Allocation and Scheduling System of Emission Testing using Optical Character Recognition

Florence Jean B. Talirongan[1], Hidear Talirongan[2], Roseclaremath A. Caroro[3], Rolysent K. Paredes[4], Kenneth Largo[5], and Jasmine C. Dico[6]

1-6 College of Computer Studies, Misamis University, Ozamiz City, Philippines

DOI: https://doi.org/10.56293/IJASR.2024.6106

IJASR 2024
VOLUME 7
ISSUE 5 SEPTEMBER - OCTOBER **ISSN: 2581-7876**

Abstract: The scheduling system helped businesses in prioritizing tasks. However, emission centers are still using the manual scheduling of clients through an on-site visit without guaranteeing any available schedule. Hence, this study developed a system that allows vehicle owners to set an online schedule for their emission testing based on real-time schedule availability. The system used a Tesseract Optical Character Recognition (OCR) engine to scan the Certificate of Registration (CR) and Official Receipt (OR) images to extract CR and OR numbers. This approach assesses the authenticity of a vehicle's ownership to avoid hoarding schedules for fraudulent reasons. Moreover, the system obtained an extraction accuracy rate of 63.33, and the stress testing on the website using loader.io resulted in 104ms average response time tested in 100 clients per second. The evaluation results show that the system provides the vehicle owners a convenient way of scheduling from the emission center without directly visiting the designated vehicle emission centers. Furthermore, the system's assessment findings reveal that it is simple to use, user-friendly, and all its functions are completely functional, allowing clients to learn to navigate the system quickly.

Keywords: scheduling, emission, tesseract, OCR

1. Introduction

The scheduling system is very useful in scheduling an appointment online, even in health care services, by designing a system suited for that type of service. For example, the secretaries and doctors of the clinics use the appointment and management under the system, and clients can communicate through the system by either using a website or mobile application (Habibi et al., 2019). In addition, there have been many projects that use a scheduling system to plan and manage their productivity. Therefore, the scheduling system is helpful in most cases, especially in the Private Emission Testing Center (PETC), as it creates stress-free scheduling in vehicle emission testing. In the Philippines, a car's registration includes a mandated emission test. This test measures the level of pollutants or fumes produced from a vehicle's exhaust manifold (Cesar, 2021).

The Private Emission Center in Jakarta, Indonesia, is still using a traditional way of scheduling clients on emission testing. The PETC of the Philippines is still using the traditional way to schedule the clients. The clients still need to go to the PETC to have manual scheduling for emission. Although traditional ways still have their advantages, adapting to the new system and technology is not bad for improving business capabilities.

Employees are primarily responsible for traditional workforce scheduling, whereas objective ability influences workforce scheduling. People felt tired and inattentive, making them uneasy to carry out operations following the regulations when facing many tasks. The system requires a computer to increase job performance efficiency and save financial resources. It is possible to produce efficient software by optimizing the software algorithm and selecting the programming language. It may create efficient software. The most common problem with the Private Emission Testing Center scheduling system is that the clients' lack of guidance when scheduling emission testing creates confusion.

The local emission testing center in Misamis Occidental is currently using manual scheduling vehicle owners' emission tests. Thus, it resulted in an unorganized way of giving priority numbers to the vehicle owners.

2. Literature Review

2.1 Scheduling Algorithm

First Come First Serve (FCFS) is a non-preemptive scheduling method often used in hardware, software, and production systems. First-in, first-out (also known as First Come First Serve) task scheduling is a scheduling technique used in many hardware, software, and production systems (Robin et al., 2016). First Come, First Served is the most straightforward scheduling algorithm available. This method emphasizes the sequential execution of operations. In other words, in this kind of schedule, the processes arrive in a specific sequence then give attention solely to the order in which they arrive, a sequential order (Teorey & Pinkerton, 2021). When employing FCFS, the processes include placing them in a queue and waiting for them to arrive. FCFS is one of the simplest scheduling systems.

