Enhancing Consumer Safety: Automated Health Risk Identification from Product Labels.

AET 002

Ashutosh Shukla
Department of Electronics and Telecommunication.
Vidyalankar Institute of Technology
Badlapur, India
ashutoshshuklaaaa@gmail.com

Anush Yadav
Department of Electronics and Telecommunication.
Vidyalankar Institute of Technology
Alibag, India
anushyadav135265@gmail.com

Harshad Dhaygude
Department of Electronics and Telecommunication.
Vidyalankar Institute of Technology
Satara, India
harshadhaygude24@gmail.com

Ankit Kumar
Department of Electronics and Telecommunication.
Vidyalankar Institute of Technology
Kharghar, India
ankkumar0204a1@gmail.com

***Abstract*—In today's fast paced world, consumers are lacing the general information about the ingredients and preservatives that are added in the eatables. Many of these ingredients possess significant health risk and potentially harming the vital organs as the time rolls [1]. Yet there is a lack of accessible tools which can address the potential risks and make consumers educated about these ingredients. To address this problem, we came up with web based platform which will make you informed about dietary that leverages Optical character Recognition (OCR) through tesseract which extracts the ingredients from the product label image. Our system uses the extracted ingredients to be analyzed with the curated database which will show ingredients name, ingredients chemical name, is it harmful, short description, affected body part and food sources in the structured format at the output window. We not only provide the solution to the arising problem but also spreading awareness about the rising concern of health risks.**

Keywords—Tesseract OCR, Ingredients, labels, health impact

I. Introduction

The Modern food Industry is heavily relying on chemical additives, preservatives, artificial sweeteners to increase the shelf life, taste and appearance. But despite the awareness the consumers are facing health issues which are affecting the potential risk to the body organs over the time. This chemicals are linked to the health loss and long term diseases. Consumers are struggling to identify the harmful ingredients but due to complex labelling and lack of ingredients information [2].

To bridge this gap between the consumer and increasing health risks we came up with the web based tool to help the consumer understand impact of food ingredients on their health. The ingredients mentioned on the product is extracted by the tesseract OCR and then it is analysed further comparing with the prepared database showing the summary of ingredients present in the product in structured format. The table will consist of ingredients name, ingredients chemical name, is it harmful, short description, affected body part and food sources.

This project was conceived from the personal need to identify the ingredients present in the eatables due to health issue and allergies. But this project is not only restricted to only one user, there are millions of people on this globe facing same problems and that problem can be solved by this platform and spread the awareness about the rising health concerns.

This paper discusses the development, implementation, and potential impact of the Ingredients Analyzer**,** emphasizing the need of the same.

II. Literature Review

A. ***Reference:*** *Bohn, K., Amberg, M., Meier, T., Forner, F., Stangl, G. I., & Mäde, P. (2022). Estimating food ingredient compositions based on mandatory product labeling. Journal of Food Composition and Analysis, 110, 104508*

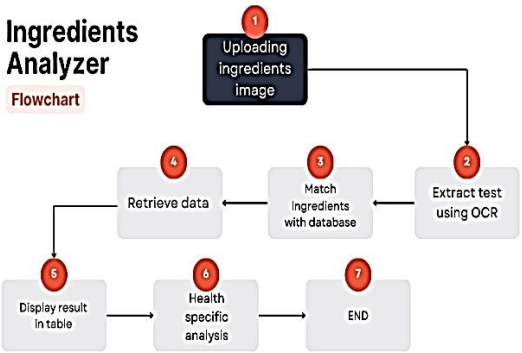
John et al. (2022) proposed an automated optimization-based approach to estimate ingredient compositions from product labels. By parsing ingredient lists and applying mathematical optimization, they achieved 76% accuracy in estimating compositions across a dataset of 1804 products. This method allows for detailed nutritional and ecological evaluations of food products [3].

B. ***Reference:*** *Gautam, H., Singh, V., Kumar, D., & Kaur, S. (2022). Product Ingredient Analysis. International Journal for Research in Applied Science & Engineering Technology, 10(VI), 618-620*

Gautam et al. (2022) developed a product-based recommendation system using natural language processing to analyze cosmetic ingredients. The system visualizes the similarity between products using t-SNE and helps users make informed decisions about cosmetics based on their chemical composition[4].

III. Flow Chart

The following flowchart gives the detailed understanding of the overall Ingredients analyzer project. Starting from uploading the ingredients label image which gets extracted by the OCR. The extracted ingredients are compared to the database retrieving data namely ingredient names, affected body part etc. The analyzed data keeps the user informed about the eatables they are consuming.



