Proceedings of Engineering and Technology Innovation, vol. 33, 2026, pp. 51-66

Dietary Supplement Products Registration Verification Application Using Thai FDA Number Images with OpenAI Vision API

Arnon Phongrusamepane, Nattapat Thongkaew, Teetach Chuenkamonrat, Chalothon Chootong,
Chatchai Kasemtaweechok*

Department of Computer Science and Information, Faculty of Science at Sriracha, Kasetsart University, Bangkok, Thailand

Received 15 September 2025; received in revised form 11 February 2026; accepted 23 February 2026

DOI: https://doi.org/10.46604/peti.2026.15693

Abstract

Thai Food and Drug Administration (FDA) registration numbers displayed on product labels are essential for verifying the authenticity of dietary supplements. This study aims to develop a Dietary Supplement Verification Application (DSVAPP) that enables consumers to upload images of product labels containing FDA registration numbers to authenticate products. The system uses the OpenAI vision Application Programming Interface (API) to extract registration numbers, thereby reducing error propagation between object detection and Optical Character Recognition (OCR). DSVAPP is evaluated using 118 images and nine combinations of object detection and OCR methods. The results demonstrate that the DSVAPP incorporating the GPT-5.2 vision model achieves superior performance compared to most evaluated methods.

Keywords: Thai FDA number, dietary supplement products verification, OpenAI vision API, optical character recognition

1. Introduction

Health and wellness awareness among Thai people has been growing rapidly, leading to increased demand for dietary supplement products. For example, in 2023, the market was valued at approximately 77.8 billion baht (about USD 2.2 billion), growing at a compound annual growth rate of 5.9% from 2024 to 2027 [1]. The key product categories include beauty supplements and sports nutrition supplements, as well as traditional vitamins, minerals, and herbal remedies [2].

The rapidly growing market has attracted more manufacturers and distributors, which in turn has increased the risk of substandard or fraudulent products entering the supply chain. The Thai Food and Drug Administration (FDA), a department under the Ministry of Public Health of Thailand, is responsible for regulating dietary supplements in Thailand. The Thai FDA holds a broad mandate covering various critical aspects of the supplement lifecycle, including the evaluation and registration of dietary supplement products before they can be legally marketed, the issuance of necessary licenses for manufacturing and importation, and the oversight of quality control standards in manufacturing processes [3]. Product validation and registration are essential tools for promoting the Thai dietary supplement industry and protecting consumers by ensuring the safety, efficacy, and overall quality of products throughout their life cycle, from raw material sourcing to consumption.

However, counterfeit and substandard pharmaceuticals are a severe public health concern in several countries [4-6]. Such practices not only deceive consumers but also pose potential health risks. The Thai FDA strives to educate consumers to check for FDA registration numbers on product labels and to verify the registration number through the Thai FDA's official website

---

* Corresponding author. E-mail address: chatchai.ka@ku.th

or application [7]. According to Thai food supplement regulations, three groups of food products must display an FDA registration number on the product label. The first group covers specially controlled foods such as beverages and foods packed in sealed containers. The second group consists of foods with specified quality standards, such as drinking water in sealed containers, dietary supplements, vegetable oil, and semi-finished foods. Finally, the third group involves foods required to have labels, such as meat products, chewing gum, candy, ready-to-eat foods, and ready-to-cook foods [8].

A 13-digit number within the FDA registration symbol (Fig. 1) must be printed on the product label. The 13-digit number is divided into five sections that identify the food production or import site (as applicable), the licensing agency, and the food sequence, according to the information on the respective certificates. Fig. 2 shows samples of Thai FDA registration numbers on a product label.

Fig. 1 FDA registration symbol and number format [8]

Fig. 2 FDA registration numbers on product label samples

Despite the growing availability of product images across e-commerce and social media platforms, there is currently no fully automated, image-based system designed for verifying Thai FDA registration numbers. Existing solutions focus primarily on text-based validation or require manual data entry, which fails to leverage the visual information present on product packaging. Manual inspection and database-based text searches rely heavily on human judgment, are time-consuming, and are prone to human error, particularly when processing large volumes of product images. Moreover, manual verification is ineffective when FDA numbers are partially obscured, distorted, or embedded within complex packaging designs. Such approaches also lack scalability and are unsuitable for real-time or large-scale monitoring of online product listings.

Optical Character Recognition (OCR) technology involves converting images of text into machine-readable text formats and has been widely adopted for many applications. However, its application still involves certain inaccuracies. Several research efforts have sought to address these issues through various methods, such as post-recognition error correction using N-gram edit distance and neural network models, including Bidirectional Long Short-Term Memory (BiLSTM), Transformer, mT5-base, and mBART-50 [9]. Other approaches included the application of transformer-based models, such as Bidirectional Encoder Representations from Transformers (BERT), the use of ensemble techniques combining multiple methods, and the exploration of semi-supervised learning paradigms [10]. Furthermore, natural language processing techniques were introduced to enhance accuracy by utilizing models such as T5 [11]. Other improvements included the optimization of Tesseract settings, advanced preprocessing techniques, and dictionary-based post-recognition correction mechanisms [12].

Another approach uses the You Only Look Once (YOLO) algorithm, which was originally designed for object detection via bounding boxes in images [13-14]. This approach has been adapted for OCR tasks to locate words or characters and, optionally, to recognize the characters using a modified YOLO architecture [15-16].

The OCR technique has been widely applied in food and drug verification tasks, primarily for extracting nutrition information, expiration dates, and product identifiers from packaging [17-18]. Early approaches focused on rule-based image preprocessing combined with traditional OCR engines, such as Tesseract, to recognize textual information. More recent studies incorporated deep learning–based OCR frameworks to improve recognition accuracy under varying lighting conditions, font styles, and complex backgrounds [19-20].

However, these methods generally require manual region selection or assume clear visibility of the target text [21]. Moreover, most existing OCR-based systems focus on interpreting text within broader visual and semantic contexts through the integration of OCR, natural language processing (NLP), large language models (LLMs), and contextual reasoning. Nevertheless, these approaches exhibit limitations in large-scale or automated compliance monitoring scenarios [22].

Other researchers applied OCR technology to pharmaceutical data extraction and verification, including the recognition of FDA-related information. For example, some research utilized an OCR model to extract text from medical prescriptions, including drug names and dosages [23-24]. In other studies, OCR was applied in automated label verification systems for food and pharmaceutical products by converting product labels into digital data that were linked to compliance databases [25-26]. These research applications enhanced security and reliability, addressing the growing threat of counterfeit products.

In addition, OCR has been used to read and verify critical information in pharmaceutical packaging, such as lot numbers and expiration dates, which are tied to FDA regulatory compliance [27-28]. Although there are limited direct methods using OCR specifically for FDA number verification, the research studies discussed above showed the potential of OCR technology in related pharmaceutical data extraction and label verification tasks. The integration of OCR with online databases is a promising direction for verifying FDA registration numbers.

Regardless of advancements in OCR technologies, there remains a lack of integrated, automated image-based verification systems specifically tailored to Thai FDA labels. Prior studies focus mainly on standalone OCR accuracy without addressing end-to-end verification, including label localization, robust text extraction, and intelligent validation against FDA records. Consequently, the research gap lies in the absence of a unified framework that combines object detection, OCR, and intelligent verification mechanisms capable of handling real-world packaging variability. This study addresses this gap by proposing a Generative AI–assisted verification framework that improves robustness, automation, and scalability compared with traditional OCR-based approaches.