FCFS scheduling is fair and straightforward in the formal sense, but the average waiting time and the average turnaround time in this scheduling are very long. The process starts when the queue gets performed first (Joshi & Goswami, 2017). Others who come first will be the ones served first. Jobs added to the queue will be at the last line of the queue. Each process gets removed from the front of the queue one at a time (Kaur, 2018). FCFS is the most appropriate among data analysis algorithms because it effectively implements data sets and data. It performs the best in a basic application that does not need complicated logic (El-Sharawy, 2021). FCFS, often known as first-come, first-served, is the common service process used when scheduling. Any process in the ready state is inline in the FCFS based on its scheduled arrival time (Siahaan, 2017).

According to the study of Krishna and company (2018), they designed a MapReduce analysis process management system that uses the BigData Process Management (BPM) engine. They found out that First Come First Serve (FCFS) is the most appropriate algorithm for data analysis and time-efficient execution of data set and storing the data after analysis (Krishna, 2018).

In the study of Tarigan and colleagues (2020), The production in scheduling applications such as CPU scheduling, automobile manufacturing systems, and job sequencing in veneer lamination lines uses the FCFS approach to organize the work in order. However, the FCFS techniques consider scheduling in the manufacturing process since scheduling with this approach carries the client's first requests (Tarigan et al., 2020).

Shanthan and Arockiam (2018) research about Spell Aware Meta Task Scheduling Algorithm (SAMHSA) for Multi-cloud found out that one of the convenient ways to arrange tasks for execution in the multi-cloud environment is by using the FCFS algorithm to arrange the jobs in sequential order (Shanthan & Arockiam, 2018).

2.2 Emission Center

The Philippines requires emission testing as part of the vehicle registration process. The test determines the number of fumes or pollution emitted from a vehicle's exhaust manifold. The primary objective of the emission test is to reduce the number of air pollutants that are detrimental to the environment. It measures the concentrations of hydrocarbons, carbon dioxide, nitrogen oxide, carbon monoxide, and other air pollutants in the atmosphere, among other things. In general, the emission test determines whether a vehicle complies with a set of emission requirements. It is also necessary for a vehicle to have a CEC, or Certificate of Emission Compliance, to be eligible for renewal of its registration. PETC stands for Private Emission Testing Center, and it is a private facility that monitors the amount of gas emitted by a motor vehicle and the composition of that gas. It determines whether the vehicle complies with the Department of the Environment and Natural Resources' emission regulations (DENR). In addition, one must pass the requirements of the Department of Trade and Industry (DTI), particularly the Bureau of Philippine Standards (BPS), and get permission from the Department of Transportation (DOT) to operate as an authorized PETC (Sanchez, 2021).

2.2.1 Status of Emission Testing Center of Misamis Occidental

The emission testing center of Misamis Occidental followed the mandatory requirement with vehicle registration renewal. The machines of the emission testing center operate well and are still functional. As part of the

government regulation, the Land of Transportation (LTO) will only accept PMVIC results rather than the PETC results.

2.3 Optical Character Recognition (OCR)

Optical Character Recognition (OCR) is a crucial processing step that aims to improve the quality of OCR documents by detecting and removing residual errors (Nguyen et al., 2019). OCR has shown to be particularly useful in computerizing physical office papers by allowing users to input sales reports into a digital. However, it might not be easy to obtain text from images because image size, style, orientation, and complex background differences. There are numerous types of OCR software available today: Desktop OCR, Server OCR, and Web OCR The accuracy of any OCR technology ranges from 71% to 98 %. Many OCR tools are currently available. However, just a few are open source and free (Patel et al., 2017).

According to the study of Jasmhed and colleagues (2020) about Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR). OCR is a technology that converts many forms of texts or photos into analyzable, editable, and searchable data. One of the first creations of OCR systems was in the 1940s. With the evolution of technology, the system grew more resilient in dealing with both printed and handwritten characters, which resulted in the commercial availability of OCR machines. OCR is helping to digitize medieval manuscripts and turn typewritten materials into digital form (Memon et al., 2020).

2.3.1 Tesseract OCR Engine

Tesseract OCR engine that is free and released under the Apache V2.0 open-source license. Tesseract is an open-source optical character recognition engine. It started as a Ph.D. research project in HP Labs, Bristol. Furthermore, HP developed and enhanced the engine between 1984 to 1994. Then, the date of modification and improvement in 1995 with greater accuracy follows. In 2005, the public opened and made Tesseract available (Nair, 2017). Tesseract is one of the most accurate open-source OCR algorithms, with Google funding its research (Cakic et al., 2020).

