Allergen detection based on food packaged products for eczema patients using optical character recognition method

Dian Palupi Rini[1], Riska Tri Mardilah[1], Muhammad Qurhanul Rizqie[1], Dwi Rosa Indah[2], Alvi Syahrini Utami[1], Jovanic Morgan[1]
[1]Department of Informatics, Faculty of Computer Science, Sriwijaya University, Palembang, Indonesia
[2]Department of Information System, Faculty of Computer Science, Sriwijaya University, Palembang, Indonesia

Article Info

Article history:

Received Dec 18, 2024
Revised Apr 5, 2026
Accepted Apr 18, 2026

Keywords:

Allergen
Detection
Eczema
Food
Optical character recognition

ABSTRACT

Eczema, also known as dermatitis, is a chronic skin condition that causes recurring episodes of dry and itchy skin. It can be managed through medication and by avoiding triggers like stress and certain foods. To help patients avoid food-related triggers, researchers conducted a study to detect allergenic food compositions in packaged products using optical character recognition (OCR) techniques, specifically open computer vision (OpenCV) and Tesseract. The study involved analyzing 120 images of food labels. The process included several steps: preprocessing the images by converting them to a text-friendly format (gray scaling, denoising, and thresholding), using Tesseract for text detection, followed by case folding and tokenization. The results showed that the system achieved an average text detection accuracy of 61.88% and an average allergen detection accuracy of 83.06%. The highest accuracy for text detection was 78.52%, and the highest accuracy for allergen detection was 100%. These findings suggest that OCR techniques can be a useful tool for helping eczema patients manage their diet and minimize flare-ups.

Corresponding Author:

Dian Palupi Rini
Department of Informatics, Faculty of Computer Science, Sriwijaya University
Palembang, Indonesia
Email: dprini@unsri.ac.id

1. INTRODUCTION

Eczema or dermatitis is a chronic skin disease characterized by recurring symptoms of dry and itchy skin [1], [2]. Eczema remains a health problem worldwide, where childhood and adults' prevalence of eczema has been increasing worldwide. The estimated prevalence of Atopic Dermatitis in Indonesia reaches around 10% of the total population, with a higher incidence rate among children and individuals in productive age [3]. The exact cause of this condition is still not fully understood [4]. Eczema itself cannot be cured but can be managed with medication and by eliminating triggering factors such as stress and certain foods [5].

Food plays a crucial role in human life as a daily essential need. However, food itself can become a boomerang for eczema patients if not properly controlled [3]. The World Health Organization (WHO) states that food is a primary necessity that must be consistently fulfilled by humans wherever they are. Additionally, proper and adequate food management is highly necessary to ensure that the consumed food provides optimal benefits to the body [6]. Therefore, to eliminate triggering allergens in food, eczema patients must pay attention to the composition of their meals especially in packaged food products [7], [8]. Thus, the detection of triggering food allergens, particularly in eczema patients, needs to be done to avoid any resulting consequences.

The topic of detecting food composition in packaged food products has been the subject of research by several scholars. These researchers have employed optical character recognition (OCR) and utilized the

Tesseract and Leptonica libraries [9]–[11]. One notable example is the development of the OCR-based safety check system of packaged food for food inconvenience patients. In developing this system, a series of processes were carried out, beginning with text recognition and concluding with data matching. The final result is a match of food ingredients with the database, which then displays a notification indicating whether the food item is safe or unsafe for the user's diet [12]. The Tesseract library was employed in the development of the Implementation of the you only look once (YOLO) algorithm and Tesseract OCR in the automatic number plate detection system, as outlined by [13]. The system utilizes 700 datasets, achieving 100% testing accuracy under conditions of sufficient lighting and a threshold of 0.5 [13]. Additionally, related research that employs OCR includes the application to provide the nutrient and allergen content in fruits and vegetables, to display allergen information in packaged food using OCR [11]. Another research is identifying ingredients of a consumer's products that define three food category entities in NER: *Halal, Haram,* and *Subhat* (doubtful). The system is built using OCR to scan the composition listed on packaged products and processed with the trained NER model [10].

Therefore, a system for detecting allergens for eczema patients is built using open computer vision (OpenCV) and Tesseract to eliminate allergy-triggering factors in packaged food products. This system reads the food composition from product packaging and searches for allergenic ingredients that can trigger allergic reactions in eczema patients. It is hoped that this system will assist eczema patients in managing and eliminating potential triggers for eczema flare-ups in food by quickly and accurately detecting allergens, thus improving their quality of life through safe food choices. Additionally, this technology acknowledges the specific needs of individuals with allergic conditions, demonstrating a commitment to supporting their overall health and well-being. As a scope of this research, the detection of food product uses Indonesian language, and this is only detection for eczema dermatitis atopic. There are some literatures that use in this research.

