Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 8 No. 6 (2024) 758 - 767 e-ISSN: 2580-0760

Ingredients Identification Through Label Scanning Using PaddleOCR and ChatGPT for Information Retrieval

Ahmad Wahyu Rosyadi[1]*, Siti Ma'shumah[2], Muhammad Qomaruz Zaman[3], Moh. Rizki Fajar[4]
[1,4]Department of Informatics Engineering, Faculty of Engineering, Universitas Qomaruddin, Gresik, Indonesia
[2,3]Department of Electrical Engineering, Faculty of Engineering, Universitas Qomaruddin, Gresik, Indonesia
[1]awrosyadi@gmail.com*, [2]sitimashumah@uqgresik.ac.id, [3]me@zamzaman.com, [4]mohrizkifajar@gmail.com

Abstract

Human health depends on choosing food ingredients that align with dietary needs and avoid allergens. However, consumers often encounter unfamiliar ingredients that require additional information. Traditionally, they search online by typing in the ingredient's name which can be time-consuming and may not yield relevant results. Therefore, a system to identify and display ingredient information is necessary. This study proposes a new system that identifies ingredients by scanning the composition label on packaging using PaddleOCR and retrieving information through ChatGPT on a smartphone. The process begins with capturing an image of the composition label. Then PaddleOCR is employed to extract text from the scanned label, enabling identification of the listed ingredients. Subsequently, ChatGPT retrieves detailed information about the desired ingredients and displays it, allowing users to easily understand the ingredients. The system's effectiveness in text recognition is assessed using the character error rate (CER). The results show robust performance by achieving an average CER of 0.14, with flat packaging reaching an impressive CER of 0.05. Additionally, the system's usability was assessed through pilot testing which received significant positive user feedback achieving a 4.37 satisfaction level on the Likert scale, particularly regarding the clarity and relevance of the ingredient information provided.

Keywords: *food identification, optical character recognition, PaddleOCR, information retrieval, ChatGPT*

How to Cite: A. W. Rosyadi, Siti Ma'shumah, Muhammad Qomaruz Zaman, and Moh. Rizki Fajar, "Ingredients Identification Through Label Scanning Using PaddleOCR and ChatGPT for Information Retrieval", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 8, no. 6, pp. 758 - 767, Dec. 2024.
DOI: https://doi.org/10.29207/resti.v8i6.6119

1. Introduction

Human health can be optimally maintained by selecting appropriate food ingredients. These foods can contain essential macronutrients such as carbohydrates, proteins, and lipids, as well as micronutrients such as minerals and vitamins that are important to keep the body in good condition [1]. It is also very important to know the food content information to avoid allergens. However, consumers may encounter unfamiliar ingredients that necessitate additional information. Conventionally, consumers resort to seeking explanations on the internet by manually typing the ingredient's name. This process can be time-consuming and may not always yield reliable or comprehensive information. Moreover, the manual search method may deter consumers from fully understanding the composition and benefits of the products they are considering. To address this, previous works have utilized built-in smartphone cameras to scan and display ingredient information of the product. One of the preventive measures is to utilize the capabilities of smartphones with built-in cameras. The camera can be used to photograph product packaging which will then be scanned to display nutrition information from the packaging.

Scanning of food packaging by utilizing a smartphone camera has been done by Maringer, M., et al in 2019 [2]. The study conducted by the authors explored the use of smartphone cameras to scan barcodes on food packaging, enabling users to identify food items and access detailed nutritional information. This approach leverages the smartphone's camera to capture the Universal Product Code (UPC) on food packaging, which can then be used to retrieve data such as energy content, macronutrient composition, and other relevant nutritional details. The study found that barcode

Received: 18-10-2024 | Accepted: 13-12-2024 | Published Online: 27-12-2024

scanning requires minimal user interaction, making it a convenient and non-intrusive tool for dietary assessment. This functionality has been shown to simplify the recording of food intake and reduce the burden typically associated with manual food logging [3-6]. However, the authors highlight limitations in the accuracy and reliability of the underlying food product databases. The study finds substantial variation in the identification rates and nutrient data accuracy across different apps. While some apps, such as MyFitnessPal, accurately identified up to 96% of scanned products, others had much lower success rates [2]. Additionally, although energy values were consistently reported, nutrient details like protein, sugar, and fat content varied widely, limiting these applications' usefulness for detailed nutrient-level dietary assessments.

Building on existing technology, NutriLens supports informed dietary choices with a comprehensive set of features. Central to NutriLens is a barcode scanning function that retrieves detailed nutritional information on packaged foods through the Open Food Facts API. Beyond barcode scanning, NutriLens also includes custom data capture for non-packaged foods, utilizing NLP to analyze user queries and provide personalized nutritional insights. This approach exemplifies a holistic dietary guidance system, empowering users to make health-conscious decisions [7]. However, not all food packaging is labeled with barcodes, and not all barcodes can be scanned to reveal ingredient information. Instead, almost all food packaging displays ingredient information that can be read and scanned by consumers.

Reliable and rapid scanning, along with text recognition, can be achieved through the Optical Character Recognition (OCR) algorithm. The OCR algorithm's ability to recognize text with minimal errors impacts the accuracy of subsequent processes, such as information retrieval (IR). Moreover, quick text recognition enhances the overall user experience of the application. Several algorithms can be used for OCR, including TesseractOCR, media, KerasOCR, EasyOCR, and PaddleOCR [8], [9]. Among these, PaddleOCR stands out as an open-source, deep-learning-based technology known for its speed and efficiency in text recognition [10]. PaddleOCR has been used to perform chip text detection, video content text extraction, and invoice text recognition [10]-[12]. The research was able to improve the accuracy of character printing quality checks with an accuracy of more than 99%. In terms of code recognition for epidemic prevention and control, PaddleOCR produces good text recognition and high accuracy as demonstrated by Ref. [13]. Then, PaddleOCR is also used to recognize the texts in the video content of the smart dot reader educational application [11]. In addition, PaddleOCR is also good for text extraction from invoices because it produces good accuracy and has the ability to adapt to various invoice layouts [12]. However, the aforementioned approaches are not capable of comprehending the text they recognize. This limitation results in an inability to correct misrecognized words. Consequently, these inaccuracies may propagate further unreliable information regarding product ingredients. Therefore, further research and investigation into these challenges is imperative to advance the reliability of OCR systems in delivering precise and contextually appropriate information.

