

MULTI-TASK DEEP LEARNING SEQUENCE PREDICTION FOR BILINGUAL HANDWRITTEN CLINICAL PRESCRIPTIONS

Ayesha Ahmad¹, Hafiz Haris Ali^{*2}, Hafiz Muhammad Hussnain², Muhammad Hussnain Butt², Nadeem Iqbal³, Abdul Razzaq¹, Fatima Bukhari⁴

¹Faculty of Computing, MNS-University of Agriculture, Multan, 60000, Pakistan

²School of Business and Technology, Indiana Wesleyan University, 8415 Georgia St, Merrillville IN 46410, USA

³School of Informatics and Creative Arts, Dundalk Institute of Technology, Ireland

⁴Department of Computing, National University of Modern Languages, Multan, 60000, Pakistan

^{*2}hafizharis.ali@myemail.indwes.edu

DOI: <https://doi.org/10.5281/zenodo.22888147>

Keywords

CNN, TrOCR, HTR, VGG Image Annotator, CER, WER

Article History

Received: 14 June 2026

Accepted: 05 September 2026

Published: 22 September 2026

Copyright @Author

Corresponding Author: *

Hafiz Haris Ali

Abstract

Handwritten Text Recognition (HTR) in medical domain documents, particularly doctors' prescriptions, remain a major challenge due to unconstrained handwriting, domain-specific terminology, and multi-script variations. To address these complexities, this study presents a specialized recognition workflow evaluated on a custom doctors' prescription dataset categorized into seven distinct structural classes (Personal Information, Medicine Name, Dosage, Diagnostic, Symptoms, Numeric Data, and Text) across English, Urdu, and Numeric scripts. Ground-truth annotation was performed using the VGG Image Annotator (VIA) and exported to JSON format for pipeline integration. The proposed methodology employs a hybrid pipeline CNN - TrOCR comprising an EfficientNet-B0 backbone coupled with a Transformer encoder for structural classification, followed by Transformer based OCR (TrOCR) for end-to-end word level text transcription. Experimental evaluation demonstrates that the classification network achieves an accuracy of 87.77% for language identification and 69.69% for structural category classification. For sequence recognition using TrOCR, the system achieves a Character Error Rate (CER) of 23.51% and a Word Error Rate (WER) of 37.91%. These findings highlight the potential of combining convolution-transformer feature extractors with dedicated sequence decoders to transcribe complex, multi-lingual medical handwriting, offering a foundation for automated clinical documentation workflows.

Introduction

As the world's healthcare systems are undergoing rapid digital transformation, the need for automated document processing and Handwritten Text Recognition (HTR) has become more vital than ever in modernizing the clinical infrastructure. The ability to seamlessly transfer these unstructured physical aspects to structured Electronic Health Record (EHR) databases is essential for creating interoperable data-driven healthcare ecosystems.

Medical prescriptions are one of the most crucial means of communication in clinical practice, and they are the official way of communicating patient diagnoses, complex drug regimes, and exact drug dosages.

Medication errors represent a common cause of adverse events in modern healthcare practice. For example, according to WHO, unsafe medication practices and medication errors were among the main reasons for avoidable damage to patients' health all around the world,

therefore, the WHO initiated the global challenge "Medication Without Harm" with a goal to reduce avoidable severe medication harm by 50%. Also, WHO reports that every 30th patient suffers from medication-related harm, while more than one-quarter of such cases is severe or life-threatening (World Health Organization, 2024). Prescription-related errors are particularly important because they occur at the beginning of the medication-use process. If the original prescription is unclear or incomplete, the risk may continue through dispensing and administration. Elliott et al. (2021) estimate that medication errors were responsible for 237 million errors and 1,708 fatalities due to definite avoidance of adverse drug events. Mulac et al. (2021) report that 0.8% of reported medication error cases in Norwegian hospitals ended up in fatalities, and 5.2% cases led to the harm of patients. It means that poor handwriting of prescriptions must be considered as a risk factor for patients' safety.

- A large portion of existing studies on medical handwriting recognition focuses exclusively on isolated Medicine Name recognition or single cropped word snippets.

