

NutriScan: Smart Food Label Analyzer for Health Awareness

Ms. Ishani H B¹, Ms. Khathija Rafida², Ms. Kadija Afra³, Ms. Lakshmi Chandran⁴, Ms. Shruthi⁵
^{1,2,3,4,5}*Department of Computer Science and Engineering Shree Devi Institute of Technology, Mangalore, Karnataka, India*

Abstract—The increasing consumption of packaged food products makes it vital for consumers to understand the ingredients and nutritional information mentioned on the label of the food product. However, food labels often contain complex ingredient name, technical nutritional information, various additives, preservatives, and other components that maybe challenging to understand for an ordinary consumer. This paper presents a Food Label Analyzer for Health Awareness a smart system intended to simplify and analyze food product labels. The proposed system utilizes Optical Character Recognition (OCR) to extract textual information contained in the image of food label and employs Machine learning (ML) and Natural Language Processing (NLP) method for ingredient and nutritional analysis. Moreover, the system extracts relevant ingredient and recognizes harmful components such as excessive sugar, unhealthy fats, preservatives, artificial additives etc., and evaluates their possible impact, providing simplified information about each ingredient and its potential side effects. In addition, the system offers personalized analysis based on user's requirements such as diabetes, hypertension, or food allergies. A health classification mechanism classifies products into Healthy, Moderate, and Avoid categories' according to the obtained information. The proposed approach is intended to help bridge gap between complex information provided on the food label and consumers understanding, leading to better food awareness and healthier eating habits

Index Terms—Food label Analysis, Machine Learning, Natural Language Processing, Nutritional Analysis, Optical Character Recognition (OCR), Personalized Health Analysis.

I. INTRODUCTION

The growing intake of packaged food has made food labels increasingly important as a source of information for consumers. Food label provide details regarding ingredients, nutritional value, additives

preservatives, allergen and other information related to the product. However, this information is often presented in a highly technical way using public ingredient name, nutritional table with very small fonts, and specialized terms such as allergen, which make it harder for the average consumer to understand and make healthier choice. Recent studies have concentrated on automating in extraction of information from food labels to OCR, ML and AI in order to close this information gap between the food label and the consumer.

Optical character recognition (OCR) to make it possible to automate the extraction of information from food labels. The information extracted from the label image can then be processed with Machine Learning (ML) and Natural Language Processing (NLP) techniques to classify ingredients, nutritional value, additives, preservatives and allergen. The most recent research has focused on applying OCR and ML together with NLP to classify the ingredient contained in food products. In addition, some studies have also invested the use of advanced ai algorithm to extract complex nutritional information from food labels.

While existing studies have concentrated on automation there is now a need to develop an easy understandable application that can process the information on a food label and highlight the relevant details for the consumer. Existing tool may process nutritional value and classify ingredients but they fail to provide information about potential harmful ingredients, additives and preservatives. Furthermore, a personalized, health centric approach based on the user's health matrix is also required to determine whether a product is healthy or not. Therefore, personalization and health-based information about the food item present in newest research direction in the field of automated food label readers.

The proposed tool, Nutri-scan, is an effort to create an intelligent system for a Food Product Label Analyzer for Healthier Awareness. The application will accept the image of food label as an input, extract the relevant information using OCR, and process the extracted information with ML and NLP to identify ingredients, and nutritional value, additives, preservatives, and potential allergen. It will then produce an easily understandable report on how the product is based on its values and ingredients in addition, the system will take into account each user's specific health needs to determine whether the product is safe to consume to deliver and detail the report on how its ingredients affect the body. This detail analysis will also solve the product into one of three categories: Healthy, Moderate or Avoid.

By offering a clear and easily understandable breakdown of the food label, the proposed system seeks to close the information gap between the consumer and the food product. The application will confirm a highly technical and complex food label into a meaningful, easy-to-understand format which helps consumer make healthier dietary choices.