Errors in text recognition result from PaddleOCR may occur due to low image quality, impacting the accuracy of subsequent processes like information retrieval. Spelling correction techniques can be applied to address these errors by adjusting misspelt words to their most probable accurate forms. Effective spelling correction not only corrects typographical errors but also ensures that the corrected word is contextually appropriate. For instance, if "brown suar" is incorrectly recognized instead of "brown sugar," an ideal correction would return "brown sugar" rather than similar yet out-of-context words like "brown star." To achieve this, ChatGPT is used in this study for spelling correction in ingredient names prior to information retrieval, leveraging its contextual understanding to provide precise and contextually relevant adjustments. ChatGPT has been successfully applied in spelling correction tasks across languages, including Chinese and English [14]. Its effectiveness in spelling correction has been demonstrated in various fields, such as education and medical writing, thanks to its ability to understand broad word contexts and its vast vocabulary built from extensive data processing [15], [16].

ChatGPT uses deep neural networks based on a large language model (LLM) to enhance the accuracy of semantic-level retrieval by enabling a profound understanding of text semantics. Moreover, its generative system enhances the flexibility and expressiveness of information retrieval by enabling the creation of precise query expressions and detailed retrieval results. Leveraging zero or few-shot learning capabilities, where models require minimal or no training data, these models reduce the dependency on extensive annotated datasets, thus simplifying complex retrieval tasks [17].

In addition to spelling correction, ChatGPT has been explored in the case of information retrieval to investigate its potential for enhancing access to relevant data and improving information accuracy across various fields, such as its role in public health, its impact on information retrieval in the era of generative AI, and the exploration of ChatGPT for next-generation information retrieval [17]. In the age of massive models, ChatGPT's generative models are bringing fresh viewpoints and approaches to the fundamental information retrieval task. The goal of IR systems is to generate pertinent information from vast volumes of textual data. The traditional IR system which uses keyword matching as its common component is gradually moving toward semantic-based retrieval with the introduction of neural networks and deep learning

[18]. Building on this foundation, this study also utilizes ChatGPT to obtain nutritional information for the desired food or beverage ingredients. The nutritional information displayed is expected to provide an explanation that is easily understood by users so that it can help users in understanding and choosing their food. Therefore, a system that can scan the composition of food packaging and display relevant information on each ingredient in the composition is needed.

In this study, a new system for food identification by scanning the composition label on the packaging with the PaddleOCR algorithm and information retrieval with ChatGPT on a smartphone is proposed. This system offers an innovative solution by integrating ChatGPT for automatic spelling correction on misspelt OCR outputs prior to information retrieval, ensuring greater accuracy. Additionally, by utilizing ChatGPT's information retrieval capabilities, the model delivers precise and relevant nutritional insights derived from detailed food label data. This advancement minimizes the need for manual error correction while broadening the nutritional information accessible to users. Consequently, this study presents a unique approach to enhancing dietary awareness, blending advanced AI-driven spelling correction and nutritional information retrieval within an easy-to-use application aimed at promoting healthier lifestyle choices.

The system was named GiziLens, derived from the words "gizi" (an Indonesian term for nutrition) and "lens". Gizi, which has a meaning similar to nutrition, is one of the most important elements in maintaining human health and is among the key information provided by this system, along with descriptions, safety for consumption, and potential disadvantages of each ingredient. The key contribution of this research includes the direct recognition of ingredients from composition label scans. This approach deviates from existing studies in Refs. [2-7], which rely on barcodes that may not always be linked to detailed ingredient information. Additionally, this system incorporates a spelling correction process via ChatGPT to rectify PaddleOCR text recognition errors prior to information retrieval. ChatGPT is also employed in the information retrieval process to provide users with relevant and easily understandable information on food ingredients. The ingredient information generated by the system can serve as a reference for users to assess whether the food is suitable for consumption. Finally, pilot testing was conducted to evaluate user experiences using a Likert scale.

The remainder of this paper is organized into three sections. Section 2 describes the proposed method, Section 3 presents the experimental results, and Section 4 provides the conclusions.

2. Research Methods

This investigation presents a fundamental implementation of optical character recognition utilizing smartphone-captured images, thereby establishing a foundational benchmark for future algorithmic enhancements and methodological optimizations. The current scope deliberately excludes sophisticated image preprocessing techniques, including low-luminance compensation, specular reflection mitigation, and geometric distortion correction.

There are several steps involved in obtaining information about each ingredient in the product composition as shown in Figure 1. First, an image containing compositional details is obtained by scanning the packaging's composition label using a smartphone camera. Then, cropping is performed to exclude irrelevant information, keeping only the composition details. PaddleOCR is used to extract text from the product composition label. Next, each of the recognized words is broken down into individual tokens, which are expected to represent the names of each ingredient. Information regarding the description, safety for consumption, nutrition, and disadvantages of the individual token is then obtained using ChatGPT. This information is displayed on the user page, allowing users to learn about the ingredients listed on the composition label.

Image cropping involves removing specific parts of an image to enhance its focus, composition, or size according to the needs. Cropping is used to obtain a good Region of Interest (ROI) by focusing on the image area that contains only the composition details as shown in Figure 1(c, d). Selecting an effective ROI reduces text recognition errors by minimizing the presence of extraneous backgrounds and false objects [19], [20].

