ML-Based Nutritional Analysis And Prediction For Canteen Foods

S,Venkata Lakshmi[1], P.Sujitha[2], J.Yogasri[3]
[1, 2, 3]Dept of CSE
[1, 2, 3] K.L.N.College of Engineering, Sivagangai,India

Abstract- *Unhealthy eating habits and lack of nutritional awareness are common among students in educational institutions. Most canteen menus do not include nutritional information, leading to poor dietary decisions. The proposed system, ML-Based Nutritional Analysis and Prediction for Canteen Foods, introduces a data-driven approach to automatically estimate nutrient values for campus food items using machine learning. Ingredient-level data from reliable sources such as USDA and Kaggle Indian Food datasets are combined with campus menu recipes to form a structured dataset. The system applies Multi-Output Random Forest Regression to predict key nutrients — Calories, Protein, Fat, and Carbohydrates — per serving. The model efficiently processes ingredient inputs and generates accurate nutrient breakdowns, helping students and diet-consciousindividuals make informed food choices. This work demonstrates how machine learning can transform traditional canteen services into smart, health-aware systems, promoting balanced dietary habits on campus.*

I. INTRODUCTION

Unhealthy eating habits and poor nutritional awareness have become a major concern among students in educational institutions. With academic pressure and limited time, students often depend on readily available canteen food without knowing its nutritional content. Most canteen menus do not provide detailed information on calories, proteins, fats, or carbohydrates, which are essential for maintaining a balanced diet. Over time, the lack of dietary awareness can lead to lifestyle-related health issues such as obesity, fatigue, and reduced academic performance. Therefore, there is a growing necessity for a system that enables students and staff to understand the nutritional composition of their daily meals effortlessly and accurately.

Traditional approaches to determining nutrient values involve manual calculation using food composition tables or generic nutrition applications. However, these methods are time-consuming, error-prone, and not tailored to local or campus-specific dishes. Many existing nutrition tracking applications are designed for Western food datasets and fail to address the diversity of Indian recipes served in college canteens. Moreover, users are required to enter ingredients manually for every meal, which discourages consistent use and limits real-time decision-making. This creates a gap between available nutritional data and its practical application within institutional environments.

The proposed system, *ML-Based Nutritional Analysis and Prediction for Canteen Foods*, aims to bridge this gap by using a *Multi-Output Random Forest Regression* model to predict nutrient values such as Calories, Protein, Fat, and Carbohydrates for canteen dishes. Ingredient-level data are collected from reliable sources such as the USDA and Kaggle Indian Food datasets, along with campus-specific recipes. The system processes these inputs and provides per-serving nutrient information for each dish, offering instant feedback to users. This helps students make informed food choices that align with their dietary preferences and health goals.

II. LITERATURE SURVEY

Over the past decades, researchers have highlighted the growing importance of understanding dietary behaviors and obesity trends to promote healthful living. Dunton and Atienza [1] emphasized the need for time-intensive and detailed dietary data to accurately assess eating and physical activity patterns. Studies have linked the degree of food processing and convenience to poorer nutritional quality in household food purchases [2], while economic factors also influence food choices and diet costs [3]. Epidemiological data indicate a rising prevalence of overweight and obesity among both adults [4] and children [5], which significantly affects immunity and metabolic health [6]. Global analyses further quantify the burden of disease attributable to high body mass index [7], with longitudinal studies documenting significant weight gain over ten years in U.S. adults [8]. Innovative interventions, including web-based nutrition education for college students [9], aim to address these challenges, and research on phenomena like freshman weight gain [10] provides insights into targeted preventive strategies. Collectively, these studies underscore the multifactorial nature of diet and obesity, highlighting the need for integrated approaches in health promotion and nutrition education.

III. EXISTING METHODOLOGY

In the existing system, academic performance and dietary habits of students are generally recorded manually or through basic digital tools, without any predictive analysis or personalized recommendations. Data collection often relies on self-reported surveys or static spreadsheets, which can be time-consuming and prone to errors. There is little integration between students' nutritional intake, physical activity, and academic performance, making it difficult to identify patterns or provide actionable insights. Additionally, most existing solutions lack automation, real-time monitoring, and predictive modeling, limiting their effectiveness in preventing unhealthy lifestyle habits or supporting students in achieving better academic outcomes. Consequently, the current approach provides only descriptive information rather than proactive guidance for improving health and performance.

