

Extraction of Transaction Information on Handwritten Notes Using Optical Character Recognition (OCR) and Named Entity Recognition (NER)

Meyry Ardianti¹, Puguh Hiskiawan^{2*}

Data Science, Faculty of Technology and Design, Bunda Mulia University
s36230015@student.ubm.ac.id¹, phiskiawan@bundamulia.ac.id²

Article Info

Article history:

Received 2026-08-20

Revised 2026-09-25

Accepted 2026-10-06

Keyword:

Handwritten Notes,
IndoBERT,
Information Extraction,
NER,
OCR.

ABSTRACT

Handwritten notes remain widely used as transaction proofs, yet extracting information from them automatically is challenging due to variations in handwriting styles, layouts, and image qualities. This research aims to develop a transaction information extraction method for handwritten notes by integrating Optical Character Recognition (OCR) and Named Entity Recognition (NER). The research stages include image preprocessing using grayscale conversion, denoising, and Contrast Limited Adaptive Histogram Equalization (CLAHE), text recognition using EasyOCR across multiple image orientations, and transaction entity identification using IndoBERT with the BIO scheme. OCR performance is evaluated using Character Error Rate (CER) and Word Error Rate (WER), while NER is evaluated using Accuracy, Precision, Recall, and F1-score. Test results show that the NER model achieved an Accuracy of 0.8889, Precision of 0.7857, Recall of 0.8462, and F1-score of 0.8148 on the test data. However, the OCR-NER integration still exhibits character recognition errors and fails to completely extract all transaction attributes, particularly PRICE and AMOUNT. These findings indicate that the OCR-NER integration holds potential for supporting transaction data digitization in handwritten notes, though improving OCR quality and using a larger dataset are necessary to enhance system accuracy and generalizability.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Sales transactions are an important activity in various business sectors, especially in micro, small, and medium enterprises (MSMEs), which still heavily rely on the use of physical notes and handwritten recapitulation as transaction proof [1]. However, manual recording and management of transaction data often lead to various problems, such as input errors, data loss, slow recapitulation processes, and low efficiency in information processing, which can ultimately burden administrative management and cause losses [2], [3]. This condition indicates the need for an automated system capable of converting transaction data from handwritten notes into structured digital information.

Advances in deep learning technology, particularly in the field of computer vision, have opened up opportunities for the automated processing and interpretation of image-based

information [4]. The application of deep learning approaches to image analysis also demonstrates potential for developing automated systems capable of processing and analyzing visual information in a more structured manner [5]. One of the technologies widely used to convert textual information within images into digital data is Optical Character Recognition (OCR), which can be applied to both printed and handwritten documents [6]. In practice, OCR can be used to recognize text in handwritten notes; however, variations in handwriting styles, layout and tilt angles, image quality, and lighting conditions can affect recognition results [7]. Therefore, advanced stages are needed so that the text resulting from OCR can be understood and classified according to its informational meaning.

One of the widely used approaches to understanding text information is Named Entity Recognition (NER). NER is a Natural Language Processing (NLP) technique that aims to

identify entity names in text and classify them into predefined categories, such as names, locations, organizations, or other categories such as monetary amounts on receipts [8]. In the context of transaction receipts, NER can be utilized to automatically extract important information such as product names, purchase quantities, unit prices, total transactions, and purchase dates. Through a comprehensive two-stage workflow, the integration between OCR and NER enables end-to-end automated information extraction; starting from text segmentation on unstructured document images to the identification of key entities, resulting in structured financial data ready to be used for analysis or integration into information systems [9]. Although printed materials have been successfully transformed into digital formats, most research related to document information extraction still focuses on printed documents or requires complex visual and computational data [6]. Recent studies have explored handwritten document understanding through the combination of handwriting recognition and semantic information extraction [10]. However, research on transaction information extraction from handwritten notes using OCR and NER remains relatively limited, particularly for Indonesian transaction notes with diverse writing characteristics. Recent studies have explored information extraction from historical handwritten records [11] and OCR with post-processing techniques for receipt documents [12]. Furthermore, other challenges lie in the scarcity of properly labeled data, difficulties in entity label adjustment, and the need to improve model accuracy in handling imperfect or noisy OCR-extracted text, particularly because larger and high-quality datasets are important for developing reliable and generalizable deep learning models [6], [10], [13].

Based on these problems, this study proposes a transaction information extraction method on handwritten notes using the integration of OCR and NER. The research stages include note image acquisition and preprocessing, text extraction using OCR, transaction entity annotation, training of a transformer-based NER model, and model performance evaluation using accuracy, precision, recall, and F1-score metrics [10], [14]. This approach is expected to produce a system capable of converting handwritten transaction documents into structured data automatically and efficiently.

The scientific contributions of this research are centered on the application and integration of existing techniques for handwritten Indonesian transaction notes rather than on introducing a new OCR or NER architecture. First, the study

constructs a transaction-specific handwritten note dataset with BIO annotations covering eight transaction entity categories. Second, the proposed pipeline combines image preprocessing with multi-orientation EasyOCR to handle handwritten notes captured in different orientations. Third, an IndoBERT-based NER model is fine-tuned using transaction-specific BIO annotations to identify relevant entities from the transcribed text. Finally, these components are integrated into a reproducible OCR-NER workflow that transforms handwritten transaction notes into structured transaction information.

The novelty of this study lies in the transaction-specific adaptation and integration of these existing components for Indonesian handwritten transaction notes. In particular, the proposed workflow combines multi-orientation OCR, transaction-specific BIO annotation, and IndoBERT-based entity recognition to transform handwritten transaction content into structured transaction information.

Overall, this research investigates the feasibility of integrating OCR and NER to support transaction data management based on handwritten notes. The proposed approach aims to transform handwritten notes into structured transaction information while identifying the limitations of OCR and NER in handling handwritten transaction data [17].

II. METHOD

The research methodology outlines the stages involved in addressing the research problem. These stages encompass the datasets, tools, materials, and methods employed throughout the research process. Each stage is carried out systematically to achieve results that align with the research objectives.

A. Overview of the Proposed Framework

This study develops a transaction information extraction framework specifically designed for Indonesian handwritten transaction notes by combining image preprocessing, multi-orientation Optical Character Recognition (OCR), and transaction-specific Named Entity Recognition (NER) [18]. This framework converts handwritten transaction note images into structured transaction information through a sequential pipeline consisting of image preprocessing, OCR, NER, transaction information extraction, and evaluation [19]. It can be seen in Figure 1 how these stages form a processing pipeline starting from the handwritten note image to producing entity predictions and structured transaction data.