PaddleOCR is an open-source OCR tool library that is comprehensive, advanced, lightweight, and highly practical and developed by Baidu. Since its release on May 14, 2020, PaddleOCR has undergone continuous optimization and updates. It offers a wide range of open-source models for language recognition, including the widely used ultra-lightweight PP-OCR models for Chinese and English OCR. Additionally, it supports over 80 other languages, such as Japanese, Korean, French, and German. PaddleOCR includes various text detection training algorithms (EAST, DB) and character recognition training algorithms (Rosetta, CRNN, STAR-Net, RARE). It also supports custom training, offers diverse prediction deployment solutions, allows for quick installation via PIP, and is compatible with Windows, Linux, macOS, and other systems [13].

PaddleOCR's structure consists of three key components: text detection, detection boxes rectify, and text recognition as shown in Figure 2. This study utilizes the PP-OCRv4 model for text recognition. The PP-OCRv4 model follows the same framework diagram as the PP-OCRv3 model but with some configuration and tool changes to improve its performance. First, the text detection component is responsible for identifying and locating areas in images that contain text and generating bounding boxes. In this component, PP-OCRv4 replaces MobileNetV3 with PP-LCNetV3 and incorporates the PFhead structure. This update reduces the model's resource consumption while maintaining performance, thereby enhancing operational efficiency [21]. The detection box rectification component then adjusts the bounding boxes around detected text areas. It applies Green's theorem to determine the orientation of the image (clockwise or counterclockwise) and rotates it accordingly to align the text in the correct direction [22].

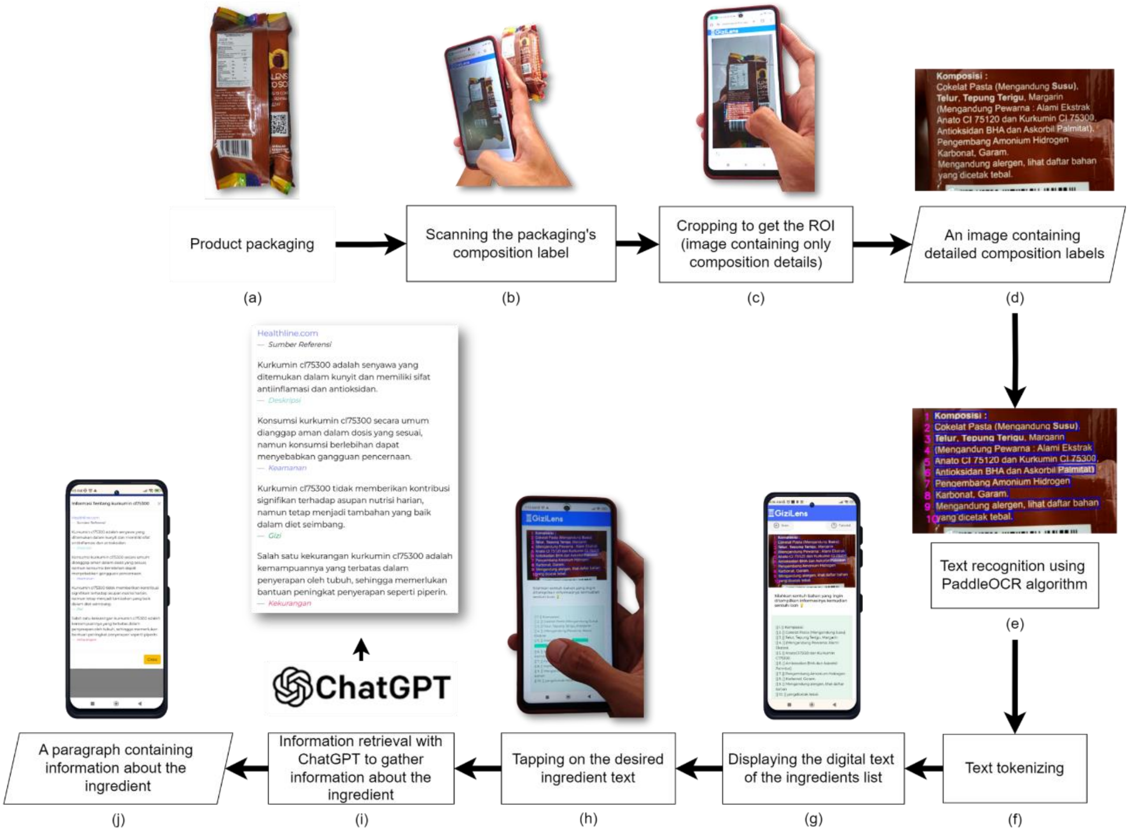


Figure 1. System Design

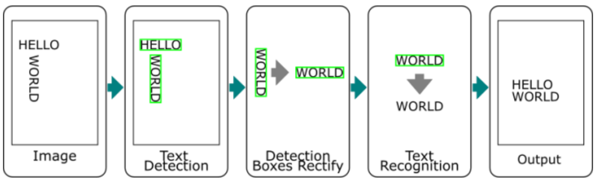


Figure 2. Working Procedure of PaddleOCR

The final component is the text recognition, which consists of two stages. In Stage 1, text areas are detected and segmented into individual characters. In Stage 2, the model recognizes each character [22]. The PP-OCRv4 recognizer is optimized using a newly developed lightweight model for text recognition SVTR_LCNet. It integrates two methods: SVTR (a method based on transformer networks) [23], and PP-LCNet (a method based on convolutional networks) [24]. This approach is taken to leverage the strengths of both accuracy and speed in text recognition.

Text tokenization is performed to break sentences into individual words, allowing each word to stand alone [25]. This step is necessary because the digital text generated from the segmentation and text recognition process using PaddleOCR is still in the form of an array based on rows in the composition. Therefore, the digital text in each row must be separated to extract a collection of independent items that make up the composition as shown in Figure 3.

