COMP S450 Applied Computing Project

DREAMS: A ONE-STOP DOCUMENT RECOGNITION, EXTRACTION AND MANIPULATION SOLUTION FOR LOGISTICS INDUSTRY

Final Year Project Report

Group name: HKMU Logistics (2023-2024)

Group members:

1155180	Eraka Yip	(BCOMPHIT1)
1265005	Timothy Chan	(BSCHCN)
1284810	Andy Lau	(BCOMPHIT1)

Project Supervisor

Dr. Voyce Li

Course Coordinator

Dr. Jeff Au Yeung

Abstract

Logistics industry is an important sector to Hong Kong economy. As a logistics solution provider, we have implemented an export and import platform that can be utilized by different stakeholders of the industry.

However, it is not possible for us to directly collect the relations between booking data and shipper information because they are considered as proprietary information. Hence, the workflow of requesting booking ownership (also called retrieve booking) by the shipper cannot be digitalized Instead, manual input is required.

In this project, we provide a one-stop solution on this challenge task. Our solution is a document processing system for shippers. Instead of manually inputting booking information one by one, shippers can upload multiple booking documents simultaneously in our solution. State-of-the-art, fine-tuned AI custom models are developed for validating booking documents which include object detection (YOLOv8), document classification and data extraction (DONUT).

Our solution is incorporated into the current logistics platform seamlessly with a digitalized, automated manner workflow that provides better user experience and reducing manual input, also data security is increased by validating more data from AI models.

Declaration

We certify that the work described in this document is original and done by us and we have utilized guidance of our supervisor in completing this project, and that the content which is not our own has been attributed and referenced properly. There should be no copyrighted content without permission to use. There should be no confidential data. We also declare that the contributions listed in Self Appraisal of Contribution are truly and correctly attributed to us.

Signature of the authors and Date

Eraka Yip (1155180) Date: 2024-04-20

Signature

Timothy Chan (1265005) Date: 2024-04-20

Signature

Andy Lau (1284810) Date: 2024-04-20

Signature

Table of Content

List of Tables

List of Figures

1. Introduction

1.1. Overview

In Hong Kong, maritime and port industry has taken the key role in supporting the tiny city as one of the most important trading hubs in the world. Based on statistics in 2022, the whole industry contributed 4.1% (equivalent of HKD 111.8 billion) of the GDP and 2.1% (about 78,400 jobs) of the total employment [1].

Within the industry, export and import services are the most important sector. Facing the challenges from neighboring cities, the government has been committed to promoting smart ports to optimize service capacity and efficiency. A further approach is to establish an automated container terminal, where terminal operations are fully digitalized and automated, i.e. without human intervention [1, p.9], [2].

As a logistics solution provider, we always support the growth of our port community. With our effort, we have implemented an export and import platform that can be utilized by different stakeholders of the industry.

Using the platform, most of the terminal operations are now digitalized and paperless. For example, in the past, manual inspection of documents was required for entry and exit at the terminal gate. This process is no longer required. Instead, the platform automatically consolidates necessary data to the terminal for review. Now, end users (i.e. truck drivers) can simply use the platform's mobile application for a variety of terminal operations, saving a lot of time and reducing human error.

However, as the platform runs, there are still some unsolved problems that hinder the smooth run of the platform. To understand the problems clearly, we first introduce the stakeholders of the platform.

TABLE 1: ROLES AND RESPONSIBILITIES OF DIFFERENT STAKEHOLDERS IN OUR PLATFORM

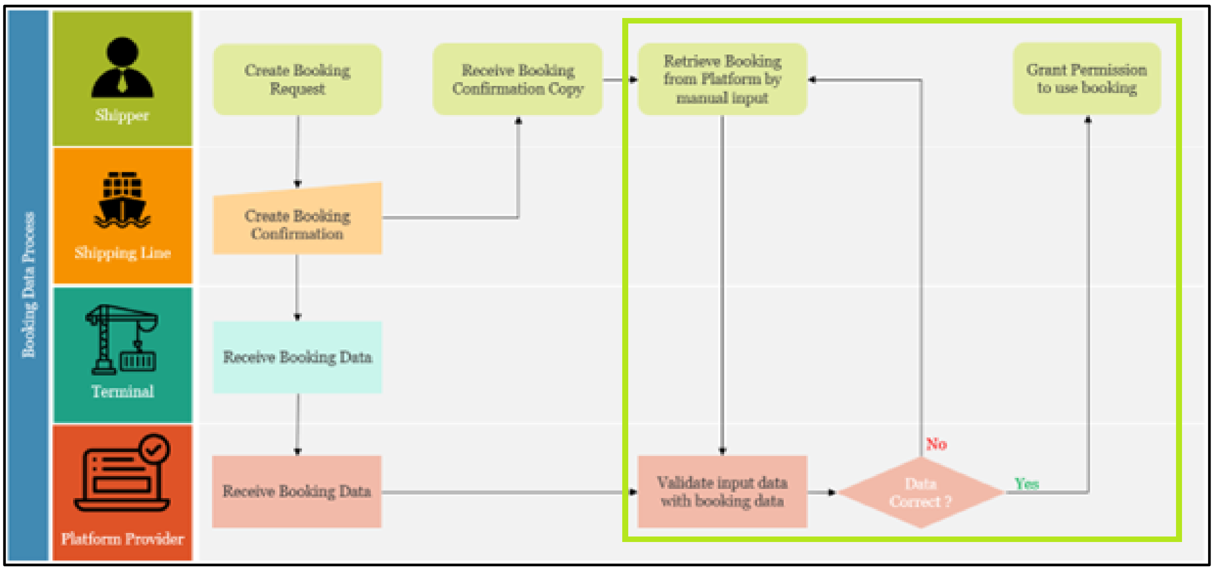
Stakeholders	Roles & Responsibilities
Terminals 	**Logistics:** – Control gate in and gate out of a port. – Provide container storage for pickup and return. – Vessel arrival / departure. **Platform:** – Provide data for the platform that shows information to end users (Shippers and trucking companies).
Shipping Lines 	**Logistics:** – Provide vessels for ocean freight shipping. – Work with terminals on available timeslots for vessel arrival / departure and containers loading / discharge. **Platform:** – Provides data for terminal operation.
Shippers 	**Logistics:** – Request bookings from shipping lines for ocean freight shipping of their goods. **Platform:** – End users of the platform. – Since booking data in the platform does not include shipper information, shippers have to input **booking number** in the booking document **manually** to prove the ownership of a particular booking.

Stakeholders	Roles & Responsibilities
	Then, shippers use the booking ownership as an access token for various automated logistics operations in the current platform (e.g. forward to trucking company, entry and exit of terminal gates).
Trucking companies 	**Logistics:** – Pick up or return containers. **Platform:** – End users of the platform (Roles can be delegated to Shippers) – Drivers can use a mobile app provided by the platform for gate-in and gate-out terminals. – Receive booking from shippers via the platform and assign drivers to pick up or return containers.
Platform provider 	**Platform:** – Establish an online export and import platform for end users (i.e. shippers and trucking companies) to complete export and import operations in a digitalized and paperless manner. – Continuous system integration and maintenance tasks of the existing platform.

[Remarks: For simplification, we assume that end users are shippers only afterwards.]

The following diagram shows the necessary operations among stakeholders in the booking retrieval procedure. The green part is the area to be discussed.

FIGURE 1: BOOKING DATA PROCESS FLOW



A shipper carries out the following steps for terminal operations using the platform.

1. A shipper retrieves ownership of booking in the platform as an access token to use our platform for automated terminal operations.
2. If the retrieval is successful, the shipper forwards the booking to the trucking company in the platform.
3. The trucking company assigns a truck tractor and driver to pick up the container from the terminal.

4. After the container has been picked up and the cargo has been packed into the container, either the shipper or the trucking company submit the declaration for the return of container to the terminal.
5. After the container is returned to the terminal, the shipper will get the trans-shipment receipt and then wait for the container loading on the vessel.

Under business logic, when ownership of booking has been retrieved by a user, other users cannot retrieve the ownership of the same booking. The reason is obvious, as the ownership cannot belong to two different shippers. It prevents multiple retrievals of the same booking for terminal operations.

The problem is that data relationship is not transitive.

In the booking data processing flow, shipper creates booking requests booking from shipping lines and shipping lines cooperates with the terminal on the data sharing for automated terminal operations in the platform.

It does not imply that the platform can own the shipper data for each booking as it is out of scope of data sharing agreement of shipping lines. Instead, these are shipping lines' customer information, which is proprietary information that should be protected from exposure to the public.

As shipping lines cannot provide shipper information to the platform, there is a loss of relation between the booking information and the shipper who owns the booking in the platform.

Under the constraints, the relationship between ownership of a booking and the corresponding shipper must be somehow manually established by the shipper side.

Currently, the platform assumes that only the owner of the booking can get the booking document and the booking number. Shipper who owns a particular booking is required to input the booking number in the booking document manually to establish the relation, linking as the owner of the booking retrieval and granting for remaining logistics processes.

There are two problems for the current practice.