Atopic eczema is a chronic inflammation of the skin that is presented in Figure 1. This inflammation will appear if it meets one of its triggers or stimulants. Atopic eczema usually begins in childhood and often continues into adulthood [14]. This disease is characterized by recurrent itchy lesions, wounds, skin thickening, and dry skin which follows a chronic relapsing and remitting course that causing discomfort for those affected. Atopic eczema has significant impact on children's quality of life due to itching, scratching, emotional distress, and sleep disturbance.

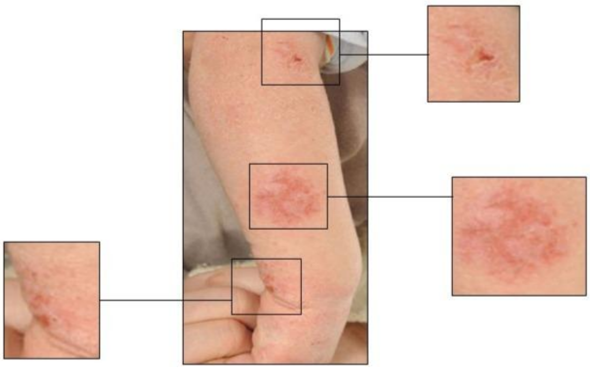


Figure 1. Child with atopic eczema [15]

Atopic eczema is an inherent predisposition where the body produces immunoglobulin E (IgE) antibodies in response to a small amount of common environmental proteins, such as pollen, dust mites, and food [16]. It is also affected by peripheral eosinophilia, and other allergic diseases [17].

Some foods that are triggers for atopic dermatitis flare-ups [18] include: cloves, custard, diacetyl, traditional Indonesian buffalo milk yogurt, wheat, ghee (clarified butter), gluten, soy sauce, cheese, soybean, nuts/beans, legumes, peanuts, casein, cream, lactalbumin, lactoferrin, lactose, lactulose, butter, citrus, milk, tomato, wheat flour, wheat, vanilla, and yogurt.

OCR is the science that possible to translate various types of documents or images into data that can be edited, analyzed and searched [19]. This is a technology that aims to identify and recognize text contained within images. OCR is intended to convert images containing written text into text data that can be read by a system.

OCR utilizes scanning technology to transform physical documents into accessible and editable digital text. After scanning, the software analyzes the image, identifying characters that require recognition. This recognition is achieved using two methods: pattern recognition, which compares examples of text to identify characters, and feature recognition, employing specific rules based on characteristics like slanted lines or curves in letters or numbers [20].

Additionally, OCR software analyzes the document's structure, dividing it into elements such as text blocks, tables, or images. Once the characters are recognized and processed, the software displays the extracted text, allowing users to easily access and edit the document without manual data entry [21]. This automation streamlines the conversion of text-based documents to digital formats, enhancing efficiency and convenience in working with digital materials.

PyTesseract is a Python library that serves as a tool for OCR, enabling the identification and reading of text within images. Tesseract can recognize various image formats such as jpeg, png, gif, bmp, tiff, and others. The output from Tesseract is the displayed text that has been recognized or read from the image [22], [23].

OpenCV is an open-source library developed by Intel, aimed at simplifying programming related to real time digital image processing. OpenCV offers a wide range of features, including identify edges, objects, people, face recognition, face tracking, face detection, Kalman filtering, and various artificial intelligence methods. Additionally, OpenCV provides various simple algorithms related to computer vision for low-level application programming interfaces (APIs) [24].

The natural language toolkit (NLTK) is a platform for Python programming that operates with human language data. It provides an easy-to-use interface and offers text processing libraries that can be used for tasks such as classification, tokenization, stemming, tagging, parsing, and semantic reasoning [25].

2. METHOD

2.1. Data collection

In this study, two types of data were used, namely primary data and secondary data. The collection of primary data involved gathering images of food composition information from packaged products obtained from LOTTE Grosir Mart Palembang, totaling 120 images. On the other hand, secondary data was collected based on the journal titled "Atopic dermatitis and food allergy: a complex interplay: what we know and what we would like to learn" [18] from the National Center for Biotechnology Information. This secondary data was then summarized and used as a reference for the development of the system. A data matching process is performed between the extracted primary data and the secondary data. This process is carried out after the pre-processing and text recognition stages on product images. The matching mechanism aims to verify whether the food ingredients contained in packaged products are safe for eczema sufferers.

2.2. Data processing

The stages of analysis in the process of developing and manually calculating accuracy in the system using OpenCV and Tesseract described in Figure 2 are discussed as follows:

At this juncture, a series of preparatory steps are undertaken to facilitate the subsequent word detection process, which is conducted using the OpenCV software. Among the processes is the conversion of the image from RGB to grayscale, which is intended to facilitate subsequent processing. Moreover, the grayscaling stage involves the removal of noise or disturbing pixels from the image. This stage is referred to as denoising. Subsequent to the completion of the denoising stage, the final stage in the data pre-processing sequence is thresholding. In this thresholding stage, the grey image derived from the previous grayscaling and denoising stages is transformed into a binary image, which is a black and white image, with the objective of facilitating the system's ability to identify words.