The current study proposes a dietary supplement verification application (DSVAPP) in which consumers input FDA label images for product authentication. The proposed system utilizes OpenAI vision Application Programming Interface (API) to extract and interpret FDA numbers from the input images based on the OCR process. Then, it queries product information in the FDA database using a crawling API. The approach is expected to provide high accuracy, precision, recall, and F1-score in verifying FDA registration numbers.

2. Application for Dietary Supplement Product Registration Verification

This paper presents an application designed to verify Thai FDA registration numbers displayed on dietary supplement product labels using a Generative AI-based approach. The proposed system aims to improve the accuracy and efficiency of product registration number validation under real-world imaging conditions. Fig. 3 illustrates the overall process flow and system architecture of the proposed framework. The workflow consists of six main steps: image acquisition, image preprocessing, number extraction, database validation, data storage, and result reporting. This structured pipeline ensures systematic verification and enhances reliability in practical deployment scenarios.

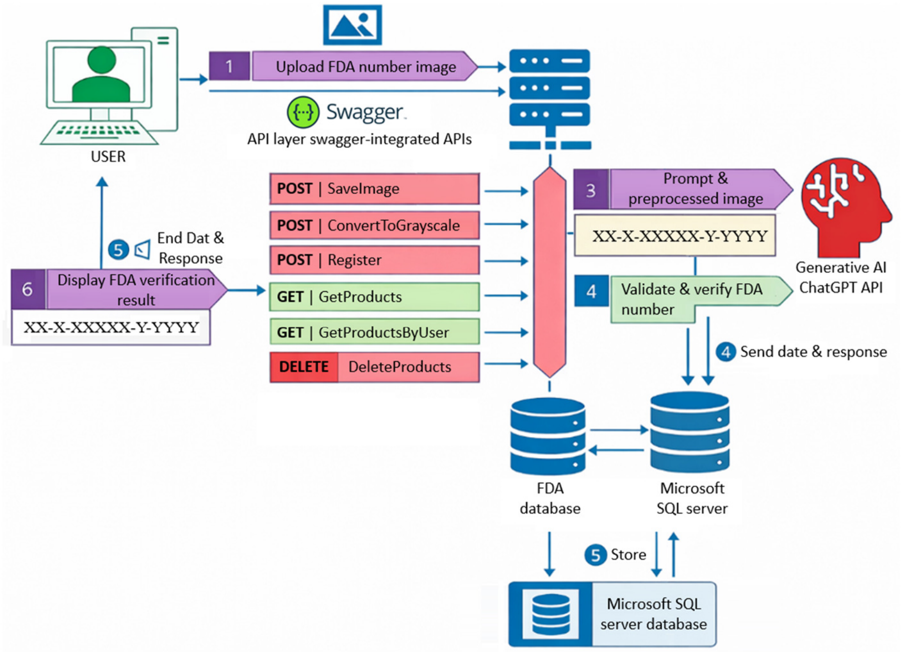
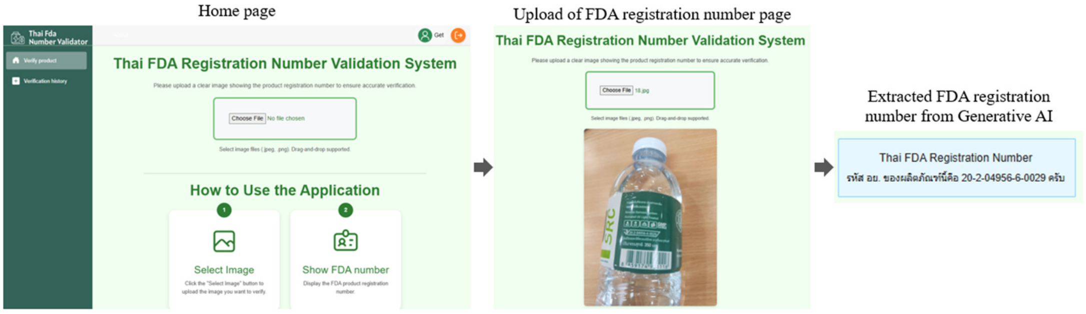


Fig. 3 Process flow and system architecture of DSVAPP

To search for product information from an image containing an FDA registration number, the user uploads the image as the initial step of the methodology. The uploaded image is submitted through the Swagger-integrated API layer, where the SaveImage endpoint stores the image in a temporary data repository. Next, the ConvertToGrayscale endpoint converts the image into grayscale format to facilitate text extraction. Subsequently, the GetProducts endpoint generates a structured prompt and sends both the prompt and the processed image to the ChatGPT vision API, which extracts the FDA registration number from the product label using OCR-based processing.

The ChatGPT API then returns the formatted FDA number to the GetProducts endpoint. This extracted number is used as a key to query product information via a web-crawling API, after which the FDA database responds with the corresponding product details. The uploaded image and retrieved product information are subsequently stored in Microsoft SQL Server. Finally, the DSVAPP displays the search results to the user. In addition, the system preserves both the uploaded image and the search results in the database, enabling users to access their search history later.

Fig. 4 shows the web interfaces provided by the web server. After the user logs in successfully, DSVAPP displays the home page, which explains the verification steps. The user then uploads a product image containing the FDA registration number, and the extracted FDA registration number is displayed in a light blue box using Generative AI, as shown in Fig. 4(a). Fig. 4(b) presents the product information retrieved from the Thai FDA database, which consists of the FDA registration number, Thai and English product names, production type, product and license status, the licensee company, location, and a link to the product page in the Thai FDA database. Fig. 4(c) shows a sample of a product license information page from the Thai FDA database.



Thai FDA Registration Number Validation System

Please upload a clear image showing the product registration number to ensure accurate verification.

Select image files (.jpeg, .png). Drag-and-drop supported.

How to Use the Application

Click the "Select Image" button to upload the image you want to verify.

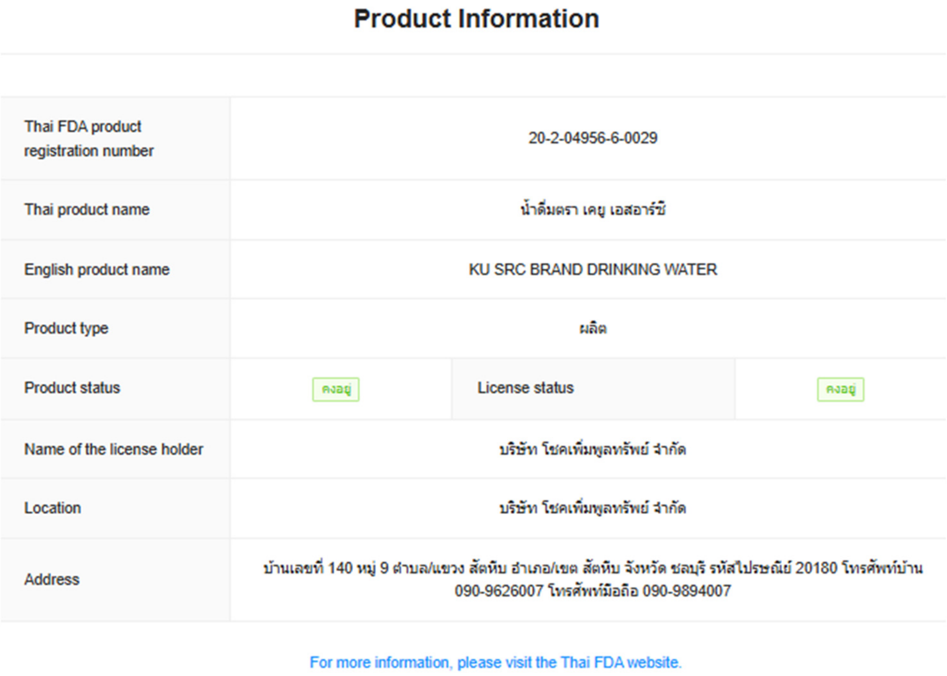
Display the FDA product registration number.