1. **Low data security**
   Booking number is sequential for each shipping line. Therefore, unauthorized users can get access to other booking information by random guessing the valid booking numbers, at least in theory. As booking information contains sensitive information of the shipper, it reduces data security of the whole platform. As a responsible company we should always protect the data integrity of our users.

2. **A non-digital procedure**
   Unlike other steps of terminal operations which have been digitalized, the booking retrieval step is manual. It requires manual input of the shipper, reducing efficiency and user experience. Also, it increases the risk of manual input errors, which may retrieve booking from other shippers without intention.

Therefore, it is necessary to design a new solution to enhance data security and reduce manual, repetitive input tasks of the shippers. Also, the new solution can provide better user experience to shippers.

We aim to solve the problems using state-of-the-art AI models for object detection, document classification and extraction of booking data under the new stage the AI wave.

1.2. Project Aim

The aim of this project is to develop an AI-based document processing system that allows shippers to use a web application to retrieve booking document in an automated and digitalized manner for terminal operations, providing better user experience, reducing manual input and enhancing data security.

1.3. Project Objectives

1. Collect data and build datasets

Booking documents from the following ten shipping lines were considered in our projects. These shipping lines represent around 75% of the total shipping lines booking documents in the platform.

TABLE 2: TEN SHIPPING LINES CLASSES AND THEIR COMPANY LOGOS

Shipping Line Class	CMA CGM	COSCO	EVERGREEN	HAPAG	MEGATOP
Company Logo	CMA CGM	COSCO SHIPPING	EVERGREEN	Hapag-Lloyd	Megatop

Shipping Line Class	ONEY	OOCL	SITC	TSLines	WANHOI
Company Logo	ONE OCEAN NETWORK EXPRESS	OOCL	SITC	TS	W

We collected one softcopy booking document in PDF format for each shipping line. Then, dummy samples were generated by our random booking generator. 200 booking documents per shipping line were generated.

These documents were converted to JPEG format and then propagated using image augmentation in python cv2 module to mimic variety and quality of actual booking documents.

They were used as the datasets for logo detection in our YOLOv8 model, document classification and data extraction in our DONUT model after labelling and distribution to corresponding training set, validation set and testing set.

2. Setup AI training environment

We prepared the basic machine learning environment, installing all necessary modules and libraries to fine-tune the AI models.

In addition, we Installed Nvidia RTX 4090 Ti in the computer and enabled GPU in the training environment. The aims were to speed up the training speed and execution speed of our AI models so that we could earn time for selecting the optimal models at the final stage.

We also downloaded the source code and weights of pre-trained models from the Github repositories and loaded the models for test run since we are not quite familiar with transformer-based models.

3. Train, test and evaluate models

There are two types of models in our proposed system. One is for classifying the document by logo detection and another one is for classifying document by document format and extracting the useful information from the documents.

These pre-trained models were fine-tuned by our custom dataset prepared in the first stage. We tested the models using our test set to evaluate the performance of our AI models in different parameters such as size of dataset, absence or presence of background image and the corresponding performance metrics such as precision, accuracy and F1 score can help us find the optimal combination.

4. Select the best models

Models with the optimal performance are selected for system implementation. Since our model aims to increase the data security and reduce the manual input, we focused on performance metrics such as F1 score and document accuracy rate (DAR) to select the best models.

5. Package and evaluate the solution

Packaged the AI models with a python web server. Also, we provided API for request from existing logistics system with file input and response with booking data for validation. A prototype that simulates an existing system that will provide web UI for shippers to upload documents with server and a database for persistent storage.

Evaluation based on functional requirements and non-functional requirements of the system which mainly include better user experience, higher data security, minimize manual input error.

1.4. Impact and Value

The success of this project can bring us a platform with enhanced data security, protecting user data from leakage. In addition, it reduces manual input and the associated human error, resulting in an increase in platform efficiency. It also provides better user experience to the end users (shippers)

With the success of our project, we hope to provide a fresh perspective to existing problems and inspire more companies to work on automation to improve productivity, accuracy and effectiveness with the use of artificial intelligence.

2. Background or Literature Review

2.1. Review of Existing or Related Solutions for the Problem

As the platform provider, we had proposed three solutions to tackle with the problems using system approach and data approach.

Since the root cause is a loss of relation between the booking and the corresponding booking owner, we aim to reconstruct the relation with the support from shipping line system. Another approach is to reconstruct the relation through provision of more information by the shippers.

Solution 1: Unique ID as the booking key

Shipping line can send a unique ID to both shipper and terminal as the booking key. The unique ID will be stored in our platform database.

This establishes a relation between the shipper and the booking which can be used to retrieve booking in the platform using the unique ID.

This solution ensures data security and keeps the workflow simple.

However, it requires each shipping line to change their own system. Since shipping lines are generally global companies, it is very difficult for them to change their system for a use case in a local company.

Also, there are so many shipping lines in Hong Kong, making this solution unrealistic in terms of system implementation.

Solution 2: An extra field for company ID in the booking request

This is a simplified version of solution 1.

Instead of system change, shipping line provides a field for shippers to input their company ID in the booking request. The data is then sent to the terminal to be used for validation.

This solution requires the shipping line to add an input field, which is less development effort than the previous solution. However, it is difficult to control if the shipper inputs the company ID incorrectly.

Solution 3: Shipper input more information

Since it is difficult to change the system of shipping line, another approach is to request shippers to input more information to retrieve the booking.

This method does not require any enhancement of the system of shipping lines and keep data security since booking request with multiple information required is very difficult to guess out correctly. However, it makes the workflow complicated, which is more difficult to retrieve a booking and against our digitalization and automation direction.

Given the existing limitations, we looked for technologies that can automatically analyze the booking document submitted by the shipper and extract the booking data that can incorporate in our current digitalized operations. These technologies can simplify the workload of end users, enhance data security and maintain our digitalization policy workable.

We searched for two supporting technologies available on the market and they gave us some insights on the project direction.

1. SaaS (Software-as-a-Service)

   Nowadays, many companies have migrated their on-premises data centers to cloud to enjoy the benefits of cloud services such as high flexibility and scalability [3]. These can help companies reduce their maintenance cost and can respond to sudden demand change quickly.

   Similarly, the big-three cloud providers – AWS, Microsoft Azure and Google Cloud Platform are also offering their intelligent document processing (IDP) solution [4]. When shippers upload documents to the cloud, the solution will analyze the document layout and extract the key-value pairs based on pre-built AI models. When necessary, custom model can be trained to achieve better performance and fit the specified use case [5].

2. AI-powered document processing software

   There is several AI-powered document processing software available in the market. This kind of software makes use of their custom AI models and optical character recognition (OCR) to analyze the document and extract the key-value pairs. Studies showed some outperform current available AI models [6].

There are several considerations when using these technologies as an enterprise solution.

a) Data governance policy

   We have to update our client's sensitive information to the public cloud for document processing, which are not allowed in IT policy of the terminals.

b) Black-box solution

   We have no idea of the details of the AI models, including the algorithm and model architectures. We also lose an opportunity to learn new technology.

c) Lack of control

   As a platform provider, we should have better control on the models used. However, in a SaaS solution, the cloud provider is responsible for providing the updates and patches.

From the review of existing solutions and related technologies, we have the following conclusions.

1. System change of shipping lines and voluntary submission of shipping information by the shippers are not feasible because of the global scale of shipping lines and increasing workload to shippers.

2. AI is the key to the success of the project. The solution can reduce the workload of shippers, validate the booking document while extracting key data that can be incorporated into current platform. Shippers no longer need to actively provide information for booking retrieval, validating the ownership as an access token to enjoy all sort of automated terminal operations.

3. As an enterprise solution, custom AI models should be developed and deployed to avoid the problems arisen from using public SaaS solution / software shown above.

We proposed a system with two AI models. One model is for object (logo) detection and another model is for document classification and extraction of targeted data. There are various pre-trained models using CNN, RNN and transformers [7,8] as our candidates.

For object detection, YOLO (You-Only-Look-Once) model is a candidate model [9]. We discovered that there were three FYPs in last year utilizing YOLO model for object detection. Previous success stories give us more confidence in applying this model in our project.

YOLO is based on Convolutional Neural Network (CNN) and the latest version is v8 (2023).

For document classification and data extraction, there are two main approaches, which are OCR and OCR-free. Optical Character Recognition (OCR) is the process to convert the different forms of text to digital editable form [12]. OCR approach requires an OCR model to get the words, and hence requires other models for data extraction. Alternatively, one OCR-free model can handle both. The representative extraction model of OCR approach is LayoutLMv3 [10], and that of OCR-free approach is DONUT [11].

Both LayoutLMv3 and DONUT are transformer-based models [8] and achieve great results in benchmark tests.