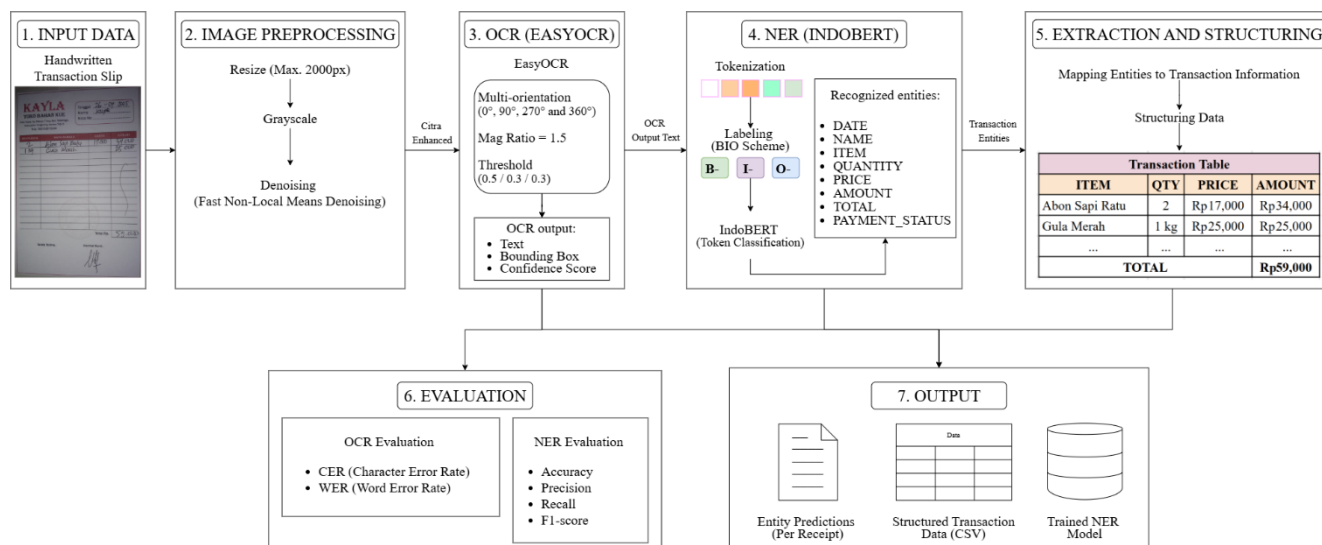


Figure 1. Framework for the transaction information extraction method from handwritten receipts based on OCR-NER.

During the OCR stage, the note image first goes through an image preprocessing process as shown in the cleaning and image quality enhancement process before being fed into the OCR model which includes resizing, grayscale conversion, denoising, and contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE) [20]. The preprocessed image serves as EasyOCR input to generate recognized text, bounding boxes, and confidence scores across multiple orientations. The quality of the text recognition results is evaluated using Character Error Rate (CER) and Word Error Rate (WER), with lower values indicating better OCR performance. The transaction text is subsequently processed using an IndoBERT-based NER approach, which utilizes an NER pipeline to identify and classify entities through fine-tuning and BIO format labeling stages, to identify and classify entities related to transaction information [21]. The text is tokenized and labeled using the BIO scheme, after which the token classification model is trained to recognize DATE, NAME, ITEM, QUANTITY, PRICE, AMOUNT, TOTAL, and PAYMENT_STATUS entities and evaluated using Accuracy, Precision, Recall, and F1-score. The OCR and NER results are then integrated to generate entity predictions per note and compile the transaction information into structured data.

B. Dataset and Data Acquisition

The research dataset consists of 10 handwritten transaction note images collected from transaction documents in Indonesian. The dataset includes 8 portrait-oriented images and 2 landscape-oriented images, with image widths ranging from 720 to 2160 pixels and heights ranging from 780 to 2874 pixels. The notes exhibit variations in handwriting characteristics, including differences in letter size, stroke thickness, writing inclination, character spacing, and writing consistency. The documents also vary in printed layouts, table structures, and paper conditions, with some images containing folds, shadows, stamps, or other background artifacts.

The image conditions also show variations in illumination and visual quality. The mean brightness values range from 98.99 to 161.06, with an overall mean of 131.82, while the contrast values range from 28.06 to 59.02, with an average of 42.68. The sharpness values, measured using the variance of the Laplacian, range from 19.47 to 1487.92, with an overall mean of 689.86, indicating differences in image clarity across the dataset.

The number of individual writers was not explicitly recorded during data collection. Therefore, writer-independent evaluation cannot be established from the current dataset. Nevertheless, the handwritten entries exhibit observable variations in writing characteristics, including letter formation, stroke patterns, writing size, and writing consistency.

The dataset is grouped into images, annotation, and output directories, where the images directory is used to store note images, annotation for storing reference text and entity annotations, and output for storing OCR results, NER predictions, trained models, and structured transaction data. Before processing, each handwritten record is paired with a corresponding ground-truth text to ensure that every document has consistent reference data, allowing the transcribed text (hypothesis) to be accurately compared against the actual text [23]. The ground-truth text was manually transcribed from each handwritten note by referring directly to the original document image. The transcription preserves the textual content of the note and serves as the reference text for OCR evaluation using CER and WER. The same verified transcription is subsequently used as the basis for tokenization and entity annotation in the NER dataset.

TABLE 1
CHARACTERISTICS OF THE HANDWRITTEN TRANSACTION NOTE DATASET.

Characteristic	Description
Number of images	10
Image orientation	8 portrait, 2 landscape
Image width	720–2160 pixels
Image height	780–2874 pixels
Mean brightness	131.82
Mean contrast	42.68
Mean sharpness	689.86
Handwriting variation	Letter size, stroke thickness, writing inclination, spacing, and consistency
Document format	Printed transaction forms with different layouts and table structures
Paper condition	White/light-colored transaction forms; some contain folds, shadows, stamps, or background artifacts
Paper type	Not systematically recorded during data collection
Number of writers	Not explicitly recorded

C. Image Preprocessing and Optical Character Recognition (OCR)

The image preprocessing stage is performed by resizing the image if its largest dimension exceeds 2000 pixels, then converting it to grayscale to simplify the pixel intensity representation before using it as OCR input. Noise reduction and image contrast enhancement are carried out using Fast Non-Local Means Denoising and Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve image quality prior to the text recognition process [24]. In the denoising process, a parameter of $h=10$, a template window of 7, and a search window of 21 are used, while CLAHE uses a clip limit of 2.0 and a tile grid size of 8×8 . The image resulting from all these processes is used as the enhanced image which becomes the input at the OCR stage.

OCR is performed using EasyOCR with Indonesian and English language configurations. In the OCR-NER integration stage, the enhanced image is processed at four orientations: 0° , 90° , 180° , and 270° . For each orientation, OCR generates recognized text and confidence scores, and the result with the highest average confidence score is selected as the final OCR output [25]. A minimum confidence threshold of 0.20 is applied to filter low-confidence detections. The recognized text from detections that meet the confidence threshold is combined into a single text sequence for subsequent processing.

To handle OCR inconsistencies, detected text segments are retained together with their bounding boxes and confidence scores, while detections below the minimum confidence threshold are excluded from the resulting transaction text. Text segments from the same document are subsequently combined and normalized to reduce inconsistencies caused by separated or merged tokens. However, no rule-based correction is applied to incorrectly recognized characters or numerical values; therefore, errors in handwritten numbers

and symbols may remain in the OCR output and propagate to the NER stage, particularly affecting PRICE and AMOUNT entities.

EasyOCR was selected as the primary OCR engine because it provides an end-to-end text detection and recognition pipeline, supports Indonesian and English language configurations, and produces bounding boxes and confidence scores that can be used to select the most reliable recognition result. These characteristics are relevant to the handwritten transaction notes in this study, where text orientation and recognition confidence vary across documents. To assess the suitability of EasyOCR, its performance was compared with Tesseract OCR using the same preprocessed handwritten note images and the same ground-truth transcriptions.