Thai FDA Registration Number

รหัส อย. ของผลิตภัณฑ์นี้คือ 20-2-04956-6-0029 ครับ

(a) Home, image upload, and FDA number extraction interface

Fig. 4 Web interface screens of DSVAPP

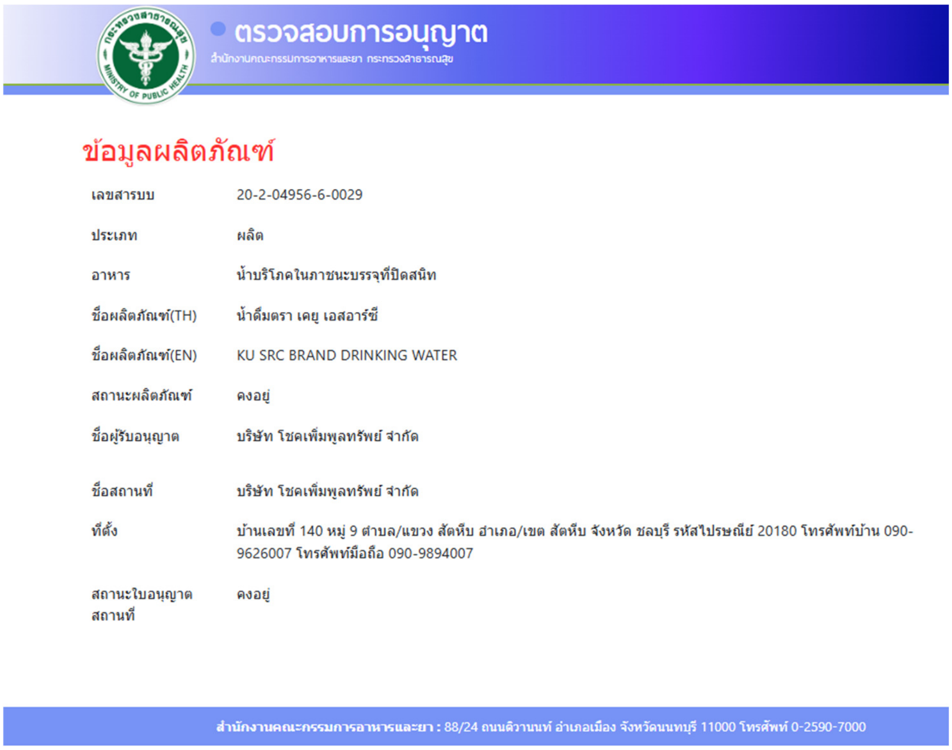


Product Information

Thai FDA product registration number	20-2-04956-6-0029		
Thai product name	น้ำดื่มตรา เคยู เอสอาร์ซี		
English product name	KU SRC BRAND DRINKING WATER		
Product type	ผลิต		
Product status	คงอยู่	License status	คงอยู่
Name of the license holder	บริษัท โชคเพิ่มพูลทรัพย์ จำกัด		
Location	บริษัท โชคเพิ่มพูลทรัพย์ จำกัด		
Address	บ้านเลขที่ 140 หมู่ 9 ตำบล/แขวง สัตหีบ อำเภอ/เขต สัตหีบ จังหวัด ชลบุรี รหัสไปรษณีย์ 20180 โทรศัพท์บ้าน 090-9626007 โทรศัพท์มือถือ 090-9894007		

For more information, please visit the Thai FDA website.

(b) Product information results page in DSVAPP



ข้อมูลผลิตภัณฑ์

เลขสารบบ	20-2-04956-6-0029
ประเภท	ผลิต
อาหาร	น้ำบริโภคในภาชนะบรรจุที่ปิดสนิท
ชื่อผลิตภัณฑ์(TH)	น้ำดื่มตรา เคยู เอสอาร์ซี
ชื่อผลิตภัณฑ์(EN)	KU SRC BRAND DRINKING WATER
สถานะผลิตภัณฑ์	คงอยู่
ชื่อผู้รับอนุญาต	บริษัท โชคเพิ่มพูลทรัพย์ จำกัด
ชื่อสถานที่	บริษัท โชคเพิ่มพูลทรัพย์ จำกัด
ที่ตั้ง	บ้านเลขที่ 140 หมู่ 9 ตำบล/แขวง สัตหีบ อำเภอ/เขต สัตหีบ จังหวัด ชลบุรี รหัสไปรษณีย์ 20180 โทรศัพท์บ้าน 090-9626007 โทรศัพท์มือถือ 090-9894007
สถานะใบอนุญาต สถานที่	คงอยู่

สำนักงานคณะกรรมการอาหารและยา : 88/24 ถนนติวานนท์ อำเภอเมือง จังหวัดนนทบุรี 11000 โทรศัพท์ 0-2590-7000

(c) Product license information page from the Thai FDA database

Fig. 4 Web interface screens of DSVAPP (continued)

Algorithm 1 presents the pseudocode of the DSVAPP framework, which is designed to extract Thai FDA registration numbers from product label images using ChatGPT and evaluate system performance. First, DSVAPP initializes the OpenAI vision API client. Next, the ground-truth annotation file is loaded, containing a list of product label images and their corresponding ground-truth FDA numbers. Each product label image is then uploaded to the OpenAI API for analysis. After the API returns either the extracted Thai FDA number or a non-detection result, DSVAPP compares the API output with the ground-truth value. If the extracted result matches the ground-truth label, y_pred is set to 1; otherwise, it is set to 0. Finally, the detection results for method m are evaluated in terms of $Precision_m$, $Recall_m$, and $F1\text{-}score_m$, while the processing times for all images are averaged to obtain the average processing time.

Algorithm 1 FDA registration label detection and evaluation

```
Initialize OpenAI vision API client
Load ground-truth annotations from CSV file
Set image directory path
Initialize empty result list $R$
for each image $I_i$ in the dataset do
    Load image $I_i$
    Encode image $I_i$ into Base64 format
    Start timer
    Send encoded image and textual prompt to ChatGPT vision model
    Receive textual response from model
    Extract FDA registration numbers using a regular expression
    Stop timer
    if the extracted FDA number matches ground-truth $G_i$ then
        $y\_pred \leftarrow 1$
    else
        $y\_pred \leftarrow 0$
    end if
    $y\_true \leftarrow 1$
    Record method name, $i$, elapsed time, $y\_true$, $y\_pred$ into $R$
end for
for each detection method $m$ in $R$ do
    Compute $Precision_m$
    Compute $Recall_m$
    Compute $F1\text{-}score_m$
    Compute average processing $Time_m$
end for
Output evaluation summary table
Send the extracted FDA number to query product information via a web-crawling API
Retrieve product information stored in Microsoft SQL Server
Display the search results to the user
```

3. Experiments

This section describes the experimental setup and evaluation protocol used to assess the proposed framework. Performance is quantitatively measured using Accuracy Rate (AR) and Zero-Result Rate (ZRR) to evaluate extraction effectiveness, as well as precision, recall, and F1-score to assess detection and recognition performance. All methods are tested under identical conditions to ensure fair comparison. Statistical analyses, including normality testing and paired significance tests, are conducted to determine whether the observed performance differences are statistically significant. These procedures provide a rigorous and reproducible evaluation of the proposed approach.