In the study of Yankey and companion (2018) titled "Automatic Number Plate Recognition System using OpenCV and Tesseract OCR Engine," they studied automatic detection of plate numbers and process it using Tesseract OCR engine. As a result, the proposed Automatic Number Plate Recognition System (ANPR) detected the majority of Ghanaian license plates. Sixty percent of the time, the recognition rate was successful, with an average of 60 percent. Furthermore, the whole picture capture to character recognition stage takes roughly 0.2 seconds to process (Agbemenu et al., 2018).

According to Cakic and colleagues (2020) a study titled "The Use of Tesseract OCR Number Recognition for Food Tracking and Tracing," the researchers used Tesseract to recognize serial numbers from products. The original test dataset performed the recognition script without any preparation, and the system correctly recognized around 62 percent of pictures with whole serial numbers. The success rate climbed to 87.5 percent after preprocessing the photos. Thus, the better the preprocessing and photos in the test dataset, the better the outcomes (Cakic et al., 2020).

2.4 Short Message Service (SMS)

SMS has grown in popularity over the previous decade. In terms of business, text messages outperform emails in terms of effectiveness, where 98% of mobile users read their SMS at the end of the day, over 80% of emails stay unread. As a result, it is simple to see why SMS has evolved into a multibillion-dollar commercial sector (Roy et al., 2020).

Mobile devices are becoming increasingly popular as they become more affordable to the public. By lowering service costs, we can provide a wide range of services. For example, short SMS (Short Message Service) is a frequently utilized communication service. This increase attracted attackers, resulting in the SMS Spam problem. SMS is one of the cheapest methods to communicate and the simplest way to conduct phishing attacks since mobile devices carry sensitive and personal information such as credit card numbers, usernames, and passwords, among other things (Choudhary & Jain, 2017).

2.4.1 MessageBird

MessageBird is a service firm that offers an Application Programming Interface (API) for sending, receiving, and controlling SMS, Voice, and Whatsapp communications. MessageBird enables the integration of Short Message Services messages into the remote-control system and the automation of user-to-receiver communication (Winderl, 2020). In addition, MessageBird provides businesses with innovative methods to route consumer calls and helps them communicate more effectively with their customers to improve the efficiency and enjoyment of interactions with clients (Effing, 2019).

3. Methods Used

3.1 System Architecture

The study only focused on the online scheduling of emission testing centers to provide guidance and hassle-free scheduling. The system architecture is in Figure 1.

