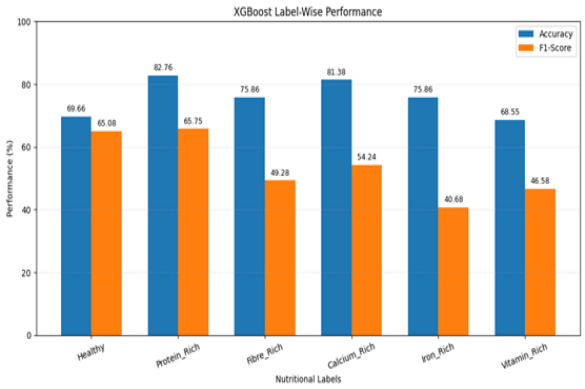


Fig. 2. Label-wise performance of the XGBoost model across six nutritional categories.

Figure 3: Label-Wise F1-Score for Random Forest

A bar chart showing the F1-Score of Random Forest for all six nutritional labels.

Caption:

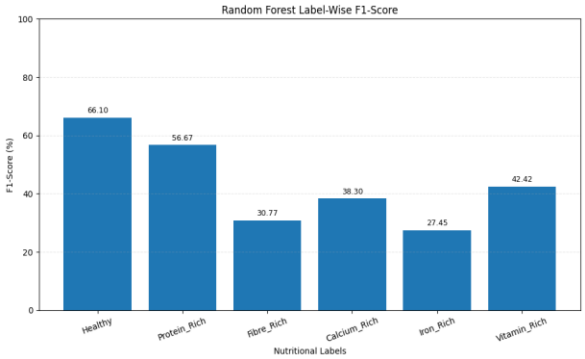


Fig. 3. Label-wise F1-Score of the Random Forest model.

Figure 4: Confusion Matrices

Confusion matrices can be presented for the individual binary nutritional labels, particularly for the XGBoost model.

Each matrix should display:

- True Positive
- True Negative
- False Positive
- False Negative

Caption:

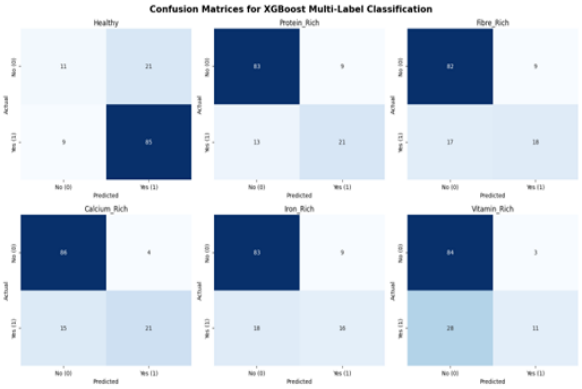


Fig. 4. Confusion matrices for the nutritional labels using the XGBoost model.

XI. DISCUSSION

Interpretation

The experimental findings show that while machine-learning algorithms are capable of identifying a variety of nutritional attributes from recipe ingredient data, their efficacy differs depending on the nutritional category. The investigated models' overall accuracy ranged from 66.69% to 75.68%, indicating that ingredient-based TF-IDF representations had valuable information for classifying nutritional recipes.

With an overall accuracy of 75.68%, XGBoost outperformed the other four algorithms. This suggests that beneficial patterns in the high-dimensional ingredient representation were captured by the boosting technique. The findings do, however, also demonstrate that not all nutritional labels perform as well when accuracy is high.

For instance, XGBoost's F1-score was lower for Iron_Rich (40.68%) and Vitamin_Rich (46.58%) than it was for Protein_Rich (65.75%) and Healthy (65.08%). As a result, certain nutritional groups benefit more from the model than others.

The findings show that nutritional classification at the recipe level is possible, however it is a multi-label problem with different levels of classification complexity across goals.

Algorithm Strengths and Weaknesses

Because its decision-making process may be depicted as a series of feature-based splits, the Decision Tree offers an interpretable baseline. This makes it helpful in comprehending the process of reaching a classification decision.

Nevertheless, out of the four models, the Decision Tree had the lowest overall accuracy (66.69%). This implies that the ensemble techniques were more successful than a single tree in capturing the patterns found in the high-dimensional TF-IDF representation.