3.1. Dataset

The dataset consists of 118 product label images randomly collected from shelves in three well-known convenience store chains. The study focuses on drinks, food, and dietary supplement product categories. Half of the dataset (59 images) is used for model development, of which 42 images are employed to train the object detection models, while the remaining images are reserved for testing. After retrieving product information from the Thai FDA database, DSVAPP stores the extracted information in a product table, as described in Table 1. The stored product information can be reused to support future user queries.

Table 1 Product information table

Columns	Data type	Sample
ProductID	INT (11)	7410745550111
RegisterationNumber	Varchar (50)	74-1-07455-5-0111
URL	Varchar (100)	https://porta.fda.moph.go.th/FDA_SEARCH_ALL/PRODUCT.aspx?fdpdtno=7410745550111
Type	Varchar (100)	Produced
THName	Varchar (255)	ผลิตภัณฑ์เสริมอาหาร
ENGName	Varchar (255)	Dietary Supplement Product
Status	Varchar (100)	Active
Licensee	Varchar (50)	บริษัท ดีอีที จำกัด
Location	Varchar (255)	DET Company Ltd.
Address	Varchar (255)	111 Moo 3 Sukhumvit Road, Chonburi
LicenseStatus	Varchar (100)	Active
CreateDate	Datetime	2025-02-20 10:25:54
UpdateDate	Datetime	2025-02-20 10:25:54

3.2. Performance comparison test

In this experiment, two versions of DSVAPP are evaluated using 118 product label images. The input images are categorized into three types: 23 boxes, 35 packs, and 60 bottles. The first version (DSVAPPV1) sends the original image directly to the OpenAI vision API to extract the formatted Thai FDA number. In contrast, the second version (DSVAPPV2) first preprocesses the original image into grayscale before sending it to the Generative AI. Two metrics are used to evaluate the performance of the two versions: AR and ZRR. The corresponding formulations are given in the following equations.

$$\text{AR} = \sum_{k=1}^{3}\left[\frac{n(C_k)}{n}\right] \tag{1}$$

where $n(C_k)$ is the number of correctly detected numbers for the package k, and n is the total number of images. The larger the AR value, the better the detection performance [29].

$$\text{ZPR} = \left(\frac{n_{zero}}{n}\right)\times 100 \tag{2}$$

where n_{zero} is the number of FDA numbers not found in the FDA database, and n is the total number of images. The smaller the ZRR value for the compared method, the better the product information retrieval performance.

Moreover, DSVAPPv1 and DSVAPPv2 are compared with nine combined models formed by integrating four object detection models with three OCR models to evaluate their performance in Thai FDA number detection. The performance of both versions is assessed using four evaluation metrics: precision, recall, F1-score, and average processing time.

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{3}$$

where true positive (TP) is the number of correctly detected and recognized Thai FDA numbers. False positive (FP) is the number of incorrectly detected regions or wrongly recognized numbers. False negative (FN) is the number of Thai FDA numbers that exist but were not detected or recognized. Precision measures the proportion of correctly detected FDA numbers among all detected instances. A higher precision indicates a lower false detection rate and better product label detection performance for the compared method.

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{4}$$

Recall measures the proportion of actual FDA numbers that are successfully detected. A higher recall indicates better Thai FDA number detection performance by reducing missed detections.

$$\text{F1-score} = 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{5}$$

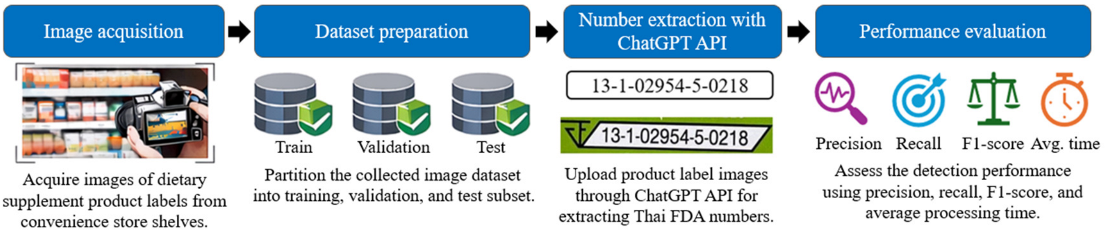
F1-score is the harmonic mean of precision and recall, providing a balanced measure of both metrics. A higher F1-score indicates better overall performance by achieving a more balanced trade-off between precision and recall. Lastly, average processing time represents the mean time required to process a single image, including both Thai FDA number detection and recognition.

3.3. *Statistical test for performance comparison*

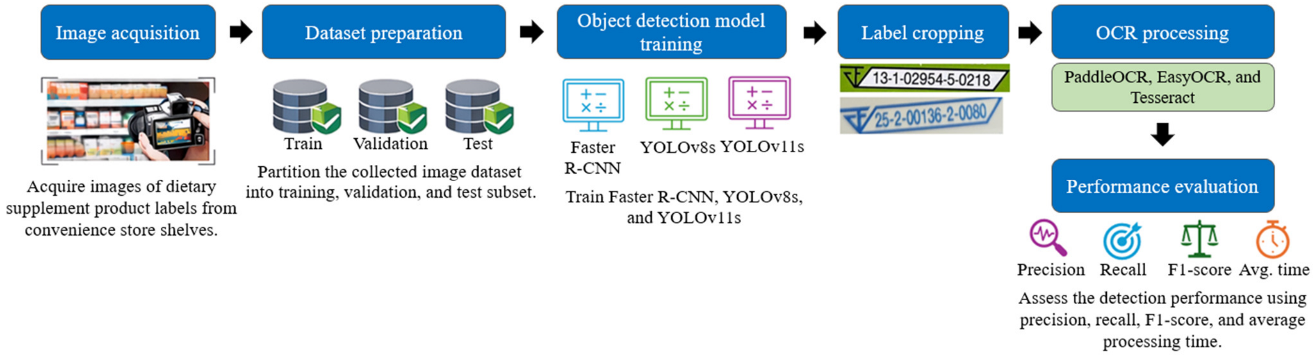
To evaluate the performance differences between the proposed DSVAPP method and the compared methods, paired statistical tests were conducted on recall and F1-score. Since all methods were evaluated on the same set of images, a paired comparison design was adopted. For each paired comparison, the normality of the performance differences was first examined using the Shapiro–Wilk test. When the normality assumption was satisfied ($p \geq 0.05$), a paired t-test was applied. Otherwise, the non-parametric Wilcoxon signed-rank test was employed.

A one-sided hypothesis test was performed to determine whether the DSVAPP method significantly outperformed the compared methods. The null hypothesis is defined as H_0: the average performance of the DSVAPP method is less than or equal to that of the paired method. The alternative hypothesis is defined as H_1: the average performance of the DSVAPP method is greater than that of the paired method. Statistical significance was determined at a confidence level of 95% ($\alpha = 0.05$).

3.4. *Performance comparison with other object detection and OCR methods*



(a) DSVAPP



(b) Compared methods

Fig. 5 Workflow for performance evaluation of DSVAPP and compared methods

To evaluate DSVAPP, Fig. 5(a) presents the overall performance evaluation workflow, while Fig. 5(b) illustrates the Thai FDA number recognition workflow for dietary supplement products based on object detection and OCR models. The detailed evaluation workflow of DSVAPP is described as follows:

(1) Acquire images of dietary supplement product labels from convenience store shelves under real-world conditions.
(2) Partition the collected image dataset into training, validation, and test subsets.
(3) Upload 26 product label images from the test subset through the ChatGPT API to extract Thai FDA numbers.
(4) Assess the performance of Thai FDA number detection and recognition using precision, recall, F1-score, ZRR, and average processing time (in seconds) as evaluation metrics.