D. OCR Evaluation

OCR evaluation is conducted by comparing the text resulting from OCR recognition with the ground truth. Character Error Rate (CER) and Word Error Rate (WER) are used to measure the error rates of the text recognition results at the character and word levels [26]. Both metrics are employed in the evaluation of handwritten text recognition by comparing recognition results against reference text (ground truth) to measure the recognition error rate.

1) Character Error Rate

CER (Character Error Rate) measures the OCR error rate at the character level based on the number of substitution, deletion, and insertion operations. CER is formulated as:

$$ER = \frac{(S_c + D_c + I_c)}{N_c} \quad (1)$$

where S_c , D_c , and I_c represent the number of characters that undergo substitution, deletion, and insertion, respectively, while N_c is the total number of characters in the ground truth. A smaller CER value indicates a lower character error rate in the OCR results.

2) Word Error Rate

Makalah WER (Word Error Rate) measures the OCR error rate at the word level based on the number of substitution, deletion, and insertion operations. WER is formulated as:

$$WER = \frac{(S_w + D_w + I_w)}{N_w} \quad (2)$$

where S_w , D_w , and I_w represent the number of words that undergo substitution, deletion, and insertion, respectively, while N_w is the total number of words in the ground truth. A smaller WER value indicates better word recognition results.

The OCR evaluation is conducted independently from the NER evaluation. The OCR output is compared directly with the corresponding ground-truth transcription using CER and WER. Therefore, the ground-truth transcription serves as the reference text for measuring the recognition performance of the OCR component.

E. NER Dataset Construction and BIO Annotation

The verified ground-truth text is used to construct the NER dataset by first tokenizing each transaction text into individual tokens. Each token is then manually assigned to one of the predefined transaction entity categories based on its semantic role in the note. The annotations follow the Beginning–Inside–Outside (BIO) scheme, where B- denotes the first token of an entity, I- denotes a subsequent token belonging to the same entity, and O denotes a token that does not belong to any entity [27]. For example, in the transaction text “26-04-2025 KAYLA 2 Abon Sapi Ratu 17.000 34.000”, “26-04-2025” is labeled B-DATE, “KAYLA” is labeled B-NAME, “Abon” is labeled B-ITEM, “Sapi” and “Ratu” are labeled I-ITEM, “17.000” is labeled B-PRICE, and “34.000” is labeled B-AMOUNT. This annotation process is applied consistently to the ground-truth text of all transaction notes to form the labeled NER dataset. The transaction entities used consist of DATE, NAME, ITEM, QUANTITY, PRICE, AMOUNT, TOTAL, and PAYMENT_STATUS, enabling the NER process to not only recognize text on the notes but also identify the informational function of each token.

TABLE 2
EXAMPLE OF BIO ANNOTATION.

Token	Label	Entity
26-04-2025	B-DATE	DATE
KAYLA	B-NAME	NAME
2	B-QUANTITY	QUANTITY
Abon	B-ITEM	ITEM
Sapi	I-ITEM	ITEM
Ratu	I-ITEM	ITEM
Rp	O	None
17.000	B-PRICE	PRICE
34.000	B-AMOUNT	AMOUNT
1	B-QUANTITY	QUANTITY
kg	I-QUANTITY	QUANTITY
Gula	B-ITEM	ITEM
Merah	I-ITEM	ITEM
25.000	B-PRICE	PRICE
59.000	B-TOTAL	TOTAL

F. IndoBERT-Based NER Model and Training

The NER model uses IndoBERT base-p1 as the base transformer model and is fine-tuned for token classification on the transaction-specific BIO-annotated handwritten note dataset. The model performs token classification after aligning the BIO labels with the subword tokens generated by the IndoBERT tokenizer [28]. The dataset is split at the note level into 80% training, 10% validation, and 10% testing data using a random state of 42. Tokenization is performed with a maximum sequence length of 256 tokens. Fine-tuning is conducted using the Hugging Face Trainer and DataCollatorForTokenClassification for 10 epochs, with a learning rate of 2×10^{-5} , a batch size of 2, and a weight decay of 0.01. The best-performing model is selected based on the

F1-score obtained on the validation set. For evaluation, two evaluation settings are distinguished. First, the standalone NER evaluation uses the verified ground-truth transcription as input and the corresponding BIO entity annotations as reference labels. This setting measures the ability of the NER model to classify transaction entities without the influence of OCR recognition errors. Second, the integrated OCR-NER evaluation uses OCR-generated text as the input to the same NER model. The results from these two settings are analyzed separately to distinguish errors originating from OCR recognition from errors produced by the NER model. The token classification training process minimizes the cross-entropy loss function as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \log P(y_i | x_i) \quad (3)$$

where N is the number of tokens used in the loss calculation, y_i is the actual label of the i -th token, and $P(y_i | x_i)$ is the model's probability of predicting the actual label based on the token representation x_i .

G. NER Evaluation and Transaction Information Structuring

Accuracy, Precision, Recall, and F1-score are used to measure the performance of the NER model in classifying transaction entities. The standalone NER evaluation is performed using the verified ground-truth transcription and its corresponding BIO annotations. Therefore, the input text and reference labels in this evaluation are independent of OCR recognition errors. The confusion matrix compares the NER predictions with the corresponding ground-truth entity labels based on True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [29].

Accuracy represents the proportion of tokens that are correctly classified by the model from all evaluated tokens. A value closer to 1 indicates that the model provides more accurate predictions across the evaluated data.

Precision measures the proportion of correctly predicted entities among all entities predicted by the model. A high precision value indicates that most entities identified by the model correspond to the correct entity labels.

Recall measures the model's ability to identify entities that are actually present in the ground-truth data. A high recall value indicates that the model is able to recognize a large proportion of the existing entities in the transaction receipts.

F1-score represents the balance between Precision and Recall. This metric is useful for evaluating the overall performance of the model when both the correctness of predictions and the ability to identify entities need to be considered. A higher F1-score indicates better overall classification performance.

In contrast, the integrated OCR-NER evaluation uses OCR-generated text as the input to the trained NER model.

This evaluation is performed separately from the standalone NER evaluation to observe the effect of OCR recognition errors on subsequent entity extraction. Consequently, the standalone NER metrics represent the classification capability of the NER model under verified text input, whereas the integrated OCR-NER results represent the performance of the complete processing pipeline under OCR-generated input.

In addition to quantitative metrics, a classification report is used to analyze the performance of each entity label, while a confusion matrix is employed to identify misclassification patterns between labels.

Transaction Information Structuring integrates OCR and NER to extract information from new receipts, using the OCR result with the highest confidence score as input for the NER process. Entity predictions are then mapped to the categories DATE, NAME, ITEM, QUANTITY, PRICE, AMOUNT, TOTAL, and PAYMENT_STATUS, with ITEM, QUANTITY, PRICE, and AMOUNT organized into a transaction table. The extraction results are saved in CSV format, alongside the entity predictions and the trained NER model.