The Random Forest

The maximum precision (83.38%) and overall accuracy (75.58%) were attained by Random Forest. Because of its ensemble of several decision trees, it can make more reliable predictions and is less reliant on the actions of a single tree.

Even though its positive predictions were reasonably accurate, its recall was only 32.27%, meaning that it was unable to detect a significant percentage of positive cases. As a result, its aggregate F1-score was only 43.62%.

SVM

SVM obtained the highest recall of 58.05% and the highest overall F1-score of 57.60%. Because SVM works well with high-dimensional feature spaces like TF-IDF representations, this is significant.

Although its accuracy of 73.29% was lower than that of XGBoost and Random Forest, it produced the highest overall F1-score due to its better precision-recall balance.

XGBoost

At 75.68%, XGBoost had the best overall accuracy. Its gradient-boosting method constructs trees in a sequential fashion so that subsequent trees can concentrate on mistakes committed by previous trees. Additionally, the model showed comparatively good label-wise performance, especially for Protein_Rich and Healthy. Iron_Rich and Vitamin_Rich, on the other hand, performed worse, indicating that the boosting strategy did not completely remove the challenges related to each nutritional group.

Important Observations

1. The difficulty of various labels varies.

Significant variations between dietary categories are revealed by the XGBoost results:

Table No.11: XGBoost Observations

Label	XGBoost Accuracy	XGBoost F1-Score
Healthy	69.66%	65.08%
Protein_Rich	82.76%	65.75%
Fibre_Rich	75.86%	49.28%
Calcium_Rich	81.38%	54.24%
Iron_Rich	75.86%	40.68%
Vitamin_Rich	68.55%	46.58%

With an F1-score of 65.75%, Protein_Rich had the best XGBoost performance. Iron_Rich, on the other hand, had the lowest F1-score (40.68%).

This suggests that not all six nutritional categories can benefit equally from the same ingredient representation in terms of predictive information.

2. Class disparity is a crucial factor to take into account.

There was some imbalance in the target distributions. For instance:

Protein_Rich: 541 negative and 184 positives

Fiber_Rich: 535 negative and 190 positives

Calcium_Rich: 542 negative and 183 positives

Iron_Rich: 539 negative and 186 positives

Accuracy by itself might not adequately characterize model performance because the positive class only accounts for about 25% of the recipes for these labels. For this reason, while evaluating the data, precision, recall, and F1-score are crucial.

3. The same model is not always identified by accuracy and F1-score.

The fact that XGBoost had the best accuracy (75.68%) and SVM had the highest F1-score (57.60%) is a significant finding.

This discrepancy shows how the evaluation goal influences model selection. XGBoost was used as the final model in this study since accuracy was chosen as the major metric. But according to the stated evaluation, SVM offers a superior mix between precision and recall, as evidenced by its higher F1-score.

4. It is not appropriate to take the findings as medical forecasts.

Nutritional categories identified by study include Healthy, Protein_Rich, Fiber_Rich, Calcium_Rich, Iron_Rich, and Vitamin_Rich. Therefore, the model does not offer medical or personalized nutritional recommendations; instead, it finds patterns based on the dataset and label-generation criteria.

Overall, the results reveal that machine learning can be used for structured recipe-level nutritional classification, but they also indicate that further work is needed to ensure strong generalization and enhance performance for the weaker nutritional categories.

XII. LIMITATIONS

Study Limitations

When evaluating the findings, it is important to take into account the following constraints of the proposed study:

1. Size of the dataset:

The dataset has 725 recipes, which is somewhat tiny when compared to large-scale food datasets like Recipe1M+ yet appropriate for a comparative machine-learning study. As a result, the models might not fully represent the variety of recipes and cuisines.

2. Unbalanced Class:

There is an imbalance in the target labels. For instance, there are much fewer positive samples than negative samples in the categories Protein_Rich, Fiber_Rich,

Calcium_Rich, Iron_Rich, and Vitamin_Rich. Recall and F1-score may be impacted by this mismatch, especially for minority classes.