Furthermore, the workflow of Thai FDA number recognition for dietary supplement products using object detection and OCR models is described as follows:

(1) Acquire images of dietary supplement product labels from convenience store shelves under real-world conditions.
(2) Partition the collected image dataset into training, validation, and test subsets.
(3) Train and validate multiple object detection models, including Faster Region-based Convolutional Neural Network (Faster R-CNN), YOLOv8s, and YOLOv11s, using 40 labeled training images.
(4) Evaluate the trained detection models on 26 test images by localizing and cropping Thai FDA number label regions.
(5) Apply three widely used OCR models: PaddleOCR, EasyOCR, and Tesseract to extract Thai FDA numbers from the cropped label images.
(6) Assess the performance of Thai FDA number detection and recognition using precision, recall, F1-score, ZRR, and average processing time (in seconds) as evaluation metrics.

3.5. Parameter settings of the compared object detection methods

In this experiment, nine models are evaluated, obtained by combining four object detection models with three OCR models. All compared methods are implemented and evaluated using Python 3 on Google Compute Engine, with 12.7 GB of memory and 107.7 GB of disk storage, via the Google Colab service. The parameter settings for each compared model are presented in Table 2.

Table 2 Parameter settings for each compared model

Models	Parameter settings
ChatGPT API	Model = [GPT-4o mini, GPT-5.2], data: image/jpeg; base64, Text = “Please check whether there is an FDA registration label in the image and describe its location in detail.”
Faster R-CNN	Learning rate = 0.005, momentum = 0.9, weight_decay = 0.0005, Epochs = 10
YOLOv8s	Image size = 640, batch = 16, Epochs = 100, lr0 = 0.01, lrf = 0.01, optimizer = SGD, momentum = 0.937, patience = 50 weight_decay = 0.0005, confidence threshold = 0.5
YOLOv11s	Image size = 640, batch = 16, Epochs = 100, lr0 = 0.01, lrf = 0.01, optimizer = SGD, momentum = 0.937, patience = 50 weight_decay = 0.0005, confidence threshold = 0.5

4. Results and Discussion

This section presents the experimental results and discusses the performance of the proposed framework in detail. First, the two versions of DSVAPP are compared to analyze system improvements. Next, DSVAPP is evaluated against other object detection and OCR-based methods to demonstrate its relative effectiveness. Statistical analyses are then conducted to validate the significance of the performance differences. Finally, the hybrid YOLOv8 and GPT-5.2-based solution is introduced and discussed, highlighting its enhanced detection and recognition capabilities.

4.1. Performance comparison between two versions of DSVAPP applications

Table 3 shows the number of detected FDA registration numbers and the corresponding AR values for DSVAPPV1 and DSVAPPV2 across each package type. A detection status of 0 indicates failure, while 1 indicates success in detecting the FDA registration number. The AR values for DSVAPPV2 were greater than those of DSVAPPV1 across all package types. The overall AR value of DSVAPPV2 (89%) was greater than that of DSVAPPV1 (84%). The additional grayscale preprocessing step in DSVAPPV2 improved detection performance by approximately 4–5%. Among the package types, bottles were the most challenging due to their curved surfaces and high reflectivity, which make image capture more difficult. Additionally, DSVAPPV2 successfully detected FDA registration numbers in all box and pack types, likely due to their flat surfaces and clearer images.

Table 3 AR for DSVAPPV1 and DSVAPPV2

Type	0	1	AR
Bottle (DSVAPPV1)	14	46	76.67%
Bottle (DSVAPPV2)	12	48	80.00%
Box (DSVAPPV1)	1	22	95.65%
Box (DSVAPPV2)	0	23	100.00%
Pack (DSVAPPV1)	3	32	91.43%
Pack (DSVAPPV2)	0	35	100.00%
Overall (DSVAPPV1)	18	100	84.75%
Overall (DSVAPPV2)	12	106	89.83%

Table 4 presents the number of products found in the Thai FDA database and the corresponding ZRR values for DSVAPPV1 and DSVAPPV2 across each package type. A retrieval status of 0 indicates failure, while 1 indicates successful retrieval from the Thai FDA database. Consistent with the AR results, DSVAPPV2 had a lower ZRR than DSVAPPV1 across all package types, with an improvement of approximately 4%. In the bottle package type, seven FDA registration numbers detected by DSVAPPV1 were not found in the Thai FDA database, compared to only four for DSVAPPV2. For the pack package type, only one FDA registration number detected by both methods was not found in the Thai FDA database. Both versions of DSVAPP accurately detected all Thai FDA registration numbers on box-type packages with flat surfaces.

Table 4 ZRR for DSVAPPV1 and DSVAPPV2

Type	0	1	ZRR
Bottle (DSVAPPV1)	7	39	15.22%
Bottle (DSVAPPV2)	3	45	6.25%
Box (DSVAPPV1)	0	22	0.00%
Box (DSVAPPV2)	0	23	0.00%
Pack (DSVAPPV1)	1	31	3.13%
Pack (DSVAPPV2)	1	34	2.86%
Overall (DSVAPPV1)	8	92	8.00%
Overall (DSVAPPV2)	4	102	3.77%

In addition, two key factors affecting the OCR performance of DSVAPPV1 and DSVAPPV2 in recognizing FDA numbers from images were the capture angle and the clarity of the numbers. The FDA number images that were horizontally aligned (Fig. 6) were correctly recognized by both DSVAPPV1 and DSVAPPV2.

Fig. 6 FDA number images with horizontal alignment

Fig. 7 shows that when FDA number images were tilted at an angle greater than 45°, the OCR process did not produce accurate results. To evaluate the effect of the capture angle on OCR performance, 17 pairs of FDA number images were tested using DSVAPPV2. In each pair, one image was tilted at an angle greater than 45°, and the other was horizontally aligned. Table 5 presents a comparison of the AR and ZRR values for these image pairs. All 17 horizontally aligned images (100%) were recognized, while only 4 out of the 17 tilted images (23.53%) were successfully recognized. Among the 17 horizontally aligned images, 16 (94.12%) were found in the Thai FDA database, while only 2 out of the 4 recognized tilted images (50%) were found.

Fig. 7 FDA number image detection images tilted at angles greater than 45°

Table 5 AR and ZRR comparison for FDA number images tilted beyond 45° and horizontally aligned

Image type	Not recognized	Recognized and not found	Recognized and found	AR	ZRR
Tilted > 45°	13	2	2	23.53%	50.00%
Horizontally aligned	0	1	16	100.00%	5.88%

In Fig. 8, the experimental results clearly demonstrated the significant negative impact of capture angle on OCR performance. In addition to the capture angle, image clarity also affects the accuracy of FDA number recognition. The clarity of the image depends on the size and thickness of the font used to display the numbers on the label. Consequently, some images tilted close to 90° were correctly detected, as shown in Fig. 9, because the numbers were captured clearly. To address this problem, DSVAPP integrates an object detection model to first localize, crop, and rotate regions containing Thai FDA registration numbers. The cropped regions are then sent to the OpenAI vision model for Thai FDA number extraction.