TABLE 3: TASKS AND CORRESPONDING CANDIDATE MODELS AND TYPES

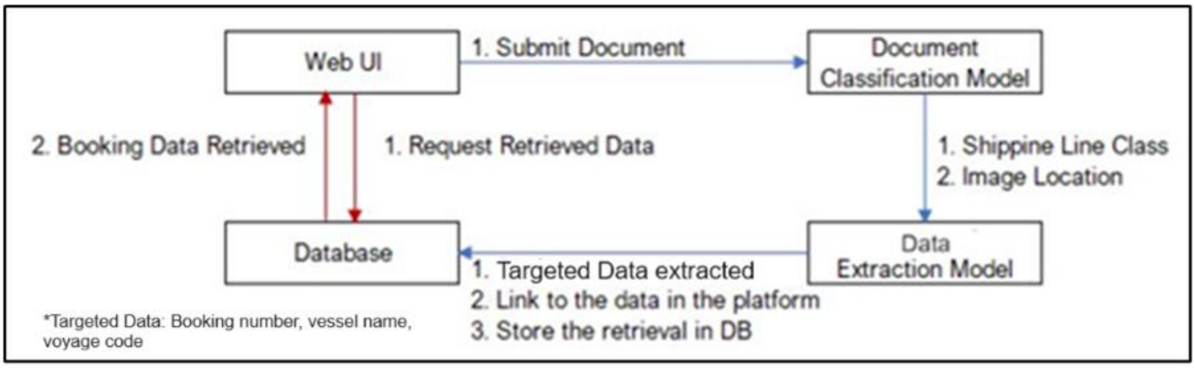
Task (s)	Models and Types
1. Object detection	YOLOv8 (based on CNNs)
2. Document classification	LayoutLMv3 (Transformer-based) DONUT (Transformer-based)
3. Data Extraction	LayoutLMv3 (Transformer-based) DONUT (Transformer-based)

2.2. Highlight of the proposed solution

The proposed solution is to develop a document processing system for extraction of the booking document uploaded by the shippers. The extracted data can be incorporated into the current platform for terminal operations.

We made a sketch on our proposed solution to facilitate discussion on system design.

FIGURE 2: SKETCH OF OUR PROPOSED SYSTEM



From the figure, the following core components are necessary to build our solution.

1. Web UI

   Shippers will upload booking documents in PDF format in a web interface that can be incorporated in the current system and view retrieved booking after submission. Therefore, it is necessary to build a web UI for submission and view retrieval purposes.

2. AI Models

   Our document processing system composes of two AI model components – A object detection model for shipping line logo classification and a document classification and data extraction model for extraction of booking number, vessel name and voyage number from booking documents.

   The object detection model aims to classify booking documents into shipping lines.

   An easy approach is to identify the company logo of the booking document, which is unique for each shipping line. There are various object detection models available on the market including the YOLO model.

   The document extraction component aims to extract the booking number vessel name and voyage number for each booking document. It requires an understanding of the text information and the document layout.

   We have fine-tuned the pre-trained models, LayoutLMv3 and DONUT, with our dataset. Finally, we chose DONUT. (The rationale is further illustrated in clause 3.3)

3. Infrastructure (Web Server / Model Server and back-end database)

   To host the web application and run the AI models, we have to set up a web server for the web application and a model server or service endpoints for calling the AI models and returning results to the web application.

   It is necessary to create a database for storage booking data retrieved by the users on the booking retrieval system.

The methodology and system design are further elaborated in Part 3.

3. Methodology

3.1. Requirements, Supporting Technologies and Technical Gap

A. Requirements

To develop a system, it is necessary to define the system requirements. It is the foundation of all software engineering activities.

The primary actor is the shipper, who uploads booking document using the web interface provided by the system for booking ownership.

The custom AI model is a supporting actor that acts a service endpoint providing document classification and extraction service for the system.

The following use case diagram shows the actors and necessary use cases:

FIGURE 3: USE CASE DIAGRAM OF THE DOCUMENT PROCESSING SYSTEM

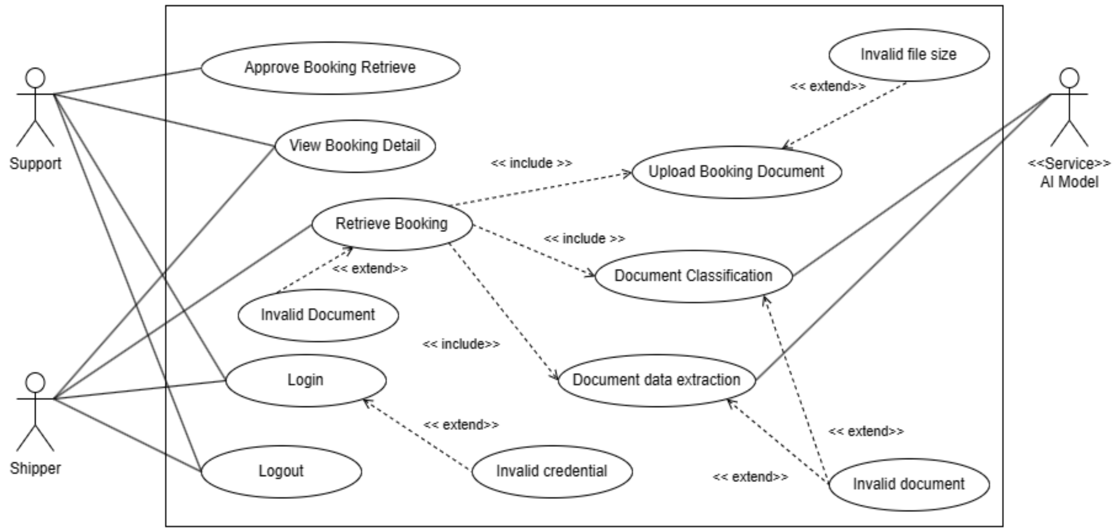


TABLE 4: USE CASE DESCRIPTIONS OF USE CASES

Use Case name	Retrieval Booking
Primary actor	Shipper
Participating actor	AI Model <<Service>>
Flow of events	1. The shipper uploads the booking document (include Upload Booking Document) to the AI model service. 2. The AI model performs document classification by the booking document. (include Document Classification). 3. The AI model performs data extraction by document classification result. (include Document Data Extraction). 4. The web UI shows the result from the AI model service. 5. The retrieved booking is linked to the corresponding booking data in the extract and import platform. 6. After retrieval and linking, the Shipper can view the booking details (use case: View Booking Detail)
Pre-conditions:	1. Shipper has logged in to the system with a valid user id and password.

Post-conditions:	1. Shipper has successfully submitted the booking document and received the predicted result on the upload panel.
Use Case name	**Upload Booking Document**
Primary actor	Shipper
Participating actor	System
Flow of events	1. The Shipper clicks the “Choose file” button in the user booking list page to upload a booking document. 2. The system prompts the Shipper to select a booking document in the local drive to upload. 3. The Shipper selects the desired booking document and clicks the “Upload” button. 4. The system validates the file extensions and file sizes. 5. The system saves the document.
Alternative Flows:	At step 4, if the uploaded booking document has an invalid format, the system will reject the uploaded document and show an error message (Invalid document) to the shipper. Then back to step 1.
Pre-conditions:	1. Shipper has logged in to the system with a valid user id and password.
Use Case name	**Document Classification**
Primary actor	AI model
Participating actor	None
Flow of events	1. The AI model converts the first page of booking document to JPEG image. 2. The AI model predicts the class of shipping line from the image. 3. The AI model validates the confidence score. 4. The AI model outputs the image and the class of the shipping line.
Alternative Flows:	At step 3, if the confidence score of the booking document is not enough. The AI model will return a fail response to the system with fail reason “Invalid Document”.
Pre-conditions:	1. The AI model receives a request with a booking document.
Use Case name	**Document Data Extraction**
Primary actor	AI model
Participating actor	None
Flow of events	1. The AI model selects the data extraction model based on the class of the shipping line. 2. The AI model applies OCR to the image. 3. The AI model use data extraction model to get target booking number. 4. The AI model validates the confidence score. 5. The AI model outputs the class of the shipping line and the booking number.
Use Case name	**Approve Booking Retrieve**
Primary actor	Support
Participating actor	None
Flow of events	1. Support goes to page “Wait for approve booking list” 2. Support checks booking’s data and upload PDF 3. Support clicks “Approve” button to approve booking 4. The retrieved booking is linked to the corresponding upload user in the extract and import platform.
Pre-conditions:	1. Support has logged in to the system with a valid user id and password.

Use Case name	View Booking Detail
Primary actor	Shipper, Support
Flow of events	1. Shipper selects the option to view booking data 2. The system displays the details of each booking, including the 3. booking number, vessel information and date of retrieval. 4. The Shipper can filter and sort the retrieved booking by different criteria. 5. After retrieval, the Shipper can log-out the system or return back to the main page.
Pre-conditions:	1. Shipper has logged in to the system with valid user id and password. 2. Shipper submitted booking for retrieval previously.

Functional requirements:

1. Shipper can login the document processing system with valid company account. Only authenticated user can access to the system.
2. Shipper can logout the document processing system after use.
3. Shipper can upload booking document in valid PDF format on the system platform.
4. The system can classify the shipping line of the booking document and extract the booking number, vessel name and voyage code as the key for retrieving booking.
5. The system shows an error message when the uploaded booking document is invalid.
6. Shippers can view the details of retrieved booking.