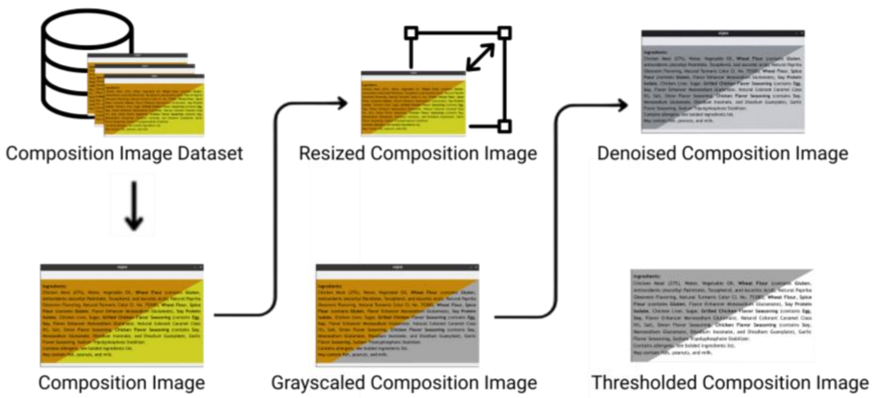


Figure 2. Image preprocessing

2.3. Main processing data

The main data processing stage is presented in Figure 3. The word detection process is conducted using the PyTesseract library. At this stage, the text on the image, which has been prepared from the previous pre-processing stage, is identified through the utilization of the image-to-data function.

Figure 3. Main process of image

2.4. Post-processing data

Post-processing data presented in Figure 4. It describes the process of word filtering is initiated and subsequently followed by tokenization utilizing the NLTK. Once the data extracted from the image has been processed, a comparison is conducted between the data and a list of allergenic food ingredients for individuals with AD.

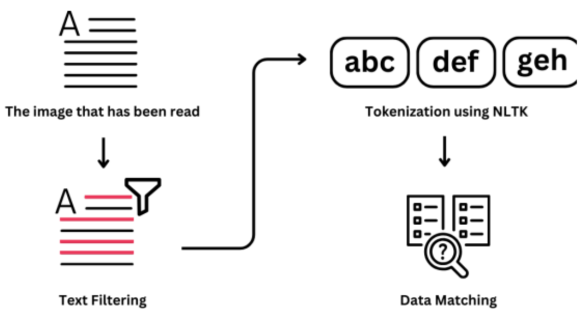


Figure 4. Post-processing image

2.5. Testing

During the testing stage, described in Figure 5, the system analyzes the image dataset stored in the local folder. The goal of this analysis is to evaluate two key metrics: the accuracy of the extracted text and the effectiveness of the data matching process.

Figure 5. Process of dataset testing

The details process are defined below:
[Step 1] Accessing local files to obtain images. An example of accessing local files is illustrated in "Figure 6".
[Step 2] Performing the preprocessing stages, namely gray scaling, denoising, and thresholding. Examples of before and after gray scaling, denoising, and thresholding are presented in "Figures 7-10".

