Personalized Allergen Recommender System: Ingredient List Scanning for Safe Food Choices

W.G.S. Subashini[1], and A.R.F. Shafana[2]

[1,2]Department of Information and Communication Technology, South Eastern University of Sri Lanka, Sri Lanka

[1]114sudheera@gamil.com, [2]arfshafana@seu.ac.lk

Abstract

Understanding every ingredient in a food is crucial for everyone who has food allergies and consumes packaged foods. However, the component or ingredient lists are unable to recognized by only reading due to the lack of education related to coding conventions among the public. Despite the availability of certain allergen recommendation applications that might aid in pinpointing ingredients, the available applications are not personalized and are not supported efficiently. This research aims to develop a personalized allergen recommender system that helps with safe food selections through ingredient list scanning. The final deliverable is a mobile application that scans the list of ingredients, classifies allergen ingredients for the consumer, and provides allergen recommendations to the user. The application can identify high-risk and the most common 25 food ingredients that can cause allergic. The Tesseract OCR has been used to extract text from the captured image of ingredients. After the text is extracted, the application compares it with a user-predefined list of allergens and based on the result provides the recommendation to the user. The developed application was able to successfully extract ingredient text and accurately identify allergenic ingredients in real-time. This study demonstrates that a AI -enhanced personalized allergen recommender system can improve user safety and awareness, supporting informed decisions about food products.

Keywords: Allergen Free Food, Allergen Food Recommendation Systems, Food Allergies, Mobile Application, Tesseract

I. INTRODUCTION

Food allergy (FA) is one of the most common issues all around the world that can lead to both major and minor health problems. FA is an exaggerated or hypersensitive immune response to substances that are generally not harmful (Jain and Zachariah, 2022). It is any negative reaction to food that is immunologically mediated and repeatable in a blinded environment (De Martinis et al., 2020; Roberts and Lack, 2003). The allergic reaction can be different from one person to other person and leads to various kinds of symptoms such as dermatitis, asthma, rhinitis, and potentially severe, fatal responses like anaphylaxis, insect sensitivity, animal danger, food, and pollen, which may result in allergies. Individuals with food allergies must carefully evaluate food ingredients to avoid potential allergic reactions.

Currently, food allergy has become a significant public health concern due to its increasing prevalence, affecting approximately 220 million people worldwide (De Martinis et al., 2020). Food allergies tend to have a greater impact on children than on adults. Children are particularly vulnerable due to factors such as their immature immune and digestive systems, lower body weight, potential for long-term complications, distinct behavioral traits, limited communication abilities, and heightened susceptibility to allergens and sensitivities (Genuneit et al., 2017). The prevalence of food allergies among children is notably higher, estimated at around 10% compared to 1%–2% in adults (Mandracchia et al., 2020). Therefore, it is crucial for both children and adults to remain vigilant about avoiding allergenic food ingredients. One effective preventive strategy is the accurate identification of food allergens, which helps minimize the risk of exposure to substances that may trigger allergic reactions (Anvari et al., 2018; Jain and Zachariah, 2022).

Consequently, understanding every ingredient in food products is essential for individuals with food allergies (FA), particularly when consuming packaged foods (Wong et al., 2016). Individuals with allergies and other health conditions tend to be especially vigilant about the ingredients they consume. This concern is heightened by the widespread consumption of packaged snacks and processed foods, which are particularly common in the diets of both children and adults (Gearhardt & Hebebrand, 2021). According to international standards in the food industry, all food packaging must clearly list the ingredients used, including food additives such as colorings and flavorings in order to ensure consumer safety and informed decision-making. Reading the ingredient list or the bar code that is available at the front or back of the packaging is a common approach to obtaining the list of ingredients (Ni Mhurchu et al., 2018). The World Health Organization (WHO) has developed the Codex Alimentarius (commonly referred to as CODEX), a comprehensive system of standards for the labeling and naming of food ingredients. According to the Codex Alimentarius Committee, some European countries utilize E numbers, while others adopt the International Numbering System (INS) for food additives (Cox et al., 2020; Holleman et al., 2021; Stolte et al., 2013). Most food companies adopt these codes as a standardized method for listing base ingredients in their products.