II. PROBLEM STATEMENT

Because of the increasing habit of buying packaged goods, the necessity of monitoring diet quality by analyzing the information on packaging has become even more crucial. In most cases, the nutritional information on packaging appears as an ingredient list along with a set of nutritional facts. Usually, these lists include many scientific terms and the symbols that typical shoppers do not find easy to understand. Consequently, the buyer is unable to gain a clear understanding of the kind of food they are purchasing. The present range of food-related smartphone applications on the market consist mainly of apps that offer basic nutritional and health rating information or nothing more than a shallow analysis of ingredients. There is a need for highly contextual information and an intuitive breakdown of harmful and beneficial ingredients, such as sugar, preservative, artificial additives, fats, and allergen.

In addition, users with particular nutritional and health condition, including diabetes, high blood pressure, and nut allergy, need more attention and detailed description.

To tackle these problems, a multifunctional and intuitive application name Nutri-Scan was developed. It makes use of model technologies including Optical Character Recognition (OCR), Machine Learning (ML) and Natural Language Processing (NLP), to extract and interpret details about ingredients, harmful additive and nutritional components.

The application scans the label of the product being purchased, breaks down the information it contains, detects harmful and beneficial ingredients and delivers a personal health insight based on the items selected. The main value of such application lies in its ability to make food scanning intuitive and to give the user highly contextual information.

III. OBJECTIVE

NutriScan's primary goal is to create an intelligent, easy-to-use system that makes food label information simpler and assist consumers in grasping the potential health effect of food products.

The specific objectives are as follows:

- 1) Food Label Extraction: To pull ingredients and nutritional details from image of food labels by means of Optical Character Recognition (OCR).
- 2) Ingredient classification: To apply machine learning method to categorize ingredients and detect potential harmful elements like added sugar, preservatives and artificial additives.
- 3) Nutritional analysis: To examine calories, carbohydrates, fat, proteins, sodium and other pertinent nutritional components
- 4) Health Impact Analysis: To offer a clearer and easily understood information regarding the potential health effects of the ingredients identified
- 5) Allergy detection: To spot common allergens such as dairy, nut, soya, egg and gluten on the food label.
- 6) Personalized analysis: To take into account user specific need such as diabetes, hypertension, and allergies when analyzing a product.
- 7) Health Classification: To sort food products into healthy moderate and avoid categories According to the information analyst user friendly result to display the analysis in straightforward format so that the user can grasp food label information without needing specialized nutritional knowledge.

IV. LITERATURE SURVEY

A. OCR for Extracting Information from Food Labels
Optical Character Recognition (OCR) is essential to automated food-label analysis, as it enables machines to read text directly from images of food labels. Recent studies have investigated the use of OCR to extract nutritional and ingredient information from food package labels. When paired with machine learning, OCR can help automate the checking and analysis of nutritional claims, reducing the need to manually read complicated labels. [1]

B. Machine Learning for Ingredient and Nutritional Analysis

Machine Learning (ML) techniques have been applied to classify the information extracted from food labels. Recent work has shown how classification algorithms can be used to assess nutritional claims and pinpoint relevant food-label information. Approaches of this kind can provide automated interpretation of ingredients and nutritional properties. [1]

C. Food Allergen Detection

Identifying food allergens is a key component of food safety. Recent studies have applied machine learning techniques such as Decision Trees, Random Forest, Support Vector Machines, and K-Nearest Neighbors to detect and classify food allergens. These approaches can show how automated systems are able to detect allergenic components and support safer food choices. [2]

D. Personalized Food Analysis

Recent food-label analysis systems have begun to incorporate user-specific information such as allergies and dietary requirements. OCR and machine learning can be combined with user profiles to deliver personalized analysis and to score products according to those requirements. This method is beneficial for consumers, as it enables them to select foods that align with their dietary restrictions. [3]

E. Nutritional Claims and Consumer Awareness

Studies on packaged food products have also examined how accurate the nutritional claims presented to consumers are regarding what the product contains. A recent survey of food products listed on Indian e-commerce platforms uncovered multiple

inconsistencies in nutritional claims, highlighting the importance of trustworthy nutritional information and consumer awareness. These findings point to the necessity of systems that assist consumers in interpreting and verifying food-label information more effectively. [4]

F. NLP for Food-Label Understanding

Natural Language Processing (NLP) can be applied to the text extracted from food labels in order to detect meaningful information within ingredient descriptions.

