AlphaOCR: Integrated Deep Learning and Optical Character Recognition for Receipt Extraction

Muhammad Haikal Iman Osman[1], Nor Samsiah Sani[2]*, Luan Xiang Wei Liu[3],
Zalinda Othman[4], Mohd Aliff Afira Sani[5], Ibrahim Zebiri[6]
Center for Artificial Intelligence Technology-Faculty of Information Science Technology,
Universiti Kebangsaan Malaysia, Bangi, 43600, Selangor, Malaysia[1,2,3,4]
Quality Engineering Research Cluster (QEREC)-Malaysian Institute of Industrial Technology,
Universiti Kuala Lumpur, Bandar Seri Alam, 81750, Johor, Malaysia[5]
Department of Computer Science-Faculty of Sciences,
20 Août 1955 Skikda University, Skikda, 21000, Skikda, Algeria[6]

***Abstract*—Receipts are vital documents that validate transactions by capturing key details such as dates, items purchased, prices, and seller information. However, accurately documenting receipts is often compromised by issues like blurring, which can result from poor image quality, physical wear, or suboptimal scanning conditions. While the Processing Key Information Extraction from Documents Using Improved Graph Learning-Convolutional Networks (PICK) deep learning model is effective for structured information extraction, it struggles with processing blurred text, leading to inaccuracies in data retrieval. To overcome these challenges, this research introduces AlphaOCR, a web-based application that integrates the PICK model with advanced Optical Character Recognition (OCR) technology. The methodology integrates OCR-based line-item recognition with PICK-based structured-field extraction to extend the range of information recovered from receipts. The deep learning component was evaluated using mean Entity Precision (mEP), mean Entity Recall (mER), mean Entity F1-score (mEF), and mean Entity Accuracy (mEA), together with a paired t-test for the entity-level mEF results. PICK-2 increased the mEF from 83.32% to 83.88%, corresponding to a modest absolute gain of 0.56 percentage points. The complete AlphaOCR configuration combining PICK-2 with EasyOCR achieved an overall accuracy of 87.96% under the evaluation setting used in this study. Among the OCR models evaluated, EasyOCR achieved a Character Accuracy Rate (CAR) of 92.04% and provided the most suitable output characteristics for integration with PICK-2 within the tested configuration. The reported results should be interpreted within the experimental scope of this study rather than as a direct performance ranking against previously published systems that used different datasets, tasks, extracted fields, and evaluation metrics. AlphaOCR provides a practical integrated workflow for receipt information extraction, although broader claims regarding robustness, scalability, and generalisability require evaluation on larger and more diverse receipt collections. This research contributes to receipt information extraction primarily through the system-level integration of graph-based key information extraction, OCR-based item recognition, output fusion, and a web-based verification workflow. Future work should evaluate the system on larger and more diverse datasets and investigate enhanced semantic understanding and additional document types.**

***Keywords*—AlphaOCR; optical character recognition; information extraction; receipt processing; deep learning**

*Corresponding author.

I. Introduction

Artificial intelligence (AI) enables computer systems to perform tasks involving pattern recognition, data interpretation, prediction, and decision-making. Machine-learning models have been applied to security-related problems, including Internet of Things threat detection and interpretable risk analysis [1]. Predictive machine-learning methods have also supported environmental monitoring by modelling complex water-quality characteristics from observed data [2]. In the financial and socioeconomic domain, classification algorithms have been used to identify household overspending patterns among different income groups [3]. Unsupervised learning techniques, including clustering, have been applied to analyse and classify student academic performance [4]. Optimisation and data-fusion approaches have further supported scientific applications such as ligand-based virtual screening [5]. Deep-learning architectures have also demonstrated strong image-classification capability in medical applications, including the automated classification of malignant and benign skin lesions [6]. Together, these studies demonstrate the adaptability of AI, machine learning, and deep learning across structured-data, predictive, optimisation, and image-analysis tasks. This broad technical capability provides a foundation for their application to document-understanding problems, although receipt information extraction introduces additional challenges related to text recognition, document layout, and field relationships.

The rapid increase in digital transactions has led to a substantial growth in the volume of digital receipts generated across industries [7]. As organisations seek to digitise and streamline their operations, the accurate extraction and processing of information from these receipts has become crucial [8]. Optical Character Recognition (OCR) technologies have been widely adopted for this purpose. Yet, they face challenges in dealing with the diverse formats, fonts, and layouts inherent to receipts [9]. These challenges often result in inaccuracies in data extraction, which can undermine the efficiency and reliability of automated systems [10].

The practical impact of these limitations is also reflected in previous receipt-processing studies. Kumar et al. reported an accuracy of 83% for larger bill receipts, corresponding to a residual error proportion of approximately 17% under their experimental setting [14]. Odd and Theologou reported an

accuracy of 72.32% for price extraction, leaving approximately 27.68% of the evaluated price fields incorrectly extracted [19]. Although these values were obtained under different experimental conditions, they demonstrate that apparently strong document-level recognition performance may still leave a substantial proportion of business-relevant fields requiring manual verification or correction. Such residual errors can propagate into expense records, transaction summaries, and downstream reporting workflows. These documented field-level errors motivate the need for a receipt-processing architecture that does more than recognise individual characters. The system must also preserve the relationships among structured receipt fields, line-item descriptions, and their corresponding prices.

This research introduces AlphaOCR, an advanced system designed to address the limitations of existing OCR methods, particularly in receipt processing. The system leverages deep learning models, specifically the PICK (Processing Key Information Extraction) models, which are enhanced through hyperparameter tuning to improve performance [11]. The study compares two versions of the PICK model: PICK-1, the original model developed by Yu et al. (2020), and PICK-2, a modified version that incorporates improvements aimed at increasing accuracy and efficiency.

Despite progress in receipt understanding, three technical limitations remain insufficiently addressed in the reviewed approaches. First, the PICK configuration used in this study extracts a fixed set of structured entities but does not directly recover line-item product descriptions and their corresponding prices. Second, conventional OCR engines can recognise characters from the receipt body, but their outputs do not necessarily preserve phrase continuity or the structural association between a product description and its price. Third, independently generated structured-field and OCR outputs require an explicit fusion and verification mechanism before they can be used as a coherent transaction record.

The unresolved research gap is therefore not limited to blur sensitivity. It concerns how graph-based structured information extraction and OCR-based line-item recognition can be integrated without losing field relationships, while allowing uncertain or incorrect outputs to be reviewed before storage and reporting. AlphaOCR addresses this gap through a dual-branch architecture in which PICK-2 extracts predefined structured entities, an OCR component processes line-item content, and a web-based verification stage consolidates and corrects the combined output.

Accordingly, this study investigates whether a graph-based structured extractor can be extended with an OCR branch and a verification layer to recover both receipt-level entities and line-item information within a unified processing workflow.

The primary objectives of this research are to compare PICK-1 and PICK-2 using multiple entity-level evaluation metrics, to examine the performance of different OCR engines when integrated with PICK-2, to evaluate the resulting AlphaOCR receipt-processing configuration, and to assess the usability of its web-based application. Rather than claiming direct superiority over previously published systems, the study characterises the performance of the proposed components under a common internal evaluation setting.

The objectives of this study are as follows:

- To introduce AlphaOCR, an integrated receipt-processing architecture that combines graph-based key information extraction through PICK-2 with OCR-based product and price recognition.
- To compare PICK-1 and PICK-2 using mEP, mER, mEF, and mEA, and to evaluate TesseractOCR, KerasOCR, and EasyOCR under the same study-specific experimental setting.
- To implement the integrated model as a web-based application supporting receipt upload, output fusion, user verification, correction, analysis, and report generation.

This article is organised into key sections: Section II reviews recent literature on receipt processing and OCR advancements. Section III outlines the methodology used in this research, detailing the experimental setup and model configurations. Section IV offers an in-depth analysis of the experimental outcomes, including model performance and usability testing. Finally, Section V summarises the findings, discusses the research implications, and Section VI concludes the study and suggests potential directions for future work. To position this architecture within existing document-processing research, the next section compares OCR-centred, layout-aware, and graph-based information-extraction approaches.

II. Related Work

In 2019, Anh Le Duc, Dung Van Pham, and Tuan Anh Nguyen conducted a study on text detection and recognition in receipts. They used the Connectionist Text Proposal Network (CTPN) algorithm for detecting text, utilising Visual Geometry Group (VGG-16), Bidirectional Long Short-Term Memory (BiLSTM), and bounding box regression to detect horizontal text effectively. They employed the Attention-based Encoder-Decoder (AED) algorithm for text recognition, which used DenseNet and OCR for validation and was particularly effective at recognising mathematical expressions and handwritten Vietnamese characters [12]. The study utilised the SROIE2019 dataset, which was divided into training, testing, and validation sets. Three variations of the CTPN algorithm were tested: CTPN alone, CTPN with preprocessing, and CTPN with preprocessing and OCR validation. The best results were achieved with the CTPN combined with preprocessing and OCR validation, yielding F1, Precision, and Recall scores of 71.9%, 72.5%, and 71.3%, respectively. While the CTPN with preprocessing alone had a slightly higher Recall of 72.3%, the combined approach outperformed across other metrics. A key limitation was the exclusion of handwritten text detection and recognition, indicating a direction for future research [13].