However, the public often struggles to interpret these codes due to a lack of awareness and education regarding the coding conventions. In the challenging context of identifying allergens and understanding food ingredients in packaged foods, numerous computational research efforts have sought digital solutions to address this issue. One commonly used technological approach is Optical Character Recognition (OCR), which enables the extraction of text from images and is employed by several existing mobile applications (Jain & Zachariah, 2022). Additionally, some applications utilize deep learning techniques to identify food items (Zhou et al., 2019). However, these systems do not specifically identify allergenic components in food products, nor do they provide allergen recommendations personalized to individual users.

The objective of this study is to design and develop a mobile application capable of identifying allergenic ingredients by scanning food ingredient lists. Specifically, the study aims to create a personalized allergen recommender system that facilitates safe food choices through automated analysis of ingredient information. The final deliverable was a mobile application capable of scanning ingredient lists and classifying potential allergens based on user-specific input. The system provides personalized recommendations by considering the user's individual allergen profile. The study primarily targeted the identification of the 25 most common and high-risk food allergens, using ingredient lists presented in English and conforming to international labeling standards defined by the Codex Alimentarius.

II. RELATED WORKS

In recent years, various technologies have been employed to develop food-related recommendation systems, utilizing machine learning, deep learning, and other computational methods to deliver personalized dietary insights, food identification, and health-related guidance. Alemany-Bordera et al. (2016) developed a web-based application using machine learning to recommend Spanish recipes personalized to user preferences. However, the system's recommendations are not based on specific food ingredients. Senapati et al (2025) built a computer- based system for food allergy detection that uses the ResNet-50 model, adapted to the Food-101 dataset, to recognize food types, validate labels, and provide nutritional information. Similarly, Rostami et al. (2022) introduced a deep learning-based dietary recommendation platform focused on diet management. Although the platform is effective in delivering user-specific meal suggestions, it does not incorporate ingredient-level personalization.

Jane et al. (2022) proposed another web-based system that assists users in identifying restaurants catering to their allergen requirements. Although helpful for individuals with food allergies, the system lacks functionality for personalized, ingredient-specific recommendations. In another study, Swain et al. (2023) developed a mobile application employing YOLO-based object detection and Convolutional Neural Networks (CNNs) to identify ingredients from food images and recommend recipes. While innovative, the

ISBN 978-955-627-161-4

system does not offer user-specific suggestions or account for allergen sensitivities.

Additionally, Gearhardt and Hebebrand (2021) explored the use of Optical Character Recognition (OCR) for ingredient analysis in a mobile application, providing general health suggestions. However, similar to other approaches, this system does not deliver personalized allergen recommendations. In 2021, Rohini et al. designed framework to identify fruits and packed food that can be include allergen nutrients by using OCR and deep learning. Across the reviewed literature, a significant gap exists that despite the availability of various systems to support food identification and health-based recommendations, generally a few offers personalized allergen detection and ingredient-specific dietary guidance personalized to individual user profiles.

III. METHODOLOGY

This section outlines the methodology for developing a personalized allergen recommendation system. The application is designed to operate on Android and iOS platforms, offering users a seamless experience in identifying potential allergens. The system development process follows several key phases, illustrated in Figure 01.

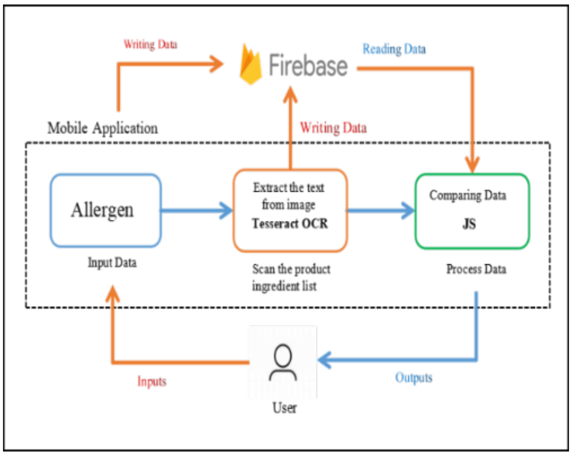

Figure 01: System Development Process

A. Text Detection

A critical step in this study involved extracting textual data from food labels to identify potential allergenic ingredients. The text extraction process was implemented using the Tesseract Optical Character Recognition (OCR) engine and followed a structured sequence of stages, including image input handling, preprocessing, and character recognition.