Processing based on NLP can assist in standardizing ingredient names, categorizing food components, and providing users with simplified explanations. Bringing together OCR, NLP, and machine learning can offer a more thorough way of making sense of complex food-label information.

G. Comparative Analysis of Existing Systems

Most existing studies concentrate on separate components, such as OCR-based extraction, verification of nutritional claims, allergen detection, or personalized dietary analysis. A few recent systems have merged OCR and machine learning with user profiles to deliver personalized food analysis. [1], [2], [3] Even so, there remains room for an integrated system that unites OCR, ML, and NLP to analyze ingredients, nutritional values, additives, preservatives, allergens, and similar elements. Nutri-Scan is built to meet the need for simplifying health-related information and for classifying products according to the characteristics identified through analysis.

V. PROPOSED METHODOLOGY

The proposed Nutri-Scan system is designed to offer a simplified form of the complex information found on food labels, giving users insight into the particular health benefits of the food item. The system operates by scanning a food label input and running a series of extraction, analysis, and personalization algorithms.

A. Food Label Image Input

The food label image is provided to the system either through a camera or by the user's upload. The image needs to contain all ingredients and nutritional information.

B. Image Preprocessing and OCR

The input image is then preprocessed to enhance its quality and get it ready for Optical Character Recognition.

This algorithm handles recognizing the characters in the image and converting them into an electronic data format.

C. Ingredient and Nutritional Analysis

The extracted information is processed and organized. Ingredients like added sugars, refined ingredients, additives, and preservatives are separated from nutritional information such as calories, carbohydrates, proteins, and others.

D. Additive and Allergen Detection

The system also scans for additives, preservatives, colors, flavors, sweeteners, and finally allergens such as gluten, dairy, nuts, soy, and eggs.

E. Health Impact and Personalization

The ingredients and nutritional information are analyzed and sorted into health impact categories. Users can also add their preferences and restrictions, such as diabetes, high blood pressure, and allergies. The system tailors the food label to the user's specifications.

F. Health Classification

Lastly, the system classifies the food item and delivers a summarized, simplified, and personalized version of the nutritional information present on the food label. The product falls into three categories: Healthy, Moderate, and Avoid.

The user receives a simplified version of the nutritional facts and gains a clearer understanding of the food label.

VI. SYSTEM IMPLEMENTATION

The Food Product Label Analyzer for Health Awareness is introduced as an intelligent agent that examines food product labels and provides simplified details about their ingredients, nutritional values, and possible health effects.

The system employs Optical Character Recognition, Machine Learning, and Natural Language Processing to examine the information it gathers.

A. Image Capture and Input

Users can scan a food product label with a mobile camera or upload a picture of the label through the system. The image that is captured or uploaded serves as the input, containing the ingredient list, nutritional information, additives, preservatives, and allergen-related details. The captured image is fed into the system as the input.

B. Image Preprocessing and OCR

The captured food-label image is first preprocessed to boost its readability, making it easier to extract textual information from it. The image undergoes cleaning, resizing, and contrast adjustments to enhance its quality. The preprocessed image is then passed through Optical Character Recognition to retrieve the textual information contained in the image. The extracted information is subsequently transformed into a machine-readable format for further processing

C. Data Processing and Ingredient Classification

The extracted text is then preprocessed to remove unwanted symbols and recognition errors. The ingredients are split apart and organized into a structured format that the system prefers. Machine Learning algorithms are then applied to categorize ingredients as healthy, moderate, or harmful. The system identifies hidden components in foods, including added sugars, preservatives, artificial additives, and unhealthy fats.

D. Nutritional and Health Impact Analysis

The system examines the nutritional information it has extracted from the food label. It recognizes healthy components such as dietary fiber, proteins, and vitamins, while pointing out unhealthy components like saturated fats, sodium, and added sugars. It clarifies to the user the possible health impact of each ingredient.