III. RESULTS AND DISCUSSION

A. Descriptive Analysis of the Transaction Note Dataset

The research dataset consists of 10 handwritten transaction note images that exhibit variations in document structure, handwriting styles, information layout, and image quality. Each note is used as input for the preprocessing, OCR, and information extraction stages using NER. In addition to the images, the research also provides ground truth in the form of text transcriptions used as a reference to evaluate the OCR results.

The entire dataset comprises 189 tokens distributed across various handwritten transaction notes. The variation in the number of tokens reflects differences in the amount and complexity of information contained in each document, providing varied input for the OCR and NER processes. The annotated tokens cover eight transaction entity categories: DATE, NAME, ITEM, QUANTITY, PRICE, AMOUNT, TOTAL, and PAYMENT_STATUS. The distribution of these entity tokens is presented in table 3.

TABLE 3
DISTRIBUTION OF TRANSACTION ENTITY TOKENS.

Entity	Token Count	Percentage (%)
DATE	13	6.91
NAME	16	8.51
ITEM	71	37.77
QUANTITY	37	19.68
PRICE	28	14.89
AMOUNT	14	7.45
TOTAL	8	4.26
PAYMENT_STATUS	1	0.53
Total entity tokens	188	100.00

Note: The complete annotation contains 189 tokens, including one token labeled O that does not belong to a transaction entity category.

The distribution shows that ITEM is the most frequent entity, accounting for 71 tokens (37.77%), followed by QUANTITY with 37 tokens (19.68%) and PRICE with 28 tokens (14.89%). NAME, AMOUNT, DATE, and TOTAL account for 16 (8.51%), 14 (7.45%), 13 (6.91%), and 8 (4.26%) tokens, respectively, while PAYMENT_STATUS is the least represented entity with one token (0.53%). The uneven distribution reflects differences in the transaction information contained in the handwritten notes and should be considered when interpreting the NER results.

The dataset is annotated using the BIO (Beginning-Inside-Outside) scheme with eight transaction entity categories, namely DATE, NAME, ITEM, QUANTITY, PRICE, AMOUNT, TOTAL, and PAYMENT_STATUS, which are mapped into 13 class labels for the token classification process. Each token is assigned a BIO label according to its corresponding transaction entity and used as ground truth for NER model training and evaluation.

B. Image Preprocessing and OCR Results

The preprocessing stage improves image quality before text recognition by resizing images exceeding 2000 pixels, converting them to grayscale, applying Fast Non-Local Means Denoising, and enhancing contrast using CLAHE (clip limit 2.0, tile grid 8×8). These processes are applied to reduce background noise and improve the visibility of handwritten characters. The resulting enhanced image provides clearer text regions for the subsequent OCR process.



Figure 2. Comparison of original, grayscale, and enhanced representations of a handwritten transaction note.

Preprocessing transforms the original image into grayscale and enhanced versions through denoising and CLAHE contrast enhancement, making characters and note structures clearer against varying backgrounds. As shown in Figure 2, the enhanced representation provides better contrast between the handwritten text and the document background while retaining the original layout of the note.

In the OCR stage, EasyOCR is configured to use both Indonesian and English. Recognition is performed on four image orientations: 0°, 90°, 180°, and 270°. For each orientation, EasyOCR generates a set of detected text segments, bounding boxes, and confidence scores. The confidence score of each detected segment is used to calculate the average confidence score for the corresponding orientation. The orientation with the highest average confidence score is selected as the final OCR result for each note. Detections with confidence scores below 0.20 are excluded before the recognized text is combined into a single text sequence. Thus, the final OCR output is not obtained by concatenating the results from all orientations, but by selecting the complete OCR result from the orientation with the highest average confidence score.

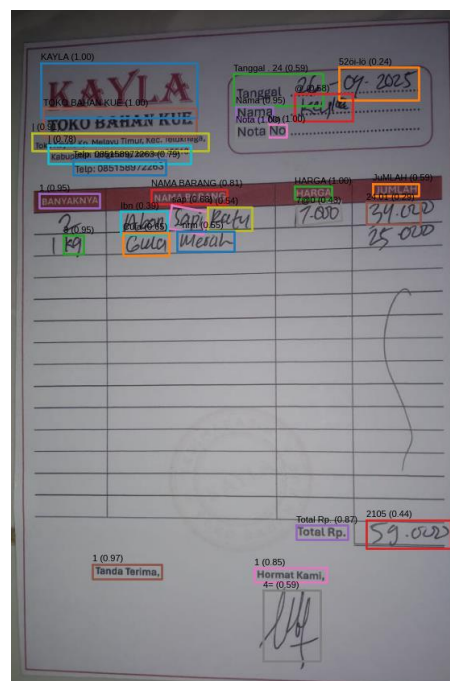


Figure 3. Visualization of OCR results with detected text regions and confidence scores.

Figure 3 shows the OCR detection results, including the recognized text regions and their confidence scores. Although OCR successfully detects most information on the notes, recognition errors still occur especially in handwriting, numbers, symbols, and inconsistent writing styles making image quality and handwriting characteristics crucial factors affecting OCR accuracy.

C. OCR Performance Evaluation

OCR evaluation is conducted using Character Error Rate (CER) and Word Error Rate (WER) by comparing the OCR output text against the ground truth. Lower values indicate better OCR performance. These metrics are used to quantify the accuracy of character and word-level recognition in the handwritten transaction notes.

TABLE 4
OCR PERFORMANCE ON HANDWRITTEN TRANSACTION NOTES.

Note	CER	WER
Nota 1	1.1039	2.0000
Nota 2	2.4322	3.5556
Nota 3	5.6563	7.8000
Nota 4	0.8190	1.2632
Nota 5	1.3280	2.5000
Nota 6	2.9294	4.4000
Nota 7	1.3482	1.9048
Nota 8	0.7647	1.2766
Nota 9	0.9855	1.5833
Nota 10	3.4070	4.0625
MEAN	2.0774	3.0346

Table 4 shows that the average CER and WER are 2.0774 and 3.0346, respectively, indicating that OCR still encounters errors in recognizing handwritten text. The best performance is achieved on note 8 with a CER of 0.7647 and a WER of 1.2766, while the lowest performance is found on note 3 with a CER of 5.6563 and a WER of 7.8000. These differences indicate that OCR performance varies depending on the writing characteristics and structure of each note.

TABLE 5
COMPARISON OF OCR METHODS ON HANDWRITTEN TRANSACTION NOTES.

OCR Method	MEAN CER	MEAN WER
EasyOCR	2.0774	3.0346
Tesseract OCR	2.5219	2.5080

Table 5 compares the performance of EasyOCR and Tesseract OCR on the same handwritten transaction note dataset. EasyOCR achieved a lower mean CER of 2.0774, while Tesseract OCR achieved a lower mean WER of 2.5080. These results indicate that EasyOCR provides better character-level recognition, whereas Tesseract OCR provides better word-level recognition. EasyOCR was retained as the primary OCR engine because it provides bounding boxes and confidence scores required for the subsequent OCR-NER integration.

CER and WER values greater than 1 are still possible because these metrics calculate substitution, deletion, and insertion operations, meaning the number of errors can exceed the total number of characters or words in the ground truth. Therefore, values exceeding unity indicate a high number of recognition errors and reflect the challenges associated with handwritten transaction data. These results show that OCR remains one of the sources of error in the information

extraction pipeline, as recognition errors can be propagated to the NER stage and affect transaction entity identification.