The process began when the user scanned a food label using the application's built-in camera functionality, which was implemented via the Expo ImagePicker module in React Native. The captured image was then passed to the OCR engine for further processing. Tesseract.js, a JavaScript wrapper for the Tesseract OCR engine, was integrated into the mobile application to enable real-time text recognition.

To ensure accurate recognition, Tesseract applied internal preprocessing techniques such as image resizing, brightness adjustment, grayscale conversion, and normalization. These processes enhanced image clarity and improved character detection performance.

The OCR engine utilized trained data models, specifically the eng.traineddata file for English, in order to accurately interpret characters. During recognition, Tesseract employed multiple OCR algorithms to analyze and convert visual text from the image into a structured string format. This extracted text was then forwarded to the next stage of processing for allergen identification and classification.

B. Text comparing and Generating Recommendation

To compare text, the application retrieved stored allergen data from Firebase, where allergen information was maintained as documents containing arrays of allergen terms. Both the extracted text from the Tesseract OCR engine and the allergen data from Firebase underwent normalization. This process included converting all text to lowercase and removing special characters and spaces to ensure consistency in the comparison.

A string-matching algorithm was applied, in which each allergen from the stored list was checked against the extracted text. A dynamic regular expression (regex) was constructed for each allergen keyword to ensure accurate detection of allergen terms within the text. For example, a regex pattern was used to identify "peanut" as a standalone word, avoiding false positives such as "peanut butter."

ISBN 978-955-627-161-4

Detected allergens were then stored in a set to prevent duplication and were displayed to the user. This approach ensured that the extracted ingredient list was systematically compared with the stored allergen data, providing users with timely feedback on potential food allergens.

C. Database Management

Firebase was utilized in this project to manage user authentication and to store allergy-related information specific to each user. The implementation included the integration of Firebase Authentication, which securely handled user login and registration processes. This ensured that users could access personalized features of the application in a secure manner. Additionally, Firebase functioned as the primary database for storing and retrieving allergenic ingredient data. This allowed the application to provide personalized allergen detection services based on individual user profiles. The use of Firebase enabled efficient and secure storage, real-time data updates, and rapid query handling for allergen-related information.

D. Mobile Application Design and Implementation

The User Interface (UI) was developed using React Native components to ensure a seamless cross-platform experience. The UI layer was designed to be intuitive and user-friendly, enabling users to navigate the application and perform necessary actions with minimal effort. By employing React Native components along with Bootstrap styling, the interface achieved visual consistency and responsiveness across various device platforms, thereby enhancing the overall user experience.

Images were captured either through the camera or selected from the gallery of the device. This dual-option functionality allowed users to either take a real-time image or choose an existing one from their photo library. The captured or selected image was then processed using the Tesseract OCR engine for text extraction. The Tesseract library was integrated into the application and configured with the eng.traineddata file to support accurate recognition of English text.

Allergic ingredient data entered by users was stored in Firebase and retrieved during the scanning process. The extracted text from the image was compared with the stored allergen data to identify potential matches. This comparison allowed the system to detect and highlight allergenic ingredients present in the scanned food label. If a match was found, the application displayed an alert message, notifying the user of the presence of specific allergens. This feature ensured that users were promptly informed about potential allergenic content in food products, promoting safer consumption decisions.

IV. RESULTS AND DISCUSSION

The development and integration of these features resulted in an application that effectively meets the needs of a personalized food allergen recommendation system. It successfully fulfilled the primary goal of the study.

The application design was evaluated against established design guidelines to ensure adherence to best practices in usability, including principles such as consistency and standards, visibility of system status, user control and freedom, error prevention, and simplicity and intuitive navigation, consistency, and responsiveness across devices. This preliminary evaluation confirmed that the system's interface and core functionalities, characterized by a clean and simple layout, consistent design language, and responsive design align well with recognized heuristics for user-friendly design. The testing also highlighted that the Optical Character Recognition (OCR) functionality, integrated with Firebase for real-time processing, performed reliably under varying conditions, enhancing user confidence in the allergen detection feature (See Figure 02 and Figure 03).

In addition, Firebase provided a seamless mechanism to store and retrieve both user-specific data, such as allergen preferences, and the global allergen database, ensuring that users received up-to-date information regardless of the device used. Any changes to allergen data or user settings were synchronized in real time across the application, enhancing the system's responsiveness and reliability.