- Most traditional Handwritten Text Recognition models are benchmarked exclusively on single language, Latin only datasets (e.g., standard English handwriting models). Urdu Nastaliq remains a profoundly underserved script within computer vision research. The lack of large-scale, annotated training datasets of handwritten medical documents has limited the development of high-accuracy models, leaving South Asian healthcare settings without reliable automated transcription tools.

- There is an immediate lack of end-to-end architectures that combine character recognition with automated entity classification to map text into structured data frames across multiple key clinical domains.

- No existing research establishes a unified processing framework designed to parse unconstrained multi-script medical text directly into distinct, structured JSON arrays across vital clinical categories.

This paper proposes a special two-stage recognition system to solve these disadvantages

for doctors' prescriptions in multi-script scenario. The VGG Image Annotator (VIA) was used to annotate input documents, which are then exported to structured JSON representations according to seven functional categories: Personal Information, Medicine Name, Dosage, Diagnostic, Symptoms, Numeric Data and Text. The proposed pipeline uses a hybrid EfficientNet-B0 backbone and Transformer Encoder for language identification and regional classification. Cropped word tokens are then fed into a TrOCR for a fine-grained sequence transcription.

Contribution:

Addressing the gaps, the main contributions of this study are as follows:

- We utilize the publicly available BHCP-DOC dataset, a real world bilingual handwritten clinical prescription benchmark introduced by kajla et al., to evaluate and benchmark our proposed handwritten text recognition (HTR) framework.

- We present an EfficientNet-B0 and Transformer framework that can effectively classify spatial document regions and determine script types: English, Urdu and Numeric.

- *TrOCR Transcription Pipeline:* We are testing an end-to-end TrOCR transcription pipeline on complex medical handwritten documents and reporting the quantitative performance of the pipeline, using Character Error Rate (CER) and Word Error Rate (WER).

Related Work:

The development of Handwritten Text Recognition (HTR) systems has progressed from classical statistical methods to deep learning architectures that integrate spatial feature extraction with sequential modeling. This section reviews related literature focusing on the transition from hybrid Convolutional Neural Network (CNN)-Recurrent Neural Network (RNN) architectures to modern Transformer-based and TrOCR models.

- ***Transformer-Based OCR (TrOCR) and Pre-trained Models:***

A major breakthrough in the field is the TrOCR model introduced by Li et al. (2023). Unlike previous hybrid models that still relied

on CNN backbones, TrOCR uses a pure Transformer architecture for both image understanding and text generation. By initializing the encoder with a pre-trained Vision Transformer (ViT) and the decoder with a pre-trained language model (such as RoBERTa or MiniLM), the system achieves state-of-the-art results. Specifically, TrOCR-Large achieved a state-of-the-art CER of 2.89% on the IAM dataset without requiring complex pre-processing or external language models. Yet the study relied on Latin script transcription.

● *Specialized Applications in Medical Prescriptions:*

Handwritten Text Recognition (HTR) research has struggled to balance unconstrained sequential transcription with document layout understanding. Early deep learning solutions focused almost on line level or word level monolingual transcription. The classic convolutional Recurrent Neural Network (CRNN) and CNN-BLSTM-CTC frameworks suffer from a critical performance. The limitation was qualified by Waly et al. (2024), who reported that unconstrained line decoder experience a sharp decline in Word Recognition Rate (WRR) dropping from 93.75% for single word inputs to a mere 18.78% for six word sentences.

Raisi et al. (2020) proposed a modified CNN-transformer model using a self attention module to capture spatial handwriting features. The flat, sequence dependent limitation is also visible in the state of the art TrOCR model introduced by Li et al. (2023). Although TrOCR achieves exceptional transcription accuracy by coupling pre-trained vision and language Transformers such as ViT and RoBERTa. Consequently, it cannot parse complex, multi-field layouts naively without relying on heavy external pre-processing pipelines. This structural gap is especially critical in the clinical domain, where handwriting variability, cursive styles, and layout variations present high risks of transcription errors. Early prescription recognition systems, such as the deep CRNN designed by Fajardo et al. (2019), were limited to basic character and digit classification of monolingual English terms, lacking the capability to dynamically segment and extract