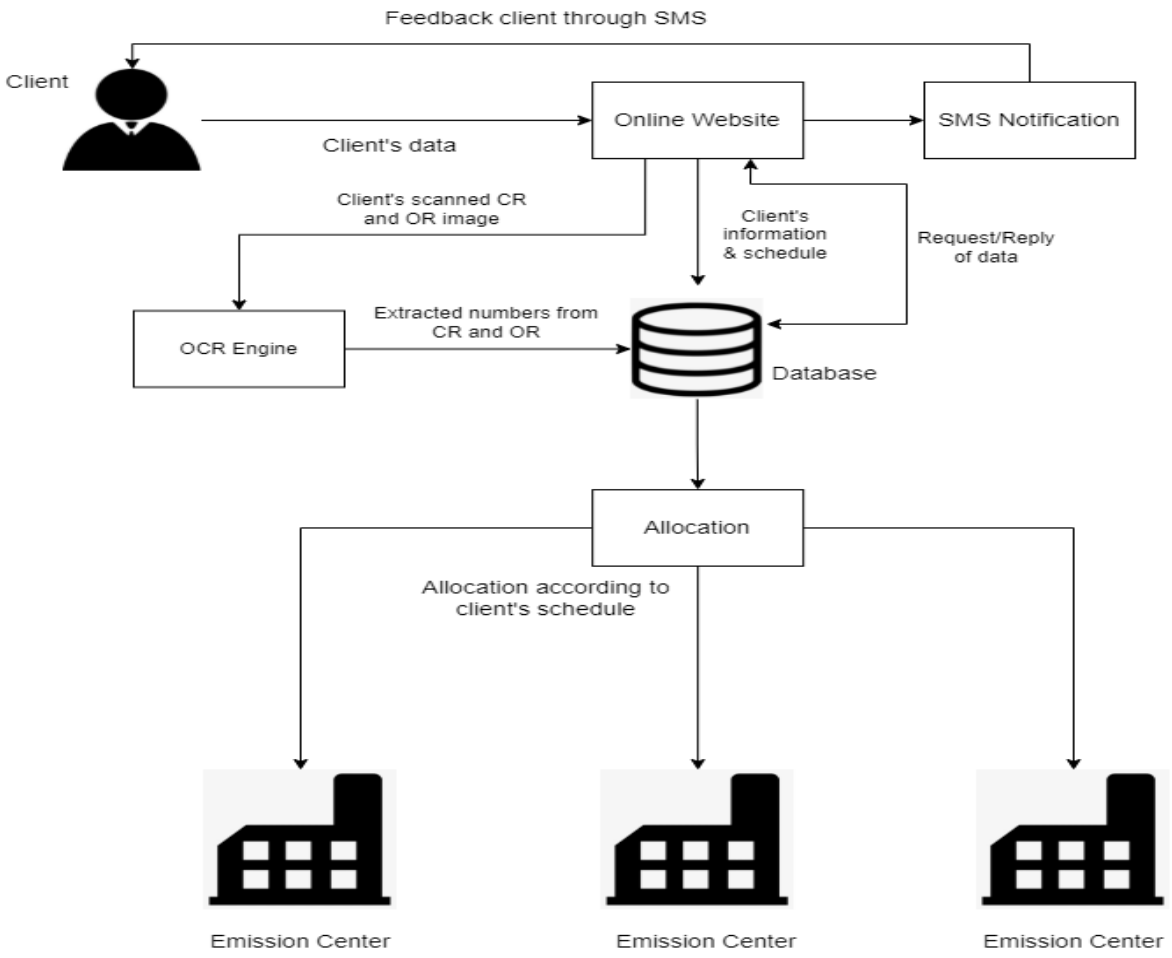

Figure 1. Architectural Design of the Proposed Study

The system only caters to online clients from the online website. The process begins when the client creates a vehicle emission schedule and then enters their personal information into the system. Next, the client will choose three different cities (Ozamiz City, Tangub City, and Oroquieta City) and select a date and time for the client's convenience. The client will also need to upload an image for the OR and CR, which will then the system will extract the CR and OR numbers using the Tesseract OCR. The process of the Tesseract OCR extraction presented in Figure 2.

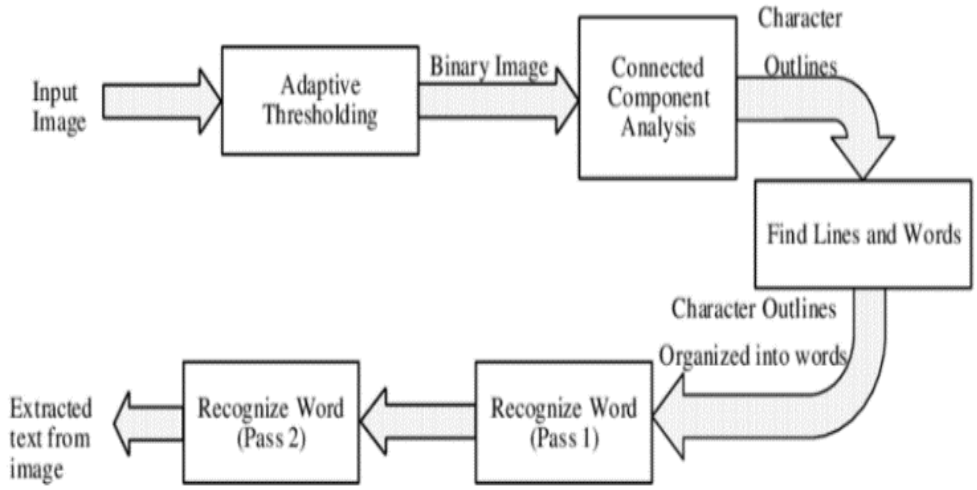

Figure 2. Tesseract OCR Diagram

This figure converts the picture to a binary form in the first stage, Adaptive Thresholding. The next phase utilizes the page layout analysis to extract the text blocks from the picture. The system recognized the character outlines recovered from the words after the baselines of each line and the separated text into words using definite and fuzzy spaces. After then, a two-pass process of text recognition begins. The first run of word recognition will be using the static classifier. Next, an adaptive classifier uses each right word as training data. Conducted the second pass through the page, the freshly trained adaptive classifier detects any phrases incorrectly detected the first time.