In 2020, Vedant Kumar and his team conducted a study that used OCR technology to extract information from receipt and bill images, making them machine-readable [14]. The study applied four main methods: removing shadows and watermarks, segmenting long invoices, recognising and extracting text, and retrieving relevant information. They used OpenCV for image processing due to its effectiveness in handling image data and Tesseract OCR for text extraction because of its ability to detect text in various languages and produce outputs closely matching the input images. Although the specific dataset details were not provided, the researchers tested their approach

on 25 bill receipts, achieving high accuracy rates—97% for small bill receipts and 83% for larger ones. Peak Signal-to-Noise Ratio (PSNR) values ranged from 38.13 to 42.57 dB. While PSNR was less pivotal in this study, the main strength of their research was the ability to extract text even in the presence of watermarks and shadows. However, a limitation was that the method was best suited for images with printed text [14].

In 2019, Xiaohui Zhao, Endi Niu, Zhuo Wu, and Xiaoguang Wang developed an OCR system for extracting key information from receipts. The study introduced grid positional mapping for CNN algorithms, utilising two versions: Convolutional Universal Text Information Extractor (CUTIE-A), which merges features from various resolutions, and CUTIE-B, which uses atrous convolution and ASPP to improve context detection [16]. They tested the system using the SROIE2019 dataset, split 75:25 for training and testing. CUTIE-A achieved AP scores of 90.8% for taxi receipts, 77.7% for food and entertainment receipts, and 69.5% for hotel receipts, with softAP scores reaching up to 97.2%. CUTIE-B performed even better, outperforming models like CloudScan and BERT. However, the study had a limitation: it lacked preprocessing steps to connect text into meaningful entities during grid mapping, relying only on existing algorithms [17].

In 2020, Weiliang Liu and colleagues developed an end-to-end neural-network method for recognising taxi receipt information without relying on conventional image-processing and OCR pipelines. The method used an SSD model with a VGG-16 backbone to identify informative regions and a specialised CNN and GRU network for character recognition. The dataset contained 1,800 taxi receipt images, and the reported cross-validation accuracy was 94.36%. The evaluation was restricted to taxi receipts from a specific company in Beijing [15]. A related application-oriented taxi receipt text-recognition system was subsequently presented by Liu et al. [18].

In 2018, Joel Odd and Emil Theologou investigated the use of OCR in machine learning for classifying receipts and extracting specific data points. They converted physical receipt datasets to digital form and trained several machine learning algorithms, comparing the performance of four: LinearSVC, MLP, Na¨ıve Bayes, and DummyClassifier. The Google Drive REST API was selected for OCR, demonstrating better accuracy than the Azure Computer Vision API [19]. The study found that LinearSVC performed best, achieving 94.07% accuracy for receipt classification and 72.32% for extracting specific data points like prices. However, the overall accuracy was still not optimal, indicating a need for more advanced methods to improve results.

In a study by Peng Zhang et al., the TRIE model achieved the highest accuracy of 93.26% for information extraction from receipts, outperforming the GCN model's 92.60%. The dataset used included SROIE2019 and a combination of Chinese text receipts and resumes. The TRIE algorithm, which integrates FPN and ResNet-50, demonstrated superior performance in this context [20].

In 2020, Yu et al. introduced the deep learning model PICK, designed for extracting key information (KIE) from documents [21]. Evaluated using the SROIE2019 dataset, PICK outperformed baseline models, achieving an impressive mEF score of 96.1%. This model integrates convolutional visual features, graph learning, sequential encoding, and CRF-based structured prediction for key information extraction [21].

In 2019, Xiaojing Liu and colleagues introduced a model for extracting information using a combined textual and visual approach [22]. The study utilised the Value-Added Tax Invoices (VATI) and International Purchase Receipts (IPR) datasets, highlighting the benefits of incorporating visual elements into information extraction. Their Bidirectional Long Short-Term Memory (BiLSTM) + Conditional Random Fields (CRF) + Graph Convolutional Network (GCN) model outperformed baseline models, achieving an F1 score of 87.3% on VATI and 83.6% on IPR [22].

In 2020, Bodhisattwa Prasad Majumder and their team demonstrated that Representation Learning models significantly outperformed Bag of Words (BoW) and Multilayer Perceptron (MLP) models in extracting information from receipts and invoices, achieving an F1 score of 83.9% on the Invoices1 and Invoices2 datasets, and 83.3% on the SROIE2019 dataset [23]. In 2022, Cheng Jian Lin and colleagues proposed the YOLOv4-s model, which was compared to YOLOv3, YOLOv4, and YOLOv5 on printed and handwritten receipt datasets. Although it was tested on a specific dataset, YOLOv4-s achieved the highest accuracy of 99.39%, showcasing its strength in minimising downsampling during image processing [24].

Finally, Chang et al. (2004) present a real-time license plate recognition method that addresses challenges like varying plate sizes, orientations, and lighting conditions [25]. Their system utilises image processing techniques to locate and segment license plates from camera-captured images, enhancing key features through edge detection and morphological operations. After detecting the license plate, the system extracts characters using template matching or neural networks. The system was evaluated for accuracy and efficiency across different conditions, demonstrating its effectiveness in overcoming the typical challenges of license plate recognition systems.

The numerical results reported in previous studies should be interpreted with caution. For example, TRIE reported an accuracy of 93.26%, while YOLOv4-s reported an accuracy of 99.39% [20], [24]. However, these values were obtained using different datasets, task definitions, document categories, extracted fields, and evaluation procedures. Similarly, the mEF reported for the original PICK study and the overall accuracy reported for AlphaOCR represent different evaluation scopes. Consequently, these reported values do not constitute a direct head-to-head comparison and should not be used to rank the methods solely according to their numerical scores.

A broader review of deep-learning-based information extraction was provided by Yang et al. [21], who surveyed the evolution from earlier feature-engineering pipelines to more recent neural architectures for extracting structured information from documents.

Table I shows a transition from OCR-centred pipelines towards layout-aware and graph-based document-understanding models. OCR-centred approaches can recover text effectively but do not inherently preserve semantic relationships among receipt fields. In contrast, structured KIE models such as PICK and TRIE model visual and contextual relationships,

TABLE I. COMPARATIVE SYNTHESIS OF RECEIPT AND DOCUMENT INFORMATION-EXTRACTION APPROACHES

Study	Architecture	Dataset	Main Strength	Reported Limitation	Remaining Gap
Le et al. [12]	CTPN, VGG-16, BiLSTM, and attention-based recognition	SROIE2019	Combines text detection and recognition for receipt images	Handwritten content was not evaluated and the pipeline contains multiple processing stages	Limited support for unified structured-field and line-item extraction
Kumar et al. [14]	OpenCV preprocessing and Tesseract OCR	25 bill receipts	Handles shadows, watermarks, and printed-text degradation	Performance decreased for larger bills and the method focused on printed text	No explicit modelling of relationships among extracted fields
Zhao et al. [16]	CUTIE-A/B with grid-based convolutional document modelling	SROIE2019 and receipt categories	Represents text and spatial layout in a two-dimensional grid	Depends on grid mapping and lacks an explicit verification workflow	Limited integration of OCR outputs with user-correctable structured records
Liu et al. [15]	SSD with VGG-16 and CNN–GRU recognition	1,800 taxi receipts	High recognition accuracy within a specific receipt domain	Restricted to taxi receipts from a specific company	Limited cross-domain generalisation and field extensibility
Odd and Theologou [19]	OCR services with conventional machine-learning classifiers	Custom receipt collection	Demonstrates receipt classification and field extraction from OCR text	Price and date extraction remained substantially less accurate than receipt classification	OCR text was not integrated with graph-based document structure
TRIE [20]	FPN, ResNet-50, text recognition, and information extraction	SROIE2019, Chinese receipts, and resumes	End-to-end text reading and information extraction	Uses task- and dataset-specific evaluation settings	No user-assisted correction or application-level output fusion
PICK [11]	CNN, graph learning, BiLSTM, and CRF	SROIE2019	Models visual, spatial, and sequential relationships among document entities	The evaluated configuration extracts a predefined entity schema	Does not directly recover product–price line-item pairs in the present implementation
AlphaOCR	PICK-2, OCR, output fusion, and web-based verification	SROIE2019 and a controlled OCR evaluation	Combines structured fields, line-item recognition, and user verification in one workflow	OCR evaluation and usability assessment remain limited in scale	Requires broader external validation and more automated line-item association

but their output schemas and evaluation settings are task-specific. AlphaOCR was therefore designed to combine the complementary strengths of these approaches: graph-based structured-field extraction, OCR-based line-item recognition, and user-assisted output verification.