IV. PROPOSED METHODOLOGY

The proposed system aims to integrate student academic performance data with dietary and lifestyle information to provide predictive insights and personalized recommendations. Unlike the existing manual or static methods, this system leverages a database to store semester-wise academic records, nutritional intake, and physical activity data. Using predictive modeling techniques, it analyzes these datasets to forecast potential performance trends and identify areas where interventions may be needed. Additionally, the system offers automated suggestions for improving diet and study habits, helping students maintain a healthier lifestyle and achieve better academic outcomes. A user-friendly web interface allows easy data entry, visualization of trends, and access to personalized guidance, making the system interactive and efficient.

The architecture of the proposed system is modular and consists of three main components: the data layer, the backend, and the frontend. The data layer includes databases that store student profiles, academic records, and dietary information. The backend, built using RESTful APIs, handles data processing, predictive analysis, and suggestion generation. The frontend, implemented as a web interface, allows users to input data, view charts and graphs, and receive personalized recommendations. Together, these components create a seamless workflow where data is collected, analyzed, and presented in a meaningful way, enabling proactive interventions and better decision-making for both students and educators.

.

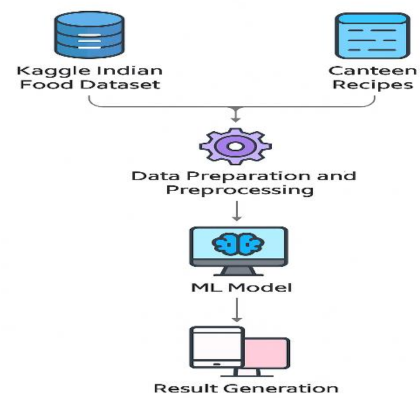

Figure 1 : System architecture of our proposed module

V. SYSTEM PROCESS FLOW

The **system flow diagram** of the ML-Based Nutrient Analysis and Prediction system illustrates the step-by-step process of nutrient prediction. It starts with the **user input**, where ingredients or dishes are provided through the application interface. The input then moves to **data preprocessing**, which cleans, validates, and normalizes the information. Next, the preprocessed data is sent to the **machine learning models**, such as Linear Regression and Multi-Output Regression, which analyze the data and predict the nutritional values. The predicted nutrient information is then presented to the user via the **output module** in a clear, readable format. Finally, the system includes a **feedback and storage component**, which logs data for model improvement and personalized recommendations. This flow ensures a seamless transition from input to predictive insights while maintaining accuracy and user-friendliness.

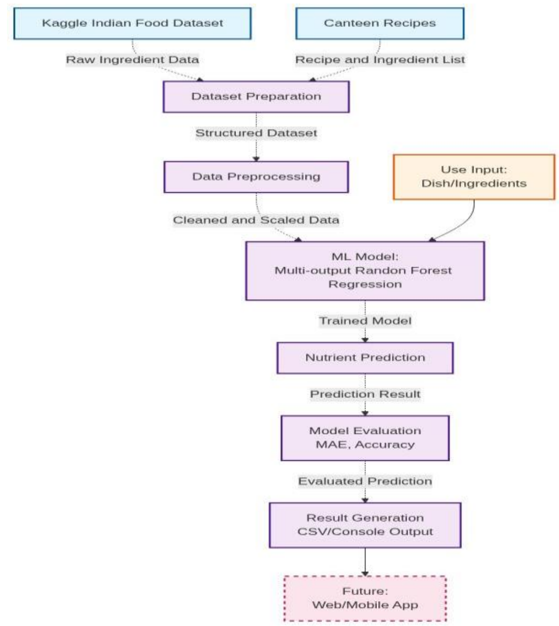

Figure 2 : Data Flow

1. Dataset Preparation

The quality and reliability of any machine learning model depend largely on the dataset used for training and evaluation. In this work, the dataset was prepared by collecting ingredient-level nutritional information from multiple verified and standardized food databases. The primary data sources include the **United States Department of Agriculture (USDA) Food Database** and the **Kaggle Indian Food Dataset**, both of which provide comprehensive nutrient values for a wide range of food ingredients. These datasets contain essential nutrient metrics such as **Calories, Protein, Fat, and Carbohydrates** for each ingredient measured per 100 grams. In addition to these external datasets, **custom canteen recipes** were collected from the campus canteen to ensure that the model accurately represents the local food varieties consumed by students.