varied clinical fields. Tabassum et al. (2021) developed a recognition system for Bangladeshi medical terms, which are often a bilingual mix of English and Bangla. Using a Bidirectional LSTM and a novel "Stroke Rotation and Parallel-shift" (SRP) data augmentation technique, they achieved 89.5% accuracy. Consequently, their framework lacked document layout awareness, rendering it unable to link related patient, drug, or dosage fields dynamically across a full prescription page. However, their framework remains a single entity extraction model optimized solely for isolated medicine names, leaving vital surrounding clinical parameters such as dosage, frequency, instructions completely unaddressed.

Material and Methods:

The primary objective of the proposed framework is to parse, classify, and transcribe unconstrained, multi-script handwritten doctors' prescriptions. The overall system operates through a two-stage pipeline: structural region and language classification using an EfficientNet-B0 backbone coupled with a Transformer Encoder, and word-level text sequence recognition using Transformer OCR (TrOCR). A two-stage deep-learning-based system for analyzing handwritten prescription images was designed. In the first stage, handwritten prescription regions are semantically and linguistically classified through a hybrid CNN + Transformer network. In the 1st stage, handwritten prescription regions undergo semantic and language classification by a hybrid CNN + Transformer network. The second step is the recognition of handwritten prescription text by a pretrained TrOCR vision-encoder-decoder model trained on cropped regions of prescription text. The entire processing pipeline consists of parsing and normalizing annotations, extracting regions using bounding boxes, filtering samples based on their quality, partitioning the data sets at the prescription level, classification with CNN-Transformer and handwritten text recognition with Transformer. The prediction of the semantic category and the language of each handwritten region is done during the classification stage, and the generation of the

corresponding textual sequence is carried out during the recognition stage.

In order to have a primary raw prescription dataset, the medical handwriting data was collected in real-time via physical collection, given that multi-script medical handwriting benchmarks are not available publicly.

Prescription images were collected from different personal networks, colleagues and private households in Pakistan that were of very high quality. All the gathered artifacts come from private practice and hospital networks around the country.

