PEN TO PILL (P2P)

Summit Machathon 6.00

Habiba Yousri - Haneen Akram - Renad Hossam - Mohammed Mostafa – Shrouk Mohamed

KEY WORDS: Prescription Digitization - Optical Character Recognition (OCR) - Medicine Extraction - YOLOv8n - Mixed-Language Recognition - BART - Handwritten Text Recognition

Abstract

Handwritten medical prescriptions are notoriously difficult to digitize, especially in regions where both Arabic and English are used interchangeably. Standard OCR systems struggle with the variability of handwriting, multilingual content, and inconsistent formatting, leading to frequent misreadings of critical information such as medicine names and dosages — a risk that can result in harmful medical errors. This project presents an end-to-end deep learning–based pipeline to extract and structure handwritten medicine data from real prescription images.The pipeline begins by using a YOLOv8 object detection model trained on a custom-labeled dataset to accurately detect handwritten regions, distinguishing between Arabic (ar_word) and English (en_word) words. These detected word-level regions are cropped and passed into a fine-tuned TrOCR model, which significantly improves recognition performance over generic OCR tools. While this output provides readable text, it may still contain misspellings or lack clear structure.To solve this, we introduce a BART sequence-to-sequence model that is trained to transform noisy OCR outputs into a standardized string format: "medicine dosage, medicine dosage, ...". This approach corrects errors and enforces ordering, enabling reliable interpretation. A final post-processing step parses the structured string into (medicine, dosage) pairs by splitting on commas and extracting tokens using basic string operations — preparing the data for downstream use in digital health records, prescription verification systems, or pharmacy applications.The proposed system demonstrates strong multilingual generalization, robustness to handwriting variation, and a practical path forward for digitizing prescriptions in real-world healthcare settings.

Introduction

In many healthcare settings, handwritten prescriptions remain the primary method of communicating treatment plans between doctors, pharmacists, and patients. However, the **illegibility of handwriting**, especially when combined with **mixed-language use (Arabic and English)**, poses a serious risk to **patient safety**. Pharmacists often struggle to decipher drug names or dosages, leading to **misinterpretations**, **delayed treatments**, or **life-threatening medication errors**.

These challenges are further complicated in regions where **digital health infrastructure is limited** and where prescriptions are not stored electronically. As a result, healthcare professionals rely heavily on manual transcription and human memory, both of which are error-prone and inefficient.

Despite advances in Optical Character Recognition (OCR), most existing systems are trained primarily on clean, typed text in single languages, and they fail when faced with the **complexities of handwritten, multilingual, and unstructured documents** like prescriptions. Arabic handwritten text adds another layer of complexity due to its **cursive nature, variable character shapes**, and **limited annotated datasets**.

The lack of accurate, automated tools for extracting essential medical information from handwritten prescriptions is a pressing issue that demands attention. This project seeks to fill that gap by leveraging AI-based techniques to detect and extract **handwritten medicine names and dosage information** from real-world prescriptions, enabling safer, faster, and more reliable digitization of healthcare records.

Materials and Tools

- **Roboflow** : Used for annotating handwritten regions in prescriptions at both block and word levels, with class labels for Arabic and English words.
- **YOLOv8 (Ultralytics)** : Used for object detection to localize handwritten word regions in prescription images.
- **TrOCR** : Used for recognizing handwritten text in cropped word images, supporting multilingual (Arabic and English) handwriting.
- **BART** : Used as a sequence-to-sequence model to correct OCR output and structure it into a standardized medicine-dosage format.
- **Python (3.x)** : Used for developing the pipeline and implementing preprocessing, model training, and post-processing logic.
- **PyTorch :** Used as the deep learning framework for training and fine-tuning TrOCR and BART models.
- **Google Colab / Kaggle Notebooks :**Used as the development and training environments for running experiments and training models.

Methodology

This project followed a multi-step pipeline to extract and structure handwritten medicine names and dosages from multilingual (Arabic-English) prescription images.