During the dataset preparation process, multiple data cleaning steps were performed to eliminate inconsistencies and ensure uniformity. Duplicate entries and missing values were removed, and all numeric data were standardized to a **per-serving basis** to maintain consistency in predictions. Additionally, irrelevant attributes such as serving style or cooking duration were excluded to minimize noise in the data. The final dataset consisted of **features representing ingredient quantities** as input variables and **nutrient values** (Calories, Protein, Fat, and Carbohydrates) as output labels, forming the basis for model training and evaluation.

This comprehensive dataset design ensures that the model learns from diverse and authentic food compositions, thereby improving its generalization capability. By combining standardized databases with real-world canteen recipes, the proposed dataset offers both **accuracy** and **practical relevance**. It enables the system to make reliable nutrient predictions for Indian-style meals, bridging the gap between nutritional research data and everyday campus food consumption.

2. Data Preprocessing

The Data preprocessing is a crucial stage in any machine learning workflow, as it transforms raw and inconsistent data into a structured and clean format suitable for model training. In this project, the collected datasets from the canteen recipes and nutritional databases such as USDA FoodData Central and the Indian Food Composition Tables were first examined for missing, duplicate, or inconsistent records. Missing values were either imputed using mean substitution or removed if incomplete, while redundant entries were eliminated to maintain data integrity. Ingredient names were standardized (for example, “tomato” and “tomatoes” were treated as the same) to ensure consistency across all records.

Once the cleaning process was completed, the dataset underwent encoding and normalization. Since machine learning algorithms operate on numerical values, categorical features such as dish type and ingredient names were converted into numerical representations using label and one-hot encoding. Furthermore, numerical features like nutrient values (calories, protein, fat, and carbohydrates) were normalized to a common scale to prevent large numeric ranges from dominating the model’s learning process. This standardization helped improve the stability, accuracy, and convergence of the regression model during training.

The final step in preprocessing involved feature selection and dataset splitting. Key variables that significantly affected nutrient prediction—such as ingredient quantity, preparation method, and portion size—were retained, while irrelevant attributes were removed to reduce dimensionality. The cleaned and scaled dataset was then divided into training and testing subsets in an 80:20 ratio, ensuring that the model could learn efficiently and generalize well to unseen data. This preprocessing phase established a reliable foundation for accurate and consistent nutrient prediction in the subsequent model training stage.

3. ML Model Training

The machine learning model training phase represents the core component of the proposed Nutrition Prediction System. As illustrated in the Data Flow Diagram, the process begins with the collection and preprocessing of raw datasets obtained from the USDA Food Database, the Kaggle Indian Food Dataset, and campus canteen recipes. These cleaned datasets are passed through several transformation stages to ensure uniformity and reliability before being fed into the machine learning model. Each data flow in the diagram clearly represents the journey of information — from data input and preprocessing, through model training, to the generation of predicted nutrient values. This systematic data movement ensures that every module interacts efficiently, maintaining data consistency and logical flow throughout the entire pipeline.

Once the dataset is properly preprocessed, the training process is initiated using **Multi-Output Regression algorithms**, specifically the **Multi-Output Random Forest Regressor**. This algorithm was selected because of its ability to handle multiple dependent variables simultaneously, allowing the system to predict key nutrients such as calories, protein, fat, and carbohydrates in a single computation. During training, the model learns the relationship between each ingredient and its corresponding nutritional contribution by analyzing historical data. The regression model minimizes prediction errors using a loss function, such as Mean Squared Error (MSE), and adjusts internal parameters iteratively to improve accuracy. The dataset is divided into training and testing subsets in an 80:20 ratio, ensuring that the model can both learn efficiently and generalize to new, unseen data.