Non-functional requirements:

1. Shipper can view the details of retrieved booking after submission of the booking document for no more than 10 seconds per each upload file.
2. The system is easy to learn. The mean training time should be less than an hour.
3. The reliability of overall system should be at least 0.7 for booking documents.

B. Supporting Technologies

To establish the system, the essential supporting technologies are as follows:

Core infrastructure

1. PHP Web Server
   Shipper accesses the platform on the Internet. We host our web application in a PHP web server with associated dependencies.

2. WSGI (Web Server Gateway Interface)
   It is necessary to have a WSGI between the web server and python application to running python application such as our AI models.

   We use the Python Flask Framework (WSGI supported) to create AI service endpoints for the web application to call.

3. MySQL database
   We also need a MySQL database for storage of booking data that has been retrieved by the shipper. Consider the workload and data storage, a simple MySQL database is sufficient.

Hardware component

1. GPU (Nvidia RTX 4090 Ti)
   Matrix multiplication and addition are fundamental for machine learning tasks such as weight updates, forward and backward passes in neural networks, and optimization algorithms.

   Nvidia RTX 4090 Ti is a powerful GPU designed specifically for high-performance computing and deep learning tasks. We use GPU to accelerate training and execution speed of AI models.

Software components

1. PHP Web MVC Framework
   We built a simple web application using the PHP Laravel Web MVC (Model-View-Controller) Framework. With the use of framework, we can develop the demo platform with minimal functions quickly, reducing both development and maintenance efforts.

   Laravel Breeze is a minimal, simple implementation of Laravel's authentication features, including login, registration and password reset.

   We apply this framework plug-in to create a simple login page for authentication of the prototype.

2. Python and its available modules, frameworks and transformer-based models
   Python is well-known for machine learning projects. The language is closed to natural language and there are so many useful modules and frameworks available on Python, including image augmentation, pdf conversion, dummy dataset generation and frameworks for AI model development such as PyTorch.

   The pre-trained models built on these frameworks can also be employed by just a few lines of codes. For example, YOLOv8 for logo detection and LayoutLMv3 for document classification and token classification. The details of modules used are included in part 3.3 Implementation Issue.

C. Technical Gap

There are hundreds of shipping lines worldwide. Booking document from each shipping line contains core information such as booking number, vessel name, voyage code, loading and discharging venues, estimated date of arrival / departure and so on. However, field positions and general layout varies from one to another.

Technical Gap in company logo detection

To classify booking documents into shipping lines, it is necessary to find out the unique characteristics of booking document for each shipping line. For example, position of a certain field, company logo, shipping line address and so on.

A proper way is to identify the company logo, which is unique for each shipping line booking document. Object detection is an area where AI models can achieve similar or even better performance than human beings. The execution speed of an object detection model is faster than models that focus on text and layout.

Therefore, the document classification component will be performed by YOLOv8, the latest version of YOLO (You-Only-Look-Once) object detection model.

YOLOv8 is user-friendly and easy to configure, train and execute. Using other characteristics such as field position involves more complicated setting and training.

Technical Gap in data extraction

The success of this project also depends on whether our system can extract the booking number for each type of booking document. There are various state-of-the-art technologies and pre-built transformer-based models available. We can fine-tune these models to fit our use case. One of the differences is whether the model applies OCR (Optical Character Recognition) for text analysis before prediction. For example, LayoutLMv3 versus DONUT.

LayoutLMv3 is a pre-trained multimodal transformer model that uses unified text and image masking. It can perform a variety of tasks including document classification, document parsing and token classification. It requires OCR for text analysis before the tasks. The analysis combines both text analysis and layout analysis. It achieves state-of-the-art performance in various tasks such as form understanding, receipt understanding, document visual question answering, and document layout analysis [10].

DONUT, on the other hand, is an OCR-free model that does not require OCR for text recognition of input image. The disadvantages of OCR are computationally expensive and any error in the OCR will propagate to the subsequent process [11].

Under our observation, similar characters, such O and 0, are easily misclassified in both OCR and OCR-free models. In OCR models, there are different post-OCR approaches to solve this problem to achieve higher accuracy. [14] Therefore, not only the text extraction models, but also additional methods should be taken to improve the performance of the system.

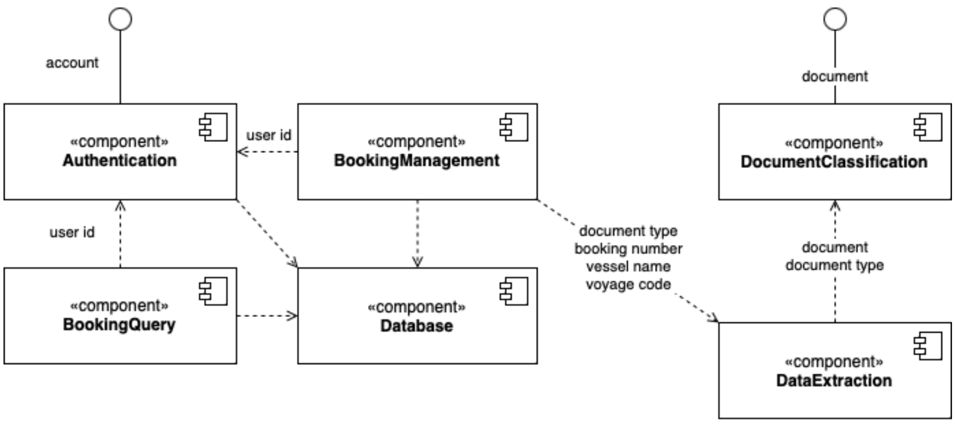
3.2. System Design

A. System components

The component diagram classifies the system into several components based on the functionality. Each component is responsible for a particular task in the system and interacts with each other to complete the system aim.

In our system design, there are six components. The interactions are shown in the following figure.

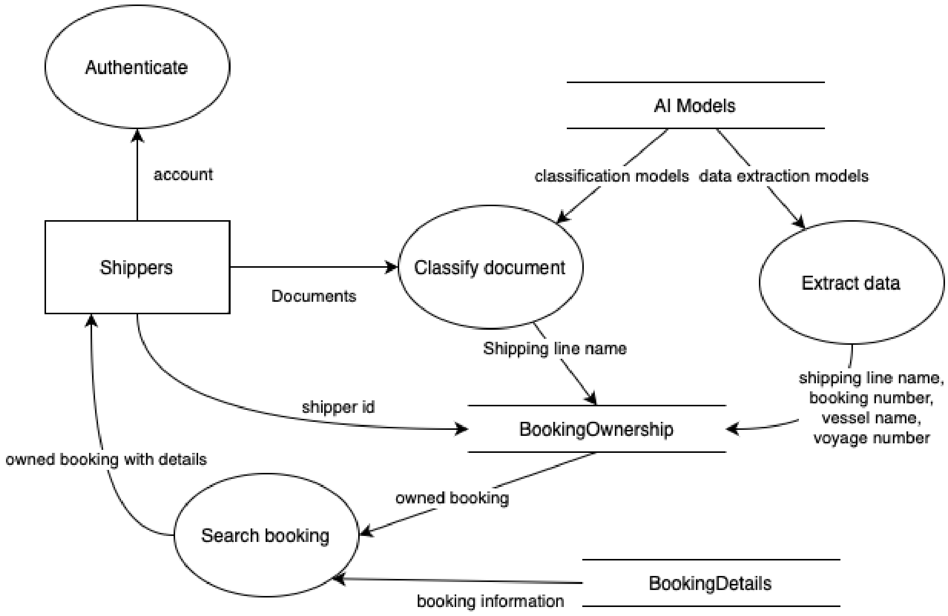
FIGURE 4: COMPONENT DIAGRAM OF THE DOCUMENT PROCESSING SYSTEM



1. The Document classification component provides an interface to receive digital booking document in PDF format.
2. The document extraction component depends on the output of the document type (i.e. shipping line class) in the document classification component.
3. The BookingQuery component is for querying booking information retrieved by the user id.
4. The BookingManagement component is for management of booking details of the user I in the system.
5. The Authentication component provides an interface to receive account credential for linkage of user id in BookingQuery component and BookingManagement components.
6. The Database component stores the information necessary for booking query, authentication and booking management.

B. Data Flow of the system

FIGURE 5: DATA FLOW DIAGRAM OF THE DOCUMENT PROCESSING SYSTEM



1. Shipper logins the web application using the company's user id and credentials.
2. Shipper obtains the access rights to upload the booking document for booking retrieval.
3. The booking document is classified into shipping lines by loading and running the AI classification model.
4. The booking document, with known shipping line class, is extracted by loading and running the AI document extraction model to extract the booking number, vessel name and voyage code.
5. The BookingOwnership is established between the booking information and the shipper's user id.
6. The key is used to extract the booking information from booking details in the database of the export and import platform.
7. The retrieved booking of the shipper is stored in the database for review afterwards.

C. System Architecture

The system architecture of the document processing system is a typical three-tier architecture, composing of the presentation tier, application tier and the data tier.