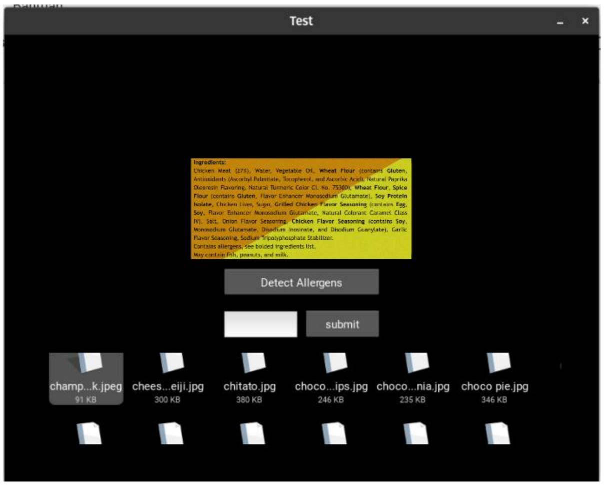


Figure 6. Accessing local files

Figure 7. Champ chicken stick original image

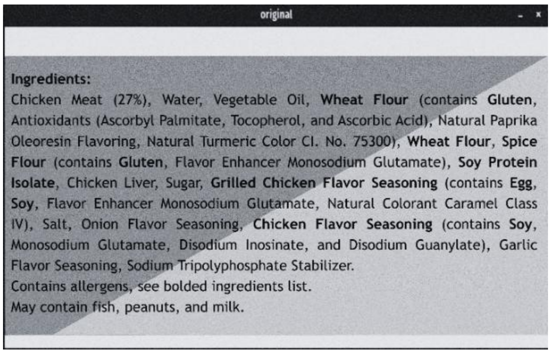


Figure 8. Champ chicken stick grayscale image

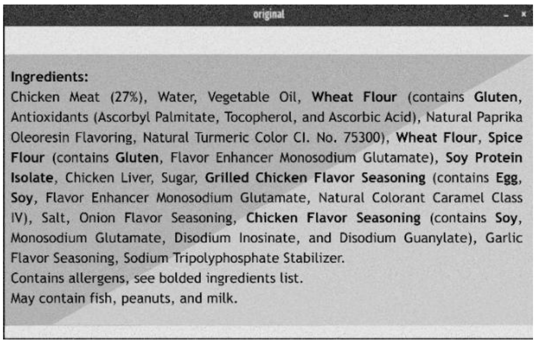


Figure 9. Champ chicken stick denoised image

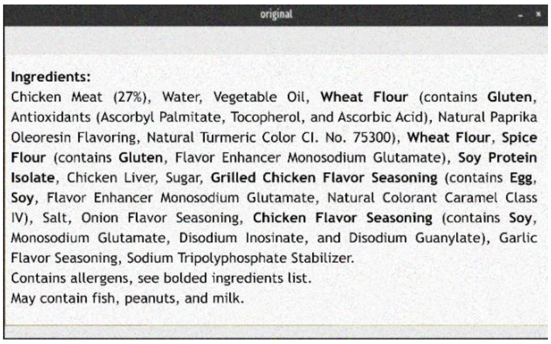


Figure 10. Champ chicken stick threshold image

[Step 3] Performing text detection on the image using Tesseract. The result of text detection on the Champ Chicken Stick image displayed in the file recognized.txt is as follows:

```
Chicken Meat (27%), Water, Vegetable Oil, Wheat Flour (contains Gluten, Antioxidants (Ascorbyl
Palmitate, Tocopherol, and Ascorbic Acid), Natural Paprika Oleoresin Flavoring, Natural
Turmeric Color CI. No. 75300), Wheat Flour, Spice Flour (contains Gluten, Flavor Enhancer
Monosodium Glutamate), Soy Protein Isolate, Chicken Liver, Sugar, Grilled Chicken Flavor
Seasoning (contains Egg, Soy, Flavor Enhancer Monosodium Glutamate, Natural Colorant Caramel
Class IV), Salt, Onion Flavor Seasoning, Chicken Flavor Seasoning (contains Soy, Monosodium
Glutamate, Disodium Inosinate, and Disodium Guanylate), Garlic Flavor Seasoning, Sodium
Tripolyphosphate Stabilizer. Contains allergens, see bolded ingredients list. May contain
fish, peanuts, and milk.
```

[Step 4] Performing the case folding stage by converting words to lowercase. The result of case folding on the Champ Chicken Stick image displayed in the file recognized.txt is as follows:

```
chicken meat 27 water vegetable oil wheat flour contains gluten antioxidants ascorbyl palmitate tocopherol and ascorbic acid natural paprika oleoresin flavoring natural turmeric color ci no 75300 wheat flour spice flour contains gluten flavor enhancer monosodium glutamate soy protein isolate chicken liver sugar grilled chicken flavor seasoning contains egg soy flavor enhancer monosodium glutamate natural colorant caramel class iv salt onion flavor seasoning chicken flavor seasoning contains soy monosodium glutamate disodium inosinate and disodium guanylate garlic flavor seasoning sodium tripolyphosphate stabilizer contains allergens see bolded ingredients list may contain fish peanuts and milk
```

[Step 5] Performing the tokenization stage. The result of tokenization on the Champ Chicken Stick image is as follows:

```
["chicken", "meat", "27", "water", "vegetable", "oil", "wheat", "flour", "contains", "gluten", "antioxidants", "ascorbyl", "palmitate", "tocopherol", "and", "ascorbic", "acid", "natural", "paprika", "oleoresin", "flavoring", "natural", "turmeric", "color", "ci", "no", "75300", "wheat", "flour", "spice", "flour", "contains", "gluten", "flavor", "enhancer", "monosodium", "glutamate", "soy", "protein", "isolate", "chicken", "liver", "sugar", "grilled", "chicken", "flavor", "seasoning", "contains", "egg", "soy", "flavor", "enhancer", "monosodium", "glutamate", "natural", "colorant", "caramel", "class", "iv", "salt", "onion", "flavor", "seasoning", "chicken", "flavor", "seasoning", "contains", "soy", "monosodium", "glutamate", "disodium", "inosinate", "and", "disodium", "guanylate", "garlic", "flavor", "seasoning", "sodium", "tripolyphosphate", "stabilizer", "contains", "allergens", "see", "bolded", "ingredients", "list", "may", "contain", "fish", "peanuts", "and", "milk"]
```