To distinguish the source of errors in the extraction pipeline, the OCR evaluation results are interpreted independently from the standalone NER evaluation. CER and WER quantify recognition errors by comparing OCR output with the ground-truth transcription, while the standalone NER metrics measure entity classification using verified text. The integrated OCR-NER results are subsequently used to observe how OCR recognition errors affect the final entity extraction.

TABLE 6
COMPARISON BEFORE AND AFTER PREPROCESSING.

Condition	CER	WER
Before Preprocessing	3.0489	3.4380
After Preprocessing	2.9802	3.3106

Table 6 compares OCR performance before and after image preprocessing. The results show that preprocessing reduced the CER from 3.0489 to 2.9802 and the WER from 3.4380 to 3.3106. These results indicate that the applied preprocessing techniques provided a measurable improvement in OCR recognition performance on the handwritten transaction notes.

TABLE 7
OCR PERFORMANCE BY IMAGE ORIENTATION.

Orientation	CER	WER
0°	2.9802	3.3106
90°	0.8497	1.5458
180°	1.1108	1.4459
270°	0.8812	1.7447

For each note, the four orientation-specific OCR results are evaluated independently before the final selection. As shown in Table 7, OCR performance varies across orientations, with 90° achieving the lowest CER and 180° achieving the lowest WER. This variation indicates that image orientation can substantially affect handwritten text recognition performance and supports the use of multi-orientation OCR to identify the most reliable recognition result.

TABLE 8
ABLATION STUDY OF IMAGE PREPROCESSING FOR OCR.

Preprocessing Configuration	CER	WER
Without Preprocessing	2.2287	3.2710
Grayscale	2.2632	3.2738
Grayscale + Denoising	2.1811	3.1879
Grayscale + Denoising + CLAHE	2.1054	3.1266

The ablation study shows that grayscale conversion alone does not improve OCR performance, with CER and WER slightly increasing compared with the configuration without preprocessing. The addition of denoising reduces the CER

from 2.2632 to 2.1811 and the WER from 3.2738 to 3.1879. The complete preprocessing configuration, consisting of grayscale conversion, denoising, and CLAHE, achieves the lowest CER of 2.1054 and WER of 3.1266, indicating that the combined preprocessing steps provide the most effective configuration for OCR on the handwritten transaction notes.

D. IndoBERT-Based NER Training and Performance

The NER model was built using a dataset containing 189 tokens from 10 notes. The dataset was split at the note level into 80% training, 10% validation, and 10% testing data using a random state of 42. This note-level split prevents tokens from the same note from being distributed across different

subsets. However, writer-level information was not available for the dataset; therefore, it could not be verified whether notes written by the same person were present in different subsets. Consequently, a strictly writer-independent evaluation could not be performed, and the current results should not be interpreted as a direct measure of generalization to unseen handwriting styles. A larger dataset containing multiple identified writers is required for writer-independent evaluation. IndoBERT base-p1 was fine-tuned for token classification on the handwritten transaction note dataset. The model was trained for 10 epochs with a learning rate of 2×10^{-5} , a batch size of 2, a maximum sequence length of 256 tokens, and a weight decay of 0.01.

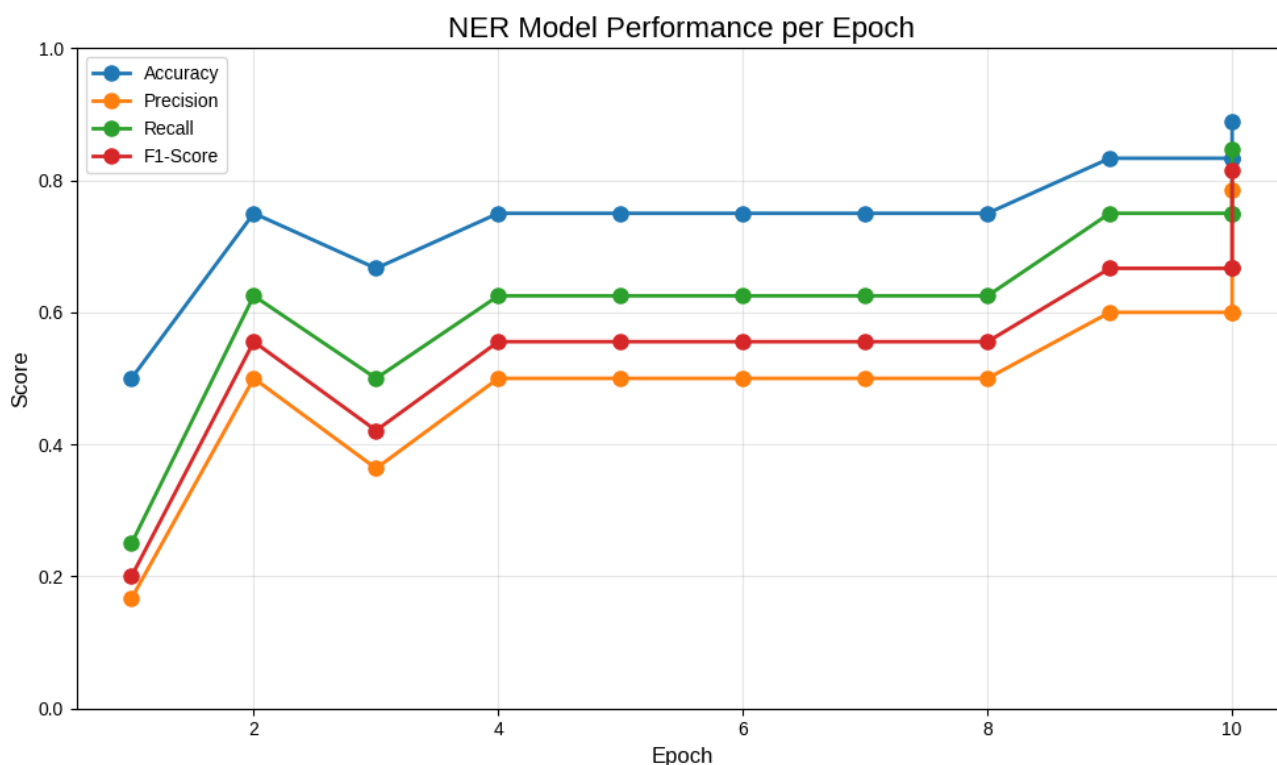


Figure 4. NER model performance across training epochs.

Model performance should be evaluated quantitatively to assess the reliability and effectiveness of the developed deep learning model [5]. The NER training results show that the training loss decreased from 2.4103 in epoch 1 to 0.0995 in epoch 10, indicating that the model progressively learned the training data. On the validation set, the highest accuracy of 0.8333, precision of 0.6000, recall of 0.7500, and F1-score of 0.6667 were achieved at epochs 9 and 10. The validation loss also decreased from 1.8695 in epoch 1 to 0.7292 at epoch 9 before slightly increasing to 0.7299 in epoch 10. These results indicate that the model's performance improved during training, although the validation metrics remained limited by the small dataset.

TABLE 9
NER MODEL PERFORMANCE.