3. Nutritional Values Missing:
There are 95 recipes with missing values for vitamin C and folate. The availability of complete nutrient profiles may be diminished by missing nutritional data, which may also have an impact on the related nutritional classification.

4. Definition of a Healthiness Label:
Rather than a widely recognized ground-truth healthiness label, the Healthy/Unhealthy classification is based on established nutritional standards and thresholds. As a result, altering the thresholds may have an impact on the model's performance and class distribution.

5. Representation of Ingredients:
Cleaned ingredient text and numerical feature representation are used to convey ingredient information. Ingredient amounts, preparation techniques, and intricate connections between elements might not be adequately represented.

6. Design of Train-Test:
There were 145 testing records and 580 training records in the sample. Due to the short size of the dataset, different train-test splits may yield different results. As a result, the results should be interpreted within the parameters of the chosen experimental configuration.

7. Minority Label Model Performance:
Certain dietary categories, especially Fibre_Rich and Iron_Rich, have lower F1-scores than the Healthy or Protein_Rich categories, according to the label-wise data. This suggests that certain minority dietary groups are more difficult for the models to recognize.

8. Generalization
Although the dataset includes recipes from a variety of cuisines, it is not guarantyd that the trained models will function as effectively on recipes from food categories or cuisines that are underrepresented in the dataset.

Methodological and Computational Scope:
Using structured recipe and ingredient data, the study focuses on Decision Tree, Random Forest, SVM, and XGBoost. The current experiment did not cover more computationally demanding methods such massive transformer-based or multimodal image-text models.

Absence of a customized health evaluation
Individual factors like age, medical history, allergies, degree of exercise, or dietary needs are not taken into account; instead, the classification is done at the recipe level. As a result, the model shouldn't be seen as offering specific food or medical recommendations.

XIII. CONCLUSION

Conclusion
In conclusion
Using structured recipe, ingredient, and nutritional data, this study created a multi-label machine learning framework at the recipe level for nutritional classification. Recipe titles, cuisine, cleansed ingredients, preparation time, and nutritional information such as calories, carbs, protein, fats, free sugar, fiber, salt, calcium, iron, vitamin C, and folate were all included in the 725 recipes in the dataset.
Data cleansing, duplicate handling, missing-value analysis, ingredient cleaning, and feature preparation were all part of the preprocessing of the dataset. While nutritional characteristics were kept as numerical predictors, ingredient data was transformed into machine-readable text features. Healthy, Protein_Rich, Fibre_Rich, Calcium_Rich, Iron_Rich, and Vitamin_Rich are the six binary target labels that were taken into consideration, enabling a single recipe to fall into several nutritional categories at once.
The same experimental setup was used to train and assess four machine learning algorithms: Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost. Accuracy, Precision, Recall, and weighted F1-Score were used to compare model performance after the dataset was split into 580 training records and 145 testing records.
XGBoost had the highest overall accuracy (75.68%) of the four models, while Random Forest had a relatively comparable accuracy (75.58%).

Additionally, XGBoost outperformed Decision Tree (51.73%) and Random Forest (43.62%) in terms of overall F1-Score (53.60%), while SVM attained 57.60%. These findings demonstrate that rather than one method being consistently better for all classification tasks, model performance varies based on the assessment parameter and nutritional label.

Performance differed among nutritional categories, according to the label-wise analysis. While categories like Fibre_Rich and Iron_Rich were more difficult to understand minority nutritional classes from the existing data, Healthy and Protein_Rich classification often produced better results.

Overall, the study shows that multi-label nutritional classification may be achieved with structured nutritional and ingredient data at the recipe level. Future food-information or recipe-recommendation systems may benefit from the results, which offer a machine-learning-based basis for comprehending the nutritional features of recipes. However, rather than serving as individualized dietary or medical recommendations, the results should be taken in light of the established nutritional thresholds and the constraints of the available information.

XIV. FUTURE SCOPE

Future Improvements

A basis for multi-label nutritional classification at the recipe level is provided by the proposed study. Future studies could take into account a number of enhancements:

1. Bigger and More Various Datasets:

Larger datasets with more recipes, cuisines, ingredients, and nutritional data may be used in future research. This may enhance the models' capacity to generalize to a greater range of foods.

