

# Ingrelize: Product Data Simplification Using Machine Learning

Ms Priyanka<sup>a</sup>, Arun Ramachandra Hegde<sup>b</sup>, Abhishek Kagawad<sup>c</sup>, Harsha M<sup>d</sup>, Abhishek A<sup>e</sup>  
<sup>abcde</sup> Department of AIML, CMR Institute of Technology, Bengaluru, India

**Abstract**—In today’s fast-paced world, consumers often find it challenging to interpret the complex and technical ingredient lists found on packaged food products. This project presents *Ingrelize*, an intelligent application designed to scan, interpret, and simplify the ingredient section on food labels. By leveraging machine learning and natural language processing techniques, the system analyzes ingredients and translates them into more understandable terms for the average user. The primary goal of this project is to enhance consumer awareness and promote healthier decision-making by demystifying the often confusing terminology used in food product labeling.

**Index Terms**—Machine Learning, Ingredient Analysis, Food Label Interpretation, Natural Language Processing, Consumer Awareness

## I. INTRODUCTION

Food is an essential part of our daily lives, and making healthy dietary choices is crucial for overall well-being. However, in today’s fast-paced world, many individuals often overlook the importance of selecting nutritious food products. This oversight may stem from various factors such as a hectic lifestyle, lack of knowledge about food contents, or budget constraints.

The rise in health-related issues such as obesity and diabetes has become a growing concern globally. According to the World Health Organization (WHO), over 1.9 billion adults were overweight in 2016, of which more than 650 million were obese. Additionally, the International Diabetes Federation (IDF) reported that approximately 537 million adults were living with diabetes in 2021, and this number is expected to rise significantly in the coming years. These alarming statistics highlight the urgent need for better consumer awareness regarding food intake and nutrition.

To address this issue, we introduce *Ingrelize* — a machine learning-powered application that simplifies and interprets the ingredients listed on packaged food products. This tool is designed to help users quickly understand what they are consuming by converting complex ingredient names into more familiar, comprehensible terms. Not only does this assist individuals in making informed and healthier food choices, but it also saves valuable time and effort.

By empowering consumers with accessible information, *Ingrelize* aims to contribute toward the development of a more health-conscious society. The application bridges the gap between technical food labeling and everyday understanding, ultimately promoting better public health outcomes.

## II. RELATED WORK

Over the past decade, mobile-based food scanner applications have emerged as tools to assist consumers in making informed dietary decisions. These applications provide nutritional insights by scanning product barcodes and interpreting ingredient labels. Several studies have assessed their impact, functionality, and technical effectiveness.

Werle et al. [1] investigated the influence of food scanner apps on consumers’ food choices. Their findings reveal that while scanner apps increase the likelihood of choosing healthy products compared to no information, they are less effective than front-of-pack (FOP) nutrition labels when it comes to real-world purchase behavior. Similarly, Cornudet et al. [2] explored conditions under which scanner apps outperform FOP labels. They identified three specific conditions: when products have low nutritional scores, when consumers are distrustful of dominant food brands, and when brand equity is average or weak. This suggests that scanner apps and FOP labels may complement one another rather than compete directly.

Maringer et al. [3] conducted a quality assessment of food product databases used by popular barcode scanning apps. Their study found significant variability in both the availability and accuracy of nutritional values across different applications, with energy values being the most reliably reported. This highlights a potential challenge for systems like *Ingrelize*, which rely on backend databases to interpret ingredient content.

From a technical standpoint, Ma et al. [4] proposed a deep learning-based approach for predicting food categories and nutrient values from ingredient statements. Their model, trained on the USDA’s Branded Food Products Database, achieved classification accuracy as high as 99% and nutrient estimation accuracy with  $R^2$  values ranging from 0.93 to 0.97. This illustrates the potential of machine learning in automating food classification and interpretation, which aligns closely with the objectives of our work.

Collectively, these studies underscore both the promise and the limitations of current food scanning and interpretation technologies. Our project builds on these insights by emphasizing the simplification and contextual translation of complex ingredient lists, aimed at improving consumer understanding and encouraging healthier food choices.