Evaluation Set	Validation	Test
Loss	0.7292	0.3671
Accuracy	0.8333	0.8889
Precision	0.6000	0.7857
Recall	0.7500	0.8462
F1-score	0.6667	0.8148

The standalone NER evaluation achieved an Accuracy of 0.8889, Precision of 0.7857, Recall of 0.8462, and an F1-score of 0.8148 on the test set. To provide a more detailed evaluation, the performance of each entity class is presented in Table 10. The reported NER metrics were calculated using the annotated ground-truth transcription and corresponding

entity labels rather than OCR-generated text. Therefore, these metrics represent the standalone NER classification performance under verified text input and do not measure the complete OCR-NER pipeline. To assess the effect of OCR errors on entity extraction, the OCR-generated text is subsequently passed to the same trained NER model in the integrated OCR-NER evaluation. However, the integrated results should be interpreted cautiously because the test set consists of only one handwritten note containing 18 tokens, resulting in limited support for several entity classes. A larger and more diverse dataset is required to obtain more reliable and generalizable NER performance [30].

TABLE 10
COMPARISON OF STANDALONE NER AND INTEGRATED OCR-NER PERFORMANCE.

Evaluation Setting	Input	Precision	Recall	F1-score
Standalone NER	Ground-truth text	0.7857	0.8462	0.8148
Integrated OCR-NER	OCR-generated text	0.0190	0.0593	0.0287

Table 10 compares the performance of the standalone NER model and the integrated OCR-NER pipeline. The standalone NER evaluation uses verified ground-truth transcription as input, thereby measuring the entity classification capability of the NER model independently of OCR recognition errors. In contrast, the integrated evaluation uses OCR-generated text as input to the same trained NER model. The substantial decrease in precision, recall, and F1-score in the integrated setting indicates that OCR errors substantially affect downstream entity extraction. However, the integrated results represent the combined performance of the OCR and NER components, and therefore the individual contribution of OCR errors and NER classification errors cannot be isolated from these results alone. This comparison demonstrates the difference between the NER model's performance under verified text input and its performance when operating on automatically recognized handwritten text.

TABLE 11
NER PERFORMANCE BY ENTITY CLASS ON THE TEST SET.

Entity	Precision	Recall	F1-score	Support
DATE	1.00	1.00	1.00	1
NAME	0.00	0.00	0.00	1
ITEM	0.80	1.00	0.89	4
QUANTITY	1.00	1.00	1.00	4
PRICE	1.00	1.00	1.00	2
AMOUNT	0.00	0.00	0.00	1
TOTAL	0.00	0.00	0.00	0
PAYMENT_STATUS	0.00	0.00	0.00	0

The class-level evaluation presented in Table 11 demonstrates variations in NER performance across entity classes. DATE, QUANTITY, and PRICE achieved an F1-score of 1.00, while ITEM obtained an F1-score of 0.89. In contrast, NAME and AMOUNT recorded an F1-score of 0.00,

indicating that these entities were not correctly identified in the test data. These findings demonstrate that aggregate evaluation metrics may not fully represent the performance of individual entity classes, particularly for entities with limited representation in the test set.

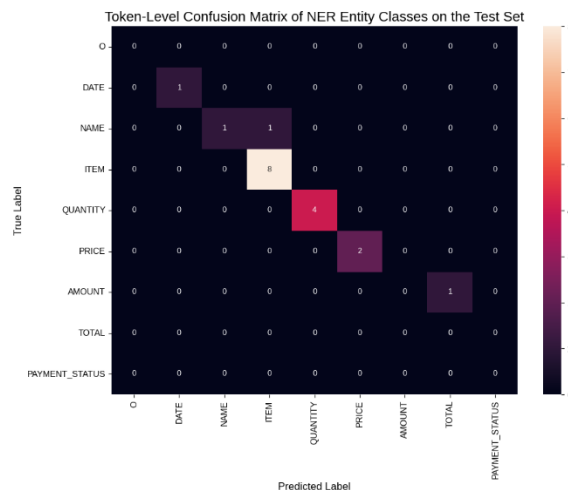


Figure 5. Token-level confusion matrix of NER entity classes on the test set.

The confusion matrix in Figure 5 provides further insight into the misclassification patterns among the NER entity classes. Both evaluated PRICE entities were correctly classified, resulting in no misclassification for this entity in the test set. In contrast, the single AMOUNT instance was misclassified as TOTAL, resulting in an F1-score of 0.00 for AMOUNT. This error indicates that the model was unable to distinguish the AMOUNT entity from the TOTAL entity in the evaluated test sample. Nevertheless, these error patterns should be interpreted cautiously because the test set contains only one handwritten note and provides only a limited number of instances for each entity class. Therefore, the observed misclassification of AMOUNT as TOTAL cannot be generalized to the overall performance of the NER model.

To further evaluate the recognition of numerical transaction attributes, an exact-match evaluation was performed for PRICE and AMOUNT after numerical normalization. The normalization removes formatting differences such as thousand separators before comparison with the ground truth. The results show an exact-match accuracy of 20.00% for PRICE and 50.00% for AMOUNT. These results indicate that numerical entities remain challenging for the NER model, particularly PRICE, where differences in the recognized numerical values prevented exact matching.

TABLE 12
EXACT-MATCH ACCURACY OF NUMERICAL ENTITIES.

Entity	Exact-Match Accuracy
PRICE	20.00%
AMOUNT	50.00%

E. Integrated OCR-NER Information Extraction Results

The OCR-NER integration stage produces final outputs in the form of transaction entities obtained from the recognized text of the notes. In testing on one of the notes, OCR produced an average confidence score of 0.6759. The OCR text was then processed by the NER model to identify information based on predetermined entity categories. The prediction results indicate that several entities, particularly ITEM and QUANTITY, were successfully recognized by the model.

However, the extraction results still have limitations in terms of information completeness and accuracy. Several ITEM entities contain characters that do not match the actual text, while the PRICE and AMOUNT information is not fully

retrieved. This indicates that errors introduced during the OCR stage can propagate to the NER stage and affect entity identification. Such errors are particularly critical for numerical transaction attributes because incorrectly recognized or fragmented digits may alter the actual monetary values. The system does not apply explicit numerical correction or arithmetic validation to PRICE and AMOUNT after OCR; therefore, these fields depend directly on the quality of the recognized text. Consequently, errors in recognizing handwritten words and numerical characters may provide incomplete or incorrect input to the NER model, making OCR quality an important factor in determining the accuracy of the final information extraction.