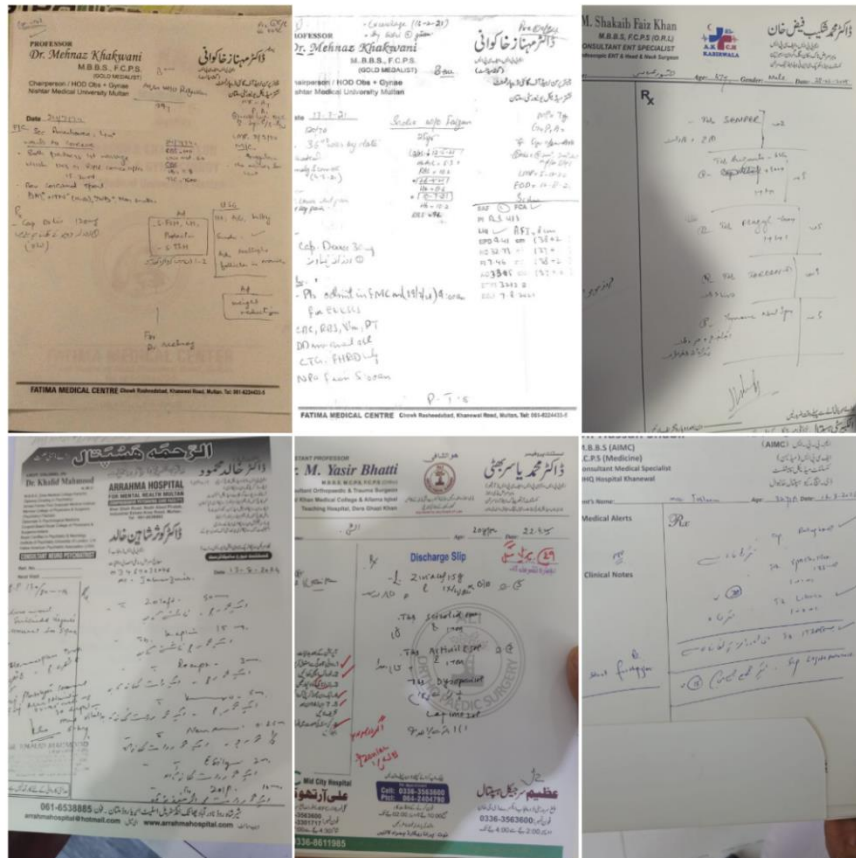


Figure 1: Prescription Variations

The entire dataset contains real-world clinical variations (figure 1) such as multi-script handwriting, where medical terms are written in English, notes in Urdu or English, and numerical doses are written in Urdu or English. The data was collected in real time with different smartphone cameras, thus the images are raw data and contain a lot of unconstrained noise, such as:

- Ambiguous lighting, background shadows and glare.
- Distortions in perspective, documents with non-standard orientations, and camera angles that are oblique.
- Paper ageing, folds, ink bleed-through and overlapping strokes.

The complete dataset is organized within a multi-tiered directory topography. The structural volume of the master corpus comprises **256 distinct prescription folders**, each having different Doctors' ID, which function as the primary organizational unit for patient-specific or clinic-specific groupings. In each unique Doctor ID folder, there are multiple images with the same handwriting of the same doctor and organization-/clinic. Each folder has 1 associated JSON file.

In order to turn this handwriting data into a structured digital medical record, a multitask annotation scheme was created. In each region of the data, there are two different target (figure 2) spaces that are examined, which are Clinical Category Prediction (7-class task, Personal

Information, Medicine Name, Numeric Data, Dosage, Diagnostic, Symptoms, Text) and Linguistic Script Identification (3-class task, Urdu, English, Numeric). Annotations for the prescription set were created (figure 3) by applying VIA tool

(annotate.officialstatistics.org/) a standard open-source interface for bounding box mapping. VIA (VGG Image Annotator) is a conventional open-source interface for mapping bounding boxes.

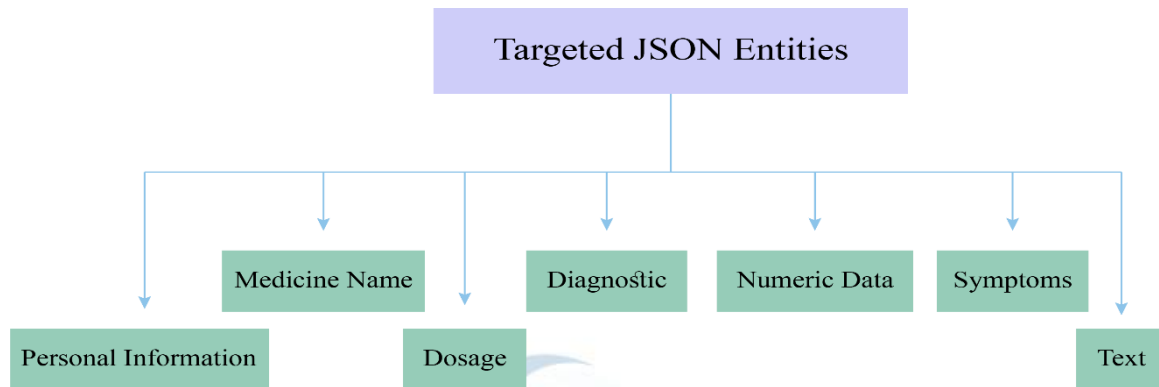


Figure 2: Targeted Entities

Annotations will be stored in one single VIA JSON file structure for each document folder. Such arrangement links each image from the

raw prescriptions with a specific annotation that describes its text regions explicitly.