[Step 6] Performing the text detection accuracy calculation stage. To obtain the text detection accuracy, the average confidence value of each recognized word is calculated. This confidence value can be seen in the 10th column of the output table displayed in Figure 11. To obtain the text detection accuracy, all values in the 10th column are summed up and then divided by the total number of words, and finally converted into percentage form.

```
level   page_num   block_num   par_num line_num   word_num   left   top width   height  conf   text
1  1  0  0  0  0  0  0  1024  425 -1
2  1  1  0  0  0  16  58  993 318 -1
3  1  1  1  0  0  16  58  993 318 -1
4  1  1  1  1  0  16  58  982 54  -1
5  1  1  1  1  1  16  58  85  32  93  Daging
5  1  1  1  1  2  117 60  70  31  96  Ayam
5  1  1  1  1  3  203 61  83  30  96  (27%),
5  1  1  1  1  4  303 63  47  28  93  Air,
5  1  1  1  1  5  368 63  94  31  92  Minyak
```

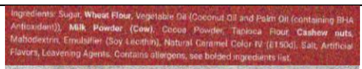
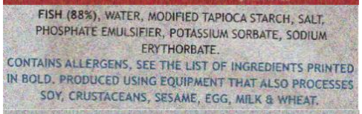
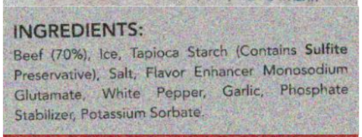
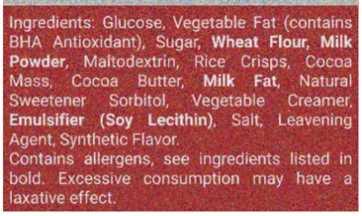
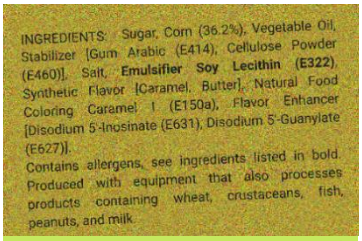
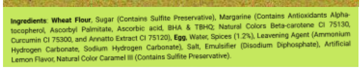
Figure 11. Confidence value of champ chicken stick image

[Step 7] Performing the stage of calculating allergen detection accuracy. To obtain the accuracy value for allergen detection, the system's iterated allergen count is divided by the manually calculated allergen count inputted by the developer in the system's interface.

3. RESULTS AND DISCUSSION

The research trial involved inputting 120 images following the testing procedure outlined. This procedure was designed to produce output results related to allergen detection in food products. In addition to testing the Main.py file, accuracy assessments were conducted for both text and allergen detection using the test.py file. After processing the 120 images, the expected output results were obtained, including a status of either Unsafe or Safe, along with the allergen information from the product packaging. Accuracy calculations for the 120 test images were also derived from the test.py file. Table 1 was the results of testing the 120 images.

Table 1. Results test images

No	Images test	Product name	Output	Accuracy of text detection (%)	Accuracy of allergen detection (%)	Note
1		Astor Chocolate	Unsafe: milk and flour	57.76	50.00	Soybeans
2		Fish Balls Bernardi	Unsafe: milk, soy, and wheat	29.19	75.00	Egg
3		Meat Balls Karawaci	Safe	67.63	100.00	Suitable
4		Beng Beng Maxx	Unsafe: milk, soybeans, and wheat	67.75	100.00	Suitable
:						
:						
119		Usagi Popcorn	Unsafe: milk, soybeans, and wheat	56.38	60.00	Butter and peanuts
120		Verkade Kruid Biscuit	Unsafe: wheat and egg	78.52	100.00	Suitable

The test images were captured under varying lighting conditions, with some experiencing issues such as uneven packaging surfaces and light reflections. Additionally, the differing packaging colors on certain products impacted the image threshold value. These three factors significantly influenced the accuracy of text detection OCR in the images. Figure 12 illustrates the text detection accuracy. From the 120 test images, an average text detection accuracy of 61.88% was achieved.

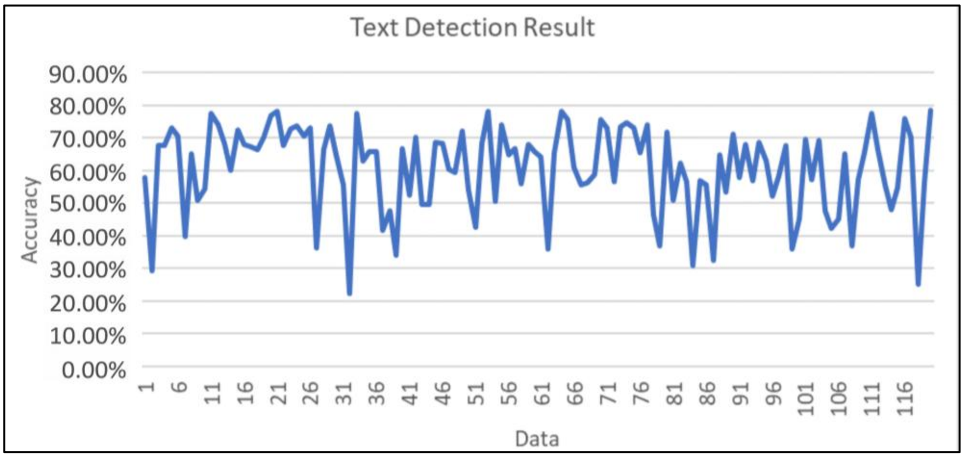


Figure 12. Accuracy result image trend of text detection

Lighting, noise, and packaging color all impact the accuracy of text readability, even though each image underwent the same pre-processing steps. These factors influence the confidence level values, which are then used to compute the overall accuracy.

