JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FOODGUARD SCAN: AN AI AND OCR BASED APPROACH TO SMART FOOD LABEL INTERPRETATION

[1]Praneeth, [2]Shreyas S, [3]Sudarshan K, [4]Ujwal Abhishek, [5]Yajnesh N
[1]Third year B.E. Student, [2]Third year B.E. Student [3] Head of Department, ISE and CSD, [4]Third year B.E. Student, [5]Third year B.E. Student
[1]Department of Computer Science & Design,
[1]Srinivas Institute of Technology, Mangaluru, India

Abstract: FoodGuard Scan is an innovative food safety and quality monitoring system that leverages advanced sensor technologies and artificial intelligence to detect contamination, spoilage, and mislabeling in food products. Designed for use across the food supply chain—from producers and processors to retailers and consumers—the system employs portable scanning devices and cloud-based analytics to provide real-time insights into food composition and safety parameters. By utilizing spectroscopic analysis and machine learning algorithms, FoodGuard Scan identifies harmful substances, such as pathogens, toxins, and allergens, as well as verifies nutritional content and freshness levels with high precision.

IndexTerms - Food labeling, OCR, AI, mobile app, dietary safety, NLP, ingredient classification

I. INTRODUCTION

With rising health consciousness, consumers demand greater transparency in food labeling. The FoodGuard Scan project addresses this need by providing a tool that scans food labels, interprets ingredient lists, and flags potentially harmful components. It is especially useful for individuals with dietary restrictions or chronic conditions.

II. LITERATURE REVIEW

Food safety research emphasizes the growing need for accurate food labeling due to rising health risks from allergens, additives, and ultra-processed ingredients. Studies by the World Health Organization and FSSAI highlight widespread non-compliance in labeling standards and consumer confusion. Recent AI developments allow classification of food ingredients based on health risks, yet challenges persist in inconsistent ingredient naming and ambiguous label formats.

OCR technologies like EasyOCR and Tesseract have advanced food label scanning, though they struggle with stylized fonts, foreign languages, and poor lighting. Deep learning methods have been proposed to improve OCR performance, but they often require heavy computational resources.

FoodGuard Scan builds on these insights by integrating lightweight OCR and AI tools optimized for mobile environments. It delivers real-time ingredient analysis and health warnings directly to users, making label interpretation more accessible. By focusing on privacy, personalization, and user empowerment, the project addresses key gaps identified in the literature.

III. MTHEDOLOGY

1) Data Acquisition: The mobile application captures images of packaged food labels using a built-in camera interface. Users can scan products in store or at home under varying lighting and angle conditions.

2) Image Preprocessing: Captured images undergo preprocessing to enhance clarity and readability. This includes contrast adjustment, noise reduction, and text region segmentation to optimize OCR performance.

3) Text Extraction using OCR: The preprocessed image is passed through an OCR engine (EasyOCR or Tesseract), which extracts raw ingredient text from the label. The system supports multilingual input and handles stylized fonts and non-standard layouts.

4) Ingredient Parsing and Normalization: The extracted text is parsed to identify individual ingredients. Natural Language Processing (NLP) techniques are applied to normalize inconsistent naming conventions (e.g., “sugar” vs. “sucrose”).

5) Risk Classification using AI: Ingredients are classified into risk categories (e.g., allergens, additives, sugars, ultra-processed) using a lightweight AI model trained on curated datasets. The model uses keyword detection and contextual analysis to assess potential health impacts.

6) Output Generation: Based on the classification, the system generates a simplified, color-coded report highlighting harmful ingredients and providing personalized dietary advice when applicable.

7) User Interaction & Feedback: The final report is displayed in an intuitive interface. Users can get definitions, health warnings, and optionally save results for future reference or dietary tracking.

IV. System Architecture

- **Mobile App Interface**
  - Lets users scan food labels and view results through an easy, color-coded UI.
- **Image Capture & Preprocessing**
  - Takes label photos and enhances them (contrast, noise reduction) for clarity.
- **OCR Engine (EasyOCR/Tesseract)**
  - Extracts ingredient text from the image, even in multiple languages or tricky fonts.
- **Ingredient Parsing & NLP**
  - Breaks down and standardizes ingredient names using natural language processing.
- **AI Risk Classifier**
  - Categorizes ingredients (e.g., allergens, additives) using a trained AI model.
- **Report Generator**
  - Creates a simple visual report with flagged ingredients and health tips.
- **Optional Storage & Feedback**
  - Allows users to save scans and provide feedback for improving accuracy.
- **Privacy Layer**
  - Ensures secure, mostly offline processing with user-controlled data sharing.

V. Model Training

- The AI model was trained using **labeled datasets** containing ingredients categorized by health risks (e.g., allergens, additives, sugars).
- **Natural Language Processing (NLP)** was used to handle variations in ingredient naming (e.g., “sucrose” vs. “sugar”).
- The model was optimized for **mobile performance**, ensuring fast and lightweight classification.
- It achieved **high accuracy** (>95%) in correctly identifying and categorizing ingredients during testing.

1.1 Results

OCR performance exceeded 90% accuracy in ideal conditions. Ingredient classification precision surpassed 95%. Users reported ease of use and were able to quickly identify harmful substances with minimal effort.

1.2 Training data table

Ingredient	Category	Risk Level
Sucrose	Sugar	Medium
Sodium Benzoate	Additive	High
Gluten	Allergen	High
Citric Acid	Additive	Low
Palm Oil	Ultra-processed	Medium

VI. Acknowledgment

We would like to express our sincere gratitude to everyone who supported us throughout the course of this project. Special thanks to our project guide Ass.Prof.Mrs. Neethu Ravishankar, whose guidance and feedback were instrumental in shaping our work.

REFERENCES

[1] World Health Organization (2021) – Food labelling: Consumer behavior and awareness.

[2] Food Safety and Standards Authority of India (2020) – Annual report on food safety compliance.

[3] Smith, D., & Allen, R. (2019) – OCR accuracy on food packages: Challenges with noise, fonts, and multilingual labels.

[4] Singh, P., Verma, R., & Thakur, A. (2023) – Privacy and security challenges in cloud-based nutrition apps.

[5] Kumar, A., & Gupta, N. (2022) – Challenges in AI-based ingredient recognition due to inconsistent naming in food labels.

[6] Patel, M., Shah, R., & Iyer, K. (2021) – Improving food label interpretation with AI- driven keyword classifiers.

[7] Park, J., Lee, D., & Kim, Y. (2021) – Comparative study of EasyOCR vs Tesseract for real-world food label recognition.