3.2 System Design

Figure 3 presents the extraction design for extracting OR and CR numbers.

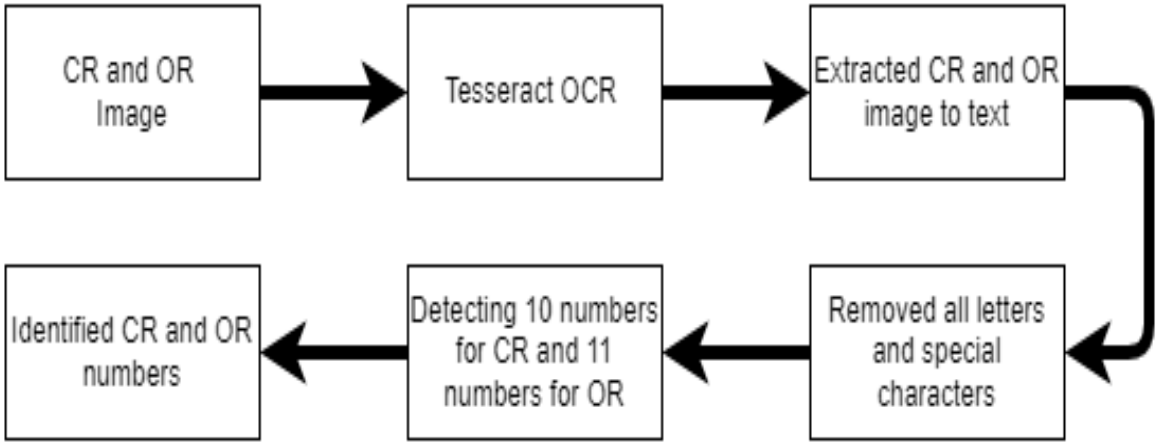

Figure 3. Extraction Design

The CR and OR number extraction process starts from the uploaded image, and afterward, the API of Tesseract OCR executes, which has pre-trained English language data. After that, the extracted text from the CR and OR undergoes filtering using a regular expression that removes all the letters and special characters from the text; only the numbers remain. After the filtering process, the system will look for ten digits for the CR and 11 digits for the OR. After the system finds the 10- and 11-digit numbers, it will identify them as CR and OR numbers.