E. Detection of Additives, Preservatives, and Allergens

The analyzer examines the food label to identify colors, flavors, sweeteners, and preservatives, and then assesses their possible effects on health. It identifies allergens such as gluten, dairy, nuts, soy, and eggs, and points out precautionary warnings like 'may contain traces of'. The analyzer assists the user in grasping how the ingredients affect them.

F. Personalized Recommendations and Health Classification

The system lets the user input their health condition, for example diabetes, hypertension, or an allergy. The health requirements are taken into account, and the system offers personalized suggestions based on the user's health status along with details about the food product. It rates the product's health benefits and labels it as healthy, moderate, or avoid, helping the user monitor their diet and make healthier choices.

VII. RESULT AND DISCUSSION

The Food Product Label Analyzer that has been implemented provides a complete solution for retrieving and analyzing the information found on food labels. The system applies OCR to process and analyze the image of the food label that is uploaded.

A. Food Label Text Extraction

From the uploaded image, the OCR module pulls out the ingredients and nutritional information of a food product. Preprocessing the image helps improve the extraction of label information. The information that is extracted is then processed so that further analysis can be carried out.

B. Ingredient and Nutritional Analysis

The system examines the key ingredients and nutritional components present in the food product. Components including added sugars, preservatives, artificial additives, unhealthy fats, calories, carbohydrates, proteins, fats, and sodium levels are extracted. This helps convert the complex information on the label into a format that is easier to understand.

C. Harmful Ingredient and Allergen Detection

The system identifies harmful food ingredients and allergens, including excess sugar, preservatives, artificial additives, unhealthy fats, gluten, dairy, nuts, soy, and eggs. This points out possible warning flags regarding a food product.

D. Personalized Health Analysis

The system provides personalized health analysis tailored to individual health needs such as diabetes, hypertension, or allergies. The extracted food information is compared with the user's health needs in order to provide specific warnings and recommendations.

E. Health Classification

On the basis of the health analysis, the system classifies the food product as Healthy, Moderate, or Avoid. This helps an average user grasp the nature of the food product without needing to interpret complex nutritional and ingredient information.

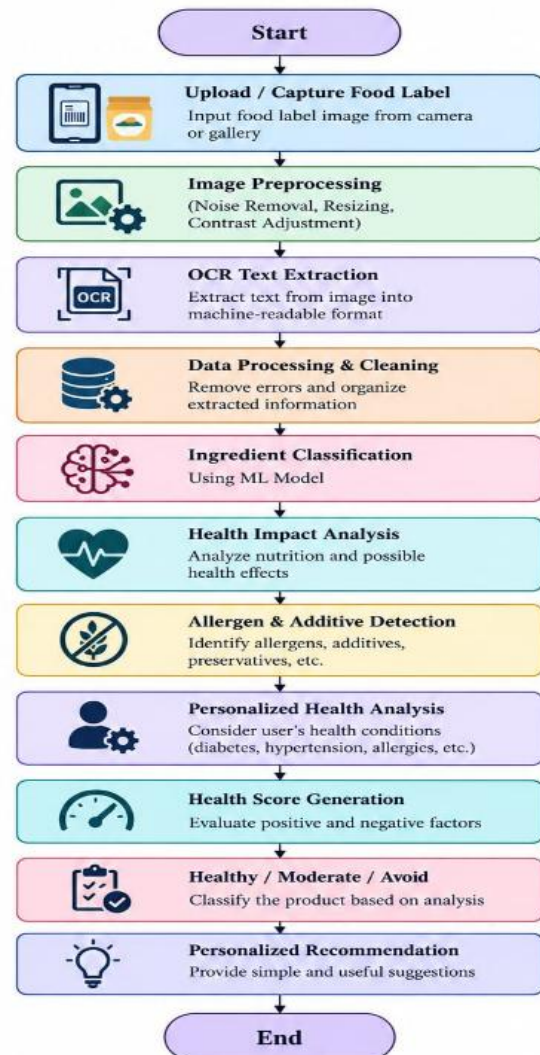


Fig. 1. Flowchart of the Proposed Food Product Label Analyzer.