When calculating accuracy for both text detection and allergen detection, each is affected by different factors. Text detection accuracy is primarily determined by the confidence level, while allergen detection accuracy is influenced by the number of actual allergens present in the image. The accuracy is particularly affected if only a few allergens are detected by the system. Figure 13 displays the average allergen detection accuracy from the 120 data points, which is 83.06%.

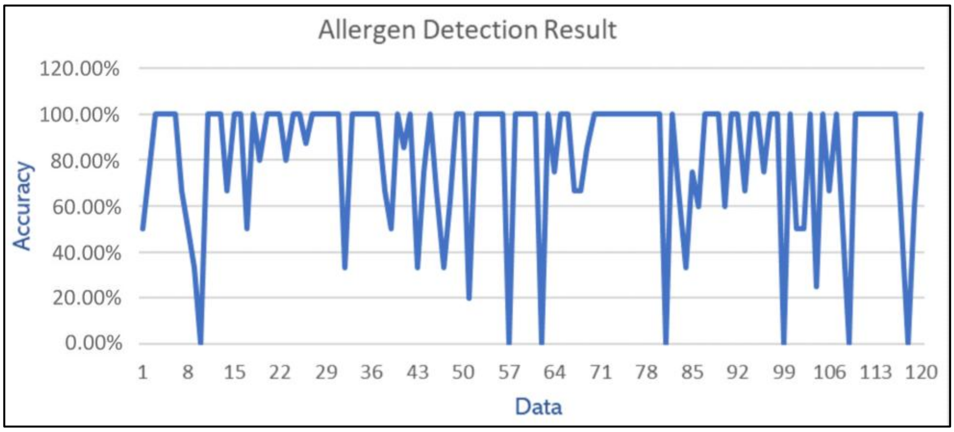


Figure 13. Accuracy result image trend of allergen detection

This indicates that the accuracy of text readability is crucial for the accuracy of allergen detection. If an allergen is not detected during the text recognition process, it will also be missed in the allergen detection step. As a result, manual verification of allergen detection is necessary to ensure the system can accurately calculate the results. The accuracy of allergen detection improves when all allergens are successfully identified during text detection. Therefore, allergen detection accuracy is closely linked to the accuracy of text detection, which in turn is influenced by factors such as noise, lighting, and packaging color.

4. CONCLUSION

The system is designed to assist eczema patients in managing their packaged food consumption through an allergen detection tool with a user-friendly interface. Users upload an image and initiate the allergen detection process by clicking the "detect allergens" button. The system uses OpenCV to convert the image into binary format, reads the text using PyTesseract, and tokenizes it with NLTK. Utilizing secondary data, the system searches for and identifies allergens, presenting the results on the main page, indicating safety status and listing the detected allergens. The system's performance was evaluated, achieving an average accuracy of 61.88% for text detection and 83.06% for allergen detection in 120 test images, with the highest accuracy being 78.52% for text detection and 100% for allergen detection. The lowest accuracy was 22.20% for text detection and 0.00% for allergen detection.

Despite achieving significant results, this study is limited by a small dataset and sensitivity to lighting conditions. However, these constraints provide a clear foundation for future research to enhance model robustness. Furthermore, the other future research will focus on integrating advanced deep learning (DL) OCR, expanding language support, and optimizing models for real-time mobile applications.

ACKNOWLEDGMENTS

The authors would like to thank the Faculty of Computer Science, Sriwijaya University for providing the laboratory facilities and technical support necessary for this research.

FUNDING INFORMATION

The author(s) received no financial support for the research, authorship, and/or publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dian Palupi Rini	✓	✓		✓		✓	✓			✓		✓		✓
Riska Tri Mardilah		✓	✓				✓		✓					
Muhammad Qurhanul Rizqie						✓			✓					
Dwi Rosa Indah						✓				✓	✓		✓	
Alvi Syahrini Utami		✓							✓					
Jovanic Morgan			✓		✓				✓		✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest associated with this research.

INFORMED CONSENT

Informed consent was obtained from all individual participants included in this study.

DATA AVAILABILITY

The primary data that support the findings of the research are available from the corresponding author, [Dian Palupi Rini], upon reasonable request. Explain the reason why the readers must request the data.