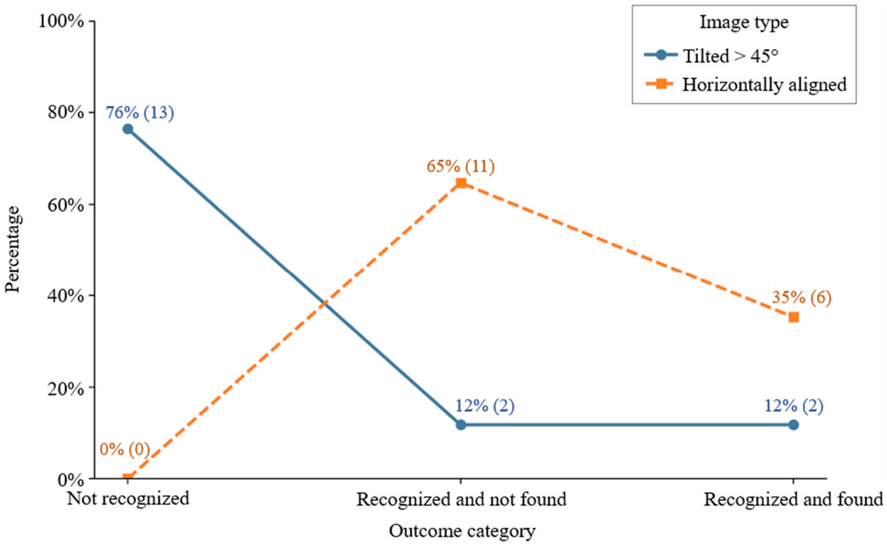


Fig. 8 Detection and database matching results for FDA number images in DSVAPPV2

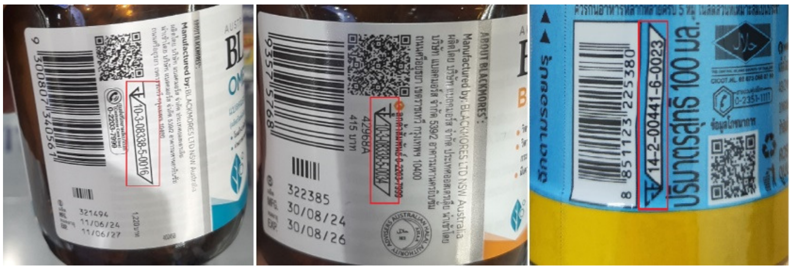


Fig. 9 FDA number images tilted at angles close to 90°

4.2. Performance results with DSVAPP and other object detection and OCR methods

The experimental results provided clear evidence that DSVAPP outperformed conventional object detection and OCR combinations across all evaluated metrics, as shown in Fig. 10. Although Faster R-CNN and YOLOv11s methods demonstrated relatively high precision, their lower recall indicated a tendency to miss valid FDA numbers, which is undesirable for Thai FDA detection applications. In contrast, YOLOv8s combined with Paddle-OCR achieved performance comparable to that of both DSVAPP versions based on the GPT-5.2 model.

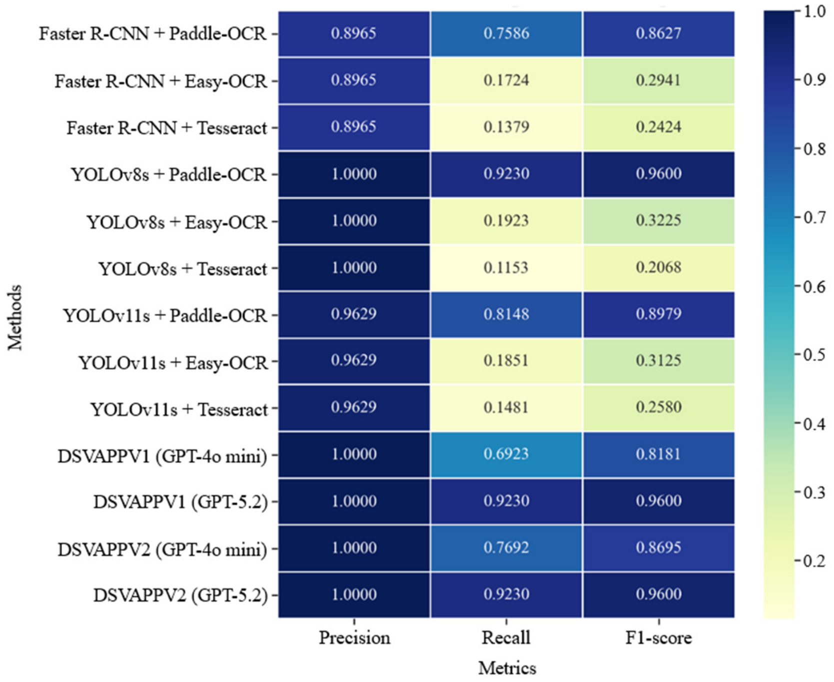


Fig. 10 Precision, recall, and F1-score for DSVAPPV1, DSVAPPV2, and 9 combined methods

In product label–based OCR tasks, where textual regions are typically small and densely packed, the choice of object detection model has a direct impact on overall recognition performance. Among the evaluated models, YOLOv8s demonstrates the strongest performance due to its deeper backbone and wider feature channels, which enable effective multi-scale feature extraction and precise localization of small text regions. In contrast, YOLOv11s prioritizes computational efficiency and inference speed, making it suitable for real-time or embedded systems. However, its reduced representational capacity limits its effectiveness in detecting very small text regions and densely packed labels, leading to a higher likelihood of missed detections in complex OCR scenarios. As a result, YOLOv11s is better suited for applications with strict hardware constraints rather than high-accuracy OCR tasks.

The experimental results indicated that the performance values of Paddle-OCR were consistently higher than those of Easy-OCR and Tesseract. Paddle-OCR achieved higher performance than Easy-OCR and Tesseract due to its advanced deep learning architecture, large-scale multilingual training data, and strong support for Thai text. Its integrated text detection and recognition pipeline enables more accurate localization and recognition of FDA numbers, especially under challenging conditions such as complex backgrounds and low image quality.

A comparison between the two versions of DSVAPP revealed that DSVAPPV2 consistently delivered incremental improvements across all evaluation metrics. The performance of the GPT-4o mini model in DSVAPPV2 was superior to that in DSVAPPV1, whereas the performance of the GPT-5.2 model remained unchanged across both versions for all metrics. Because GPT-5.2 combines visual perception with semantic reasoning, it can identify Thai FDA numbers even when the text is distorted, curved, or embedded in complex layouts, and can distinguish FDA numbers from visually similar numeric strings. Overall, DSVAPP maintained a favorable balance between precision and recall, resulting in higher F1-scores that indicate more reliable and stable performance.

In this experiment, the processing time of all compared methods was divided into two components: cropping time, which refers to the time required to localize and crop Thai FDA regions using object detection models, and OCR time, which denotes the time required to recognize Thai FDA numbers from the cropped regions using OCR models. DSVAPP does not include cropping time, as it retrieves Thai FDA numbers directly from the OpenAI vision API after the product label image is submitted. In Fig. 11, the Tesseract OCR–based models exhibited the fastest OCR processing time.

Consequently, YOLOv8 and YOLOv11 combined with Tesseract achieved the lowest overall processing times, ranking first and second, respectively. Among all DSVAPP variants, DSVAPPV1 with the GPT-4o mini model achieved the lowest processing time. Due to the additional grayscale preprocessing step, DSVAPPV2 incurred a higher processing time than DSVAPPV1. When comparing the processing times of the two highest-performing methods, YOLOv8 combined with Paddle-OCR achieved a lower processing time than DSVAPPV1 with the GPT-5.2 model.