PICK was selected as the structural information extraction component of AlphaOCR not because it achieved the highest reported numerical result among all cited approaches, but because its architecture was specifically designed for key information extraction from visually rich documents. Its combination of convolutional feature extraction, graph-based relationship modelling, sequential encoding, and structured prediction enables it to represent spatial and contextual relationships among receipt fields.

The contribution of AlphaOCR is therefore positioned as a system-level integration rather than a claim of state-of-the-art extraction accuracy. PICK-2 extracts the predefined structured entities of company, address, date, and total, whereas the OCR component extends the output to product-level text and price information. The combined system further provides output fusion, user verification, correction, storage, and reporting within a web-based workflow. A direct comparison with TRIE, YOLOv4-s, or other external systems would require those models to be reimplemented and evaluated on the same data partition, extracted fields, preprocessing procedures, and evaluation metrics. The comparison identifies the need for an architecture that combines structured-field modelling, line-item OCR, output fusion, and user verification. The following section describes how these components were implemented in AlphaOCR.

III. METHODOLOGY

TThe AlphaOCR system was designed to support the extraction of structured fields and line-item information from receipts. Fig. 1 illustrates the development process, highlighting the integration of three primary components: the PICK deep learning model, the OCR model, and the web-based application. The methodology involves five phases: PICK model development, OCR extraction and PICK integration, output fusion, Post-Processing and Web Application Deployment, and AlphaOCR System Evaluation.

A. PICK Model development

The PICK component comprises three main modules: an encoder capable of interpreting the layout and text in the input image, a graph module that receives embeddings from the encoder to explore and extract hidden features in the receipt embeddings, and a decoder that identifies phrases with the highest probability values.

1) Encoder: The encoder in the PICK model uses Convolutional Neural Networks (CNNs) to extract hierarchical features from the receipt images. These features include spatial and visual patterns, such as text blocks, lines, and characters. The encoder transforms the raw image data into feature maps, a foundation for further analysis. The feature maps represent the various components of the receipt, allowing the model to capture the structure and layout of the document. This is crucial for identifying key regions containing important information, such as the receipt's header, body, and footer.

2) Graph module: After encoding, the extracted features are passed to the graph module, which employs Graph Learning-Convolutional Networks (GLCN) to model the relationships between different text regions. Each node corresponds to a text region in this graph representation, and edges represent the spatial and logical relationships between these regions. The graph module is essential for understanding the layout of the receipt and how different elements are interconnected. For example, it helps the model determine the

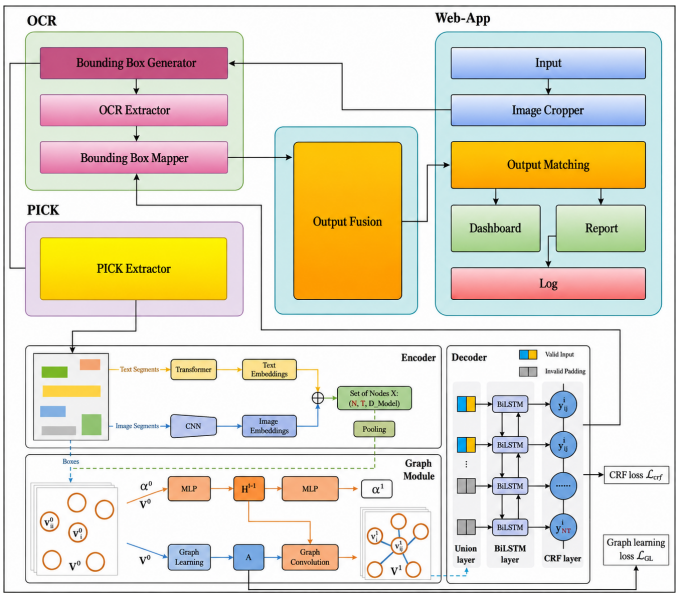

Fig. 1. AlphaOCR model development.

relationship between a product name and its corresponding price or between the company name and the total amount. By learning these relationships, the model can more accurately interpret the structure of the receipt and extract relevant information.

3) Decoder: The decoder stage of the PICK model integrates Bidirectional Long Short-Term Memory (BiLSTM) networks and Conditional Random Fields (CRF) to generate the final structured output. BiLSTM networks capture sequential dependencies between text regions, ensuring that the model accurately reflects the order and flow of information on the receipt. For instance, it helps ensure that product names are correctly associated with their prices and that dates are properly linked to their respective transactions. The CRF layer then takes these sequential outputs and applies structured prediction to make coherent and contextually accurate decisions about the content of the receipt. This structured prediction allows the model to extract key fields, such as the company name, address, date, and total amount, with high precision and consistency.

The development of AlphaOCR began with the construction, adaptation, and training of the PICK deep learning model. This model was trained on the publicly available SROIE2019 dataset, which consists of 973 purchase receipts from Malaysia. The SROIE2019 dataset comprises three key components: images, bounding boxes, and entities, all of which are essential for model training. The images, provided in *.jpg format, often include varying degrees of blurring, reflecting real-world challenges such as poor image quality. The bounding boxes, stored in *.csv format, contain coordinates that enclose the text within the image, and the entities, saved in *.json format, represent the crucial information the model must extract from the receipt. In this dataset, the bounding box files also store the text bounded by these coordinates, allowing the model to effectively learn how to recognise and extract key information despite potential blurring. These components are commonly referred to as img (image), box (bounding box),


tan woon yann

BOOK TA .K (TAMAN DAYA) SDN BHD
789117-W
NO.5, 55,57 & 59, JALAN SAGU 18,
TAMAN DAYA,
81100 JOHOR BAHRU,
JOHOR.

Document No : TD01167104
Date : 25/12/2018 8:13:39 PM
Cashier : MANIS
Member :

CASH BILL

CODE/DESC		PRICE	Disc	AMOUNT
QTY		RM		RM
9556959040116	KF MODELLING CLAY KIDDY FISH			
1 PC	*	9.000	0.00	9.00

Total : 9.00
Rour ding Adjustment : 0.00
Round :d Total (RM): 9.00

Cash 10.00
CHANGE 1.00

9.00

GOODS SOLD ARE NOT RETURNABLE OR EXCHANGEABLE
货物出门，恕不退还或交换
商品门后，恕拒退换，概新！

THANK YOU
PLEASE COME AGAIN !

Fig. 2. Example of an image file from the SROIE2019 dataset with ID "000".

and key or entity (entity). Fig. 2, 3, and 4 show examples of image files, bounding boxes, and entities for the SROIE2019 dataset with the ID "000".

B. OCR Extraction and PICK Integration

Once the models are trained, the AlphaOCR system integrates the outputs of the PICK and OCR models to produce a comprehensive extraction of receipt information:

1) OCR extraction: The OCR component consists of three main modules: a bounding box generator that creates boundary boxes around phrases on the receipt, an OCR extractor for extracting text from the receipt's body, and a bounding box mapper that maps the boundary boxes to the areas identified by the OCR model. The PICK deep learning model is specifically designed to extract four key pieces of information from a receipt: the issuing company, the company address, the date of issuance, and the total payment made. The OCR model extracts two additional important details, namely the product and price. Initially, the receipt is processed by the PICK deep learning model to extract the four main pieces of information.

	A	B	C	D	E	F	G	H	I	J	K
1	72	25	326	25	326	64	72	64	TAN WOON YANN		
2	50	82	440	82	440	121	50	121	BOOK T A. K(TAMAN DAYA) SDN BHD		
3	205	121	285	121	285	139	205	139	783417-W		
4	110	144	383	144	383	163	110	163	NO.53 55	57 & 59	JALAN SAGU 18
5	192	169	299	169	299	187	192	187	TAMAN DAYA		
6	162	193	334	193	334	211	162	211	81100 JOHOR BAHRU		
7	217	216	275	216	275	233	217	233	JOHOR.		
8	50	342	279	342	279	359	50	359	DOCUMENT NO.: TD01167104		
9	50	372	96	372	96	390	50	390	DATE:		
10	165	372	342	372	342	389	165	389	25/12/2018 20:13		
11	48	396	117	396	117	415	48	415	CASHIER:		
12	164	397	215	397	215	413	164	413	MANIS		
13	49	423	122	423	122	440	49	440	MEMBER:		
14	191	460	298	460	298	476	191	476	CASH BILL		
15	30	508	121	508	121	523	30	523	CODE/DESC		
16	200	507	247	507	247	521	200	521	PRICE		
17	276	506	306	506	306	522	276	522	DISC		
18	374	507	441	507	441	521	374	521	AMOUNT		
19	69	531	102	531	102	550	69	550	QTY		
20	221	531	247	531	247	545	221	545	RM		
21	420	529	443	529	443	547	420	547	RM		
22	27	570	137	570	137	583	27	583	9.55694E+12		
23	159	570	396	570	396	584	159	584	KF MODELLING CLAY KIDDY FISH		
24	77	598	113	598	113	613	77	613	1PC		
25	138	597	148	597	148	607	138	607	'		
26	202	597	245	597	245	612	202	612	9		
27	275	598	309	598	309	612	275	612	0		
28	411	596	443	596	443	613	411	613	9		
29	245	639	293	639	293	658	245	658	TOTAL:		
30	118	671	291	671	291	687	118	687	ROUNDING ADJUSTMENT:		
31	408	669	443	669	443	684	408	684	0		
32	86	704	292	704	292	723	86	723	ROUND D TOTAL (RM):		
33	401	703	443	703	443	719	401	719	9		
34	402	748	441	748	441	763	402	763	CASH		
35	205	770	271	770	271	788	205	788	10		
37	412	772	443	772	443	786	412	786	CHANGE	1	
38	97	845	401	845	401	860	97	860	GOODS SOLD ARE NOT RETURNABLE OR		
39	190	864	309	864	309	880	190	880	EXCHANGEABLE		
40	142	883	353	883	353	901	142	901	...		
41	137	903	351	903	351	920	137	920	...		
42	202	942	292	942	292	959	202	959	THANK YOU		
43	163	962	330	962	330	977	163	977	PLEASE COME AGAIN!		
44	412	639	442	639	442	654	412	654	9		