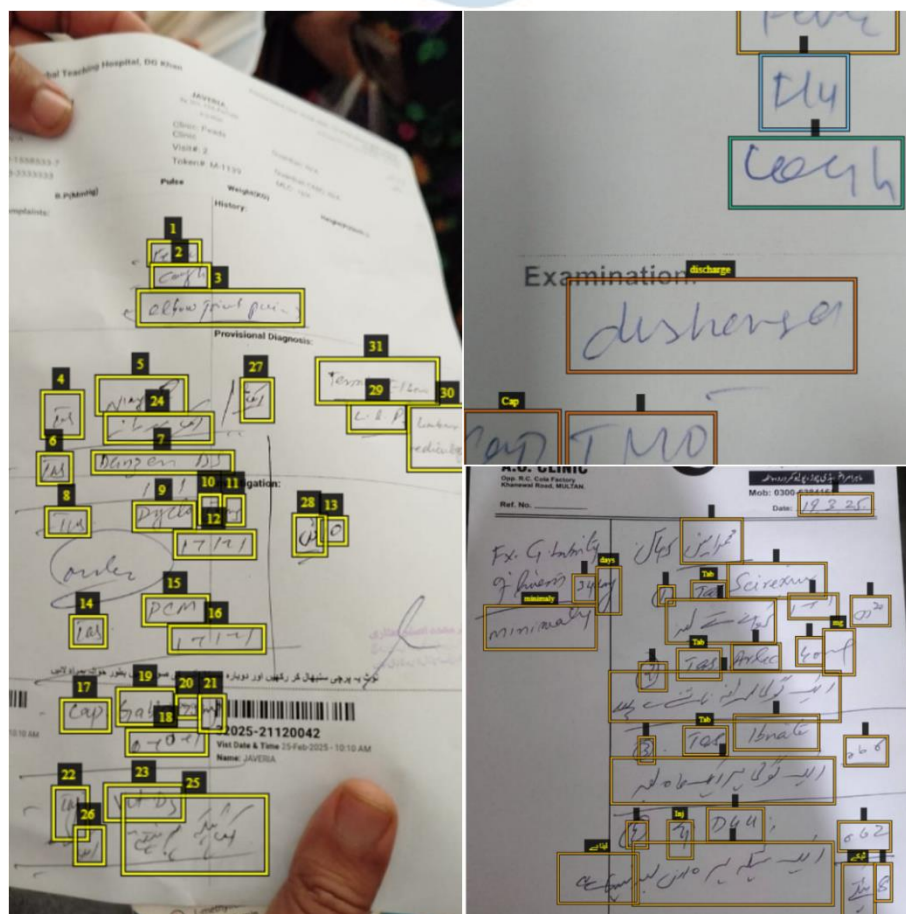


Figure 3: Annotated Prescriptions

Within the VIA JSON schema, each annotated prescription sheet is represented as a distinct object key containing two main primary data arrays (Shape attributes, region attributes).

The annotation structure consisted of image metadata, annotated regions, geometric attributes of a shape and region attributes. A rectangular window was used to represent each of the annotated regions, where the dimension and spatial location of the window was described by its x-y coordinates and width-height. After the initial annotation-processing stage, 17,254 annotated regions were found, with four duplicate annotation records being removed. The annotations that were studied belonged to seven semantic categories: Text, Numeric Data, Medicine Name, Dosage, Personal Information, Diagnostic and Symptoms as shown in figure 3.

Spelling and naming differences were translated to seven standard categories to uniform the annotation terminology. Normalization was done for variants of diagnostic, dosage, symptom and medicine-name labels. Additionally, semantic relabeling was done based on the text in the annotations. Medication-form expressions like tablet, capsule, syrup, injection, drop, cream, ointment, gel or lotion were differentiated from the real names of the medicines. Symptoms and diagnoses were also used to normalize the corresponding semantic categories, as well as to generate lexical patterns. This operation caused 263 category reassignment operations. All language labels were standardized as: English, Urdu, Numeric, Mixed, and Unknown. After performing the normalization of the annotation, there were 12,727 English, 3,281

Urdu, 1,234 Numeric, 11 Unknown, and 1 Mixed annotated regions.