REFERENCES

[1] S. Traidl, L. Roesner, J. Zeitvogel, and T. Werfel, "Eczema herpeticum in atopic dermatitis," *Allergy*, vol. 76, no. 10, pp. 3017–3027, Oct. 2021, doi: 10.1111/all.14853.

[2] Y. Tokura, M. Yunoki, S. Kondo, and M. Otsuka, "What is 'eczema'?," *The Journal of Dermatology*, vol. 52, no. 2, pp. 192–203, Feb. 2025, doi: 10.1111/1346-8138.17439.

[3] A. Das and S. Panda, "Role of elimination diet in atopic dermatitis: Current evidence and understanding," *Indian Journal of Paediatric Dermatology*, vol. 22, no. 1, pp. 21–28, 2021, doi: 10.4103/ijpd.IJPD_88_20.

[4] W. Liu, J. Cai, C. Sun, Z. Zou, J. Zhang, and C. Huang, "Time-trends for eczema prevalences among children and adults from 1985 to 2015 in China: a systematic review," *BMC Public Health*, vol. 22, no. 1, Jul. 2022, doi: 10.1186/s12889-022-13650-7.

[5] Y. Schonmann *et al.*, "Atopic eczema in adulthood and risk of depression and anxiety: A population-based cohort study," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 8, no. 1, pp. 248-257.e16, Jan. 2020, doi: 10.1016/j.jaip.2019.08.030.

[6] J. Fanzo, A. L. Bellows, M. L. Spiker, A. L. Thorne-Lyman, and M. W. Bloem, "The importance of food systems and the environment for nutrition," *The American Journal of Clinical Nutrition*, vol. 113, no. 1, pp. 7–16, Jan. 2021, doi: 10.1093/ajcn/nqaa313.

[7] M. Rotaru, G. Iancu, and I. Matran, "Importance of food in the control of inflammation in atopic dermatitis," *Experimental and Therapeutic Medicine*, vol. 20, no. 6, p. 1, Oct. 2020, doi: 10.3892/etm.2020.9336.

[8] A. Fiocchi *et al.*, "Food labeling issues for severe food allergic patients," *World Allergy Organization Journal*, vol. 14, no. 10, Oct. 2021, doi: 10.1016/j.waojou.2021.100598.

[9] A. Parkavi, T. B. N. Shetty, V. A. Raj, S. B. Upadhyaya, and R. Thairani, "Android application for food label recognition to ensure safe food consumption based on user allergen information leveraging OCR," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–6, doi: 10.1109/ICCCNT56998.2023.10307054.

[10] D. Khairani, D. A. Bangkit, N. F. Rozi, S. U. Masruroh, S. Oktaviana, and T. Rosyadi, "Named-entity recognition and optical character recognition for detecting halal food ingredients: Indonesian case study," in *2022 10th International Conference on Cyber and IT Service Management (CITSM)*, Sep. 2022, pp. 01–05, doi: 10.1109/CITSM56380.2022.9935966.

[11] B. Rohini, D. M. Pavuluri, L. N. Kumar, V. Soorya, and J. Aravinth, "A framework to identify allergen and nutrient content in fruits and packaged food using deep learning and OCR," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2021, pp. 72–77, doi: 10.1109/ICACCS51430.2021.9441800.

[12] N. A. P. Kamis and O.-K. Shin, "OCR-based safety check system of packaged food for food inconvenience patients," *Journal of Digital Contents Society*, vol. 21, no. 6, pp. 1025–1032, Jun. 2020, doi: 10.9728/dcs.2020.21.6.1025.

[13] P. P. Reddy, P. S. Shruthi, P. Himanshu, and T. Singh, "License plate detection using YOLO v8 and performance evaluation of EasyOCR, PaddleOCR and tesseract," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jun. 2024, pp. 1–6, doi: 10.1109/ICCCNT61001.2024.10725878.