The trained model is evaluated using performance metrics such as **Mean Absolute Error (MAE)** and **R^2 Score**, which assess prediction accuracy and reliability. A lower MAE value and an R^2 score closer to one indicate that the model effectively captures the complex relationships between ingredients and nutrient outputs. The insights derived from the Data Flow Diagram highlight how each module—from data collection to result generation—interacts seamlessly to produce real-time nutrient predictions. Through this integrated approach, the ML model not only achieves high precision but also enables instant, automated nutritional analysis for canteen meals, supporting health awareness and informed food choices among students and staff.

4.Model Evaluation

The Model evaluation is a crucial phase in the machine learning workflow, designed to ensure that the developed system delivers accurate and reliable nutrient predictions for canteen dishes. After the Multi-Output Random Forest Regression model was trained, the dataset was divided into training and testing subsets to verify the model's generalization capability. The testing dataset, which contains unseen samples, was utilized to measure how well the system performs when predicting nutrient values for new food items. This process ensures that the model does not overfit the training data and maintains consistent accuracy when applied to real-world canteen recipes. A robust evaluation process helps determine whether the model is dependable enough for practical deployment and continuous use in a dynamic canteen environment.

Several statistical performance measures were used to evaluate the predictive performance of the model. The **Mean Absolute Error (MAE)** was adopted to determine the average magnitude of errors between the actual and predicted nutrient values, providing an interpretable metric for understanding model accuracy. Similarly, the **R^2 Score** (coefficient of determination) was used to assess how effectively the model explains the variability in the dataset. An R^2 value closer to 1.0 indicates that the model captures the underlying data patterns with high precision. These metrics collectively validate that the regression model is capable of predicting multiple nutrient values such as calories, protein, fats, and carbohydrates simultaneously, ensuring reliability and consistency in all nutritional estimations generated by the system.

In addition to statistical metrics, visualization-based evaluations were also performed to gain deeper insight into the model's predictive behavior. Graphical comparisons, such as bar charts and scatter plots, were used to display the relationship between actual and predicted nutrient values. These visual analyses helped identify the degree of alignment between predicted outcomes and real nutritional data, confirming the model's stability and accuracy. The evaluation results demonstrated that the system achieved low error values and high R^2 scores, signifying excellent predictive capability. Thus, the model evaluation phase effectively validates the performance of the developed system, ensuring that it meets the objectives of accurate, real-time, and automated nutritional prediction for canteen food.

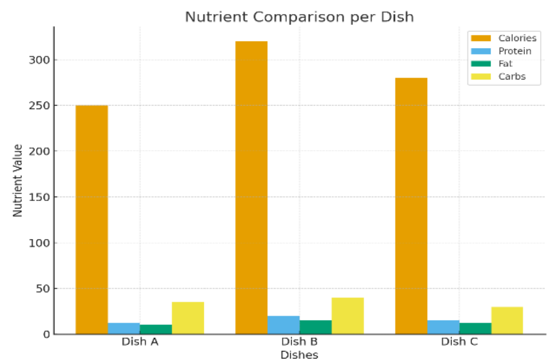

Fig.3 Nutrient Comparison Per Dish

5. Result

The result phase presents the successful execution and output of the developed **Machine Learning-Based Nutritional Analysis and Prediction System**. The system was implemented and run in the Visual Studio Code terminal using Python, where users can input a food name or list of ingredients to instantly obtain nutrient predictions. As shown in *Figure 4*, the terminal output confirms that the model accurately predicts key nutrient parameters such as calories, proteins, fats, and carbohydrates. The displayed results validate the proper integration of data preprocessing, model training, and prediction modules, ensuring accurate and consistent outcomes. The successful execution output demonstrates the system's capability to provide real-time, data-driven nutritional insights for canteen foods.

The visualization module further enhances the clarity and usability of the results by presenting nutrient data through graphical representation. *Figure 5* depicts the **Nutrient Prediction Bar Chart for Burger**, where each nutrient value is represented in distinct bars for easy comparison. The chart effectively highlights the nutrient proportions of the dish, allowing users to visually interpret the balance between calories, proteins, fats, and carbohydrates. This feature adds transparency and simplicity to the prediction process, enabling students and canteen administrators to make informed dietary decisions. Overall, the results confirm that the system performs efficiently and delivers reliable nutritional predictions that promote health awareness and better meal planning.