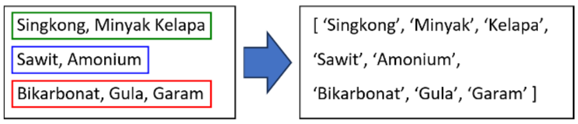


Figure 3. Text tokenizing

Information retrieval (IR) is the process of searching for relevant information from a large dataset, typically in text form, with the goal of answering specific questions or fulfilling a user's information needs. In recent years, numerous advancements have been made in the field of IR, one of which is the development of LLM like ChatGPT. This technology has introduced significant innovation to IR, as LLMs can understand and interpret text in complex contexts, enabling them to generate relevant and detailed responses to user queries. For instance, ChatGPT utilizes deep learning techniques to analyze and process vast amounts of text data, offering more accurate and user-friendly retrieval results compared to traditional IR methods [13].

The deep learning of ChatGPT enhances the accuracy of semantic-level retrieval by enabling a profound understanding of text semantics. Moreover, its generative system enhances the flexibility and expressiveness of information retrieval by enabling the creation of precise query expressions and detailed retrieval results. Leveraging zero or few-shot learning capabilities makes models require minimal or no training data and reduces the dependency on extensive annotated datasets, thus simplifying complex retrieval tasks [17].

Consequently, ChatGPT has been widely used for information retrieval because it offers relevant information based on a vast dataset or documents in nearly every aspect of daily life. ChatGPT's conversational capability and user-friendly interface contribute to its popularity, allowing it to respond effectively to a wide range of questions. It leverages deep learning techniques in analyzing and processing extensive text data to deliver highly accurate and relevant responses based on existing information. This results in more precise retrieval outcomes that align closely with user needs, surpassing traditional information retrieval methods [26].

With its wide range of features, ChatGPT can be a valuable resource for obtaining information across various fields, including public health. Here are some examples of how it can be utilized in public health: (1) offering information on public health concerns, including infectious diseases and environmental health risks; (2) responding to inquiries about strategies for health promotion and disease prevention [27]. In addition, information about food or beverages to be consumed can also be provided. This includes descriptions, safety for consumption, nutritional content, and potential deficiencies of the food or its ingredients. Such information is crucial for consumers, as it helps them avoid products containing hazardous substances.

In this study, ChatGPT is used to perform both spelling correction and information retrieval because it has several advantages over other IR technologies. First, ChatGPT demonstrates shorter task completion times compared to Google, as it provides immediate, dialogue-based responses that eliminate the need for users to navigate through multiple links as in traditional search engines. Second, users find ChatGPT's information to be clearer and of higher quality, thanks to its concise answer summaries. In contrast, Google typically directs users to various sources, which require further interpretation. However, a primary limitation of ChatGPT is its potential to present inaccurate or unverifiable information due to limited training data, whereas Google excels at fact verification through external sources [28]. This limitation can be mitigated by prompting ChatGPT to generate responses based on reliable sources.

The information generated from ChatGPT will be processed and displayed to the user when they tap on the desired ingredient text. Upon tapping, the system sends the prompt in Indonesian, *"Correct the spelling of the term if there is any mistake. Based on a reliable source, provide a description, safety for consumption, nutritional value, and disadvantages of the term in 4 sentences. Include a link to the reference source."* to retrieve information from ChatGPT as shown in Figure 1(h,i,j).

In terms of prompt design, the ChatGPT inquiry is designed to be concise, aiming to minimize the cost of API usage. The design focuses on eliciting four essential pieces of information: a brief definition of the ingredient, a summary of its safety for consumption, and specific details regarding its nutritional value and potential disadvantages. The requirement that answers be limited to four sentences constrains the scope of creativity, compelling ChatGPT to provide only the most pertinent facts and avoid irrelevant information that might arise in longer responses. This approach not only ensures concise and relevant answers but also helps to reduce API costs from the answers. Additionally, the inclusion of reference sources can increase the trustworthiness of the provided answers and allow users to verify the information. The combined efforts of the prompt thus ensure cost-effective and reliable information retrieval while maintaining a high level of accuracy and relevance. It is important to note that the inquiry is designed to provide a first glance or brief information; therefore, it should not be used as the final decisive answer regarding the consumption of food containing the ingredient in question.

3. Results and Discussions

This research aims to produce a system that can be used to scan the composition of food packages and can display information on each ingredient in the composition. The food information generated by the system can be used as a reference by users to determine whether the food to be consumed is good for consumption or not.

This system is designed as a web-based application so that users do not need to install it on their smartphones. With just an internet connection, users can easily access the application on any device. By being web-based, the application can be accessed on various devices and platforms without being limited to a particular type of operating system. Moreover, this approach enables feature updates and bug fixes to be implemented directly on the server, so users always get the latest version without having to update the application manually. The system runs on a locally hosted server with 4 CPU cores and 30 GB of RAM, ensuring robust performance especially in scanning and text recognition processes.

The scanning feature of the composition label on the packaging is equipped with text recognition using

PaddleOCR PP-OCRv4 and three pre-trained models with different functions: 1) en_PP-OCRv3_det_infer is used to detect areas or regions containing English text in images; 2) en_PP-OCRv4_rec_infer is used to convert the detected text images into digital text; 3) ch_ppocr_mobile_v2.0_cls_infer is used to classify text images, distinguishing between different types of text, such as headings, paragraphs, or tables. The performance of text recognition with PaddleOCR was evaluated using three hundred OCR document images. The ground truth image was obtained from cropping process and then was manually transcribed.

There are several stages that are conducted to test this system. First, the composition label on the packaging is scanned and cropped using a smartphone camera. The image is then processed using PaddleOCR to obtain the digital text of the composition label. The text generated by PaddleOCR is evaluated using the character error rate (CER) to determine the percentage of characters incorrectly transcribed by the text recognition model. Second, the Usability of the system will be tested using pilot testing to obtain information about application performance, ease of use, and the effectiveness of the application in helping users understand product composition information.

To evaluate the effectiveness of the PaddleOCR algorithm in text recognition on packaging composition labels, we use various types of packaging surfaces including flat surfaces (such as smooth cardboard boxes), curved surfaces (like cans or snack packaging), reflective surfaces (plastic or aluminum packaging that reflects light), and textured surfaces (such as paper packaging or other materials with surface texture) as shown in the Figure 4.