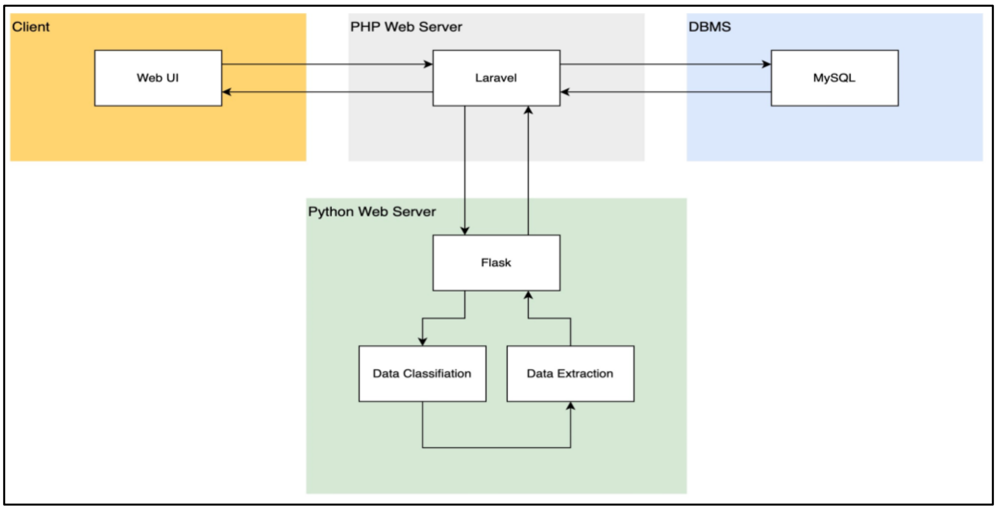
The presentation tier is responsible for the user interface and communication with the application. It is used to display the booking information and allow the users to upload the booking documents.

The application tier is responsible for processing, rule checking and notification by the application. There are two servers in the application tier. One is the web server for running PHP Laravel Web application. Another one is to host python application for document classification and document extraction, serving as a service endpoint to the system.

The data tier is responsible for storage, retrieval and manage of persistent objects or information. All data retrieved successfully will be saved in the database for the user to review after uploading.

The benefits of a three-tier architecture is that each tier can be developed separately without impacting other tiers.

FIGURE 6: SYSTEM ARCHITECTURE OF THE DOCUMENT PROCESSING SYSTEM



3.3. Implementation

To implement the prototype of the document processing system, we use the following programming languages and run-time dependencies:

TABLE 5: VARIOUS COMPONENTS FOR SYSTEM IMPLEMENTATION

Component	Name of the dependencies / technologies
Web Server	PHP Web Server
AI Model Server / Services endpoint	Python Flask framework (WSGI supported), backed up by Nvidia A100 to execute the models.
Database management system (DBMS)	MySQL
Web Application	Laravel (Web MVC) Framework
Programming languages	PHP (for core web application) Python (for machine learning / AI tasks)
Libraries	Pdf2image (for conversion of pdf to jpeg) cv2 (for image augmentation) ultralytics (for YOLOv8 training and production run) Pytesseract (OCR Engine)

	PyTorch (Machine learning framework)
Pre-trained AI models	YOLOv8 (Logo detection) DONUT** (Document classification and data extraction)

**Rationale to choose DONUT as our Data Extraction model

In phase 1, we decided to use an OCR-dependent approach to extract the booking number in the documents. First, we used tesseract to get words and their location. Hence, the OCR result is used to classify the type of documents by LayoutLMv3ForSequenceClassification model and extract booking number by a LayoutLMv3ForTokenClassification model per type of documents.

After that, we planned to improve data security. Other information in the documents, voyage code and vessel name are considered to verify the documents. Also, we increased the type size of documents from 4 to 10. Our old method is hard to support this update because OCR performs poorly in extracting the information in the table, and many voyage codes and vessel names are stored in the table.

Therefore, a representative OCR-free model DONUT is used. By implementing DONUT, we speed up the training and processing time since OCR requires a lot of computing resources. Besides, the number of models dropped to 1, which simplified the execution of the model.

3.4. Evaluation

User experience Evaluation

TABLE 6: USER EXPERIENCE EVALUATION

Evaluated Task	Contents
Usability	We conduct an AB test for evaluating improvement in user experience of our new system. 20 individuals are invited to participate in the test. Each participant is asked to retrieve 10 bookings in the two systems separately. After that, they will complete a questionnaire, which is shown below. **Success Criterion:** the average score of the 4 criteria should be greater than 3.
Data Protection Capability	
Efficiency	
Error Reduction Capability	

The aim of the AB test is to obtain a comprehensive understanding of how our new system improves the user experience in four different areas.

The comparative system is a newly developed mock system, which simulates the real existing booking retrieval system.

This system closely mirrors the user experience of the actual booking retrieval system. To retrieve a booking, the users are required to follow a specific process. First, they should select a shipping line in a dropdown list. Then, they should enter the booking number of the booking document twice to ensure that the users are entering the correct booking number. This comparative system ensures that the result of AB test reflects the actual user experience improvement in our new system.

After testing, the participants are invited to complete a questionnaire to collect their feedback. The questionnaire required participants to rate their agreement in four categories, which are convenience, data protection, operating speed, and minimizing manual error. The rating scale is from 1 (strongly disagree) to 5 (strongly agree).

We aim to achieve an average score of over 3 marks, indicating that most participants agree that our new system has improved the user experience.

FIGURE 7: QUESTIONNAIRE FOR AB TESTING

The purpose of this survey is to learn about the user experience of our proposed system. This survey is voluntary and confidential. That means that if you do not feel like answering the questions you do not have to, and if you feel like stopping part way through that is ok, you can stop at any time. It also means that your name will not be attached with your answers and [if appropriate] only the person interviewing you will know your answers.

Please rate your agreement with the following parameters of our new Booking Confirmation Notes submission platform compared to the old system.

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Easier to use	○	○	○	○	○
Help protecting data	○	○	○	○	○
More efficient	○	○	○	○	○
Help reducing manual error	○	○	○	○	○

Logo Detection Model Performance Evaluation

TABLE 7: LOGO DETECTION MODEL (YOLOV8) PERFORMANCE EVALUATION

Evaluated Task	Contents
F1 score & Accuracy	**Testing Set:** We prepare 200 booking documents for statistical analysis, with 20 booking documents from each class of shipping lines. In addition, 10 empty booking document are included as the background / empty set.. **Metrics:** Precision, recall, accuracy and F1-score will be assessed for the custom AI models. **Success Criterion:** The accuracy of the model should be greater than 0.90 and the F1-score of the model must be greater than 0.80.

Data Extraction Model Performance Evaluation

TABLE 8: DATA EXTRACTION MODEL PERFORMANCE EVALUATION

Evaluated Task	Contents
Field-level F1 score	**Testing Set:** We prepare 100 booking documents for statistical analysis, with 10 booking documents from each class of shipping lines. In addition, 10 non-booking documents are included. **Metrics:** precision, recall, F1 score **Success Criterion:** The F1-score of the model must be greater than 0.80.
Document Accuracy Rate (DAR)	**Testing Set:** We'll prepare 100 extra booking documents for statistical analysis, with 10 booking documents from each class of shipping lines. **Success Criterion:** the DAR should be larger than 0.70.

Field-level F1 score is a metric that determines whether the extracted field is exactly matched to the accepted value. The equations of calculating the precision, recall and F1 score are: [13]

$$Precision = \frac{The\ number\ of\ exact\ field\ matches}{The\ number\ of\ the\ detected\ field}$$

$$Recall = \frac{The\ number\ of\ exact\ field\ matched}{The\ number\ of\ the\ ground\ truth\ fields}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

The F1 score ranges from 0 to 1, with a score closer to 1 indicating a higher level of accuracy in field extraction.

Besides, Document Accuracy Rate (DAR) is another important metric used to evaluate the performance of a data extraction model. DAR measures the document that all fields are extracted correctly. According to a study, the DAR of DONUT is 0.74 and that of LayoutLM is 0.581. [13] Since we are using DONUT to extract data, it is reasonable to set the success criterion as 0.7. This means that we aim for at least 70% of the documents to have all fields accurately extracted in our data extraction model.

Overall Evaluation

TABLE 9: OVERALL EVALUATION

Evaluated Task	Contents
Reliability	**Testing Set:** We prepare 100 booking documents for statistical analysis, with 10 booking documents from each class of shipping lines. **Metrics:** rate of success in retrieving valid booking document **Success Criterion:** over 0.7
Efficiency	**Testing Set:** We prepare 100 booking documents for statistical analysis, with 10 booking documents from each class of shipping lines. **Metrics:** Average processing time **Success Criterion:** less than 10s per document

The ability of a system to recognize valid booking documents accurately is a fundamental aspect of its capability. A high rate of success in retrieving valid booking documents indicates that the system is able to both classify the type of documents and extract related information accurately.

Also, the efficiency of the system is measured by the average processing time. The processing time per document should be lower than 10s.

General users shall be able to learn within 1 hours provided that there are only a few operations. Most of the tasks are guided with hints and messages. (The training time is not included in the AB testing but easier to use is a point to evaluate.)