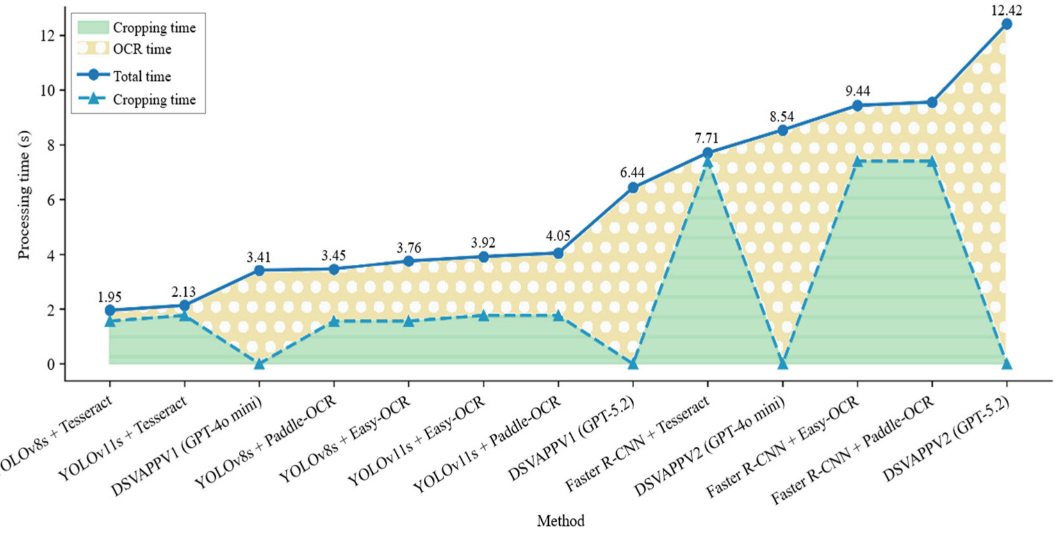


Fig. 11 Cropping and OCR time for DSVAPPV1, DSVAPPV2, and 9 combined methods

4.3. Statistical performance comparison

Table 6 Shapiro–Wilk, paired t-test, and Wilcoxon signed-rank test for recall and F1-score comparison

DSVAPP method	Compared method	Shapiro–Wilk *p*-value	Selected test	p-value	Result
DSVAPPV1 (GPT-5.2)	YOLOv8s + Paddle-OCR	0.0033 (Reject H_0)	Wilcoxon signed-rank (greater)	0.625	Accept H_0
	YOLOv11s + Paddle-OCR	0.0048 (Reject H_0)		0.046	Reject H_0
	Faster R-CNN + Paddle-OCR	0.0321 (Reject H_0)		0.002	Reject H_0

Based on the Shapiro–Wilk normality test, the performance differences between DSVAPPV1 (GPT-5.2) and all other methods were not normally distributed ($p < 0.05$). Therefore, the Wilcoxon signed-rank test (one-sided, greater) was selected for all paired comparisons. The statistical test results indicated that the null hypothesis was accepted for the comparison between DSVAPPV1 (GPT-5.2) and YOLOv8s + Paddle-OCR ($p = 0.625$), suggesting that the recall and F1-score performance of DSVAPPV1 was not significantly greater than that of YOLOv8s + Paddle-OCR.

In contrast, the null hypothesis was rejected for the comparisons between DSVAPPV1 (GPT-5.2) and YOLOv11s + Paddle-OCR ($p = 0.046$), as well as between DSVAPPV1 (GPT-5.2) and Faster R-CNN + Paddle-OCR (p = 0.002). These results demonstrated that DSVAPPV1 (GPT-5.2) achieved a statistically significant improvement in recall over both YOLOv11s + Paddle-OCR and Faster R-CNN + Paddle-OCR at the 0.05 significance level. Overall, the hypothesis tests confirmed that DSVAPPV1 (GPT-5.2) significantly outperformed most of the compared object detection and OCR-based methods in terms of recall and F1-score, except for YOLOv8s + Paddle-OCR, where no statistically significant difference was observed, as shown in Table 6.

4.4. *Hybrid YOLOv8 and GPT-5.2-based solution*

Because of the high performance achieved by YOLOv8s and DSVAPPV1 (GPT-5.2), the combination of YOLOv8s with the GPT-5.2 vision model represents an effective solution for Thai FDA number extraction from product label images. In this framework, YOLOv8s first localizes and crops regions containing Thai FDA numbers. Subsequently, DSVAPP forwards the cropped regions to the GPT-5.2 vision model for Thai FDA number extraction. This hybrid approach leverages the high processing speed of YOLOv8s for region localization and cropping, while the GPT-5.2 vision API is used as an additional OCR component to handle challenging or complex FDA regions.

Table 7 presents the performance results of the combined YOLOv8s and DSVAPPV1 (GPT-5.2) models. The results demonstrated that YOLOv8s + DSVAPPV1 (GPT-5.2) achieved 2–3% higher performance across all evaluation metrics. The use of YOLOv8s-cropped regions slightly improved Thai FDA number detection performance with GPT-5.2. Moreover, this approach reduced processing time by approximately a factor of two compared with the stand-alone DSVAPPV1 (GPT-5.2).

Table 7 Performance of YOLOv8s and DSVAPPV1 (GPT-5.2) combination models

Compared method	Precision	Recall	F1-score	Avg. time
YOLOv8s + DSVAPPV1 (GPT-5.2)	1.0	0.9516	0.9752	3.7144

However, DSVAPP still produced errors in Thai FDA number detection when image quality issues, such as motion blur and poor illumination, accounted for a substantial portion of failures and negatively affected both detection and recognition stages. Fig. 12 illustrates representative failure cases in Thai FDA number detection and recognition. These errors are mainly caused by low character–background contrast, slanted or decorative text layouts, and image blur. Such conditions led to inaccurate region localization by YOLOv8s and produced cropped regions with incomplete or distorted characters, which negatively affected the recognition performance of the GPT-5.2 vision model.

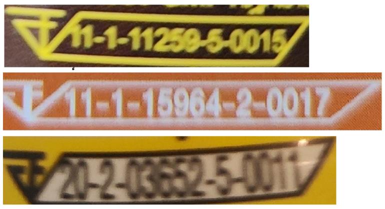


Fig. 12 Failure cases in Thai FDA number detection and recognition

Finally, these failure cases indicate that although the proposed YOLOv8s + DSVAPPV1 (GPT-5.2) framework improves overall performance, accurate Thai FDA number extraction remains sensitive to region quality, contrast conditions, and text geometry. Addressing these challenges may require enhanced image preprocessing, rotated or polygon-based text detection, and fine-tuned OCR models trained on stylized and degraded Thai FDA number samples.

5. Conclusions

This study presented DSVAPP, an application designed to verify Thai FDA registration numbers on dietary supplement product labels using vision–language models. The framework leveraged the OpenAI vision API with OCR capability to extract Thai FDA numbers directly from full product label images without requiring a separate cropping stage. Experimental results demonstrated that GPT-4o mini and GPT-5.2 achieved effective end-to-end extraction under real-world conditions, while hybrid integration with YOLOv8 further enhanced performance. The findings confirmed the feasibility of deploying vision–language models for regulatory verification tasks in practical retail environments.

The main conclusions of this study are summarized as follows:

(1) End-to-end extraction using GPT-based vision models simplified the system architecture by eliminating the object detection–OCR pipeline and reducing error propagation.

(2) Both GPT-4o mini and GPT-5.2 achieved high recall in detecting Thai FDA registration numbers, and demonstrated robustness against complex label layouts, diverse fonts, curved text, distortion, and varying image quality.

(3) Hybrid approaches integrating YOLOv8 with GPT-5.2 provided improved recognition accuracy and enabled flexible trade-offs between processing speed and performance.

(4) Vision–language models were able to function effectively as standalone solutions or as complementary components in regulatory verification systems.

(5) Current limitations included restricted dataset size and diversity, as well as performance degradation under low contrast, blurred images, stylized fonts, and slanted text.

Future work will focus on improving detection robustness under challenging imaging conditions, optimizing hybrid model coordination for more stable recognition performance, and extending the framework to a mobile application with text-to-speech functionality to enhance accessibility and usability, particularly for elderly users.