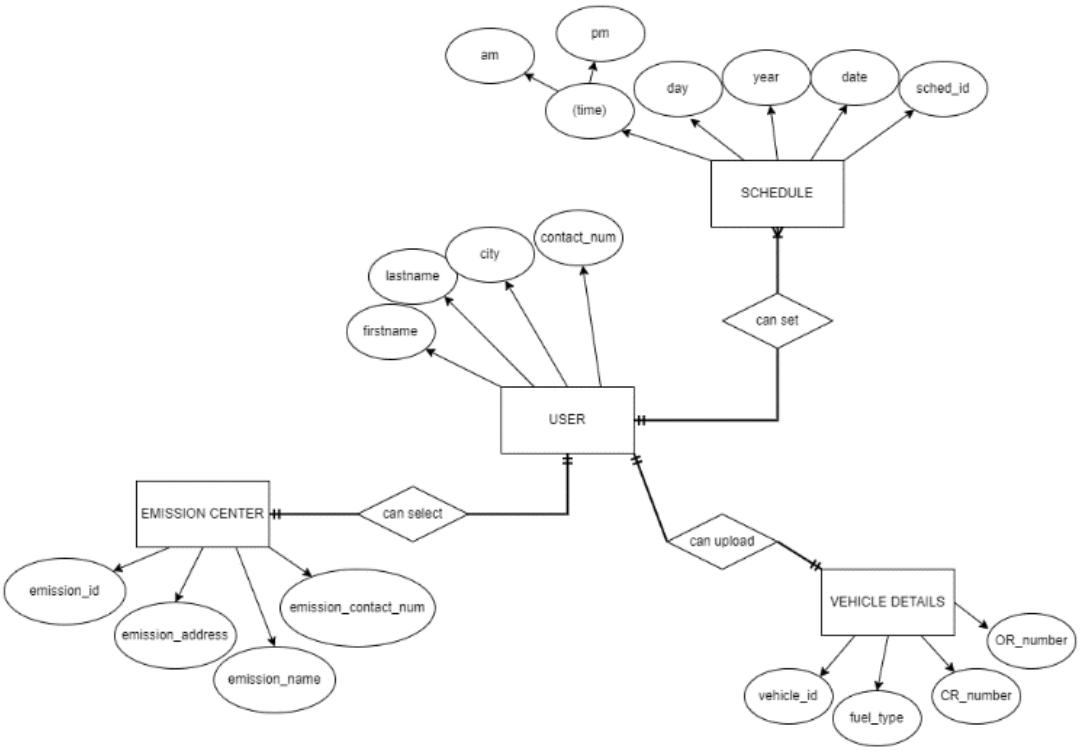

Figure 4. Entity-Relationship Diagram

The entity-relationship diagram presented in Figure 4 portrays the design of the system's database. It represents all the entities and attributes that correspond to their involvement within the database. For example, the user can set a schedule multiple times as long as the CR and OR numbers are different for each schedule. The user can also select one preferred emission center and select and upload CR and OR photos for the OCR extraction.

The researchers made a diagram to assist the data flow in the system. By doing this, the researchers will have management concerning the development of the system. The first diagram is the Top-Level Data Flow Diagram, which contains the data flow presented in Figure 5.