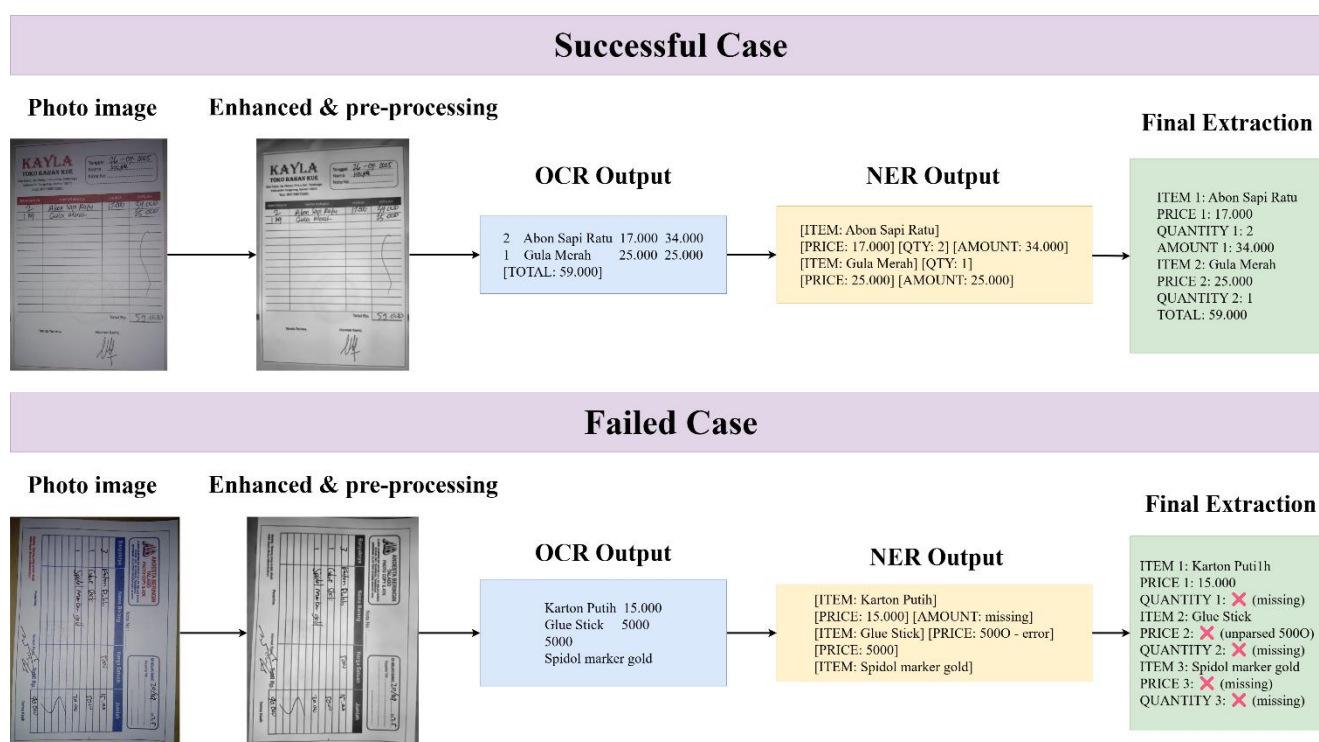


Figure 6. Relatively successful and failed cases of the OCR-NER pipeline.

Figure 6 illustrates a relatively successful case and a failed case of the complete OCR-NER pipeline. In the relatively successful case, the preprocessing stage produces a clearer representation of the handwritten note, allowing OCR to recognize the transaction text sufficiently for the NER model to identify relevant entities, including ITEM, QUANTITY, PRICE, and AMOUNT. In the failed case, OCR produces incorrect or incomplete characters, particularly in numerical information, which affects subsequent NER predictions and results in incomplete extraction of PRICE and AMOUNT. These cases demonstrate that errors introduced at the OCR stage can propagate to the NER stage and directly affect the final transaction information extraction.

Overall, the OCR-NER pipeline has successfully integrated text recognition and entity identification processes

to generate transaction information. However, the quality of the extraction results is still influenced by the OCR text quality, meaning that improving the OCR capability in handling variations of handwriting and numerical characters is necessary to enhance pipeline performance.

IV. CONCLUSION

This research successfully developed a transaction information extraction pipeline for handwritten notes by integrating Optical Character Recognition (OCR) and Named Entity Recognition (NER). The research stages included image preprocessing, text recognition using EasyOCR, and transaction entity identification using IndoBERT with the BIO scheme. Evaluation results showed that the NER model achieved an Accuracy of 0.8889, Precision of 0.7857, Recall

of 0.8462, and F1-score of 0.8148 on the test data. These results indicate that the model was able to classify the evaluated transaction entities, although the test set consisted of only one handwritten note containing 18 tokens, limiting the reliability of the evaluation results. Furthermore, the absence of a writer-independent evaluation limits the assessment of the model's generalization to unseen handwriting styles.

The results demonstrate the feasibility of integrating OCR and NER for handwritten transaction information extraction as a proof-of-concept. In the OCR-NER integration stage, the system was able to generate several transaction entities, particularly ITEM and QUANTITY, yet OCR character errors and incompletely extracted PRICE and AMOUNT attributes were still found. The OCR results also showed an average confidence score of 0.6759, indicating that text recognition quality remains a major factor affecting the extraction process. Thus, the main advantage of this approach lies in the integration of OCR and NER into a single processing workflow, while the primary limitations stem from handwriting recognition quality and a restricted dataset size. Future research should focus on improving handwriting recognition and numerical entity extraction, using larger and more diverse datasets, and conducting end-to-end evaluation to obtain a more comprehensive assessment of the complete OCR-NER pipeline.

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to Bunda Mulia University, particularly the Faculty of Technology and Design, for providing research funding, academic support, and facilities necessary for the completion of this research. The author also extends appreciation to the academic supervisor for the valuable guidance, feedback, and support throughout the research process and preparation of this article. Finally, the author gratefully acknowledges the author's family for their prayers, encouragement, and continuous support throughout the research and completion of this article.