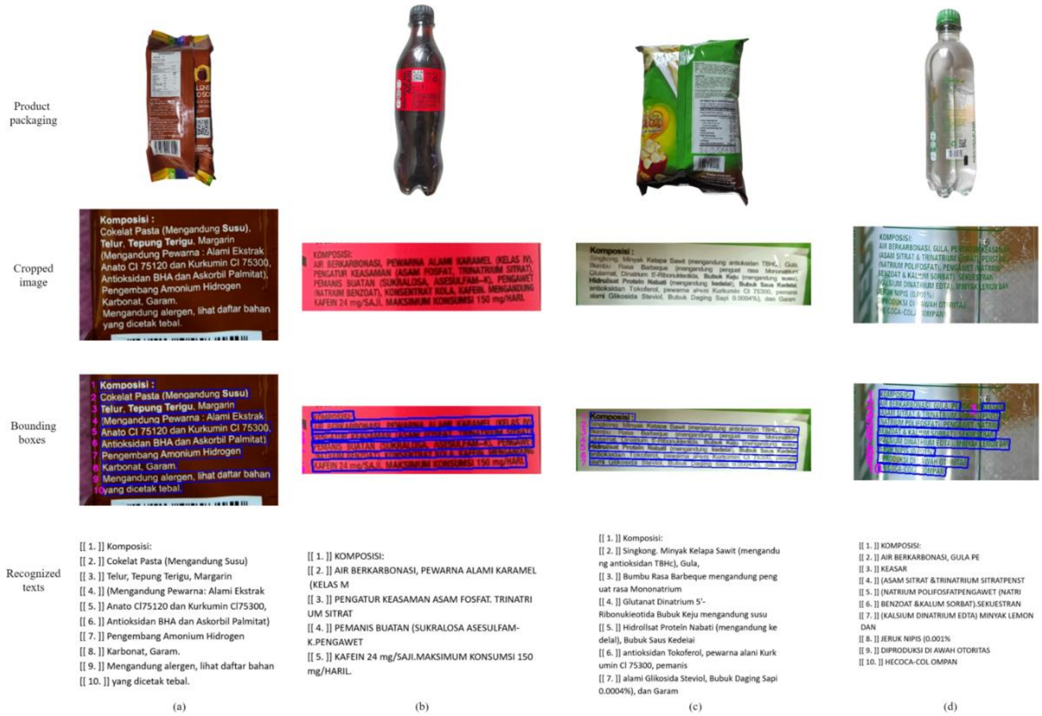


Figure 4. Various types of packaging surfaces (a) flat surfaces, (b) curved surfaces, (c) reflective surfaces, and (d) textured surfaces

3.1 Character Error Rate

$$CER = \frac{s+i+d}{t} \tag{1}$$

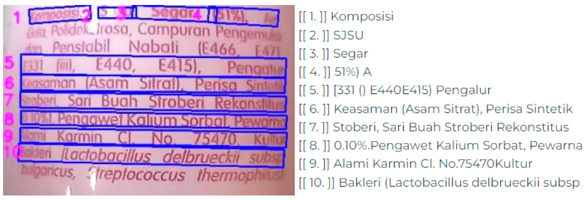
The Character error rate (CER) is a metric used to calculate the error in a text recognition model which counts the character manipulations (insertion, deletion, substitution) to map one string to another which is defined by Equation 1 [29]. CER is calculated using the Levenshtein distance which counts the number of substitutions s, insertions i, and deletions d between the predicted string and the reference string then divides that total by the count of characters in the reference string t.

Based on the CER results across various packaging surfaces, it is evident that PaddleOCR performs best on flat surfaces with a CER of 0.05, indicating high accuracy. This is because there is no physical distortion affecting the text on a flat surface as shown in Figure 4(a) and Table 1.

Table 1. Performance of PaddleOCR

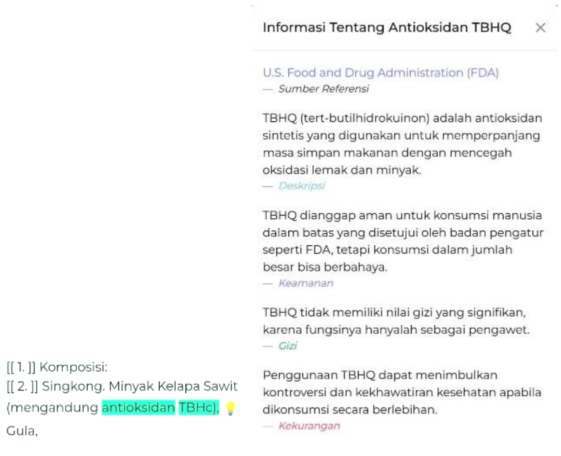
Image type	Average Number of Character Error Rate (CER)	Substitutions	Insertions	Deletions
Flat	0.05	3.97	11.84	1.92
Curved	0.12	10.25	45.87	3.83
Reflective	0.22	10.54	35.29	7.58
Textured	0.18	7.48	38.65	4.76
Average	0.14	8.06	32.91	4.52

However, its performance decreases on curved, reflective, and textured surfaces. Curved surface packaging, such as cans and bottles, presents a unique challenge for PaddleOCR because of the distortions introduced by the curvature of the surface. When packaging has a curved surface, text lines tend to vary in height, with the edges often being lower than the center. This distortion changes the geometry of the text, making it difficult for PaddleOCR to determine the correct bounding box for each character or word. As shown in Figure 5, three lines on the composition label remain undetected with no bounding boxes generated. Additionally, this distortion often causes certain characters to appear out of focus or narrower around the edges, leading to substitution errors. For example, the word "pengatur" may be misidentified as "pengalur" due to the altered character shapes. However, while the text recognition may produce minor errors, the proposed GiziLens consistently provides reliable ingredient information. This capability is attributed to ChatGPT's ability to perform spelling corrections prior to generating responses, as illustrated in Figure 6. For instance, the misrecognized word "antioksidan TBHc," which should be "antioksidan TBHQ," is accurately corrected by ChatGPT, leading to the provision of correct information about antioksidan TBHQ.