1. **Handwritten Region Detection**
   - **Objective :** Isolate only the handwritten text from printed content and layout elements in the prescription.
   - Steps:
     1. Used **Roboflow** to manually annotate word-level bounding boxes in the dataset.Each word was labeled as either ar_word (Arabic) or en_word (English).
     2. Trained a YOLOv8 object detection model to recognize and localize handwritten regions.
     3. Applied the trained YOLO model on prescriptions to extract bounding boxes around handwritten words.
2. **Duplicate Box Removal**
   - **Objective:** Eliminate overlapping or redundant detections from YOLO output.
   - Implemented a custom Non-Maximum Suppression (NMS) algorithm using an Intersection over Union (IoU) threshold of 0.8.
   - Retained bounding boxes with the highest confidence scores while suppressing overlapping boxes exceeding the IoU threshold.
   - This step ensures that each handwritten word is represented by a single bounding box.
3. **Word-Level Cropping**
   - Cropped each detected handwritten word from the prescription based on YOLO's bounding boxes.These cropped word images became input samples for the OCR model.
4. **Box Sorting (Top-to-Bottom, Left-to-Right)**
   - bounding boxes were sorted using their top-left coordinates, prioritizing vertical (y) position and then horizontal (x) position. This spatial ordering emulates the natural reading sequence and ensures coherent text reconstruction
5. **Text Recognition (TrOCR)**
   - **Initial Attempt:** Tried off-the-shelf OCR tools (like EasyOCR), but they produced inconsistent results e.g., Arabic words interpreted as English and vice versa.
   - **Improved Approach:**
     - Preprocessed word images:
       - Grayscale conversion
       - Contrast enhancement
       - Resize to fixed height (64px), maintaining aspect ratio
       - Converted back to RGB for model compatibility
   - **Fine-tuned TrOCR**, a transformer-based OCR model, on the cropped images and their correct transcriptions.
   - This significantly improved word recognition for both Arabic and English.
6. **Text Reconstruction**
   - Combined all Tr-OCR-extracted words for a single prescription into one string.
7. **Text Correction & Structuring**
   - TrOCR output was often noisy (misspellings, lack of structure).
   - Trained a mBART sequence-to-sequence model **using:**
     - **Input**: Raw OCR output (noisy string)
     - **Target**: Clean, structured format→ "medicine dosage, medicine dosage, ..."
   - The BART model corrected spelling and imposed consistent formatting.
8. **Post-Processing and Pair Extraction**
   - Parsed the structured BART output to extract (medicine, dosage) pairs.
   - Split the output by commas
   - Then split each segment by the first space character.
   - Stored the resulting data as tuples for downstream use in digital health systems or for scheduling patient dosages.

Deployment

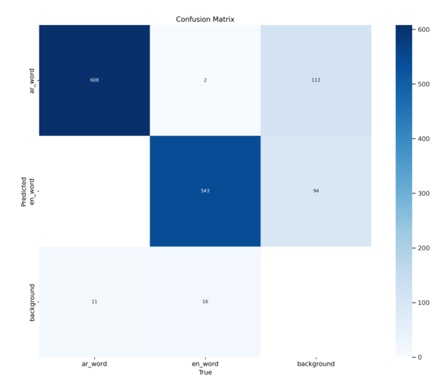
- **FastAPI-based application** provides an API for extracting medicine names and dosages from prescription label images.
  For Front-end :
  - Programming Languages : **JavaScript**
  - Frameworks : ReactJS, BootStrap
- **For Back-end:** Programming Languages : **Python**
  - **CORS Middleware**
  - **Global Variables**
  - **API Endpoints**