As a future work, formal usability testing involving actual users will be conducted to assess the effectiveness, efficiency, and user satisfaction of the application in real-world scenarios. In addition, the future work focus on refining OCR capabilities to improve text extraction accuracy and expanding the database of allergens to cover a wider range of products. These future efforts aim to enhance the application's overall performance and user experience, ensuring it becomes a more reliable and comprehensive tool for allergen detection and food safety.

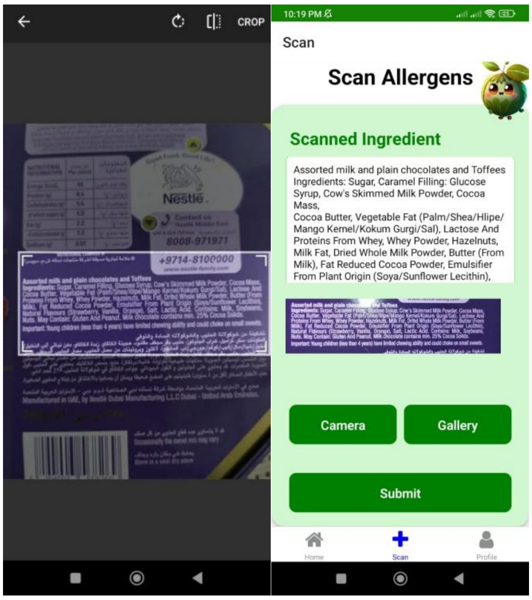

Figure 02: Interface scanning ingredients

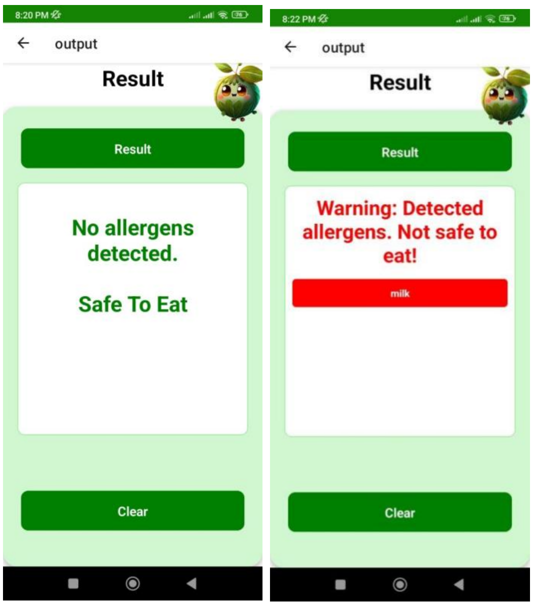

Figure 03: Interface displaying results

V. CONCLUSION

This study demonstrates the development and application of a personalized food allergy recommendation system that uses advanced technologies to enhance dietary safety. By enabling users to scan food labels and receive real-time allergen alerts based on their individual profiles, the application supports informed and safer food choices. Its user-friendly design ensures accessibility across varying levels of technical proficiency, facilitating easy navigation of key features such as allergen scanning, profile management, and personalized recommendations.

However, this research has certain limitations. The accuracy of the OCR depends heavily on image quality, font variations, and label clarity, which can affect to correct allergen detection. The allergen database used in this study was limited to 25 common allergens and may not cover region-specific or newly emerging allergens. In addition, testing was carried out in controlled conditions, and broader real-world validation with diverse food products through a usability analysis is still required.

Future work will focus on improving the accuracy of OCR and expanding the allergen database to include a broader spectrum of products, thereby increasing the system's reliability and coverage. This research highlights the significant potential of technology to address public health challenges associated with food allergies, offering a scalable solution to empower users in managing their dietary risks. Continued development and user-centered enhancements will further establish the application as an essential tool in promoting safe eating practices for individuals with food allergies.

REFERENCES

Alemany-Bordera, J., María, S., Camara, J. and Inglada, V. (2016) 'Bargaining agents-based system for automatic classification of potential allergens in recipes', ADCAIJ, 5(2), pp. 43–51. doi:10.14201/adcaij2016524351.

Anvari, S., Miller, J., Yeh, C.Y. and Davis, C.M. (2018) 'IgE-mediated food allergy', Clinical Reviews in Allergy & Immunology, 57(2), pp. 244–260. doi:10.1007/s12016-018-8710-3.