(a) (b)

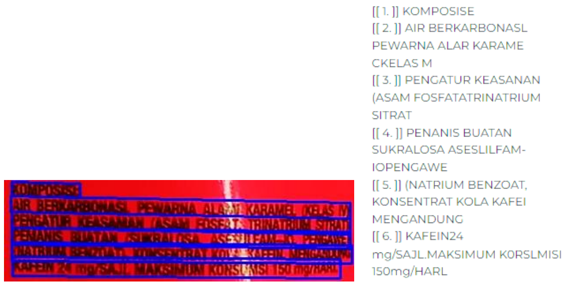
Figure 5. Curved Packaging: (a) Bounding Boxes and (b) Recognized Text.



(a) (b)

Figure 6. Information retrieved from ChatGPT:
(a) touched ingredient and (b) information retrieved.

Curved surfaces can also cause inconsistencies in character scale and affect detection algorithms based on straight-line orientation, which are commonly used in OCR systems. Therefore, PaddleOCR tends to produce more insertion and substitution errors on curved surfaces, resulting in an increase in CER of up to 0.12 on this type of surface. This challenge can be overcome through more sophisticated preprocessing, such as perspective correction to straighten text before it is processed by OCR.



(a) (b)

Figure 7. Reflective Packaging: (a) Bounding Boxes and (b) Recognized Text.

PaddleOCR's performance on reflective packaging decreases due to light reflections on the surface which can obscure the text as shown in Figure 7. Plastic and aluminium packaging has more errors than other surfaces with a CER of 0.22. This is because light reflections easily disrupt the packaging surface, causing text to appear blurred and creating high-contrast areas that confuse the OCR algorithm. PaddleOCR is sensitive to such reflections because it depends on contrast gradients between text and background for character recognition. For instance, shadows or glare on reflective surfaces create significant light intensity variations, potentially causing the OCR to miss characters entirely or insert extraneous ones (insertion errors). This challenge shows the importance of light handling through contrast enhancement techniques or special filters to reduce reflections on shiny surfaces.

(a) (b)

Figure 8. Textured Packaging:
(a) Bounding Boxes and (b) Recognized Text.

Additionally, textured packaging reduces PaddleOCR's performance due to the non-uniform background or high grayscale variation as illustrated in Figure 8. These surfaces not only impair the contrast between text and background but also introduce random gradient variations and patterns that resemble small characters. When OCR tries to recognize text, these gradient intensities and text-like patterns can lead to character misrecognition or false detections. As a result, PaddleOCR may inaccurately identify text boundaries, causing segmentation errors that contribute to substitution and deletion errors. The OCR system might mistake patterns for extra characters or fail to detect actual ones, reducing recognition accuracy. This challenge highlights the need for OCR fine-tuning to better handle varied or textured backgrounds, along with noise reduction filters before OCR processing.

Sometimes the food packaging scanned by the user is of poor quality, curved, reflective, or textured surfaces. This may occur because the packaging may be made of plastic, there are light reflections on the surface, or the shape of the packaging is curved. Therefore, improvements in preprocessing are needed to reduce distortion and reflection effects, such as using perspective correction and contrast enhancement. Additionally, fine-tuning the OCR model for food packaging datasets with a wider variety of surfaces will also be necessary.

To further assess the reliability of the answers provided by ChatGPT, the information is cross-checked with the reference links provided. A sample of 100 responses inquiring about ingredient information is collected. The results indicate that 99 responses included a reference link, while one response did not. Evaluating the correctness of the given answers in comparison to professional opinions is beyond the scope of this research and warrants a separate, comprehensive study. However, upon examining the information provided by ChatGPT, it was confirmed that the information can be found in a reliable source. The design of GiziLens is not intended to provide definitive answers but rather to offer preliminary information about ingredients, which may provide keywords to assist users in further exploring and understanding the ingredients. Within this scope of usage limitations, it is demonstrated that GiziLens, with the assistance of ChatGPT, provides answers with reliable reference links 99% of the time.

3.2 Usability Evaluation Using Pilot Testing

Pilot testing is an initial phase in research or product development that involves a small-scale trial to assess the feasibility, effectiveness, and functionality of an intervention or system prior to full-scale implementation. Its primary aim is to identify and resolve potential issues such as technical, procedural, or user-related problems, to ensure smooth execution in the final deployment. By involving a smaller group of users, pilot testing generates valuable insights that guide necessary adjustments. It allows for early identification of challenges and resulting in higher quality outcomes during larger-scale rollouts [29,30].

Usability of the system will be tested using Pilot testing to obtain information about application performance, ease of use, and the effectiveness of the application in helping users understand product composition information. We decided to recruit 27 participants from diverse backgrounds to evaluate the system's acceptability and feasibility and to test the study procedures in preparation for a larger trial in the future as shown in Table 2.

We assessed participants' experiences and perceptions of GiziLens using Likert scale response options and open-ended questions. The pre-test questionnaire was filled out by participants before using the application, while the post-test questionnaire was completed afterwards. To ensure a thorough evaluation, the post-test also gathered qualitative feedback on any difficulties encountered and suggestions for improvement, offering insights into user interaction patterns and overall satisfaction [31]-[33]. This approach allowed us to assess not only immediate reactions but also longer-term perceptions of the application's usability and effectiveness.

Table 2. Demographic Distribution of Respondents (n = 27)

Variable	Category	Frequency	Percentage
Gender	Female	12	44.40%
	Male	15	55.60%
Age	16 - 25	17	62.96%
	26 - 35	5	18.52%
	> 36	5	18.52%
Education	Senior High School	12	44.50%
	Bachelor	10	37%
	Postgraduate	5	18.50%

Based on the result of the Pre-test shown in Table 3, respondents tend to read the composition label only sometimes or rarely even though the majority of them agree that information about food or product composition is very important for them. This illustrates that respondents' awareness of the importance of food composition information is not balanced with efforts or tools that can provide the desired information easily. This study aims to solve this problem by providing a system that allows users to easily scan composition labels and display the relevant information easily.