Fig. 3. Example of a bounding box file from the SROIE2019 dataset with ID "000".


```
{
   "company": "BOOK TA .K (TAMAN DAYA) SDN BHD",
   "date": "25/12/2018",
   "address": "NO.53 55,57 & 59, JALAN SAGU 18, TAMAN DAYA, 81100 JOHOR BAHRU, JOHOR.",
   "total": "9.00"
}
```


Fig. 4. Entities for the SROIE2019 dataset with the ID "000".

Afterwards, the receipt is cropped and focused solely on the body section to extract information related to the product, quantity, and product price. This technique simplifies the complex post-processing of text by narrowing the focus to the body section, which is crucial since receipts are generally divided into three main sections: header, body, and footer.

2) PICK extraction: Simultaneously, the PICK extractor module processes the original receipt image to extract key information. During this process, the encoded feature maps generated by the CNNs are analysed by the graph module to understand the spatial layout of the receipt. The graph module identifies and models the relationships between different text regions, such as the connection between the header (e.g., company name) and the footer (e.g., total amount). The BiLSTM networks within the decoder then analyse the sequence of text regions to maintain logical consistency. At the same time, the CRF layer finalises the structured output, accurately extracting fields such as the company name, address, date, and total amount.

3) Bounding box mapping: The bounding box mapper module examines the extracted results, visually mapping the identified text areas on the receipt. This step ensures accurate identification and localisation of information.

C. Output Fusion

Following the extraction process, the Output Fusion module merges the data from the PICK and OCR models into a single, coherent output. This fusion process is essential for integrating the structured data extracted by the PICK model with the textual information extracted by the OCR model. The module aligns and combines the different data points to create a unified dataset that is both comprehensive and easy to interpret. The result is a complete and accurate representation of the receipt's content, ready for further processing or display.

D. Post-Processing and AlphaOCR Web Application Deployment

The post-processing phase is crucial in refining the fused output to ensure the extracted information is accurate and user-friendly. This final, polished data is presented to users through a web-based application, the primary interface for interacting with the processed information.

In the Output Matching and Feedback stage, the system presents the extracted data via the web application's output matching module. Users are encouraged to review and correct the extracted text, especially for crucial details such as the company name, address, date, amount paid, and product information. This feedback mechanism is essential for continuously improving the system's accuracy.

The refined data is then displayed on an Analysis Dashboard within the web application, allowing users to download the extracted information. The dashboard includes interactive features such as filtering, sorting, and exporting, allowing users to manipulate the data according to their needs.

In terms of Data Storage and Access, all extracted and refined information is consolidated and securely stored in a singular log file. This file is accessible to the administrator through the application's backend, ensuring the data is archived securely and available for future reference or analysis.

The AlphaOCR web application consists of seven main modules: the input module, which receives receipt images; the image cropper, which trims the body section of the receipt; the output fusion module, which integrates the extraction results from both OCR and PICK into a single output; the output matching module, which corrects and aligns the extracted text with its appropriate category (such as company name, company address, date, total amount paid, product, or product price); the dashboard, which displays an analysis of the extraction results; the report module, which allows users to download the processed information; and the log module, which stores cumulative extraction data accessible by the admin through the application's backend.

The AlphaOCR web-based application was developed using Python, leveraging key modules such as Streamlit, which is particularly effective in presenting a deep learning model as a user-friendly web application. Fig. 5 illustrates the design of the AlphaOCR User Guide interface, offering users clear instructions and guidance on using the application. Fig. 6 displays the design of the AlphaOCR Home interface, where users can upload receipt images and initiate the extraction process. Finally, Fig. 7 showcases the design of the AlphaOCR Extractor interface, where users can view, review, and interact with the extracted data.

Fig. 5. Design of the AlphaOCR user guide interface.

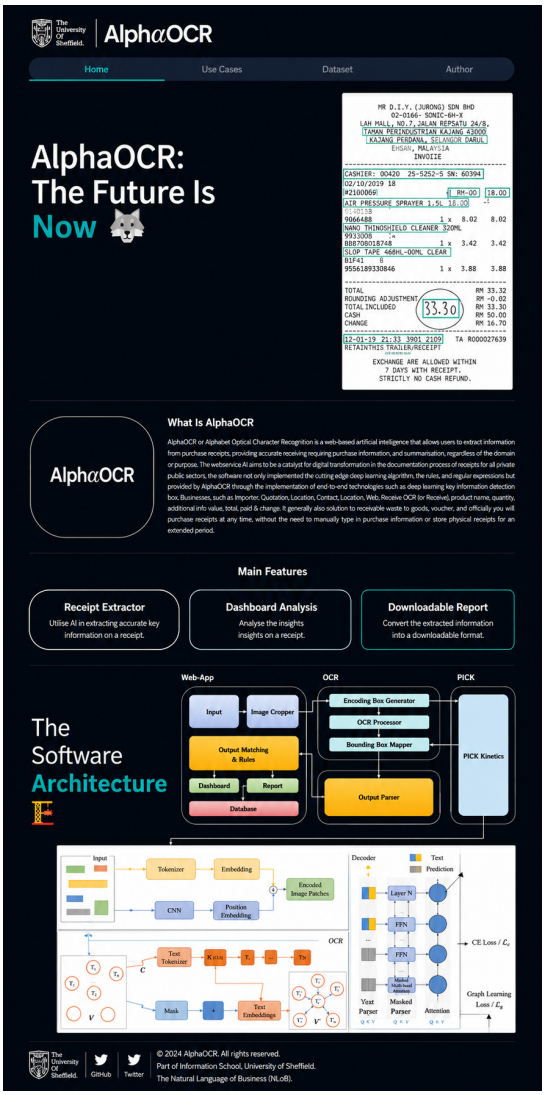

Fig. 6. Design of the AlphaOCR home interface.

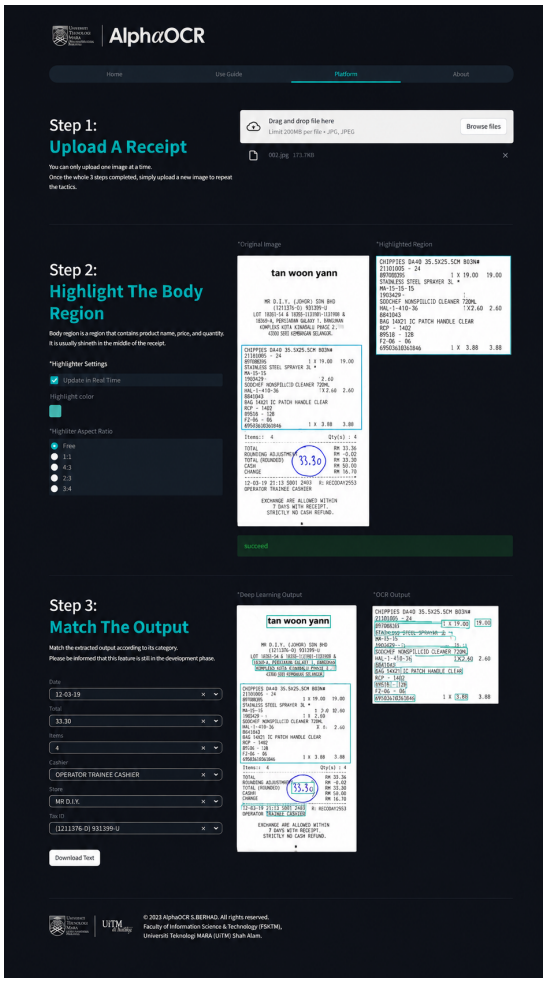

Fig. 7. Design of the AlphaOCR extractor interface.