Senapati, B., Chapra, N., Rao, D.D., Anand, M.R., Das, D., Goldar, S.S.,and Tiwari, S. (2025) 'Transfer learning approach to identify food allergy', International Journal of Intelligent Systems and Applications in Engineering, pp. 2414-2421

Cox, S., Sandall, A., Smith, L., Rossi, M. and Whelan, K. (2020) 'Food additive emulsifiers: a review of their role in foods, legislation and classifications, presence in food supply, dietary exposure, and safety assessment', Nutrition Reviews, 79(6). doi:10.1093/nutrit/nuaa038.

De Martinis, M., Sirufo, M.M., Suppa, M. and Ginaldi, L. (2020) 'New perspectives in food allergy', International Journal of Molecular Sciences, 21(4), p. 1474. doi:10.3390/ijms21041474.

Gearhardt, A.N. and Hebebrand, J. (2021) 'The concept of "food addiction" helps inform the understanding of overeating and obesity: YES', American Journal of Clinical Nutrition, 113(2), pp. 263–267. doi:10.1093/ajcn/nqaa343.

Genuneit, J. et al. (2017) 'Overview of systematic reviews in allergy epidemiology', Allergy, 72(6), pp. 849–856. doi:10.1111/all.13123.

Holleman, B.C. et al. (2021) 'Poor understanding of allergen labelling by allergic and non-allergic consumers', Clinical & Experimental Allergy, 51(10), pp. 1374–1382. doi:10.1111/cea.13975.

Jain, J.P. and Zachariah, U.E. (2022) 'Allergen-free food dataset and comparative analysis of recommendation algorithms', IEEE ICMACC. doi:10.1109/icmacc54824.2022.10093402.

Jane, M., De, G., Ruiz, C., Sunga, C.F. and Samonte, D. (2022) 'AlleRT: food recommender web application with allergy filtration'. doi:10.46254/eu05.20220370.

Mandracchia, F., Llaurado, E., Tarro, L., Valls, R. and Solà, R. (2020) 'Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the mobile app rating scale (MARS)', JMIR mHealth and uHealth, 8(9), p. e18339. doi:10.2196/18339.

Ni Mhurchu, C., Eyles, H., Jiang, Y. and Blakely, T. (2018) 'Do nutrition labels influence healthier food choices? Analysis of label viewing behaviour and subsequent food purchases in a labelling intervention trial', Appetite, 121, pp. 360–365. doi:10.1016/j.appet.2017.11.105.

Roberts, G. and Lack, G. (2003) 'Food allergy and asthma—what is the link?', Paediatric Respiratory Reviews, 4(3), pp. 205–212. doi:10.1016/S1526-0542(03)00058-7.

Rohini, B., Pavuluri, D.M., Naresh Kumar, L., Soorya, V. and Aravinth, J. (2021) 'A framework to identify allergen and nutrient content in fruits and packaged food using deep learning and OCR', Proceedings of the 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, pp. 72–77. doi:10.1109/ICACCS51430.2021.9441800.

Rostami, M., Oussalah, M. and Farrahi, V. (2022) 'A novel time-aware food recommender-system based on deep learning and graph clustering', IEEE Access, 10, pp. 52508–52524. doi:10.1109/access.2022.3175317.

Stolte, S., Steudte, S., Schebb, N.H., Willenberg, I. and Stepnowski, P. (2013) 'Ecotoxicity of artificial sweeteners and stevioside', Environment International, 60, pp. 123–127. doi:10.1016/j.envint.2013.08.010.

Swain, M., Manyatha, A.R., Dinesh, A.S., Sampatrao, G.S. and Soni, M. (2023) 'Ingredients to recipe: a YOLO-based object detector and recommendation system via clustering approach', IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/10073769

Wong, M., Ye, Q., Kai, Y., Pang, W.M. and Kwan, K.M. (2016) 'A mobile adviser of healthy eating by reading ingredient labels'. doi:10.1007/978-3-319-58877-3_4.

Zhou, L., Zhang, C., Liu, F., Qiu, Z. and He, Y. (2019) 'Application of deep learning in food: a review', Comprehensive Reviews in Food Science and Food Safety, 18(6), pp. 1793–1811. doi:10.1111/1541-4337.12492.