REFERENCES

- [1] E. L. Afriana and D. F. Suyatno, "Development of a Point of Sale (POS) System for the Digitalization of Sales Transaction Recording Based on AI (Case Study: PT XYZ)," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 7, no. 3, pp. 351–363, 2026, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/JEISBI/>
- [2] E. Dethier, D. R. Kern, G. Stevens, and A. Boden, *Making Order in Household Accounting - Digital Invoices as Domestic Work Artifacts*, vol. 33, no. 4. Springer Netherlands, 2024. doi: 10.1007/s10606-024-09495-w.
- [3] J. Sutrisno, Y. Vient, S. D. Wijaya, N. Setiawan, D. Novytsari, and R. Nattassha, "Perancangan Aplikasi Inventaris Web dengan Rekomendasi Stok AI untuk Toko Kelontong Mandiri," *J. Digismantech (Jurnal Progr. Stud. Bisnis Digit.)*, vol. 5, no. 1, 2025, doi: 10.30813/digismantech.v5i1.8396.
- [4] P. Hiskiawan, C.-C. Chen, and Z.-K. Ye, "Processing of electrical resistivity tomography data using convolutional neural network in ERT-NET architectures," *Arab. J. Geosci.*, vol. 16, no. 10, pp. 1–14, 2023, doi: 10.1007/s12517-023-11690-w.
- [5] P. Hiskiawan, M. K. Sari, R. E. Siregar, N. A. Valencia, and T. P. Wijayanti, "A Deep Learning Data Fusion Approach for Prostate MRI Zonal Segmentation Using Dual-Channel U-Net with T2 and ADC Images," in *2026 International Conference on Current Research in Artificial Intelligence and Data Science (ICCRAIDS)*, Banyumas, Indonesia: IEEE, 2026, pp. 1–7. doi: 10.1109/ICCRAIDS67816.2026.11519671.
- [6] S. V. Mahadevkar, S. Patil, K. Kotecha, L. W. Soong, and T. Choudhury, "Exploring AI-driven approaches for unstructured document analysis and future horizons," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00948-z.
- [7] Q. Zhang, F. Liu, and W. Song, "IMTLM-Net: improved multi-task transformer based on localization mechanism network for handwritten English text recognition," *Complex Intell. Syst.*, vol. 11, no. 1, pp. 1–18, 2025, doi: 10.1007/s40747-024-01713-8.
- [8] Y. Zhang and G. Xiao, "Named Entity Recognition Datasets: A Classification Framework," *Int. J. Comput. Intell. Syst.*, vol. 17, no. 1, 2024, doi: 10.1007/s44196-024-00456-1.
- [9] L. Cooper and F. Richardson, "Automatic Receipt Segmentation and Entity Recognition using OCR and Named Entity Recognition (NER)," pp. 1–13, 2023.
- [10] O. Tüselmann and G. A. Fink, "Neural models for semantic analysis of handwritten document images," *Int. J. Doc. Anal. Recognit.*, vol. 27, no. 3, pp. 245–263, 2024, doi: 10.1007/s10032-024-00477-8.
- [11] J. Andrés, C. Wall, S. Tarride, M. Coustaty, A. H. Toselli, and E. Vidal, "The PARES Database: Information Extraction over Historical Parish Records," *Int. J. Doc. Anal. Recognit.*, vol. 28, no. 3, pp. 377–389, 2025, doi: 10.1007/s10032-025-00531-z.
- [12] V. Choudhary, P. Guha, G. Pau, and S. Mishra, "Comparative Analysis of Open-Source OCR Frameworks for Receipt Data Extraction and Post-Processing," *Environ. Sustain. Indic.*, vol. 26, pp. 1–10, 2025.
- [13] P. Hiskiawan, Y. M. Geasela, H. Heryanto, E. Stephanie, M. Ardianti, and F. A. Sukarno, "Trustworthy Data Science Framework for Non-Invasive Nutritional Screening Using Computer Vision," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia: IEEE, 2025, pp. 1743–1748. doi: 10.1109/ICIMCIS68501.2025.11327268.
- [14] F. Wolf and G. A. Fink, "Self-training for handwritten word recognition and retrieval," *Int. J. Doc. Anal. Recognit.*, vol. 27, no. 3, pp. 225–244, 2024, doi: 10.1007/s10032-024-00484-9.
- [15] Z. Yan *et al.*, "DocExtractNet: A novel framework for enhanced information extraction from business documents," *Inf. Process. Manag.*, vol. 62, no. 3, p. 104046, 2025, doi: 10.1016/j.ipm.2024.104046.
- [16] V. Choudhary, P. Guha, G. Pau, and S. Mishra, "An overview of smart agriculture using internet of things (IoT) and web services," *Environ. Sustain. Indic.*, vol. 26, no. January, p. 100607, 2025, doi: 10.1016/j.indic.2025.100607.
- [17] J. R. Ilagan and J. B. Ilagan, "Graph-partitioning entity resolution for resolving noisy product names in OCR scans of retail receipts," *Procedia Comput. Sci.*, vol. 239, no. 2023, pp. 338–345, 2024, doi: 10.1016/j.procs.2024.06.180.
- [18] A. Takano, T. C. H. Cole, and H. Konagai, "A novel automated label data extraction and data base generation system from herbarium specimen images using OCR and NER," *Sci. Rep.*, vol. 14, no. 1, pp. 1–8, 2024, doi: 10.1038/s41598-023-50179-0.
- [19] A. Rombach and N. Mehdiyev, "Beyond Accuracy: Understanding Model Confidence in Key Information Extraction with Conformal Prediction," *Int. J. Doc. Anal. Recognit.*, vol. 29, no. 2, pp. 551–563, 2026, doi: 10.1007/s10032-026-00572-y.
- [20] A. Sarhan *et al.*, "Egyptian car plate recognition based on YOLOv8, Easy-OCR, and CNN," *J. Electr. Syst. Inf. Technol.*, vol. 11, no. 1, pp. 1–27, 2024, doi: 10.1186/s43067-024-00156-y.
- [21] N. U. Muchtar, D. Arifianto, and R. Umlasari, "Analisis Kinerja Transformer Untuk Named Entity Recognition (NER) Menggunakan IndoBERT Pada Transkrip Video Politik Berbahasa Indonesia," *Discov. J. Ilmu Pengetah.*, vol. 10, no. 2, pp. 180–199, 2025, doi: 10.33752/discovery.v10i2.10133.

- [22] A. De Nardin, S. Zottin, C. Piciarelli, G. L. Foresti, and E. Colombi, "In-domain versus out-of-domain transfer learning for document layout analysis," *Int. J. Doc. Anal. Recognit.*, vol. 28, no. 2, pp. 161–175, 2025, doi: 10.1007/s10032-024-00497-4.
- [23] C. A. Romein, A. Rabus, G. Leifert, and P. B. Ströbel, "Assessing advanced handwritten text recognition engines for digitizing historical documents," *Int. J. Digit. Humanit.*, vol. 7, no. 1, pp. 115–134, 2025, doi: 10.1007/s42803-025-00100-0.
- [24] S. Chakraverti, P. Agarwal, H. S. Pattanayak, S. P. S. Chauhan, A. K. Chakraverti, and M. Kumar, "De-noising the image using DBST-LCM-CLAHE: A deep learning approach," *Multimed. Tools Appl.*, vol. 83, no. 4, pp. 11017–11042, 2024, doi: 10.1007/s11042-023-16016-2.
- [25] P. Härtinger and C. Steger, "Adaptive histogram equalization in constant time," *J. Real-Time Image Process.*, vol. 21, no. 3, pp. 1–9, 2024, doi: 10.1007/s11554-024-01465-1.
- [26] M. Mayr *et al.*, "Data-efficient handwritten text recognition of diplomatic historical text," *Multimed. Tools Appl.*, vol. 84, no. 32, pp. 39107–39122, 2025, doi: 10.1007/s11042-024-20545-9.
- [27] X. Chen, W. Zhang, and S. Pan, "CLESER: Context-Based Label Knowledge Enhanced Span Recognition for Named Entity Recognition," *Int. J. Comput. Intell. Syst.*, vol. 17, no. 1, pp. 1–13, 2024, doi: 10.1007/s44196-024-00595-5.
- [28] S. C. Sreenivas, S. Chowdhury, and M. Masum, "Enhancing Clinical Named Entity Recognition via Fine-Tuned BERT and Dictionary-Infused Retrieval-Augmented Generation," *Electron.*, vol. 14, no. 18, pp. 1–19, 2025, doi: 10.3390/electronics14183676.
- [29] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min.*, vol. 16, no. 1, pp. 1–23, 2023, doi: 10.1186/s13040-023-00322-4.
- [30] P. Hiskiawan, T. P. Wijayanti, H. Heryanto, S. Everlin, S. A. Yasodhara, and D. Alexander, "Mel-Frequency Cepstral Coefficients and Neural Networks for Indonesian Traditional Music Recognition," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia: IEEE, 2025. doi: 10.1109/ICIMCIS68501.2025.11327436.