Rectangular regions of individual samples were extracted from the prescription images as per the annotated bounding boxes. A valid image dimension was used to define the boundaries of the crop considering an annotation with upper-left point (x,y), width w, and height h.

$$x_1 = \max(0, x), \quad y_1 = \max(0, y), \\ x_2 = \min(W, x+w), \quad y_2 = \min(H, y+h),$$

Where W is the width of the source image and H is the height of the source image. Before saving the cropped regions, a bounding-box quality control was used. Only regions were kept that met requirements in terms of their size and area.

$$w > 6, \quad h > 10, \quad wh > 300.$$

Regions that did not meet these criteria were not included. The 17,254 candidate cropped regions were reduced to 16,896 valid cropped regions and 358 cropped regions were discarded. This yielded a total of 4,967 Text samples, 3,539 Medicine Name samples, 3,501 Numeric Data samples, 3,199 Dosage samples, 674 Personal Information samples, 525 Diagnostic samples and 491 Symptom samples.

Dataset Partitioning:

For reasons of minimizing the risk of information leakage from multiple regions from the same prescription, the datasets were not partitioned at the level of individual crops but instead at the prescription-folder level. The 256 prescription folders were randomly shuffled, with a fixed random seed of 42, and split into train, validation and test sets. The desired partition is shown in Table 1 .

Table 1: Dataset Train/Test/Val Split

Partition	Prescription folders	Cropped samples
Training	179	10,430
Validation	38	3,114
Testing	39	3,352
Total	256	16,896

Image Preprocessing and Data Augmentation:

Prescription areas were cropped and resized to 224×224 for semantic and language classification. Training was done with data

augmentation techniques such as colour and geometric transformations. Brightness and contrast of 0.25 were used with *ColorJitter* for color variability. The translations were upto 5%

in both directions, the scaling ± 0.05 , and random rotation was applied within $\pm 8^\circ$. A fill value of 255 was used for geometric transformations.

These images were then transformed to tensors and normalized with

$\mu=[0.485,0.456,0.406]$

and $\sigma=[0.229,0.224,0.225]$.

Presents the end-to-end structural flow of the proposed multi-task classification pipeline. The

architecture unifies localized convolutional feature maps with sequential global self-attention layers to simultaneously resolve linguistic and semantic tasks without structural duplication. The unified network operates in three macroscopic phases: deep multi-scale feature extraction, sequential token-space mapping, and multi-task projection (figure 4).

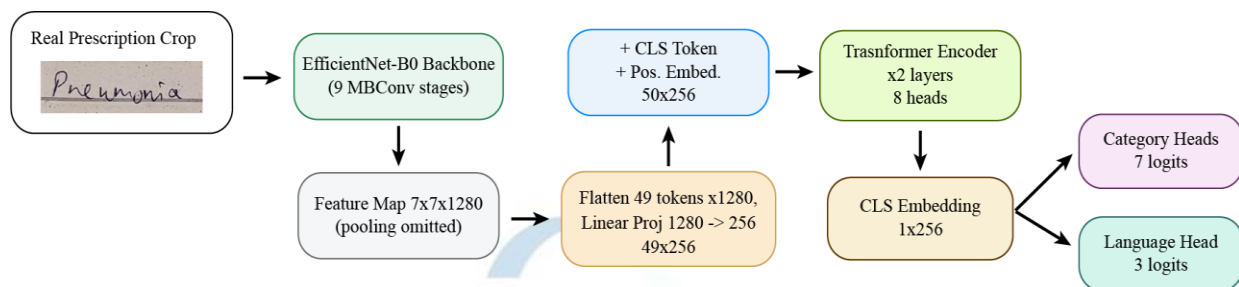


Figure 4: EfficientNet-B0 + Transformer Architecture

Multi Task learning objective: The model was optimized jointly for semantic category and language classification. For each task, cross entropy loss with label smoothing of 0.05 was used. To mitigate class imbalance during training, the classes were not sampled, but class weighting was used.

Class-Imbalance Handling: The distribution of semantic categories was greatly skewed, with fewer samples of Diagnostic, and fewer of Symptoms, and Personal Information, than of the more common Text, Medicine Name, Numeric Data, and Dosage categories. To alleviate the impact of this imbalance, during training a WeightedRandomSampler was used. The frequencies of each category were used to determine sampling weights as follows:

$$w_i = \min \left(\sqrt{\frac{N_{max}}{N_i}}, 6 \right) (i)$$

where N_i represents the number of training samples in category i , and N_{max} denotes the largest category count (i).