E. AlphaOCR System Evaluation

The deep learning component of AlphaOCR was evaluated by comparing PICK-1 and PICK-2 using four entity-level metrics: mean Entity Precision (mEP), mean Entity Recall (mER), mean Entity F1-score (mEF), and mean Entity Accuracy (mEA). These metrics were examined across the extracted entities and during both the training and testing phases. In addition to the descriptive comparison, a paired t-test was applied to the paired entity-level mEF values to assess whether the observed difference between PICK-1 and PICK-2 was statistically significant. Because statistical significance does not by itself indicate the practical magnitude of an improvement, the absolute difference in mEF between the two models was also reported and interpreted together with the loss values and the consistency between training and testing performance.

The OCR comparison was conducted using 20 selected phrases from one standardised receipt image. The same phrases were processed using TesseractOCR, KerasOCR, and EasyOCR to provide a consistent within-image comparison. This controlled evaluation was intended to support OCR-engine selection for the AlphaOCR prototype rather than to serve as a general benchmark across diverse receipt collections.

TABLE II. EVALUATION PHASES AND MODELS

Component	Evaluation Phase	Model
Deep Learning Model	Evaluating the performance of two PICK models through multiple assessment metrics.	PICK-1, PICK-2
OCR Model	Controlled phrase-level comparison for OCR-engine selection.	TesseractOCR, KerasOCR, EasyOCR
Web-based Application	Conducting usability testing on the developed application.	AlphaOCR

In addition to the technical evaluations, a formative usability assessment was conducted to identify interface, navigation, and interaction issues in the AlphaOCR web application. The assessment involved 15 participants organised into three groups of five. The small-group evaluation was intended to support prototype refinement. It was not designed to provide a statistically representative estimate of usability or satisfaction for the broader population of intended users. The resulting evaluation design examined the structured extractor, the OCR engines, the integrated configurations, and the formative user interaction. The next section reports the corresponding experimental and usability results.

IV. RESULTS

The evaluation phase was crucial for assessing the effectiveness of the AlphaOCR system. This phase involved evaluating the performance of the PICK models, comparing different OCR models, and conducting usability testing for the web-based application, as summarised in Table II. Table II outlines the evaluation phase of a system consisting of three key components: a deep learning model, an OCR model, and a web-based application. During the evaluation of the deep learning model, the performance of two models, PICK-1 and PICK-2, was assessed using multiple metrics to determine their effectiveness in tasks such as entity recognition and information extraction. In the OCR model evaluation, The three OCR engines were compared using the controlled phrase-level evaluation protocol described in Section III-E. This phase focused on measuring the accuracy and efficiency of these models in text extraction. Lastly, the web-based application was subjected to usability testing to evaluate its interface, user experience, and overall functionality, with AlphaOCR being the specific application under review. This comprehensive evaluation process ensured that each system component was thoroughly tested for performance, accuracy, and user-friendliness.

A. Deep Learning Model PICK

The deep learning component of the AlphaOCR system was tested using two versions of the PICK model. The first version, PICK-1, is the original model developed by Yu et al., which served as the baseline for comparison. The second version, PICK-2, is a modified iteration of the original model, incorporating hyperparameter tuning designed to enhance performance.

As summarised in Table III, the two versions of the PICK model differ in backbone depth and activation function while sharing the same optimiser. PICK-1, which uses the ResNet-50 architecture, the Adam optimiser, and the ReLU activation function, served as the baseline model for performance comparison. PICK-2, the enhanced version, utilises the deeper ResNet-152 backbone, still employs the Adam optimiser, but incorporates the Leaky ReLU activation function with the aim of improving model performance. These variations in architecture and activation functions were intended to optimise the model's accuracy and effectiveness in the deep learning tasks of the AlphaOCR system.

TABLE III. PARAMETER SETTINGS FOR PICK-1 AND PICK-2

Model	ResNet	Optimizer	Activation Function
PICK-1	ResNet-50	Adam	ReLU
PICK-2	ResNet-152	Adam	Leaky ReLU

Fig. 8 shows a comparison of four key evaluation metrics, including mean Entity Precision (mEP), mean Entity Recall (mER), mean Entity F1-score (mEF) and mean Entity Accuracy (mEA) across different entities for the PICK-1 and PICK-2 models, divided into training and testing phases. The green colour scheme represents the PICK-1 model, with darker green indicating training results and a lighter green for testing results. Similarly, the orange scheme is used for the PICK-2 model, with darker orange for training and lighter orange for testing.

The observations from Fig. 8 reveal that PICK-1 generally exhibits higher precision during training, particularly in the "Date" entity. However, PICK-2 demonstrates more consistent precision during testing, notably in "Company" and "Total." Both models maintain stable values across entities when examining recall, with PICK-1 showing slightly higher recall during training. In the testing phase, PICK-2 matches or surpasses PICK-1 in recall for the "Date" and "Company" entities, indicating better generalisation. The F1-score trends align with precision and recall, where PICK-1 performs better during training, especially for "Date," but PICK-2 displays a more balanced F1-score during testing. Accuracy remains high for both models across all entities, with PICK-1 showing slightly better accuracy during training. However, PICK-2 demonstrates less variability between training and testing, suggesting it may generalise better in real-world scenarios. Fig. 9 compares GL Loss, CRF Loss, and Total Loss values between the PICK-1 and PICK-2 models across both training and testing phases. The colour scheme remains consistent, with green representing PICK-1 and orange representing PICK-2, using lighter shades for testing results. The GL Loss remains nearly identical between training and testing for both models, indicating consistent global loss minimisation. However, when examining CRF Loss, PICK-2 consistently shows lower values than PICK-1 in both phases, with a slight increase observed during testing, suggesting greater efficiency in handling the conditional random fields (CRF) component.

The Total Loss comparison highlights a significant distinction between the two models, with PICK-2 consistently achieving lower Total Loss during the training and testing phases. The increase in Total Loss from training to testing is more pronounced in PICK-1, indicating that PICK-2 is better at maintaining performance when transitioning from training to testing. Overall, the analysis of both figures suggests

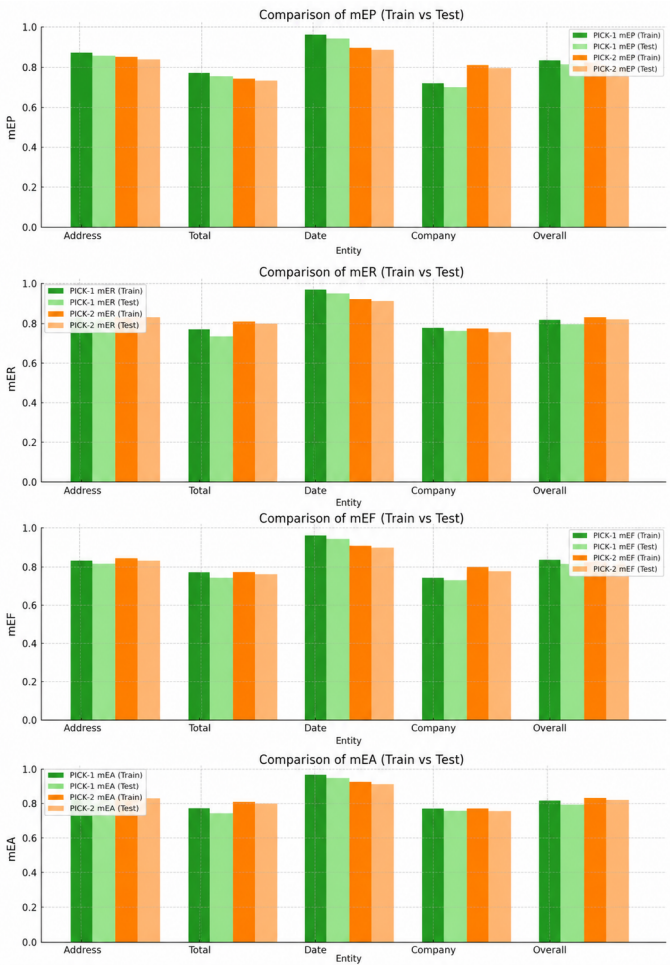

Fig. 8. Comparison of mEP, mER, mEF, and mEA for PICK-1 and PICK-2 during training and testing.

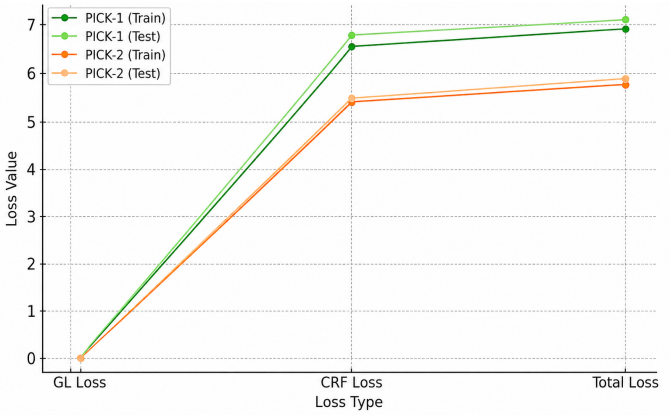

Fig. 9. Comparison of GL loss, CRF loss, and total loss for PICK-1 and PICK-2 during training and testing.

that while PICK-1 may have a slight edge in some training metrics, PICK-2 demonstrates more consistent and reliable performance, particularly during the testing phase. These results suggest that PICK-2 achieved more stable optimisation under the present experimental setting. However, evaluation on independent receipt datasets is required before drawing conclusions about its suitability for real-world deployment.