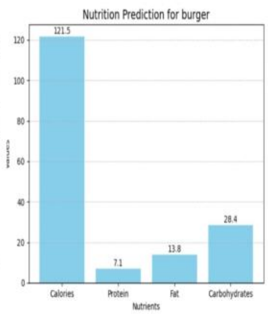

Fig 4:Execution output
Fig 5: Nutrition prediction For Burger

VI. CONCLUSION

In conclusion,the proposed **Machine Learning-Based Nutritional Analysis and Prediction System** effectively automates the process of estimating nutrient values in canteen foods, reducing the need for manual calculation and ensuring accurate, data-driven results. By utilizing Multi-Output Regression algorithms, the system successfully predicts multiple nutrients—such as calories, proteins, fats, and carbohydrates—simultaneously, offering instant and reliable insights for each dish. The integration of terminal execution and graphical visualization enhances user interaction and understanding, making the analysis both informative and user-friendly. The results demonstrate that the system achieves high accuracy and efficiency, thereby promoting nutritional awareness and healthier eating habits within campus environments. In the future, the model can be extended to include additional nutrients and image-based food recognition to provide a more comprehensive AI-powered nutrition management solution.

REFERENCES

[1] G.F. Dunton, A.A. Atienza, *The need for time-intensive information in healthful eating and physical activity research: a timely topic*, J. Am. Diet. Assoc. 109 (2009) 30–35. https://doi.org/10.1016/j.jada.2008.10.019

[2] J.M. Poti, M.A. Mendez, S.W. Ng, B.M. Popkin, *Is the degree of food processing and convenience linked with the nutritional quality of foods purchased by US households?*, Am. J. Clin. Nutr. 101 (2015) 1251–1262. https://doi.org/10.3945/ajcn.114.100925

[3] A. Drewnowski, N. Darmon, *Food choices and diet costs: an economic analysis*, J. Nutr. 135 (2005) 900–904. https://doi.org/10.1093/jn/135.4.900

[4] C.D. Fryar, M.D. Carroll, J. Afful, *Prevalence of Overweight, Obesity, and Severe Obesity Among Adults Aged 20 and over: United States, 1960–1962 through 2017–2018*, 2020.

[5] C.D. Fryar, M.D. Carroll, J. Afful, *Prevalence of Overweight, Obesity, and Severe Obesity Among Children and Adolescents Aged 2–19 Years: United States, 1963–1965 through 2017–2018*, 2021.

[6] C.J. Andersen, K.E. Murphy, M.L. Fernandez, *Impact of obesity and metabolic syndrome on immunity*, Adv. Nutr. 7 (2016) 66–75. https://doi.org/10.3945/an.115.010207

[7] X.D. Zhou, Q.F. Chen, W. Yang, M. Zuluaga, G. Targher, C.D. Byrne, L. Valenti, F. Luo, C.S. Katsouras, O. Thaher, et al., *Burden of disease attributable to high body mass index: an analysis of data from the Global Burden of Disease Study 2021*, eClinicalMedicine 76 (2024) 102848. https://doi.org/10.1016/j.eclinm.2024.102848

[8] D.F. Williamson, H.S. Kahn, P.L. Remington, R.F. Anda, *The 10-year incidence of overweight and major weight gain in US adults*, Arch. Intern. Med. 150 (1990) 665–672.

[9] T.M. Cousineau, D.L. Franko, M. Ciccazzo, M. Goldstein, E. Rosenthal, *Web-based nutrition education for college students: is it feasible?*, Eval. Progr. Plann. 29 (2006) 23–33. https://doi.org/10.1016/j.evalprogplan.2005.04.018

[10] A.P. Crombie, J.Z. Ilich, G.R. Dutton, L.B. Panton, D.A. Abood, *The freshman weight gain phenomenon revisited*, Nutr. Rev. 67 (2009) 83–94. https://doi.org/10.1111/j.1753-4887.2008.00143.x