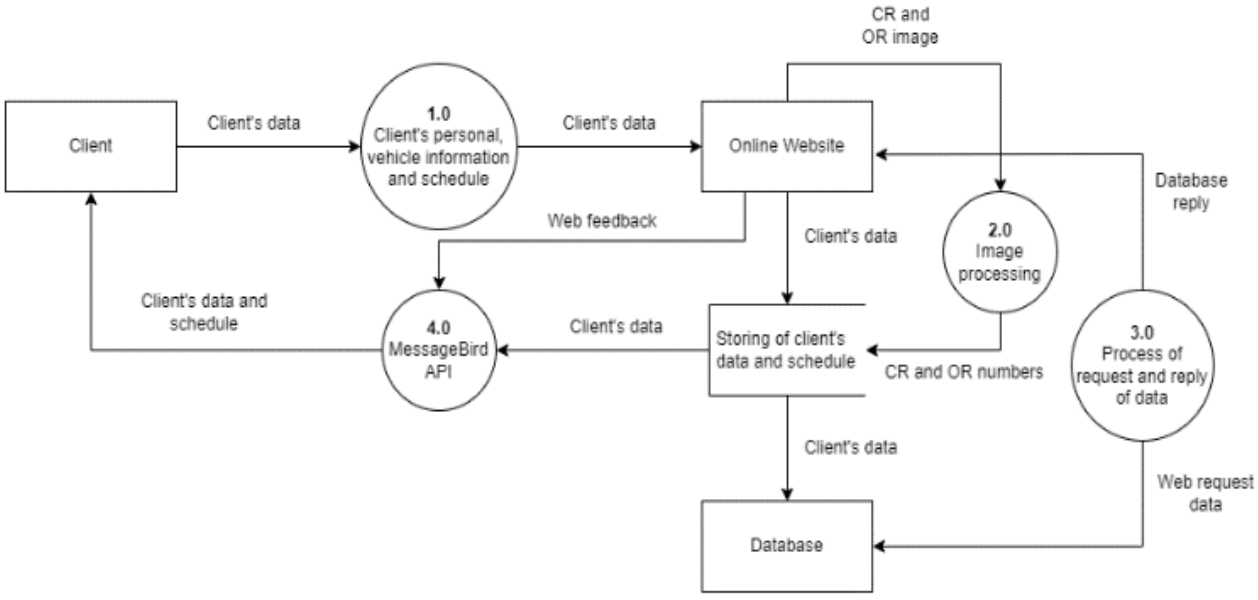

Figure 5. Level 1 Data Flow Diagram

The level 1 data flow diagram provides a rundown of how the data flow through the system allowing the researchers to have a good view of the system when it comes to the process of data. The process starts when the client sets a schedule and selects a date and time to upload pictures or scanned CR and OR images to verify vehicle ownership. The system disregards storing the uploaded photo, but it will extract and save the numbers of the CR and OR. Furthermore, the system displays the scheduled information online after the client submits the filled-up scheduled form. The client will receive a text message regarding the scheduled vehicle emission. The client will also receive an appointment code generated by the current year and a number that automatically increment by one for every scheduled client, acronym of the website, last three letters of the client's last name, and last three digits of the client's mobile number separated by a dash. The client will also view their current schedule using the appointment on the online website.

4. Results and Discussions

4.1 Accuracy Testing of OCR Engine

Table 1. CR and OR results

No. of Tests	CR Status	OR Status	Remarks
Test 1	✓	✓	✓
Test 2	✖	✓	✖
Test 3	✓	✖	✖
Test 4	✖	✖	✖
Test 5	✖	✓	✖
Test 6	✓	✖	✖
Test 7	✖	✓	✖
Test 8	✓	✓	✓
Test 9	✓	✓	✓
Test 10	✖	✓	✖
Test 11	✖	✖	✖
Test 12	✓	✓	✓
Test 13	✖	✓	✖
Test 14	✖	✓	✖
Test 15	✓	✖	✖
Test 16	✓	✓	✓
Test 17	✓	✓	✓
Test 18	✖	✓	✖
Test 19	✖	✓	✖
Test 20	✖	✓	✖
Test 21	✓	✓	✓
Test 22	✖	✓	✖
Test 23	✓	✓	✓
Test 24	✓	✓	✓
Test 25	✓	✓	✓
Test 26	✓	✓	✓
Test 27	✓	✓	✓
Test 28	✖	✖	✖
Test 29	✖	✓	✖
Test 30	✖	✓	✖

Based on the 30 sets conducted, 12 pairs of scanned CR and OR generated the data correctly, equivalent to 40% accuracy, while three pairs have incorrect OR, 12 pairs for incorrect CR, and three pairs of incorrect CR and OR. As a result, the total images scanned by the system are 60, and the system correctly identified 38 images by the Tesseract OCR engine, thus resulting in 63.33% accuracy. However, the accuracy of 63.33% is still lacking as CR

and OR determine the validity of the vehicle ownership. Therefore, there is still a chance of 36.67% not recognizing the CR, OR, or both, which can lead to the inability to set a schedule for vehicle emission testing.

5. Conclusions and Recommendations

The scheduling system helped businesses in prioritizing tasks. However, emission centers are still using the manual scheduling of clients through an on-site visit without guaranteeing any available schedule. Hence, this study developed a system that allows vehicle owners to set an online schedule for their emission testing based on real-time schedule availability. The system also used a Tesseract Optical Character Recognition (OCR) engine to scan the Certificate of Registration (CR) and Official Receipt (OR) image to extract CR and OR numbers. The main purpose of using OCR for the uploaded OR and CR is to assess the authenticity of a vehicle's ownership.

Moreover, the system obtained an extraction accuracy rate of 63.33. Furthermore, the stress testing on the website using loader.io resulted in 104ms average response time tested in 100 clients per second. The evaluation results show that the system provides the vehicle owners a convenient way of scheduling from the emission center without directly visiting the designated vehicle emission centers. The system's assessment findings reveal that it is simple to use, user-friendly, and all its functions are completely functional, allowing clients to learn to navigate the system quickly.