To assess the statistical significance of the difference between PICK-1 and PICK-2, a paired t-test was conducted using the paired entity-level mEF values produced by the two models. The test yielded a p-value of 0.042, indicating that the observed difference was statistically significant at the 5% level. Nevertheless, the mEF increased from 83.32% for PICK-1 to 83.88% for PICK-2, representing an absolute improvement of only 0.56 percentage points. Therefore, the result indicates a statistically detectable but practically modest improvement. It should not be interpreted as a large increase in extraction performance. The main observed benefit of PICK-2 was its more consistent testing behaviour and lower CRF and total loss values, rather than a substantial increase in mEF alone.

B. Optical Character Recognition (OCR) Model

The OCR models were evaluated based on their accuracy in correctly extracting twenty selected phrases from a standardised receipt image. The performance of TesseractOCR, KerasOCR, and EasyOCR was assessed using the Character Accuracy Rate (CAR) metric. The phrase extraction accuracy values are calculated by determining the ratio of correctly extracted characters to the total number of characters in the tested phrase. The average phrase extraction accuracy value can be obtained by calculating the ratio of phrase extraction accuracy to the total number of tested phrases. These values can be converted into percentages by multiplying the obtained value by one hundred. This method is known as the CAR. The two equations below show the calculation method for both values, considering symbols but not white spaces or case sensitivity.

- Phrase Extraction Accuracy: For each tested phrase i, the Character Accuracy Rate was calculated as:

$$\mathrm{CAR}_i = \frac{C_i}{N_i}, \tag{1}$$

where, C_i is the number of correctly recognised characters and N_i is the total number of characters in the corresponding reference phrase. Symbols were included, whereas white spaces and letter case were excluded from the comparison.

- Average Phrase Extraction Accuracy: The average Character Accuracy Rate across the M tested phrases was calculated as:

$$\overline{\mathrm{CAR}} = \frac{1}{M}\sum_{i=1}^{M}\mathrm{CAR}_i. \tag{2}$$

Fig. 10 shows that TesseractOCR achieved the highest character-level accuracy, followed closely by EasyOCR, whereas KerasOCR obtained a substantially lower CAR under the present evaluation setting. Although TesseractOCR achieved the highest CAR, its output more frequently separated words belonging to the same product phrase. EasyOCR achieved comparable character-level accuracy while preserving phrase grouping more consistently. Because phrase continuity was important for matching product descriptions with their

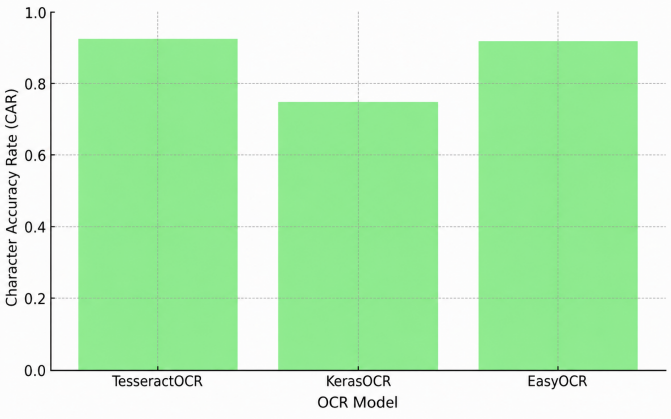

Fig. 10. Comparison of Character Accuracy Rate (CAR) of OCR models.

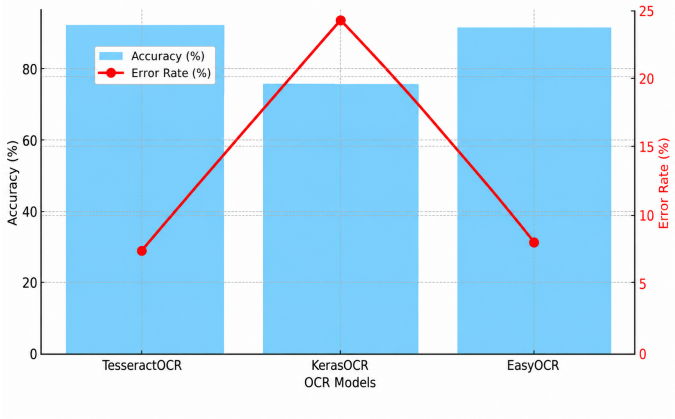

Fig. 11. Comparison of OCR model accuracy and error rates.

corresponding prices, EasyOCR was selected for downstream integration with PICK-2. This selection reflects compatibility with the AlphaOCR processing workflow and should not be interpreted as a general claim that EasyOCR outperforms all other OCR engines across different receipt types and image conditions.

Paired t-tests were conducted using the phrase-level CAR values to examine differences among the three OCR engines under the same controlled input condition. The results indicated that TesseractOCR and EasyOCR produced similar character-level performance, whereas KerasOCR obtained lower accuracy. Because the observations were derived from the same receipt image, the statistical comparisons describe within-image differences and do not establish population-level superiority across independent receipt collections. Furthermore, Fig. 11 presents the performance of three Optical Character Recognition (OCR) models—TesseractOCR, KerasOCR, and EasyOCR—by comparing their Accuracy (%) and Error Rate (%).

The blue bars represent accuracy, indicating the percentage of correctly recognised characters by each model. Higher Accuracy values signify better character recognition performance. The red line, marked with data points, represents the Error Rate, which denotes the percentage of characters each model failed to recognise accurately. A lower Error Rate is preferable, as it reflects fewer recognition errors by the model.

TesseractOCR and EasyOCR demonstrate high accuracy rates, exceeding 90%, corresponding to relatively low Error Rates. TesseractOCR exhibits an Error Rate just above 7%, while EasyOCR's Error Rate is slightly higher, though still under 8%. These results suggest that both models are highly effective in character recognition, with minimal errors. In contrast, KerasOCR shows a significantly lower accuracy rate, around 75%, coupled with an Error Rate of approximately 25%. This indicates that KerasOCR is less reliable than the other two models, making more frequent mistakes in character recognition.

Overall, Fig. 11 illustrates the expected inverse relationship between character accuracy and error rate. The results were used to support OCR-engine selection for the current AlphaOCR prototype. Broader conclusions require evaluation

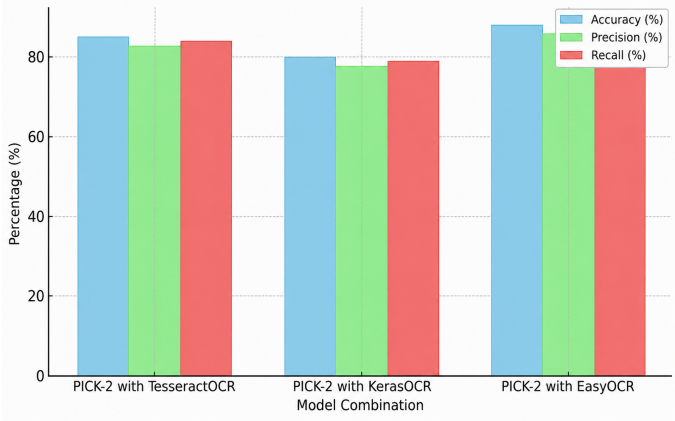

Fig. 12. Comparative performance metrics of PICK-2 with different OCR models.

using independent receipts with greater variation in layout, font, language, and image quality.

C. Combined Performance of PICK-2 with OCR Models

To evaluate the effectiveness of the AlphaOCR system, the PICK-2 model was combined with each OCR method to assess accuracy in processing receipts. Fig. 12 shows that PICK-2 combined with EasyOCR achieved the highest accuracy, precision, and recall among the three evaluated configurations, with values of 87.96%, 88.10%, and 87.80%, respectively. PICK-2 combined with TesseractOCR ranked second, while the KerasOCR configuration obtained the lowest values across the three metrics. These findings indicate that the choice of OCR engine influenced the performance of the integrated AlphaOCR configuration under the present experimental setting.

The combination of PICK-2 with EasyOCR achieved the highest accuracy at 87.96%, indicating that this pairing was the most effective in correctly extracting information from receipts. This configuration also demonstrated a strong balance between precision (88.10%) and recall (87.80%), reflecting its ability to make accurate predictions and capture a high proportion of the relevant information in the dataset.

In comparison, the combination of PICK-2 with TesseractOCR achieved a slightly lower accuracy of 85.72%. Its recall

of 86.30% was also lower than the 87.80% achieved by the EasyOCR configuration, while its precision was 84.50%.

The pairing of PICK-2 with KerasOCR performed the least effectively, with an accuracy of 78.34%. Both precision (76.80%) and recall (79.90%) were significantly lower than those achieved by the other combinations, suggesting that KerasOCR struggled to accurately and consistently extract the correct information.