IV. METHODOLOGY

1. Image Uploading and Preprocessing.
   - Users upload the product label using flask [2].
   - The systems accept only valid image formats (PNG, JPG, JPEG)

2. Optical Character Recognition.
   - The system employs Tesseract OCR, an open-source tool which extracts the ingredients from the uploaded image.
   - The extracted text is preprocessed, removing unnecessary characters before making the database to search for the same
   - If no ingredients match, then it generates the prompt saying “Ingredient not found”

3. Frontend of website.
   - Making the frontend ready for sharing the image file and user friendly
   - Writing the code for HTML, CSS and JavaScript.

4. Database Integration.
   - The extracted text is processed further to identify ingredients from the label.
   - A MySQL database stores detailed information about ingredients, including:
     - Ingredient Name
     - Chemical Composition
     - Is it harmful? (Yes/ No)
     - Description of its health effects
     - Organs affected
     - Percentage risk factor
     - Common food sources.
   - SQL queries are executed to match extracted ingredients with those which are stored in the database [5].

5. Data Retrieval and Analysis.
   - The system fetches the ingredient details from the database and formats the result in the table format.
   - The table shows all the above mentioned as the columns and different ingredients as the rows of the table.

6. Web interface and Result Display.
   - The frontend developed using HTML, CSS and JavaScript, present the analyzed data in an intuitive format.
   - If there is no match found in the database, the system notifies the user that ingredient yet to be documented [1].

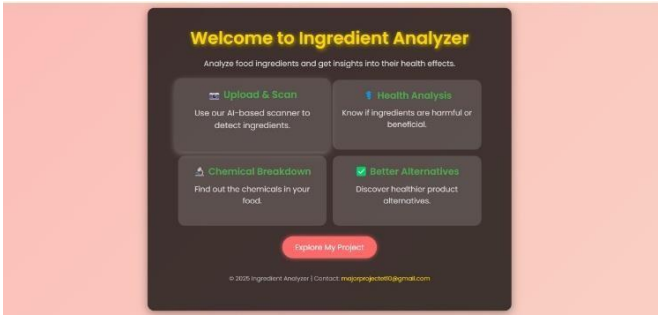
7. Error handling and security.
   - The application implements error handling to manage missing or unclear image inputs.
   - SQL queries are executed securely to prevent SQL injection vulnerabilities.

V. DEMONSTARTION & RESULT

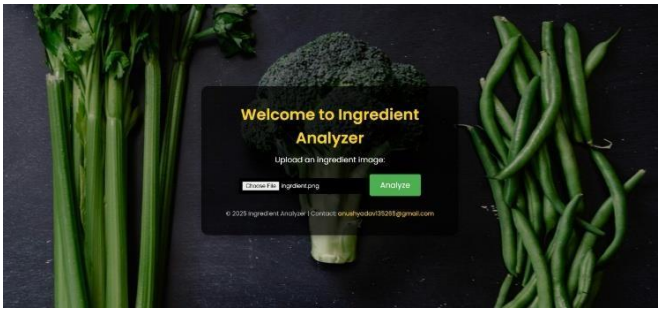
In this part we will showcase how the actual project is working and give the output step by step.


Ingredients: Enriched Corn Meal (Corn Meal, Ferrous Sulfate, Niacin, Thiamin Mononitrate, Riboflavin, Folic Acid), Vegetable Oil (Corn, Canola, and/or Sunflower Oil), Cheese Seasoning (Whey, Cheddar Cheese [Milk, Cheese Cultures, Salt, Enzymes], Canola Oil, Maltodextrin [Made from Corn], Natural and Artificial Flavors, Salt, Whey Protein Concentrate, Monosodium Glutamate, Lactic Acid, Citric Acid, Artificial Color [Yellow 6]),

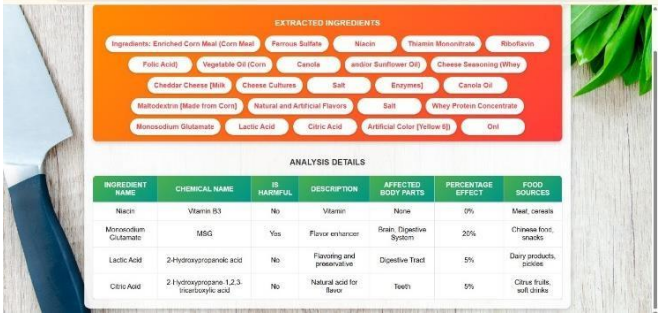
This is the label that we used while doing this project and then this text is extracted and compared[11][12].



The above image shows the very first page of website that has user friendly tabs for more analysis of the ingredients.



Here the user will be able to add the image files, and all the ingredients are extracted and compared with predefined database.



The above image shows two sections viz. Extracted ingredients and Analysis details. The structured table can be seen into the analysis part and can be easily interpreted[10].

What if the ingredient that was extracted is not present in the predefined database, then the system will inform the user that “No matching ingredient found” or “Ingredient not yet defined”[5].