### III. METHODOLOGY

The core objective of *Ingrelyze* is to simplify the complex and often technical ingredient lists found on packaged food products using Machine Learning and Natural Language Processing (NLP). The system is designed to extract the ingredients from product labels, analyze them, and present a user-friendly interpretation to the end user. The methodology involves four major stages: image preprocessing, text extraction, ingredient parsing, and interpretation using a machine learning model.

#### A. Image Preprocessing and Text Extraction

The system begins by capturing an image of the ingredient section on a food product using a mobile device. To enhance Optical Character Recognition (OCR) accuracy, preprocessing techniques such as contrast adjustment, noise reduction, and grayscale conversion are applied. Tesseract OCR is utilized to extract raw text from the processed image. This text is then segmented to isolate individual ingredient entries.

#### B. Ingredient Parsing and Standardization

Post text extraction, ingredient terms are cleaned and parsed. This involves removing duplicates, correcting spelling errors, and expanding abbreviations (e.g., “E322” is expanded to “Lecithin”). A dictionary of common food additives and their meanings is referenced during this process. Additionally, named entity recognition (NER) techniques are used to identify allergens, preservatives, colorants, and nutritional tags (e.g., “high in sodium”).

#### C. Machine Learning-Based Classification

To assess the health impact of the extracted ingredients, a supervised machine learning approach is adopted. Each ingredient is categorized into one of three classes based on its health implication: *safe*, *moderate*, or *avoid*. Several classification algorithms were evaluated, including Random Forest, Naive Bayes, and Support Vector Machines (SVM). Among them, the Multi-Layer Perceptron (MLP) model provided the best performance, achieving an accuracy of 96% on the validation set.

The MLP model was trained on a curated dataset containing over 100,000 labeled food ingredients sourced from public databases and health authority guidelines. Input features included the ingredient’s name, chemical properties, frequency of usage in processed foods, known health impacts, and regulatory classifications such as FDA status.

#### D. User Interface and Simplified Output

The final output is displayed through a user-friendly mobile interface. Each scanned ingredient is presented along with a simplified description (e.g., “Xanthan Gum – a common thickener, safe in small quantities”) and a color-coded badge (green, yellow, or red) indicating its health impact. This ensures that users, regardless of their background, can make informed food choices within seconds.

### IV. CONCLUSION

This paper presented *Ingrelyze*, a machine learning-powered mobile application designed to simplify complex food ingredient labels and promote healthier consumer decision-making. By integrating OCR, NLP, and classification algorithms such as Random Forest and Multi-Layer Perceptron, the system successfully interprets ingredient lists and presents user-friendly explanations.

The application addresses the growing need for accessible nutritional transparency, especially in an era where health issues such as obesity, diabetes, and food-related illnesses are on the rise. Through real-time scanning and intelligent analysis, *Ingrelyze* empowers users to understand what they consume, thereby contributing to a more health-conscious society.

Future enhancements may include support for multiple languages, personalized dietary recommendations based on health conditions, and integration with government-approved nutritional guidelines. The proposed system can also be extended to support regional food databases and allergen detection for sensitive individuals.

Overall, *Ingrelyze* demonstrates the potential of artificial intelligence to bridge the gap between food science and everyday consumer awareness.

### REFERENCES

- [1] C. O. Werle, C. Gauthier, A. P. Yamim, and F. Bally, “How a food scanner app influences healthy food choice,” *Appetite*, vol. 200, 2024.
- [2] C. Cornudet, M.-e. Laporte, F. Berger-Remy, B. Parguel, and J.-l. Richet, “When food scanner apps outperform front-of-pack nutrition labels,” *Psychology & Marketing*, 2025.
- [3] M. Maringer *et al.*, “Food identification by barcode scanning in the netherlands: a quality assessment of labelled food product databases underlying popular nutrition applications,” *Public Health Nutrition*, vol. 22, no. 7, pp. 1215–1222, 2018.
- [4] P. Ma, Z. Zhang, Y. Li, N. Yu, J. Sheng, H. K. McGinty, Q. Wang, and J. K. Ahuja, “Deep learning accurately predicts food categories and nutrients based on ingredient statements,” *Food Chemistry*, vol. 395, p. 133243, 2022.