Overall, PICK-2 combined with EasyOCR was the best-performing configuration among the three combinations evaluated under the present experimental setting. This result supports its selection for the AlphaOCR prototype but does not establish that it is the most reliable configuration across all receipt layouts and image-quality conditions.

To determine whether these differences in accuracy were statistically significant, paired t-tests were conducted comparing the performance of PICK-2 combined with each OCR model. The t-test between PICK-2 with TesseractOCR and PICK-2 with EasyOCR resulted in a p-value of 0.045, indicating that the improvement in accuracy with EasyOCR is statistically significant at the 5% level. The comparison between PICK-2 with KerasOCR and PICK-2 with EasyOCR yielded a p-value of 0.001, confirming a highly significant difference in favour of EasyOCR. Lastly, the comparison between PICK-2 with TesseractOCR and PICK-2 with KerasOCR produced a p-value of 0.003, demonstrating that TesseractOCR performs significantly better than KerasOCR when paired with PICK-2. These t-test results support the conclusion that PICK-2 combined with EasyOCR is the most effective configuration for the AlphaOCR system.

D. Web-Based Application Usability Testing

The usability evaluation of the AlphaOCR web application was designed as a formative assessment of the developed prototype. A total of 15 participants were divided into three groups of five. The small-group design was used to expose the prototype to users with different technical backgrounds and age profiles and to identify prominent interface, navigation, and interaction issues.

The evaluation was not designed to provide a statistically representative estimate of usability or user satisfaction for the broader population of receipt-processing users. Therefore, the findings were interpreted as preliminary feedback for prototype refinement rather than population-level validation of the application.

The usability sessions were conducted on July 3 and July 4, 2023. Group 1 consisted of Generation Z participants with programming experience. Group 2 consisted of participants without programming experience, while Group 3 consisted of Generation Y participants with programming experience. These groups enabled the prototype to be examined from different user perspectives; however, no population-level inference was made from the differences among the three groups.

During each 30-minute session, participants explored the application, completed a receipt-processing task, and responded to a questionnaire concerning the interface, instructions, navigation, layout, task-completion experience, and extraction errors. The collected responses were used to identify practical usability issues and areas requiring interface refinement.

TABLE IV. REPRESENTATIVE FEEDBACK FROM THE FORMATIVE USABILITY GROUPS

Question	Feedback	Analysis
What do you think as you view web-based AI?	"The web interface is simple, attractive, and accessible."	Most respondents highlighted the cosmetic aspects of the application.
How was the experience of using the product to complete this task?	"Simple and easy to use."	Four out of five respondents found the application easy to use.
What are your thoughts on the instructions provided?	"Clear and straightforward instructions."	All respondents agreed that the instructions were clear and easy to understand.
How easy or difficult was it to navigate?	"Easy to navigate."	All respondents found the navigation process easy and logical.
What are your thoughts on the design and layout?	"Minimalistic design and well-structured layout."	The design and layout were seen as strengths of the application.
How long did it take the user to complete this task?	"Less than 5 minutes to complete the task for one receipt."	Most respondents completed the task in under 6 minutes.
How would you describe your overall experience with the product?	"Refreshing and easy to use."	All respondents described the application as easy to use and a refreshing experience.
What did you like the most about using this product?	"The simplicity and clean layout."	Three respondents highlighted the application's ability to extract information from receipts.
What did you like the least?	"The errors encountered during extraction."	Two respondents faced errors during extraction.
What, if anything, surprised you about the experience?	"The AI's ability to read various fonts."	All respondents appreciated the AI's ability to extract information.
What, if anything, caused you frustration?	"The errors during extraction."	Three respondents found the errors during extraction frustrating.

The feedback summarised in Table IV indicates that participants in the tested sample generally expressed positive views regarding the interface, instructions, and navigation. However, extraction errors were identified as a source of frustration for several participants. These observations should not be generalised to the wider intended user population. Broader usability validation would require a larger and more representative sample that includes users who routinely process receipts, such as accounting personnel, expense-management staff, and business administrators.

V. DISCUSSION

The comparison between PICK-1 and PICK-2 showed that the architectural and activation-function modifications applied to PICK-2 produced a statistically significant but modest improvement in entity-level performance. The mEF increased from 83.32% to 83.88%, representing an absolute gain of 0.56 percentage points. Accordingly, the result does not demonstrate a substantial increase in extraction accuracy. The more notable benefit of PICK-2 was its comparatively consistent behaviour between training and testing and its lower CRF and total loss values. These observations suggest that the modified configuration may provide more stable optimisation under the present experimental setting. However, additional evaluation on independent and more diverse receipt datasets is required

before stronger conclusions regarding generalisation can be drawn.

Although the overall accuracy of the PICK-2–EasyOCR configuration was 87.96%, this value should not be directly ranked against the 93.26% reported for TRIE or the 99.39% reported for YOLOv4-s. The studies differ in dataset composition, document type, extracted fields, task formulation, and evaluation metric. The present study therefore does not claim that AlphaOCR outperforms these external methods. Its contribution lies instead in combining structured entity extraction, item-level OCR, output fusion, and user-assisted verification within an integrated receipt-processing application.

The OCR comparison showed that TesseractOCR and EasyOCR achieved similar character-level performance, while KerasOCR obtained a lower CAR. EasyOCR was selected because its phrase grouping was more compatible with the downstream product-and-price matching process, despite TesseractOCR achieving a slightly higher character-level score. Because the OCR evaluation was based on phrases obtained from a single receipt image, these findings should be interpreted as preliminary and specific to the present experimental setting. Future work should evaluate the OCR engines using multiple independent receipts with different layouts, fonts, languages, resolutions, and image-degradation levels.

When combining the PICK-2 model with the various OCR models, the results showed that PICK-2 paired with EasyOCR achieved the highest accuracy of 87.96%. Within the configurations evaluated in this study, the PICK-2–EasyOCR combination achieved the highest overall accuracy. This finding supports its selection for the AlphaOCR prototype but does not establish superiority across other receipt datasets, layouts, or image-quality conditions. These results indicate that the choice of OCR engine affected the performance of the integrated system. In the present evaluation, EasyOCR provided phrase grouping that was more compatible with the downstream product-and-price extraction process. The formative usability evaluation provided preliminary feedback on the interface and identified extraction errors as a source of frustration for several participants. Within the configurations evaluated in this study, the PICK-2–EasyOCR combination achieved the highest overall accuracy. This finding supports its selection for the AlphaOCR prototype but does not establish superiority across other receipt datasets, layouts, or image-quality conditions. The results also indicate that OCR phrase grouping influenced the performance of the downstream product-and-price extraction process. The formative usability evaluation identified generally positive perceptions of the interface, instructions, and navigation among the tested participants. It also showed that extraction errors caused frustration for several users. These findings provide preliminary guidance for interface refinement rather than population-level validation.

Overall, AlphaOCR achieved an accuracy of 87.96% under the experimental setting used in this study. This result demonstrates the feasibility of the integrated workflow but does not establish general performance across broader receipt layouts and image conditions. While the combination of the PICK-2 model and EasyOCR proved to be highly effective, there remains potential for further enhancements. Future research should focus on expanding the system's capabilities to include additional fields, integrating advanced Natural Language Processing (NLP) techniques to handle text variations better, and improving the system's scalability to process larger and more diverse datasets across different industries and regions.

A. Limitations and Failure Modes

Several limitations constrain the interpretation and generalisability of the reported results. First, PICK-1 and PICK-2 were evaluated using the SROIE2019 receipt collection. Although this dataset contains Malaysian receipts with different layouts, performance on the same dataset does not establish generalisation to receipts from other countries, languages, vendors, or acquisition devices. The 0.56-percentage-point improvement in mEF was also modest, indicating that the deeper ResNet-152 backbone and Leaky ReLU activation did not produce a large increase in entity-level extraction accuracy.

Second, the OCR comparison was limited to phrases obtained from one standardised receipt image. The evaluation therefore did not systematically measure performance across different blur levels, font sizes, languages, lighting conditions, receipt layouts, or physical degradation patterns. Phrase fragmentation was observed in the TesseractOCR output, while the lower CAR obtained by KerasOCR indicated greater character-recognition difficulty under the tested condition. EasyOCR was selected because its phrase grouping was more compatible with the downstream workflow, not because it was established as universally superior.

Third, errors in the OCR stage can propagate into line-item construction. Incorrect characters, fragmented phrases, missed prices, or incorrect body-region selection can prevent a product description from being associated with the correct price. The current verification interface reduces this risk by allowing users to correct the output, but this also means that the final workflow is not fully automatic and its reliability partly depends on user review.

Fourth, the formative usability assessment involved only 15 participants and was intended to identify interface and interaction issues. It cannot provide a population-level estimate of usability or satisfaction. Extraction errors reported during the usability sessions further indicate that recognition quality directly influences user confidence and task experience.