VI. Future Scope

The Ingredients Analyzer can have significant enhancements to improve accuracy, like integration of Machine learning [5] models to understand the toxicity of the ingredients extracted after a certain amount of training. Adding more ingredients to the database means increasing the database will make the system covering large number of products e.g. eatables, cosmetics, skin appliers etc. Adding multi-language [2] OCR to understand ingredients from different languages. Subscription for the dietary plan can be included if the business part is concerned.

VII. Conclusion

The Ingredients Analyzer serves as a powerful tool for promoting transparency in food consumption by identifying potential health risks associated with various food ingredients. By leveraging Tesseract OCR, database-driven analysis, and an intuitive web-based interface, the system effectively extracts and analyzes ingredient lists from product labels. The structured output provides users with essential insights, including ingredient names, chemical compositions, affected organs, and potential harm[7][8][9].

The Ingredients Analyzer bridges the gap between consumers and ingredient transparency, enabling users to proactively assess the impact of food products on their health. This initiative marks a significant step toward fostering health-conscious decision-making in an era where processed and chemically enhanced foods dominate the market [6].

VIII. References

1. D. Gajadur, G. Bekaroo and D. Roopowa, "Investigating the acceptance of Augmented Reality for raising awareness on potentially harmful ingredients present in consumable products," 2020 3rd International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM), Balaclava, Mauritius, 2020, pp. 40-46, doi: 10.1109/ELECOM49001.2020.9297013. keywords: {Sodium;Color;Tools;Market research;Mobile applications;Augmented reality;Diseases;ProdiScan;Augmented Reality;Health;Non- communicable Diseases;Ingredients;Consumable Products},

2. M. Kartiwi, T. S. Gunawan, A. Anwar and S. S. Fathurohmah, "Mobile Application for Halal Food Ingredients Identification using Optical Character Recognition," 2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), Songkhla, Thailand, 2018, pp. 1-4, doi: 10.1109/ICSIMA.2018.8688756. keywords: {halaal;ingredients;food;mobile application;text recognition},

3. **Reference:** Bohn, K., Amberg, M., Meier, T., Forner, F., Stangl, G. I., & Mäde, P. (2022). Estimating food ingredient compositions based on mandatory product labeling. Journal of Food Composition and Analysis, 110, 104508

4. **Reference:** Gautam, H., Singh, V., Kumar, D., & Kaur, S. (2022). Product Ingredient Analysis. International Journal for Research in Applied Science & Engineering Technology, 10(VI), 618-620

5. V. C P, A. D, D. D. Kedilaya, S. S. Gondkar and S. Halhalli, "Online Analysis of Ingredient Safety, Leveraging OCR and Machine Learning for Enhanced Consumer Product Safety," 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), Namakkal, India, 2024, pp. 1-6, doi:

10.1109/AIMLA59606.2024.10531558. keywords: {Image processing;Optical character recognition;Decision making;Consumer products;Machine learning;Natural language processing;Product safety;Ingredient analysis;consumer safety;OCR;NLP;informed choices;responsible consumerism},

6. M. Daud, J. A. Jalil, I. M. A. G. Azmi, S. F. Ismail and S. Safuan, "'Unsafe' nutraceuticals products on the Internet: The need for stricter regulation in Malaysia," 2017 5th International Conference on Cyber and IT Service Management (CITSM), Denpasar, Indonesia, 2017, pp. 1-5, doi: 10.1109/CITSM.2017.8089269. keywords: {Drugs;Social network services;Monitoring;Safety;Production;Diseases;

7. Vaclavik, Vickie A., Elizabeth W. Christian, and Tad Campbell. Essentials of food science. Vol. 42. New York: Springer, 2008.

8. Tantibanchachai, Chanapa. Teratogens. Arizona State University. School of Life Sciences. Center for Biology and Society. Embryo Project Encyclopedia.| Arizona Board of Regents, 2014.

9. Winter, Ruth. A consumer's dictionary of food additives: Descriptions in plain English of more than 12,000 ingredients both harmful and desirable found in foods. Crown Archetype, 2009.

10. Rajadurai, M., and P. Stanely Mainzen Prince. "Preventive effect of naringin on lipid peroxides and antioxidants in isoproterenol-induced cardiotoxicity in Wistar rats: biochemical and histopathological evidences." Toxicology 228.2-3 (2006): 259-268.

11. Panduri, Vijayalakshmi, Sigmund A. Weitzman, Navdeep S. Chandel, and David W. Kamp. "Mitochondrial-derived free radicals mediate asbestos-induced alveolar epithelial cell apoptosis." American Journal of Physiology-Lung Cellular and Molecular Physiology 286, no. 6 (2004): L1220-L1227.

12. Mandel, Dror, Yoav Littner, Francis B. Mimouni, Zahava Stavarovsky, and Shaul Dollberg. "Increased serum potassium and intraventricular hemorrhage revisited." The Israel Medical Association Journal: IMAJ 6, no. 2 (2004): 91-94.