2. Better Normalization of Ingredients:

To more precisely handle synonyms, spelling variants, ingredient quantities, units, and preparation techniques, more sophisticated ingredient preprocessing can be used.

3. Richer Nutritional Databases:

Additional nutritional characteristics including cholesterol, trans fat, saturated fat, various vitamins and minerals, serving size, and other pertinent nutritional metrics can be included in future datasets.

4.Better Management of Class-Imbalance:

To enhance the identification of minority nutritional groups, methods including SMOTE, class weighting, stratified sampling, and other resampling techniques might be examined.

5. Optimization of Hyperparameters:

In order to find better model configurations, future studies can systematically adjust hyperparameters using techniques like Grid Search, Random Search, or Bayesian Optimization.

6. Cross-Checking

Future research can use K-fold cross-validation to generate more reliable model performance estimates and lessen reliance on a single train-test split.

REFERENCES

[1] M. Bolaños, A. Ferrà, and P. Radeva, “Food Ingredients Recognition through Multi-label Learning,” in *New Trends in Image Analysis and Processing – ICIAP 2017*, LNCS, pp. 394–402, 2017, doi: 10.1007/978-3-319-70742-6_37.

[2] W. Wang, L.-Y. Duan, H. Jiang, P. Jing, X. Song, and L. Nie, “Market2Dish: Health-aware Food Recommendation,” 2020.

[3] S. S. Chaudhari, S. Rajarajeswari, C. Prasad, and A. L. Rane, “Intelligent Computing Approaches for Nutrition-aware Food Recommendation System: Concepts, Techniques and Trends,” International Journal of Pattern Recognition and Artificial Intelligence, 2026.

[4] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, “A Survey on Food Computing,” ACM Computing Surveys, vol. 52, no. 5, Art. 92, 2019, doi: 10.1145/3329168.

[5] K. Bora, “Controlled Recipe Generation: Adapting Food Recipes to Meet Dietary Restriction,” CEUR Workshop Proceedings, 2025.

[6] D. Tsolakidis, L. P. Gymnopoulos, and K. Dimitropoulos, “Artificial Intelligence and Machine Learning Technologies for Personalized Nutrition: A Review,” Information, 2024.

[7] M. Chen, X. Jia, E. Gorbonos, C. T. Hoang, X. Yu, and Y. Liu, "Eating Healthier: Exploring Nutrition Information for Healthier Recipe Recommendation," Information Processing & Management, vol. 57, no. 6, Art. 102051, 2020, doi: 10.1016/j.ipm.2019.05.012.

[8] K. Neha, P. Sanjan, H. Hariharan, S. Namitha, J. Jyoshna, and A. B. Prasad, "Food Prediction Based on Recipe Using Machine Learning Algorithms," in Proc. 2023 2nd Int. Conf. Augmented Intell. Sustainable Syst., pp. 411–416, 2023, doi: 10.1109/ICAISS58487.2023.10250758.

[9] R. Ruede, V. Heusser, L. Frank, A. Roitberg, M. Haurilet, and R. Stiefelhagen, "Multi-Task Learning for Calorie Prediction on a Novel Large-Scale Recipe Dataset Enriched with Nutritional Information," 2020/2021.

[10] A. Morales-Garzón, A. Quiñones-Muñoz, K. Gutiérrez-Batista, and M. J. Martin-Bautista, "A Multimodal Deep Learning Framework for Nutritional Estimation and Health-Oriented Recipe Analysis," Multimedia Systems, vol. 32, no. 2, Art. 153, 2026, doi: 10.1007/s00530-025-02151-3.

[11] M. Yamaguchi, M. Araki, K. Hamada, T. Nojiri, and N. Nishi, "Development of a Machine Learning Model for Classifying Cooking Recipes According to Dietary Styles," Foods, vol. 13, no. 5, Art. 667, 2024, doi: 10.3390/foods13050667.