Finally, the study did not systematically evaluate inference latency, computational cost, concurrent-user capacity, or large-scale deployment performance. Future work should therefore include external datasets, controlled image-degradation tests, multilingual receipts, automatic product–price association, confidence-aware output fusion, and system-level scalability measurements. Taken together, these findings and limitations support a cautious interpretation of AlphaOCR as a modular proof of concept rather than a universally validated receipt-processing solution. The Conclusion summarises the broader insights derived from this integration.

VI. Conclusion

This study examined how graph-based structured information extraction and OCR-based line-item recognition can be combined within a verified receipt-processing workflow. The results show that extending a structured KIE model with an OCR branch can broaden the output schema beyond receipt-level fields such as company, address, date, and total to include

product-and-price information. However, successful integration depends not only on character recognition accuracy but also on phrase continuity, spatial organisation, and the ability to verify uncertain output.

The comparison between PICK-1 and PICK-2 provides a second insight. Increasing backbone depth and changing the activation function produced more stable optimisation and lower loss values, but only a modest improvement in mEF. This suggests that greater architectural complexity alone may not substantially improve receipt understanding when the remaining errors originate from layout variation, OCR uncertainty, field association, or dataset limitations.

The OCR comparison further indicates that the engine with the highest character-level score is not necessarily the most suitable component for an integrated document-intelligence system. EasyOCR was selected because its phrase grouping was more compatible with downstream product-and-price processing, whereas TesseractOCR achieved a slightly higher CAR but more frequently fragmented product phrases. This finding highlights the importance of evaluating OCR engines according to downstream structural requirements rather than character accuracy alone.

AlphaOCR therefore contributes a modular design principle for document intelligence: specialised structured-field extraction, line-item recognition, output fusion, and human verification can be combined without treating any single model as sufficient for the complete task. The verification layer is particularly important when residual recognition errors can affect business-relevant fields.

The current findings remain limited by the modest PICK-2 gain, the controlled single-image OCR comparison, the small formative usability sample, and the absence of external and scalability evaluation. Future document-intelligence systems should investigate confidence-aware fusion, automatic product which price association, multilingual and degraded receipts, external datasets, and end-to-end architectures that preserve both character accuracy and document structure while reducing the need for manual correction.

ACKNOWLEDGMENT

This research was funded by the Universiti Kebangsaan Malaysia (Grant code: DPK-GPS-JORDAN-2024-018).

CONTRIBUTION

Muhammad Haikal Iman Osman: Methodology, Software, and Writing—original draft preparation. Nor Samsiah Sani: Funding acquisition, supervision, and writing—review and editing. Luan Xiang Wei Liu: Methodology, Writing—review and editing. Zalinda Othman: Conceptualization and writing—review and editing. Mohd Aliff Afira Sani: Writing—review and editing. Ibrahim Zebiri: Software and writing—review and editing.

ETHICS DECLARATIONS

Conflict of Interest/Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Ethical and Informed Consent for Data Used: The data used in this study are available from the following resources in the public domain: https://rrc.cvc.uab.es/?ch=13

REFERENCES

[1] Dobrojevic, M., Zivkovic, M., Chhabra, A., Sani, N. S., Bacanin, N., & Amin, M. M. (2023). Addressing internet of things security by enhanced sine cosine metaheuristics tuned hybrid machine learning model and results interpretation based on shap approach. *PeerJ Computer Science*, 9, e1405.

[2] Suwadi, N. A., Derbali, M., Sani, N. S., Lam, M. C., Arshad, H., Khan, I., & Kim, K. I. (2022). An optimized approach for predicting water quality features based on machine learning. *Wireless Communications and Mobile Computing*, 2022(1), 3397972.

[3] Othman, Z. A., Bakar, A. A., Sani, N. S., & Sallim, J. (2020). Household overspending model amongst B40, M40 and T20 using classification algorithm. *International Journal of Advanced Computer Science and Applications*, 11(7).

[4] Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering analysis for classifying student academic performance in higher education. *Applied Sciences*, 12(19), 9467.

[5] Holliday, J., Sani, N., & Willett, P. (2018). Ligand-based virtual screening using a genetic algorithm with data fusion. *Match: Communications in Mathematical and in Computer Chemistry*, 80(3).

[6] Bassel, A., Abdulkareem, A. B., Alyasseri, Z. A. A., Sani, N. S., & Mohammed, H. J. (2022). Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. *Diagnostics*, 12(10), 2472.

[7] Hänninen, M. (2019). Review of studies on digital transaction platforms in marketing journals. *The International Review of Retail, Distribution and Consumer Research*, 30, 164–192. https://doi.org/10.1080/09593969.2019.1651380

[8] Kandyba, V., Kushnir, O., Bredikhin, V., & Khoroshylova, I. (2023). Study of machine learning tools and algorithms for recognition and digitalisation of sales receipts. *Municipal Economy of Cities*. https://doi.org/10.33042/2522-1809-2023-6-180-7-11

[9] Nguyen, T., Jatowt, A., Coustaty, M., & Doucet, A. (2021). Survey of PostOCR Processing Approaches. *ACM Computing Surveys (CSUR)*, 54, 1–37. https://doi.org/10.1145/3453476

[10] Memon, J., Sami, M., & Khan, R. (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

[11] Yu, W., Lu, N., Qi, X., Gong, P., & Xiao, R. (2020). PICK: Processing Key Information Extraction from Documents using Improved Graph Learning Convolutional Networks. In *2020 25th International Conference on Pattern Recognition (ICPR)*, 4363–4370. https://doi.org/10.1109/ICPR48806.2021.9412927

[12] Le, A. D., Pham, D. V., & Nguyen, T. A. (2019). Deep learning approach for receipt recognition. In *Future Data and Security Engineering: 6th International Conference, FDSE 2019, Nha Trang City, Vietnam, November 27–29, 2019, Proceedings 6* (pp. 705–712). Springer International Publishing.

[13] Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021). Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions. *IEEE Access*, 9, 72894–72936. https://doi.org/10.1109/ACCESS.2021.3072900

[14] Kumar, V., Kaware, P., Singh, P., Sonkusare, R., & Kumar, S. (2020). Extraction of information from bill receipts using optical character recognition. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 72–77). IEEE. https://doi.org/10.1109/ICOSEC49089.2020.9215246

[15] W. Liu, X. Yuan, Y. Zhang, M. Liu, Z. Xiao, and J. Wu, "An End to End Method for Taxi Receipt Automatic Recognition Based on Neural Network," in *Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chongqing, China, 2020, pp. 314–318, doi: 10.1109/ITNEC48623.2020.9084712.

[16] X. Zhao, E. Niu, Z. Wu, and X. Wang, "CUTIE: Learning to Understand Documents with Convolutional Universal Text Information Extractor," arXiv preprint arXiv:1903.12363, 2019.

[17] Ha, H. T., & Horák, A. (2022). Information extraction from scanned invoice images using text analysis and layout features. *Signal Processing: Image Communication*, 102, 116601. https://doi.org/10.1016/j.image.2021.116601

[18] W. Liu, X. Yuan, Y. Zhang, M. Wu, H. Du, and Y. Cui, "An Automatic Taxi Receipt Text Recognition Application System," in *Proceedings of the 2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Paris, France, 2020, pp. 1–6, doi: 10.1109/BMSB49480.2020.9379905.

[19] J. Odd and E. Theologou, "Utilize OCR Text to Extract Receipt Data and Classify Receipts with Common Machine Learning Algorithms," Bachelor's thesis, Department of Computer and Information Science, Linköping University, Linköping, Sweden, 2018, ISRN LIU-IDA/LITH-EX-G–18/043–SE.

[20] P. Zhang, Y. Xu, Z. Cheng, S. Pu, J. Lu, L. Qiao, Y. Niu, and F. Wu, "TRIE: End-to-End Text Reading and Information Extraction for Document Understanding," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1413–1422, https://doi.org/10.1145/3394171.3413900

[21] Yang, Y., Wu, Z., Yang, Y., Lian, S., Guo, F., & Wang, Z. (2022). A survey of information extraction based on deep learning. *Applied Sciences*, 12(19), 9691.

[22] X. Liu, F. Gao, Q. Zhang, and H. Zhao, "Graph Convolution for Multimodal Information Extraction from Visually Rich Documents," in *Proceedings of NAACL-HLT*, 2019, pp. 32–39, https://doi.org/10.18653/v1/N19-2005

[23] B. P. Majumder, N. Potti, S. Tata, J. B. Wendt, Q. Zhao, and M. Najork, "Representation Learning for Information Extraction from Form-like Documents," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 6495–6504, https://doi.org/10.18653/v1/2020.acl-main.580

[24] Lin, C. J., Liu, Y. C., & Lee, C. L. (2022). Automatic receipt recognition system based on artificial intelligence technology. *Applied Sciences*, 12(2), 853. https://doi.org/10.3390/app12020853

[25] Chang, S. L., Chen, L. S., Chung, Y. C., & Chen, S. W. (2004). Automatic license plate recognition. *IEEE Transactions on Intelligent Transportation Systems*, 5(1), 42–53. https://doi.org/10.1109/TITS.2004.825086