References

1. Agbemenu, A., Yankey, J., & Addo, E. (2018). An Automatic Number Plate Recognition System using OpenCV and Tesseract OCR Engine. International Journal of Computer Applications, 180(43), 1–5. https://doi.org/10.5120/ijca2018917150
2. Agbemenu, A., Yankey, J., & Addo, E. (2018). An Automatic Number Plate Recognition System using OpenCV and Tesseract OCR Engine. International Journal of Computer Applications, 180(43), 1–5. https://doi.org/10.5120/ijca2018917150
3. Cakic, S., Popovic, T., Sandi, S., Krco, S., & Gazivoda, A. (2020). The Use of Tesseract OCR Number Recognition for Food Tracking and Tracing. 2020 24th International Conference on Information Technology, IT 2020, February. https://doi.org/10.1109/IT48810.2020.9070558
4. Cesar, M. (2021). Emission Testing Centers: Everything you need to know. https://philkotse.com/car-maintenance/philippine-emission-test-centers-guide-8139
5. Choudhary, N., & Jain, A. K. (2017). Towards filtering of SMS spam messages using machine learning based technique. Communications in Computer and Information Science, 712, 18–30. https://doi.org/10.1007/978-981-10-5780-9_2
6. Effing, J. (2019). Obtaining venture capital funding as a young entrepreneur. http://essay.utwente.nl/80055/
7. El-Sharawy, E. E. (2021, January). (PDF) A Review on the CPU Scheduling Algorithms: Comparative Study. https://www.researchgate.net/publication/351330519_A_Review_on_the_CPU_Scheduling_Algorithms_Comparative_Study
8. Habibi, M. R. M., Mohammadabadi, F., Tabesh, H., Vakili-Arki, H., Abu-Hanna, A., & Eslami, S. (2019). Effect of an Online Appointment Scheduling System on Evaluation Metrics of Outpatient Scheduling System: a before-after Multicenter Study. Journal of Medical Systems, 43(8). https://doi.org/10.1007/s10916-019-1383-5
9. Joshi, A., & Goswami, S. (2017). Modified Round Robin Algorithm by Using Priority Scheduling. 10(6), 1543–1549.
10. Kaur, A. (2018). Different Task Scheduling Algorithms in Cloud Computing. International Journal of Latest Trends in Engineering and Technology, 9(3), 217–223. https://doi.org/10.21172/1.93.37
11. Krishna, M. V. (2018). BigData Processing using First Come First Served (FCFS) Algorithm. 7(7), 83–87.
12. Memon, J., Sami, M., Khan, R. A., & Uddin, M. (2020). Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR). IEEE Access, 8, 142642–142668. https://doi.org/10.1109/ACCESS.2020.3012542

13. Nguyen, T. T. H., Jatowt, A., Coustaty, M., Nguyen, N. Van, & Doucet, A. (2019). Deep statistical analysis of OCR errors for effective post-OCR processing. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, 2019-June (1), 29–38. https://doi.org/10.1109/JCDL.2019.00015
14. Patel, C., Patel, A., & Patel, D. (2017). Optical Character Recognition by Open-source OCR Tool Tesseract: A Case Study. International Journal of Computer Applications, 55(10), 50–56. https://doi.org/10.5120/8794-2784
15. Robin, R., Scheduling, J., & Sanyal, J. (2016). Comparison of Dynamic Time Slice Allocation- Based First Come First Serve and Priority Based. 4(3), 177–179. https://doi.org/10.17148/IJIREEICE.2016.4346
16. Roy, P. K., Singh, J. P., & Banerjee, S. (2020). Deep learning to filter SMS Spam. Future Generation Computer Systems, 102, 524–533. https://doi.org/10.1016/j.future.2019.09.001
17. Sanchez, H. (2021). 6 FAQs about emission test in the Philippines. https://philkotse.com/safe-driving/6-faqs-about-emission-test-in-the-philippines-4931
18. Shanthan, B. J. H., & Arockiam, L. (2018). Spell Aware Meta Task Scheduling Algorithm (SAMTSA) for Multicloud. October.
19. Siahaan, A. P. U. (2017). Comparison Analysis of CPU Scheduling FCFS, SJF and Round Robin. 4(3), 124–131. https://doi.org/10.31227/osf.io/6dq3p
20. Tarigan, U., Siregar, I., & Sari, R. M. (2020). Implementation of first come first served in door leaf production scheduling. IOP Conference Series: Materials Science and Engineering, 725(1). https://doi.org/10.1088/1757-899X/725/1/012023
21. Teorey, T. J., & Pinkerton, T. B. (2021). A Comparative Analysis of Disk Scheduling Policies. Communications of the ACM, 15(3), 177–184. https://doi.org/10.1145/361268.361278
22. Winderl, M. (2020). Report Data Workflow and Remote-Control Concept for IoT-based Smart Farming Project Irrimode by DHI WASY GmbH (associated partner) Study Project.