[12] A. Lakshmanarao, G. Obulreddy, V. M. L. Priya, V. K. V. Ganesh, and M. Jayaram, "Diet and Recipe Recommendation System Using Machine Learning and Deep Learning Models," in Proc. 8th Int. Conf. Intelligent Sustainable Syst., pp. 1522–1527, 2026, doi: 10.1109/ICISS67859.2026.11454009.

[13] A. Nandeppanavar, M. Kudari, P. Bammigatti, and K. Vakkund, "A Machine Learning-Based Food Recommendation System with Nutrition Estimation," International Journal of Data Analysis Techniques and Strategies, vol. 16, no. 4, pp. 487–507, 2024, doi: 10.1504/IJDATS.2024.142485.

[14] "Composing Recipes Based on Nutrients in Food in a Machine Learning Context," Neurocomputing, vol. 415, pp. 382–396, 2020, doi: 10.1016/j.neucom.2020.08.071.

[15] T. Naravane and I. Tagkopoulos, "Machine Learning Models to Predict Micronutrient Profile in Food After Processing," Current Research in Food Science, vol. 6, Art. 100500, 2023, doi: 10.1016/j.crfs.2023.100500.

[16] P. Ma et al., "Application of Machine Learning for Estimating Label Nutrients Using USDA Global Branded Food Products Database (BFPD)," Journal of Food Composition and Analysis, vol. 100, Art. 103857, 2021, doi: 10.1016/j.jfca.2021.103857.

[17] P. Ma et al., "Deep Learning Accurately Predicts Food Categories and Nutrients Based on Ingredient Statements," Food Chemistry, vol. 391, Art. 133243, 2022, doi: 10.1016/j.foodchem.2022.133243.

[18] Q. Thames et al., "Nutrition5k: Towards Automatic Nutritional Understanding of Generic Food," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 8903–8911, 2021, doi: 10.1109/CVPR46437.2021.00879.

[19] J. Marin et al., "Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images," IEEE Trans. Pattern Anal. Mach. Intell., 2019, doi: 10.1109/TPAMI.2019.2927476.

[20] "AI-based digital image dietary assessment methods compared to humans and ground truth: A systematic review," 2023.

[21] "Advancements in Using AI for Dietary Assessment Based on Food Images: Scoping Review," Journal of Medical Internet Research, 2024.

[22] S. Wang et al., "A Recommender System for Healthy Food Choices Based on Integer Programming," Applied Mathematics, Modeling and Computer Simulation, 2022, doi: 10.3233/ATDE221062.

[23] "Personalised Healthy Food Text Recommendations through Fuzzy Linguistic Variables: A Generative AI-Based Approach," 2024.

[24] A. Morales-Garzón, K. Gutiérrez-Batista, and M. J. Martín-Bautista, "Adaptafood: An Intelligent System to Adapt Recipes to Specialised Diets and Healthy Lifestyles," Multimedia Systems, 2025.

[25] A. Morales-Garzón, P. Santos Peinado, R. Morcillo-Jimenez, K. Gutiérrez-Batista, and M. J. Martin-Bautista, "Service-oriented Multi-

platform for Food Computing: A Mobile Application for Recipe Adaptation to Nutrition Behaviours (AI2Cuisine)," Expert Systems with Applications, vol. 281, Art. 127603, 2025, doi: 10.1016/j.eswa.2025.127603.

[26] K. Bora, "Controlled Recipe Generation: Adapting Food Recipes to Meet Dietary Restriction," CEUR Workshop Proceedings, 2025.

[27] B. K. Rai et al., "Classifying Food Ingredients Using Machine Learning on Nutritional and Biochemical Data," Discover Food, vol. 5, Art. 382, 2025, doi: 10.1007/s44187-025-00661-7.

[28] "Precision Nutrition: A Systematic Literature Review," 2021.

[29] D. Tsolakidis, L. P. Gymnopoulos, and K. Dimitropoulos, "Artificial Intelligence and Machine Learning Technologies for Personalized Nutrition: A Review," Information, 2024.

[30] X. Teng, W. Min, C. Liu, and H. Li, "Multimodal Information Fusion for Food Nutrition Estimation: A Survey," Information Fusion, Art. 104621, 2026, doi: 10.1016/j.inffus.2026.104621.