4. Results

4.1. Improvement

4.1.1. Improving Data Security

In our new system, there are two ways to enhance data security. Firstly, by incorporating detection of company logos and analysis of the extracted booking information, we can determine the validity of booking documents. Detection of company logos helps us to differentiate actual booking documents from fake or unauthorized ones with incorrect company logo. Also, reliability and consistency can be assessed by analysis of the booking information extracted. This analysis involves exact match to the booking number and verifying the accuracy of vessel name and voyage number. The inconsistencies will indicate the document as invalid.

Also, we provide authentication to the shipper, adding an extra layer of security. Only registered shippers can access our system.

Compared to the real booking retrieval system, the cost of unauthorized access to booking information increased. The unauthorized access required individuals to compromise a shipper account and create a fake booking document with accurate information. Therefore, our new system can effectively protect the booking information from unauthorized access.

4.1.2. Reducing manual error

Our system has reduced manual errors by allowing users to upload documents instead of manually typing the booking numbers. After the booking documents are uploaded, the booking number will be extracted immediately. This minimizes the probability of manual errors occurring during data entry. Also, our system eliminates the need for users to recognize the format of different types of booking documents, identify the location of the booking number, and manually input the booking numbers twice. Therefore, overall accuracy and efficiency improved by reducing the reliance on manual input.

4.1.3. Innovation in logistics industry

Hong Kong logistics industry is stills relies on manual procedures. By the success of our system, we hope to provide a fresh perspective to existing problems and inspire more companies to work on automation to improve productivity, accuracy and effectiveness.

4.2. Evaluation Result

4.2.1. Logo Detection Result

Based on official documentation of YOLOv8, the following metrics are commonly used for evaluation of model performance, depending on objectives [15]:

- mAP: Suitable for a broad assessment of model performance.
- IoU: Essential when precise object location is crucial.
- Precision: Important when minimizing false detections is a priority.
- Recall: Vital when it's important to detect every instance of an object.

- F1 Score: Useful when a balance between precision and recall is needed.

As our YOLOv8 fine-tuned model recognizes the shipping line logo for classification of booking document, we are not concerned about the benchmark test or the exact position of shipping line logo.

Instead, we should take care about the problem of false detection and getting all documents correctly because our project aims to increase data security and reduce manual input. F1 score seems to be the most appropriate metric for evaluation.

To facilitate the evaluation, the final version of YOLOv8 model was trained with the following dataset and distribution. Originally, there were 50 documents per each class. 10% of the documents were used in test set, equivalent to only 5 documents, which may not be fair for comparison purpose.

Given that there are many classes, to show the F1-score better, we increase the proportion of testing set. The summary of dataset distribution is as follows:

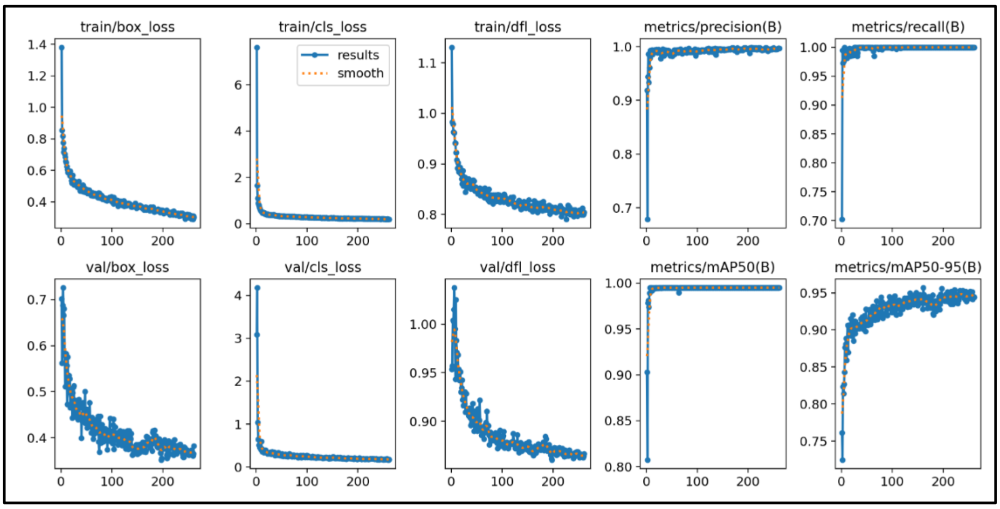
FIGURE 8: SUMMARY OF THE DATASET DISTRIBUTION IN LOGO DETECTION.

Shipping Line Class	Logo	Training	Validation	Testing	Size of class
CMA CGM		70	20	20	**110**
COSCO		70	20	20	**110**
EVERGREEN		70	20	20	**110**
HAPAG		70	20	20	**110**
MEGATOP		70	20	20	**110**
ONEY		70	20	20	**110**
OOCL		70	20	20	**110**
SITC		70	20	20	**110**
TSLines		70	20	20	**110**
WANHOI		70	20	20	**110**
BACKGROUND		35	10	10	**55**
Total Size of dataset		**735**	**210**	**210**	

Background dataset is used to reduce the false positive ratio. The size is 50% as a class.

After an epoch of 261, the training stopped because there is no better model with a successive 50 epochs. The overall training result is summarized as the following figure.

FIGURE 9: SUMMARY OF THE TRAINING METRICS



The confusion matrix and normalized confusion matrix are as follows:

FIGURE 10: CONFUSION MATRIX OF THE OBJECT (LOGO) DETECTION MODEL

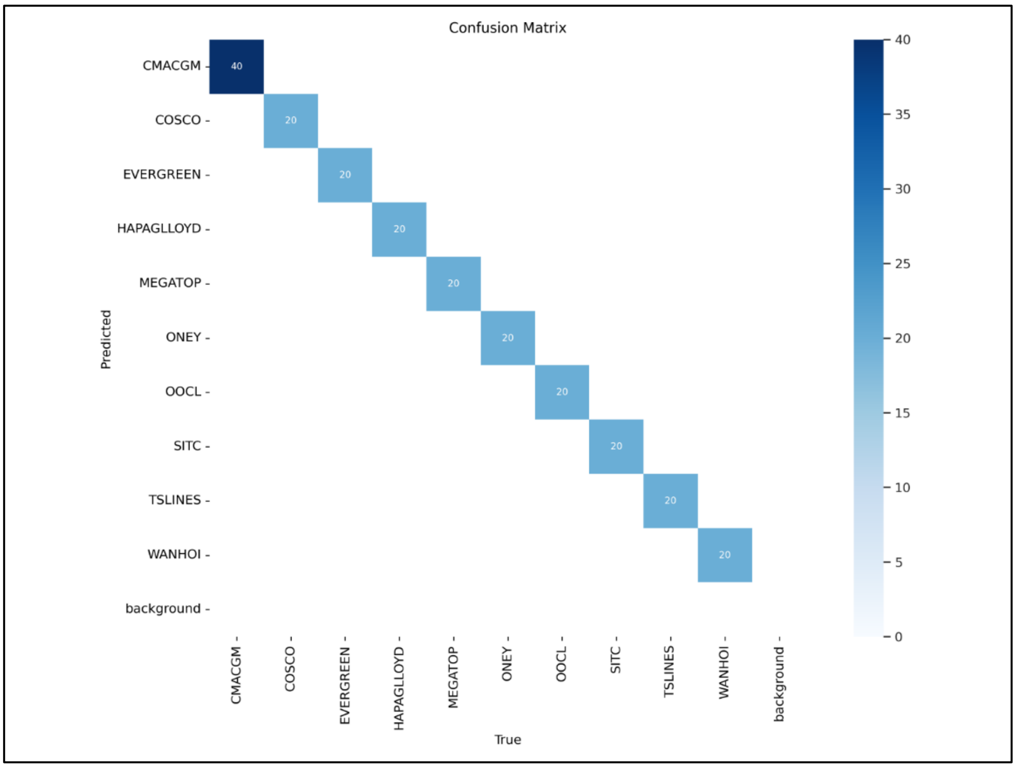
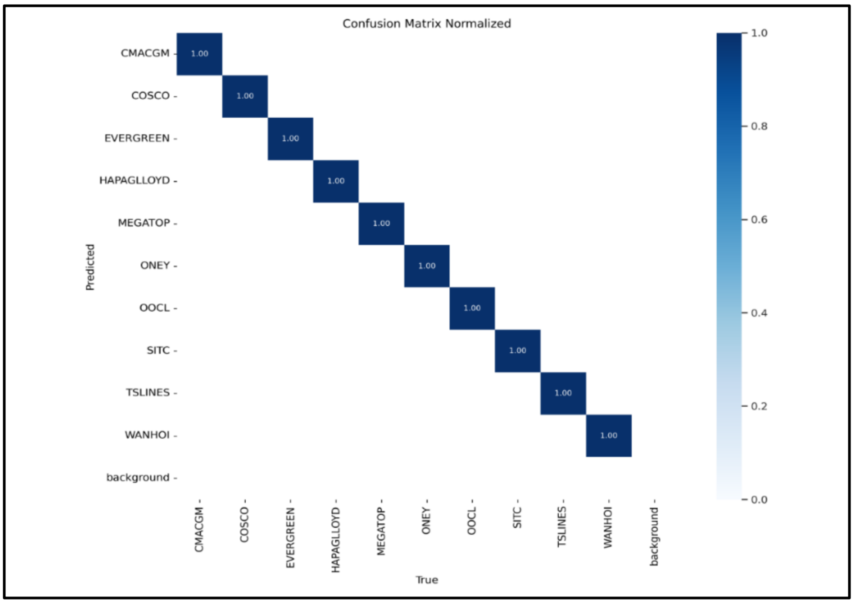