[14] T. Bieber *et al.*, "Atopic dermatitis: pathomechanisms and lessons learned from novel systemic therapeutic options," *Journal of the European Academy of Dermatology and Venereology*, vol. 36, no. 9, pp. 1432–1449, Sep. 2022, doi: 10.1111/jdv.18225.
[15] S. J. Brown, "Atopic eczema," *Clinical Medicine*, vol. 16, no. 1, pp. 66–69, Feb. 2016, doi: 10.7861/clinmedicine.16-1-66.
[16] P. Kage, J. Simon, and R. Treudler, "Atopic dermatitis and psychosocial comorbidities," *JDDG: Journal der Deutschen Dermatologischen Gesellschaft*, vol. 18, no. 2, pp. 93–102, Feb. 2020, doi: 10.1111/ddg.14029.
[17] A. C. Chong, K. Visitsunthorn, and P. Y. Ong, "Genetic/environmental contributions and immune dysregulation in children with atopic dermatitis," *Journal of Asthma and Allergy*, vol. 15, pp. 1681–1700, Nov. 2022, doi: 10.2147/JAA.S293900.
[18] N. Papapostolou, P. Xepapadaki, S. Gregoriou, and M. Makris, "Atopic dermatitis and food Allergy: A Complex Interplay What We know and What We Would Like to Learn," *Journal of Clinical Medicine*, vol. 11, no. 14, Jul. 2022, doi: 10.3390/jcm11144232.
[19] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access*, vol. 8, pp. 142642–142668, 2020, doi: 10.1109/ACCESS.2020.3012542.
[20] S. Faizullah, M. S. Ayub, S. Hussain, and M. A. Khan, "A Survey of OCR in Arabic Language: Applications, Techniques, and Challenges," *Applied Sciences*, vol. 13, no. 7, Apr. 2023, doi: 10.3390/app13074584.
[21] C. Thorat, A. Bhat, P. Sawant, I. Bartakke, and S. Shirsath, "A detailed review on text extraction using optical character recognition," in *ICT Analysis and Applications*, 2022, pp. 719–728, doi: 10.1007/978-981-16-5655-2_69.
[22] A. Chadha, S. Kashyap, M. Gupta, and V. Kumar, "License plate recognition system using OpenCV & PyTesseract," *CSI Journal of Computing*, vol. 3, no. 3, pp. 31–35, 2020.
[23] S. Dome and A. P. Sathe, "Optical Charater Recognition using Tesseract and Classification," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Mar. 2021, pp. 153–158, doi: 10.1109/ESCI50559.2021.9397008.
[24] D. A. Blubaugh, S. D. Harbour, B. Sears, and M. J. Findler, "OpenCV and Perception," in *Intelligent Autonomous Drones with Cognitive Deep Learning*, Berkeley, CA: Apress, 2022, pp. 327–361, doi: 10.1007/978-1-4842-6803-2_8.
[25] M. Wang and F. Hu, "The application of NLTK library for python natural language processing in corpus research," *Theory and Practice in Language Studies*, vol. 11, no. 9, pp. 1041–1049, Sep. 2021, doi: 10.17507/tpls.1109.09.

BIOGRAPHIES OF AUTHORS

Dian Palupi Rini holds a Doctor of Computer Science degree from Universiti Teknologi Malaysia, in 2017. She also received her B.Sc. (Mathematics) from Sriwijaya University, Indonesia. and M.Sc. (Computer Science) from Gadjah Mada University, Indonesia in 2000 and 2003, respectively. She is currently a lecturer at Computer Science Faculty, Sriwijaya University, Indonesia. Her research includes meta-heuristics, global optimization, machine learning (ML), deep learning (DL), data mining (DM), fuzzy systems, and artificial neural network (ANN). She can be contacted at email: dprini@unsri.ac.id.

Riska Tri Mardilah holds a Bachelor of Computer Science degree from Sriwijaya University, Indonesia, graduating cum laude in 2023 with a GPA 3.74. Awarded the Beasiswa Unggulan by the Ministry of Education and Culture in 2020. Her research interests include natural language processing (NLP), OCR, mobile development, data-driven strategies, and digital transformation (DT). She can be contacted at email: riskamardilah@gmail.com.

Muhammad Qurhanul Rizqie holds Doctoral degree in Biomedical Engineering from Universiti Teknologi Malaysia, Malaysia in 2021. He also received her Bachelor degree (Computer Science) from Universitas Sriwijaya, Indonesia. and Master degree (Electrical Engineering) from Institut Teknologi Bandung, Indonesia in 2009 and 2013, respectively He is currently a lecturer at the Faculty of Computer Science, Sriwijaya University. His current research interest is application of computer science in biomedical field, especially in medical image processing and medical data. He can be contacted at email: qurhanul.rizqie@ilkom.unsri.ac.id.

Dwi Rosa Indah holds her Master of Engineering degree from Bandung Institute of Technology, Indonesia in 2014. She also received her Bachelor of Engineering degree from Gajah Mada University, Indonesia in 2003. She is currently a lecturer at the Faculty of Computer Science, Sriwijaya University. His research interests include DM, UI/UX, and governance. She can be contacted at email: indah812@unsri.ac.id.

Alvi Syahrini Utami holds a Master of Computer Science degree from Gadjah Mada University, Indonesia in 2005. She also received her B.Sc. (Mathematics) from Sriwijaya University, Indonesia in 2002. She is currently a lecturer at the Faculty of Computer Science, Sriwijaya University, Indonesia. Her research interest include ML, pattern recognition, and data science. She can be contacted at email: alvisyahrini@ilkom.unsri.ac.id.

Jovanic Morgan is currently a student in the Informatics Engineering program at the Faculty of Computer Science, Sriwijaya University, Indonesia. His research includes ML, DL, and DM. He can be contacted at email: Jovanicmorgan7@gmail.com.