F. Discussion

Summary

The analysis makes it clear that combining OCR, Machine Learning, and efficient data processing can change how users view and understand food labels. In this application, the system successfully extracted the key information needed to provide an in-depth analysis of a given food product. The extracted information covers ingredients, nutritional

information, allergens, and a personalized health analysis tailored to the user's health needs. The project report also highlighted real-time feedback and personalized recommendations as essential elements in developing such an application.

VIII. CONCLUSION

The proposed NutriScan: Food Product Label Analyzer for Health Awareness provides an intelligent and user-friendly solution for understanding complex food labels and improving health awareness among consumers. By integrating technologies such as Optical Character Recognition (OCR), Machine Learning (ML), and Natural Language Processing (NLP), the system is designed to extract, analyze, and simplify ingredient and nutritional information from packaged food products.

The system identifies potentially harmful components such as hidden sugars, preservatives, unhealthy fats, and artificial additives, and provides simplified information about their possible effects on the human body. It also detects common allergens and supports personalized analysis based on user health conditions such as diabetes, hypertension, and allergies.

The proposed system classifies food products as Healthy, Moderate, or Avoid, making the analyzed information easier for users to understand. Personalized recommendations further help users evaluate food products according to their individual requirements.

Overall, Nutri-Scan aims to bridge the gap between complex food-label information and consumer understanding. By providing simplified analysis, health-related insights, and personalized recommendations, the system supports better food awareness and more informed dietary choices. The project can also serve as a foundation for further development in intelligent nutrition analysis and AI-based health-support systems.

REFERENCES

- [1] S. P. Siddique Ibrahim, L. Karthikeyan, B. Pitti, T. K. Tejaswi, K. L. Sahithi, and V. N. Sai, "Automating nutritional claim verification: The role of OCR and machine learning in enhancing food label transparency," in *Proc. 2024 Int. Conf. on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, 2024, pp. 1164–1171, doi: 10.1109/ICICNIS64247.2024.10823177.
- [2] H. Shaukat, A. Sultan, and H. Salahuddin, "Enhancing food safety: A machine learning approach for accurate detection and classification of food allergens," *Journal of Computing & Biomedical Informatics*, 2024.
- [3] A. A. Adedeji, P. V. Priyesh, and A. A. Odugbemi, "The magnitude and impact of food allergens and the potential of AI-based non-destructive testing methods in their detection and quantification," *Foods*, vol. 13, no. 7, Art. no. 994, 2024, doi: 10.3390/foods13070994.
- [4] V. Prema, R. Devika, S. Kavaya, S. Shahitha Banu, and G. Shanmugapriya, "Food label analyzer with voice-based safety on web interface: A deep learning framework for multilingual (Tamil, Hindi, English) label understanding and accessible communication," 2025.
- [5] "Food label analysis for ingredient extraction and prediction," in *Proc. 2025 IEEE 7th Int. Conf. on Computing, Communication and Automation (ICCCA)*, 2025, doi: 10.1109/ICCCA66364.2025.11325614.
- [6] "Extract nutritional information from bilingual food labels using large language models," *Journal of Imaging*, vol. 11, no. 8, Art. no. 271, 2025, doi: 10.3390/jimaging11080271.
- [7] S. Sree, S. Chinnaiyan, B. Palanisamy, and S. P. Sivakumar, "Evaluating nutrient claim accuracy of packaged food products in Indian e-commerce platforms," *Discover Food*, vol. 5, Art. no. 326, 2025, doi: 10.1007/s44187-025-00597-y.
- [8] D. Ujalambkar, K. Kamat, G. Kulkarni, P. Khedkar, and K. Korpe, "FoodX: An AI-powered system for packaged food ingredient analysis and personalized health recommendations," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2025, doi: 10.22214/ijraset.2025.70940.
- [9] T. M. E. Saputara *et al.*, "Multimodal fine-grained grocery product recognition using image and OCR text," *Machine Vision and Applications*, vol. 35, no. 4, Art. no. 79, 2024, doi: 10.1007/s00138-024-01549-9.
- [10] B. Koroušić Seljak *et al.*, "NutriGreen image dataset: A collection of annotated nutrition, organic, and vegan food products," *Frontiers in Nutrition*, 2024.