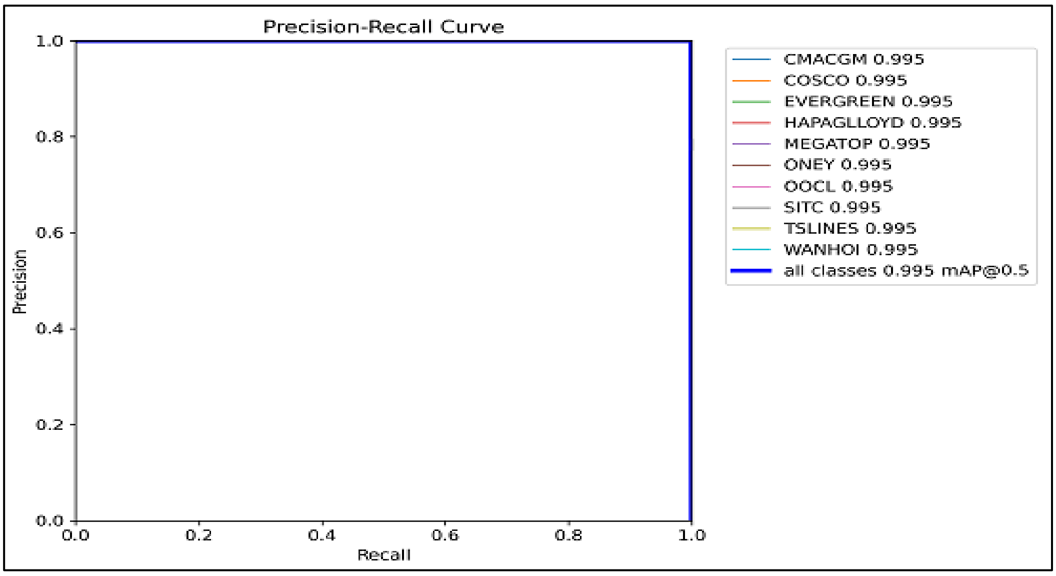
FIGURE 11: NORMALIZED CONFUSION MATRIX OF THE OBJECT (LOGO) DETECTION MODEL



Given the fixed document layout for each shipping line, our YOLOv8 model can recognize all test set sample class correctly. On the other hand, for a particular class, the model can recognize all samples without any mistakes made. It gives the perfect case in the Precision-Recall curve. It is an easy task to recognize the logo given the logo is not a moving object and the layout in a document are more or less the same, except the booking details.

YOLOv8 is famous for detection of moving object within short lag of time. so it gives an excellent performance in this part.

FIGURE 12: PRECISION-RECALL CURVE OF THE OBJECT (LOGO) DETECTION MODEL



Since the document processing solution involves two models, the accuracy data will affect the overall performance of the model and determine whether the solution is satisfactory when compared with the objectives.

Although with 100% accuracy on the testing set, to safeguard, let's assume that the accuracy of the YOLOv8 model is 0.90.

4.2.2. Data Extraction Result

In our project, there are three fields to be detected in every document, which are booking number, vessel name and voyage number, and one additional field to detect the type of the documents.

As we mentioned earlier, we used field-level F1 score and document accuracy rate to evaluate data extraction model.

TABLE 10: SUMMARY OF PERFORMANCE METRICS FOR THE DONUT MODEL

Evaluation Dataset	Mean precision	Mean recall	Mean F1 score	DAR
Total 100 (10 per shipping line)	0.8886	0.8886	0.8886	0.6090

x	Type	Booking number	Vessel name	Voyage number
Accuracy(x)	1.0	0.8636	1.0	0.6909

Since all the booking documents in our evaluation dataset correctly detected a total of 4 fields or detected nothing in other documents, the value of precision, recall and F1 score are the same.

The result shows that our data extraction model performs well in field-level F1 score, which is higher than our exception 0.8.

About the accuracy, the DAR result is lower than our target of 0.7. The result shows us that this is hard to extract all the information from booking documents correctly. The details of field-level accuracy show that type and vessel name achieve 100% correct. This is because their value comes from a countable list, but the booking number and voyage number are random values that do not have understandable pattern. However, more research is required to understand why the accuracy of voyage number is much lower than that of booking number.

4.2.3. Overall Evaluation Result

After combining the object detection models and data extraction model, we can identify the rate of success in retrieving valid booking documents.

TABLE 11: SUMMARY OF OVERALL EVALUATION RESULT

Evaluated Task	Evaluation Result
Rate of success in retrieving valid booking documents	0.75
Average processing time per document	6s

After combining the two AI models to one system, the rate of success in retrieving valid booking documents is 0.75, which is greater than expected. The 25% documents will be passed to our admin panel and wait for manual handling. This result also shows that there is improvement in reducing manual operation.

Besides, our system performs well in efficiency. The average processing time per document is 6s, which is much lower than we expected.

4.3. Implementation Issue

4.3.1. Errors in classifying similar characters in data extraction model

As we mentioned in technical gaps, similar characters are easily misclassified. This is a significant issue to us because we need to collect the exact match booking number. If one character in the booking number is detected incorrectly, the whole booking will be determined as unfound booking. That explains why sometimes the field-level accuracy cannot be achieved 100%.

The additional fields added after phase one, which are vessel name, and voyage number, dropped the acceptance rate because of this. Therefore, we turn our acceptance from fully correct to considering their similarity. In our final design, the booking is valid if:

1. Booking number must be found in database (exact match)
2. Similarity between the extracted vessel name and voyage number, and the corresponding data in the database should be not less than 0.8

In the evaluation dataset, the similarity of all records is higher than 0.8. Therefore, we successfully reduce the impact of minor misclassification in similar characters by implementing the new method.

4.3.2. Quality of dataset

There are differences between the actual booking document and the booking document in our dataset. In the beginning, we collected booking documents from various shipping line companies. Then, we converted the pdf file document to docx and created dummy datasets based on the docx template

However, the dummy did not accurately represent actual data. Minor differences in font style, spacing, or other formatting aspects can impact the effectiveness and reliability of our system when working with actual booking documents. Therefore, before deploying our solution into production, real actual documents should be required to achieve better performance.

5. Conclusion

Our project results show that our new booking retrieval system can improve the issues we mentioned. In the past, the shipper received a booking document from the shipping line and entered the booking number in our platform for further actions. This method exposes the booking information to all shippers and easily occurs manual error. In our new system, the shippers upload their received booking documents to our platform. The booking information from the shipping line is protected by incorporating detection of company logos and analysis of the extracted booking information. Also, the effectively automated process reduces human error.

5.1. Limitations of solutions and methods

Our product does not support all the shipping lines in Hong Kong. There are over 60 active shipping lines in Hong Kong, but our product supports only 10 of them. However, more classes may result in lower accuracy in all our AI models and increase the complexity of training.

In addition, AI models cannot provide 100% certain results, which rely on humans to do the final checking of uploaded documents. At the same time, human verification conflicts with one of our important purposes - reducing manual operation. We should make better in striking balance between automation and human involvement. Through active learning, the AI models can learn from mistakes with the aid of human verification. In the long term, the frequency of human verification will reduce with the model run. Still, we cannot guarantee 100% certainty for our solution.

For the mismatch of characters such as 0 and O, a particular solution is to make use of regular expression to ensure the result has a particular format (e.g. all digital characters, all non-digit characters, booking number starting with ABC and then 9 non-zero digits…etc.). We can also implement error correction script if there is only one or two mistakes in the extracted booking data. However, we still cannot eliminate all mismatch cases.

5.2. Future works

First, the rate of success in retrieving bookings is the most essential feature in our system. As we mentioned before, actual booking documents can increase the reliability of our AI models. If we have enough resources, it is better to collect shippers' booking documents for research and development. Also, we will continue to look for different methods to increase the rate of success, such as document pre-handling processors or editing layers in the AI models.

Also, this is important to increase the number of supporting document formats. Since there are at least 60 types of booking documents from various shipping line companies. To completely replace manual to automatic, this is better to support all the formats.

To scale up the project easily, the next step is to provide an admin console for adding new document formats. Since there are over 50 formats of booking documents that have not been supported yet and the format of the document can be changed suddenly, non-developer should be able to manage the formats after development.

The implementation of regular expression and error detection script can help mitigate the condition of mismatch of characters.

References

[1] Transport and Logistics Bureau, The Government of the Hong Kong Administrative Region, *Action Plan on Maritime and Port Development Strategy*. Hong Kong: Transport and Logistics Bureau, Dec. 2023.