Table 3. Pre-test Results Summary

Question	Mean	Standard Deviation
How often have you bought snacks in the past month?	3.48	1.05
Do you usually read the composition labels on food packaging?	3.11	0.93
How important is food composition information to you?	4.19	0.92
How comfortable is your comfort level in using new applications on your smartphone?	3.89	0.70
Have you ever used a similar application?	1.33	0.55

The proposed system was successfully developed and tested by respondents. After testing, respondents completed a post-test questionnaire to share their experiences using the system. The post-test results in Table 4 indicate that the application received generally positive feedback from users, with high average scores across most aspects. Users found the system very helpful for finding food composition information with a score of 4.67 and a standard deviation of 0.48 showing consistency in their evaluations. The information provided by the system was also rated as clear and easy to understand with a mean score of 4.30. Additionally, users considered the information relevant to their needs that was reflected by a score of 4.33. The ease of use and relevance of the information may be key factors influencing users' decisions to consume or purchase food. Overall, many users were satisfied with the system (4.37). User satisfaction is further highlighted by the high mean score of 4.15 for the statement that they would use the system before consuming or buying food in the future. In addition, they are also willing to recommend it to others (4.56).

However, there are some areas that need improvement, particularly in the speed of processing label scan results, which received the lowest score of 3.85 with a standard deviation of 0.95, indicating user dissatisfaction in this aspect. These lags are due to the process of uploading and storing scanned images in the server database prior to the recognition process. In addition, a score of 3.78 for the question related to whether the application meets user expectations indicates that there are expectations that have not been fully met by the system. Overall, the system shows strong potential with some aspects that need improvement, especially in terms of speed and optimizing the user experience.

Table 4. Post-test Results Summary

Question	Mean	Standard Deviation
How easy is this app to use?	4.37	0.69
Are you having trouble scanning product composition labels?	4.37	0.74
How fast does the app process label scan results?	3.85	0.95
Is the composition information displayed after scanning clear and easy to understand?	4.30	0.67
How relevant is the information displayed to your needs?	4.33	0.68
How helpful is it to search for food composition information through this app?	4.67	0.48
How was your experience with scanning snack composition labels using this application?	4.07	0.83
Does the application's interface make it easy for you to understand the displayed composition information?	3.96	0.65
How would you rate the overall navigation and interface of the app?	4.26	0.71
How important is the composition information displayed to help you make a purchasing decision?	4.52	0.58
Do you feel this app meets your expectations?	3.78	0.51
Would you recommend this app to others?	4.56	0.51
Will you use this app before consuming or purchasing food in the future?	4.15	0.46
How satisfied are you overall with this app?	4.37	0.49

4. Conclusions

This paper presented a novel system for ingredient identification by scanning composition labels on packaging using the PaddleOCR algorithm and leveraging ChatGPT for information retrieval on a smartphone. The system demonstrates strong text recognition performance by achieving an average Character Error Rate (CER) of 0.14, with particularly impressive results on flat packaging with an average CER of 0.05. Despite occasional recognition errors, ChatGPT consistently provides relevant information with reliable reference links 99% of the time. This is achieved through its ability to correct spelling errors in ingredient names based on context before retrieving information, ensuring greater precision. Furthermore, user feedback has been highly positive, with satisfaction scores reflecting appreciation for the system's ability to deliver clear, relevant, and helpful ingredient information. Overall, the system shows great potential to help users make the right food choices with further enhancements needed to optimize the user experience. Future research efforts should prioritize enhancements in preprocessing techniques such as implementing perspective correction and contrast enhancement to mitigate distortion and reflection effects. Moreover, fine-tuning the OCR model using a more diverse dataset of food packaging surfaces will be preferable. Additionally, addressing the processing speed of label scans represents an important area for future development, with the aim of improving the overall efficiency and responsiveness of the system.

References

[1] B. Ozhan and B. Tugrul, "Analysis of Turkish cuisine flavors network," *Int. J. Food Sci. Technol.*, pp. 908–915, 2023, doi: 10.1111/ijfs.16849.

[2] M. Maringer, N. Wisse-Voorwinden, P. V. t. Veer, and A. Geelen, "Food identification by barcode scanning in the Netherlands: A quality assessment of labelled food product databases underlying popular nutrition applications," *Public Health Nutr.*, vol. 22, no. 7, pp. 1215–1222, 2019, doi: 10.1017/S136898001800157X.

[3] A. G. Arens-Volland, L. Spassova, and T. Bohn, "Promising approaches of computer-supported dietary assessment and management-Current research status and available applications.," *Int. J. Med. Inform.*, vol. 84, no. 12, pp. 997–1008, Dec. 2015, doi: 10.1016/j.ijmedinf.2015.08.006.

[4] M. Rusin, E. Arsand, and G. Hartvigsen, "Functionalities and input methods for recording food intake: a systematic review.," *Int. J. Med. Inform.*, vol. 82, no. 8, pp. 653–664, Aug. 2013, doi: 10.1016/j.ijmedinf.2013.01.007.

[5] R. Steele, "An overview of the state of the art of automated capture of dietary intake information.," *Crit. Rev. Food Sci. Nutr.*, vol. 55, no. 13, pp. 1929–1938, 2015, doi: 10.1080/10408398.2013.765828.

[6] C. Byrd-Bredbenner and C. A. Bredbenner, "Assessing the home food environment nutrient supply using mobile barcode (Universal Product Code) scanning technology," *Nutr. Food Sci.*, vol. 40, no. 3, pp. 305–313, 2010, doi: 10.1108/00346651011044005.

[7] P. H. Gosavi, P. Kawale, P. Kadam, and A. Manjhi, "Nutrilens : Barcode Scanner," vol. 13, no. 4, pp. 90–97, 2024.

[8] C. Li *et al.*, "Dive into OCR," 2022.