Acknowledgments

This research was financially supported by the Faculty of Science at Sriracha and the Kasetsart University Research and Development Institute (KURDI), Kasetsart University, Bangkok, Thailand.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] S. T. Thota, "Thailand Probiotic Food Supplements Market Share Analysis, Demand, Trends, Industry and Market 2024-2027," https://www.datamintelligence.com/research-report/thailand-probiotic-food-supplements-market, 2024.

[2] D. Kalman, "Dietary Supplements: Popular vs. Validated?" https://www.researchgate.net/publication/385682129_Asia_Fitness_Conference_2024_Dietary_Supplements_Popular_vs_Validated, 2024.

[3] S. Limsuwan, N. Awaeloh, P. Na-Phatthalung, T. Kaewmanee, and S. Chusri, "Exploring Antioxidant Properties of Standardized Extracts from Medicinal Plants Approved by the Thai FDA for Dietary Supplementation," Nutrients, vol. 17, no. 5, article no. 898, 2025.

[4] N. Hamelin, S. Nwankwo, and R. El Hadouchi, "'Faking Brands': Consumer Responses to Counterfeiting," Journal of Consumer Behaviour, vol. 12, no. 3, pp. 159-170, 2013.

[5] L. Mhando, M. B. Jande, A. Liwa, S. Mwita, and K. J. Marwa, "Public Awareness and Identification of Counterfeit Drugs in Tanzania: A View on Antimalarial Drugs," Advances in Public Health, vol. 2016, no. 1, article no. 6254157, 2016.

[6] P. Karungamye, "Counterfeit and Substandard Drugs in Tanzania: A Review," Forensic Science International: Reports, vol. 7, article no. 100302, 2023.

[7] C. Inthong, S. Lerkiatbundit, S. Mekruksavanich, and P. Hanvoravongchai, "Assessing Dietary Supplement Misinformation on Popular Thai E-Marketplaces: A Cross-Sectional Content Analysis," Thai Journal of Pharmaceutical Sciences, vol. 48, no. 4, article no. 4, 2024.

[8] Food Division, "Food Labeling and Examples," https://food.fda.moph.go.th/for-entrepreneurs/food-labeling-and-examples, accessed in 2025. (In Thai)

[9] T. Aung, Y. K. Thu, and M. N. Oo, "myOCR: Optical Character Recognition for Myanmar Language with Post-OCR Error Correction," 19th International Joint Symposium on Artificial Intelligence and Natural Language Processing, pp. 1-6, 2024.

[10] A. Mishra, J. Sikdar, and S. U. Kumar, “Deep Learning-Based Optical Character Recognition for Robust Real-World Conditions: A Comparative Analysis,” 15th International Conference on Computing Communication and Networking Technologies, pp. 1-7, 2024.

[11] A. Singh, S. Jangra, and G. Aggarwal, “EnvisionText: Enhancing Text Recognition Accuracy through OCR Extraction and NLP-Based Correction,” 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 47-52, 2024.

[12] V. Atmakuri and M. Dhanalakshmi, “Improving Text Recognition in Natural Scenes Using Optimized OCR Techniques,” 3rd International Conference on Automation, Computing and Renewable Systems, pp. 943-950, 2024.

[13] Y. T. Hsiao, J. S. Sheu, and H. Ma, “Efficient Object Detection and Intelligent Information Display Using YOLOv4-Tiny,” Advances in Technology Innovation, vol. 9, no. 1, pp. 42-49, 2024.

[14] Y. Liu, J. Zhu, and H. Ma, “Improving the Vehicle Small Object Detection Algorithm of Yolov5,” International Journal of Engineering and Technology Innovation, vol. 15, no. 1, pp. 57-67, 2025.

[15] H. Moussaoui, N. E. Akkad, M. Benslimane, W. El-Shafai, A. Baihan, C. Hewage, et al., “Enhancing Automated Vehicle Identification by Integrating YOLO v8 and OCR Techniques for High-Precision License Plate Detection and Recognition,” Scientific Reports, vol. 14, article no. 14389, 2024.

[16] R. Shashidhar, A. S. Manjunath, R. S. Kumar, M. Roopa, and S. B. Puneeth, “Vehicle Number Plate Detection and Recognition Using YOLO- V3 and OCR Method,” IEEE International Conference on Mobile Networks and Wireless Communications, pp. 1-5, 2021.

[17] A. K. Kavita, V. T. Nileshsingh, S. Khonde, P. Marodkar, and H. Jadhao, “AI Powered Nutritional Label Analysis Using Vision Transformer and GNN for Ingredient Standardization and Regulatory Compliance,” 7th International Conference on Artificial Intelligence and Speech Technology, pp. 1-5, 2025.

[18] S. I. SP, L. Karthikeyan, P. Bhoomika, K. Tejaswi, L. S. Khande, and N. S. Vemishetti, “Automating Nutritional Claim Verification: The Role of OCR and Machine Learning in Enhancing Food Label Transparency,” International Conference on IoT Based Control Networks and Intelligent Systems, pp. 1164-1171, 2024.

[19] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, et al., “EAST: An Efficient and Accurate Scene Text Detector,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 2642-2651, 2017.

[20] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, “Character Region Awareness for Text Detection,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9357-9366, 2019.

[21] S. Long, J. Ruan, W. Zhang, X. He, W. Wu, and C. Yao, “TextSnake: A Flexible Representation for Detecting Text of Arbitrary Shapes,” Computer Vision – ECCV 2018: 15th European Conference, pp. 19-35, 2018.

[22] N. Khan, A. A. H. Ab Rahman, S. M. Rizvi, I. Y. Alshareef, M. N. Marsono, M. P. Bakht, et al., “Systematic Literature Review of Machine Learning Models and Applications for Text Recognition,” IEEE Access, vol. 13, pp. 177647-177670, 2025.

[23] A. Kumar, A. Goyal, B. K. Rai, and S. Sharma, “OCR Based Medical Prescription and Report Analyzer,” AIP Conference Proceedings, vol. 2424, no. 1, article no. 070006, 2022.

[24] M. Ponnuru, S. P. Ponmalar, A. Likhitha, T. B. Sree, and G. G. Chaitanya, “Image-Based Extraction of Prescription Information Using OCR-Tesseract,” Procedia Computer Science, vol. 235, pp. 1077-1086, 2024.

[25] S. K. J. Urumarudappa, C. Tungphatthong, J. Jaipaew, N. Pornputtapong, D. Pakdeesattayapong, S. Vimolmangkang, et al., “Development of a DNA Barcode Library of Plants in the Thai Herbal Pharmacopoeia and Monographs for Authentication of Herbal Products,” Scientific Reports, vol. 12, article no. 9624, 2022.

[26] P. K. Pal and M. Kule, “A New Approach to Pharmaceutical Product Verification Using Barcode and QR Code,” Proceedings of International Conference on Frontiers in Computing and Systems, pp. 223-229, 2023.

[27] S. Kavin and C. P. Shirley, “OCR-Based Extraction of Expiry Dates and Batch Numbers in Medicine Packaging for Error-Free Data Entry,” 7th International Conference on Circuit Power and Computing Technologies, pp. 278-283, 2024.

[28] P. Goyal, N. Goyal, P. Singh, N. Mittal, N. Jindal, and K. Kaur, “Pharmaceutical Drugs Expiry Date Tracking: A Visionary Approach,” Concurrency and Computation: Practice and Experience, vol. 34, no. 28, article no. e7358, 2022.

[29] K. P. Sinaga and M. S. Yang, “Unsupervised K-Means Clustering Algorithm,” IEEE Access, vol. 8, pp. 80716-80727, 2020.