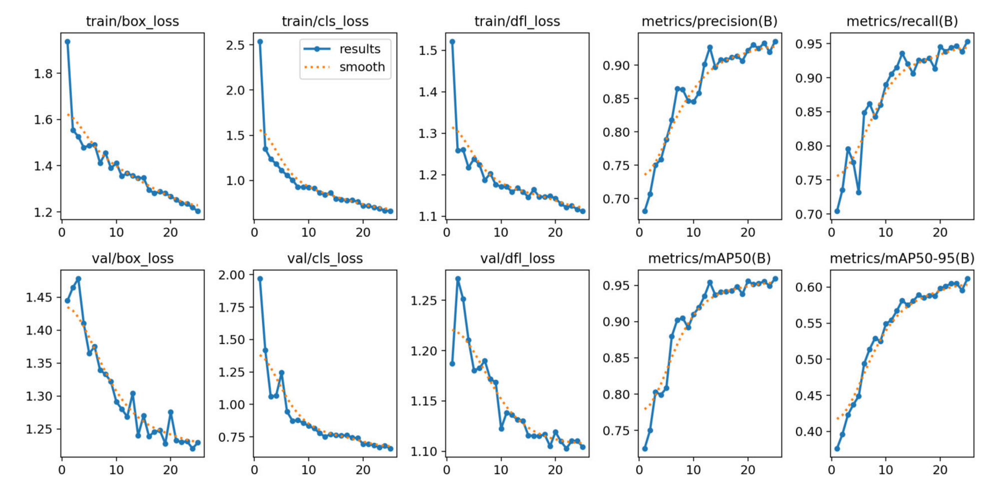
Results

- **Handwritten Word Detection (YOLOv8):**

YOLOv8				
Class	Precision	Recall	mAP50	mAP50-95
Ar_word	0.944	0.947	0.970	0.615
En_word	0.929	0.957	0.949	0.611
All	0.936	0.952	0.960	0.613

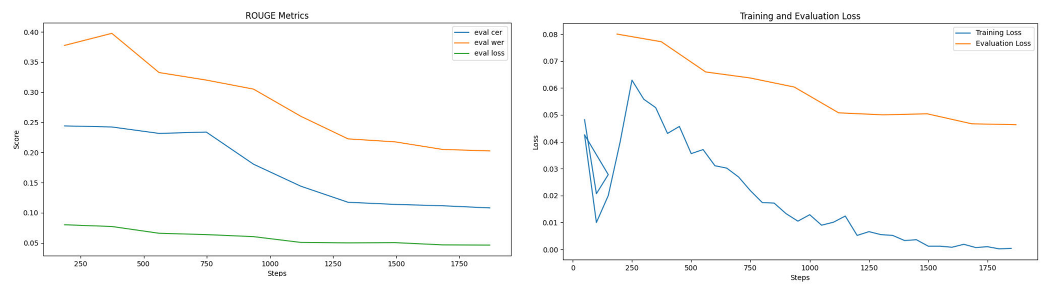
- **ar_word** was correctly predicted 608 times, misclassified 113 times as background.
- **en_word** was correctly predicted 543 times, misclassified 94 times as background.





Key observations from the plots:
- **Loss curves** (box, cls, dfl) show smooth convergence.
- **Precision and Recall** improved steadily throughout training.
- **mAP50 and mAP50-95** increased consistently, reaching:
  - mAP50: ~0.96
  - mAP50-95: ~0.61

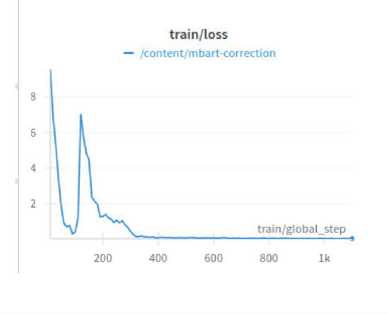
- **Text Recognition (TrOCR)**
  - **Train samples: 1496, Eval samples: 374 -> this is a ~80/20 split**



Text Recognition (TrOCR)	
Exact match accuracy	80 %
Character-Level accuracy	83.34 %

- **mBART**

ROUGE Scores for mBART		
Metric	Meaning	Value
rouge1	Overlap of unigrams (single words) between predicted and reference	0.5786
rouge2	Longest common subsequence (captures sentence-level structure)	0.5734
rougeL	Same as rougeL but adapted for multi-sentence summaries	0.5711



1. **Ordering Model:** Reorders and corrects the raw output from TrOCR.
2. **Medicine Extraction Model:** Extracts the names of medicines from the reordered text.
3. **Dosage Extraction Model:** Extracts the corresponding dosages from the same reordered text.

Future Work

Future Work:
- **Medicine Dictionary Integration**: Incorporate a verified medicine database to further improve text correction, reduce ambiguity, and support error-tolerant fuzzy matching.
- **Dosage Normalization:** Standardize dosage formats (e.g., converting "2x/day" or "twice daily" into unified terms).
- **Prescription Digitization System:** Integrate the pipeline into a full mobile/web system to assist pharmacists, caregivers, and healthcare providers.

Acknowledgement

We gratefully acknowledge the invaluable support and constructive feedback provided by the **STP team**, which significantly contributed to the progression of this project. We also extend our appreciation to the **teaching assistants at Ain Shams University** for their technical guidance and critical insights throughout the development process.