[9] K. V Ujwal Karanth, A. T. Sujan, Y. R. Thanay Kumar, S. Joshi, K. P. Asha Rani, and S. Gowrishankar, "Breaking Barriers in Text Analysis: Leveraging Lightweight OCR and Innovative Technologies for Efficient Text Analysis," pp. 359–366, 2024, doi: 10.1109/icacrs58579.2023.10404305.

[10] H. Tao and C. Mingyang, "Character Recognition and Printing Quality Inspection of Chips under Packaging Film," *Proc. - 2022 Int. Conf. Comput. Eng. Artif. Intell. ICCEAI 2022*, pp. 139–143, 2022, doi: 10.1109/ICCEAI55464.2022.00037.

[11] Z. Hu, X. He, Y. Lin, Y. Zhao, Z. Liang, and J. Wu, "Intelligent Point Reader Based on Machine Vision and Speech

Processing," *2023 IEEE Int. Conf. Electr. Autom. Comput. Eng. ICEACE 2023*, pp. 975–983, 2023, doi: 10.1109/ICEACE60673.2023.10442012.

[12] J. Krishna Manipatruni *et al.*, "Leveraging Artificial Intelligence for Simplified Invoice Automation: Paddle OCR-based Text Extraction from Invoices," *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 9, 2023, [Online]. Available: www.ijisrt.com

[13] D. Zhang and Y. Li, "Research and Application of Health Code Recognition Based on Paddle OCR under the Background of Epidemic Prevention and Control," *J. Artif. Intell. Pract.*, vol. 6, no. 1, pp. 9–16, 2023, doi: 10.23977/jaip.2023.060102.

[14] K. Li *et al.*, "Eval-GCSC: A New Metric for Evaluating ChatGPT's Performance in Chinese Spelling Correction," 2023, [Online]. Available: http://arxiv.org/abs/2311.08219

[15] A. Tlili *et al.*, "What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education," *Smart Learn. Environ.*, vol. 10, no. 1, 2023, doi: 10.1186/s40561-023-00237-x.

[16] M. Koo, "The Importance of Proper Use of ChatGPT in Medical Writing," *Radiology*, vol. 307, no. 3, p. 7945, 2023, doi: 10.1148/RADIOL.230312.

[17] Y. Huang and J. X. Huang, "Exploring ChatGPT for next-generation information retrieval: Opportunities and challenges," *Web Intell.*, vol. 22, no. 1, pp. 31–44, 2024, doi: 10.3233/WEB-230363.

[18] M. Pan, J. Wang, J. X. Huang, A. J. Huang, Q. Chen, and J. Chen, "A probabilistic framework for integrating sentence-level semantics via BERT into pseudo-relevance feedback," *Inf. Process. Manag.*, vol. 59, no. 1, p. 102734, 2022, doi: https://doi.org/10.1016/j.ipm.2021.102734.

[19] A. W. Rosyadi and N. Suciati, "Median filter for transition region refinement in image segmentation," *JUTI J. Ilm. Teknol. Inf.*, vol. 16, no. 2, pp. 158–170, 2018.

[20] A. W. Rosyadi and N. Suciati, "Image Segmentation Using Transition Region and K-Means Clustering," *IAENG Int. J. Comput. Sci.*, vol. 47, no. 1, 2020.

[21] H. Zhang *et al.*, "Identification of Illegal Outdoor Advertisements Based on CLIP Fine-Tuning and OCR Technology," *IEEE Access*, vol. 12, no. July, pp. 92976–92987, 2024, doi: 10.1109/ACCESS.2024.3424258.

[22] F. Year, P. Final, P. T. Hieu, and V. Van Nghiep, "A Deep Learning Model for Helmet Detection and Automatic License Plate Recognition," no. December, 2023.

[23] Y. Du *et al.*, "SVTR: Scene Text Recognition with a Single Visual Model," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 1, no. c, pp. 884–890, 2022, doi: 10.24963/ijcai.2022/124.

[24] C. Cui *et al.*, "PP-LCNet: A Lightweight CPU Convolutional Neural Network," 2021, [Online]. Available: http://arxiv.org/abs/2109.15099

[25] A. Z. Arifin, M. Z. Abdullah, A. W. Rosyadi, D. I. Ulumi, A. Wahib, and R. W. Sholikah, "Sentence extraction based on sentence distribution and part of speech tagging for multi-document summarization," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 16, no. 2, pp. 843–851, 2018, doi: 10.12928/TELKOMNIKA.v16i2.8431.

[26] A. Rao *et al.*, "Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study," *J. Med. Internet Res.*, vol. 25, pp. 1–9, 2023, doi: 10.2196/48659.

[27] S. S. Biswas, "Role of Chat GPT in Public Health," *Ann. Biomed. Eng.*, vol. 51, no. 5, pp. 868–869, 2023, doi: 10.1007/s10439-023-03172-7.

[28] R. Xu, Y. (Katherine) Feng, and H. Chen, "ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience," *SSRN Electron. J.*, no. July, 2023, doi: 10.2139/ssrn.4498671.

[29] G. Leifert, R. Labahn, T. Gruning, and S. Leifert, "End-to-end measure for text recognition," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, no. Ld, pp. 1424–1431, 2019, doi: 10.1109/ICDAR.2019.00-16.

[30] W. Suryn, *Software Quality Engineering: A Practitioner's Approach*, vol. 9781118592. 2014. doi: 10.1002/9781118830208.

[31] C. M. Barnum, *Usability Testing Essentials: Ready, Set…Test!* 2020. doi: 10.1016/B978-0-12-816942-1.01001-7.

[32] Y. Sari, M. Arafah, and Novitasari, "Evaluasi Usability Sistem Informasi Akademik Dosen Menggunakan User Experience Questionnaire dan Heuristic Walkthrough," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 247–253, 2021, doi: 10.29207/resti.v5i2.3022.

[33] Darussalam and G. Arief, "Jurnal Resti," *Resti*, vol. 1, no. 1, pp. 19–25, 2018.