[2] DollarAndSense Business, *Guide To Tuas Mega Port – World's Largest Fully Automated Port In Our Backyard,* 2023. [Online]. Available: https://dollarsandsense.sg/business/guide-to-tuasmega-port-worlds-largest-fully-automated-port-in-our-backyard [Assessed Feb. 2, 2024].

[3] Sydle One, *13 Advantages of Cloud Computing*, 2022. [Online]. Available: https://www.sydle.com/blog/advantages-of-cloud-computing-6154ae3d0e94dd1b0fa01fd6 [Assessed Jan. 16, 2024]

[4] Google Cloud, *Compare AWS and Azure services to Google Cloud*, 2023. [Online]. Available: https://cloud.google.com/docs/get-started/aws-azure-gcp-service-comparison [Assessed Jan. 20, 2024]

[5] Google Cloud, *Document AI*, 2023. [Online]. Available: https://cloud.google.com/document-ai?hl=en#key-features [Assessed Jan. 20, 2024]

[6] Ajay Bhatt, "Document Automation Using Artificial Intelligence", *IJRASET*, Volume 10 Issue IX, 2022.

[7] A. Vaswani, *et al*. "Attention is all you need." *Advances in neural information processing systems*, vol. 30, 2017

[8] Hugging Face, "How do Transformer work?", 2021. [Online]. Available: https://huggingface.co/learn/nlp-course/chapter1/4 [Assessed Jan. 20, 2024]

[9] Deci AI, *The History of YOLO Object Detection Models from YOLOv1 to YOLOv8,* 2023 [Online]. Available: https://deci.ai/blog/history-yolo-object-detection-models-from-yolov1-yolov8 [Assessed Feb 02, 2024]

[10] Huang, Y, et al. "Layoutlmv3: Pre-training for document ai with unified text and image masking." *Proceedings of the 30th ACM International Conference on Multimedia*. 2022.

[11] Kim, G, et al., "OCR-free Document Understanding Transformer", European Conference on Computer Vision, Cham: Springer Nature Switzerland, 2022.

[12] N. Yasin, I. Siddiqi, M. Moetesum, and S. A. Rauf, "Transformer-based neural machine translation for post-OCR error correction in cursive text," Document Analysis and Recognition – ICDAR 2023 Workshops, pp. 80–93, 2023. doi:10.1007/978-3-031-41501-2_6

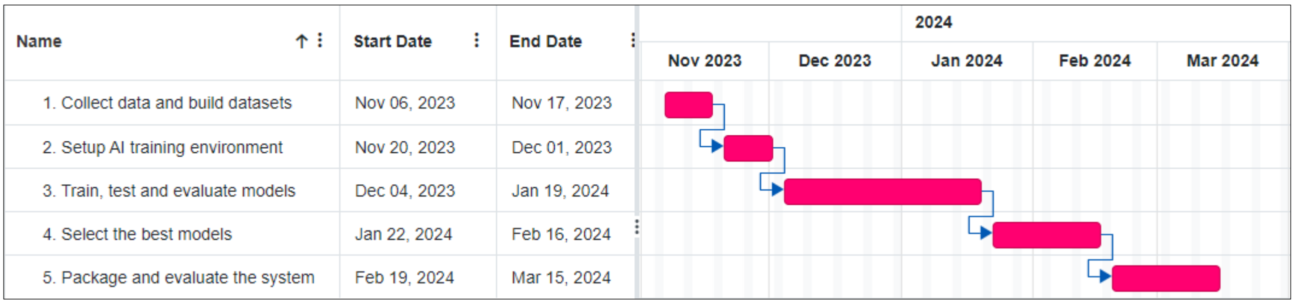
[13] M. Dhouib, G. Bettaieb, and A. Shabou, "DocParser: End-to-end OCR-free information extraction from visually rich documents," *Lecture Notes in Computer Science*, pp. 155–172, 2023. doi:10.1007/978-3-031-41734-4_10

[14] T. T. Nguyen, A. Jatowt, M. Coustaty, and A. Doucet, “Survey of post-OCR processing approaches,” *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–37, Jul. 2021. doi:10.1145/3453476

[15] Ultralytics, *YOLO Performance Metrics*, 2024. [Online]. Available: https://docs.ultralytics.com/guides/yolo-performance-metrics/ {Assessed Apr. 15, 2024]

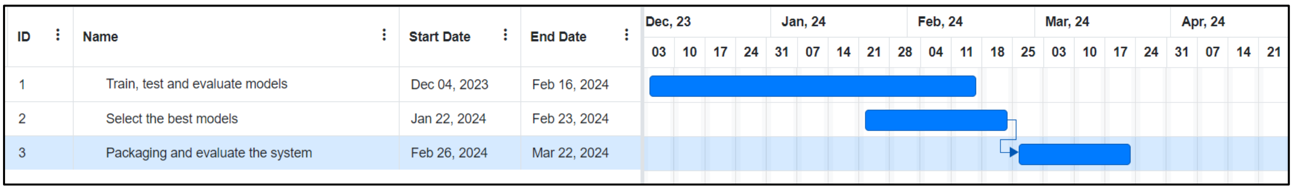
Appendix A. Team Members' Roles and Responsibility (Final)

Gantt Chart



Name	Start Date	End Date
1. Collect data and build datasets	Nov 06, 2023	Nov 17, 2023
2. Setup AI training environment	Nov 20, 2023	Dec 01, 2023
3. Train, test and evaluate models	Dec 04, 2023	Jan 19, 2024
4. Select the best models	Jan 22, 2024	Feb 16, 2024
5. Package and evaluate the system	Feb 19, 2024	Mar 15, 2024

Revised Gantt Chart in interim report



ID	Name	Start Date	End Date
1	Train, test and evaluate models	Dec 04, 2023	Feb 16, 2024
2	Select the best models	Jan 22, 2024	Feb 23, 2024
3	Packaging and evaluate the system	Feb 26, 2024	Mar 22, 2024

Roles

Roles	Member(s)	Remarks
Team coordinator	Eraka Yip	Manages the project in general and keeps records, reports, and other documents in order, and prepare the submission of reports
Team members (Roles: Designer, Programmer, Tester and System Analyst)	Eraka Yip, Timothy Chan, Andy Lau	All team members will participate in designing, coding, testing and deploying the system.

Responsibilities and Task Assignment

Tasks	Responsible Member(s)	Target Date
Collect data and build dataset	Eraka Yip, Timothy Chan, Andy Lau	Nov 17, 2023
Setup AI training environment	Eraka Yip, Timothy Chan, Andy Lau	Dec 01, 2023
Train, test and evaluate models	Eraka Yip, Timothy Chan, Andy Lau	Jan 19, 2024 →Feb 16, 2024
Select the best models	Eraka Yip, Timothy Chan, Andy Lau	Feb 16, 2024 →Feb 23, 2024

Package and evaluate the system	Eraka Yip, Timothy Chan, Andy Lau	Mar 15, 2024 →Mar 22, 2024

Tasks breakdown

1. Collect data and build datasets (Completed on schedule, Nov 17, 2023)
    a. Collection of samples of booking documents from ten shipping companies.
    b. Propagate the dummy dataset to 200 booking documents per company using Python script (aka random booking generator).
    c. Conversion of PDF to JPEG for the first page of each document and carry out image augmentation to expand the dataset to proper quantities, which are used to mimic variations and document quality in daily scenarios.
    d. Labelling, annotating and grouping the images into training set, validation set, and testing set for document classification component and document extraction component of the system.

2. Setting up the AI training environment (Completed on schedule, Dec 01, 2023)
    a. Setting up the environment to utilize GPU for training our AI model. (Install libraries and modules of Python, loading of pretrained models and test the functionality)
    b. Searching for some techniques that may speed up the training and execution speed. Finetuning the hyperparameters during the training and detection part.

3. Train, test and evaluate models (Completed on schedule, Feb 16, 2024)
    a. Train the YOLOv8 object detection model for logo detection with different set of custom dataset and batch number and evaluate the performance metrics from the results of the testing dataset.
    b. Evaluate the YOLOv8 object detection model versus LayoutLMv3 in document classification tasks.
    c. Train, test and evaluate LayoutLMv3 token classification model to extract booking number, vessel name and voyage number.
    d. Train, test and evaluate DONUT data extraction model to extract booking number, vessel name and voyage number.

4. Select the best models (Completed on schedule, Feb 23, 2024)
    a. Select the best model of YOLOv8 model for shipping line classification based on logo detected. Performance metrics are evaluated with different sizes of dataset, number of epochs and different sizes of the model itself.
    b. Select the best data extraction model from LayoutLMv3 and DONUT. Finally, DONUT model was solely chosen.

5. Package and evaluate the system (Completed on schedule, Mar 22, 2024)
    a. Complete the Web UI for shippers to login and retrieve booking by uploading booking documents.
    b. Complete the Python backend server to host the AI models.
    c. Evaluate the whole system and record the results.
    d. Summarize all statistics and conclusion in the final year project.