Model Optimization and Training: The CNN-Transformer model was optimized using AdamW with a learning rate of 1×10^{-4} and weight decay of 1×10^{-4} . Training was configured for a maximum of 40 epochs with a batch size of 32. A cosine-annealing learning

rate schedule was used with $T_{max} = 40$, And a minimum learning rate of 1×10^{-6} .

Classification Evaluation: This trained model was tested on the held-out test set of 3,352 cropped prescription regions. During the testing, the model was put in evaluation mode and the computation of the gradient was turned off. In both prediction tasks, the maximum classification logic was used to predict the class. The accuracy, precision, recall, F1-score, macro-average F1-score, weighted F1-score and confusion matrices were used to measure the performance. Confusion matrices were also analyzed by calculating the class-specific prediction behaviour by row-normalized percentages. The overall accuracy of the model for semantic-category classification was 69.69% and the macro-F1 score of the model was 53.57%. The accuracy for language classification was 87.77% and macro-F1 score was 71.26%.

Handwritten Text Recognition: The handwritten text recognition was done with the pre-trained model of handwritten recognition, microsoft/trocr-base-handwritten. It is a vision-encoder-decoder Transformer model, which was fine-tuned with cropped handwritten prescription areas and their transcriptions. The recognition dataset consisted of 16,896 cropped handwritten text regions and ground

truth text. Samples without paths to images and/or labels for transcription were dropped. The data set has been randomly split into a training set of 13,517 examples and a validation set of 3,379 examples in order to generate an 80:20 split, using a random seed of 42. The longest sequence length was defined as 64 tokens.

The experiment was based on a full, end-to-end training workflow, from metadata parsing and creation of image-text pairs, to the splitting of the dataset, image pre-processing, text pre-processing, model fine-tuning, handling the checkpoints, evaluation based on the metrics, and finally model export. Cropped handwritten text regions were used for input, and ground-truth transcriptions for target sequences. Experiment with the uploaded training pipeline is done on TrOCR's processor VisionEncoderDecoderModel using cropped metadata, sequence-to-sequence training, CER/WER evaluation, and checkpoint exporting as checkpoint-best-TrOCR.

The model was initialized from the pre-trained checkpoint microsoft/trocr-base-handwritten. This checkpoint was selected because it is already trained for handwritten text recognition, making it a suitable starting point for adaptation to the target dataset. Training was

conducted using Hugging Face's Seq2SeqTrainer using the explicit operational parameters outlined below:

- **Training Framework:** Hugging Face Seq2SeqTrainer
- **Training / Validation Partition:** 13,517 Training Samples / 3,379 Validation Samples
- **Per-Device Training Batch Size:** 2
- **Gradient Accumulation Steps:** 8
- **Effective Global Batch Size:** 8
- **Learning Rate (AdamW):** 4e-5
- **Weight Decay Rate:** 0.01
- **Total Training Target:** 4 Epochs
- **Target Metric Evaluation Strategy:** Checkpoint selection optimized for lowest validation Character Error Rate (CER)

The TrOCR model was tuned for handwritten text recognition, and the fine-tuned model showed progressive improvement in the recognition of handwritten text during training. The transcription performance (table 02) has improved over the four epochs of training, with the training loss as well as the validation loss going steadily down, and the character error rate (CER) and word error rate (WER) going down throughout the fine-tuning process.

Table 2: TrOCR Performance

Epoch	Training Loss	Validation Loss	CER	WER
1	4.1409	0.8153	0.3520	0.4955
2	2.5275	0.7037	0.2828	0.4425
3	1.4161	0.6419	0.2630	0.4155
4	0.6147	0.6155	0.2351	0.3791

Results & Discussion:

The baseline multi-task network concurrently evaluates clinical document region parsing (Task A: 7 Categories) and linguistic script identification (Task B: 3 Scripts). Task A

achieved a global test accuracy of 69.69% (where $N = 3,352$). The heatmap of the confusion matrix shows the recall of the respective categories shown in figure below.

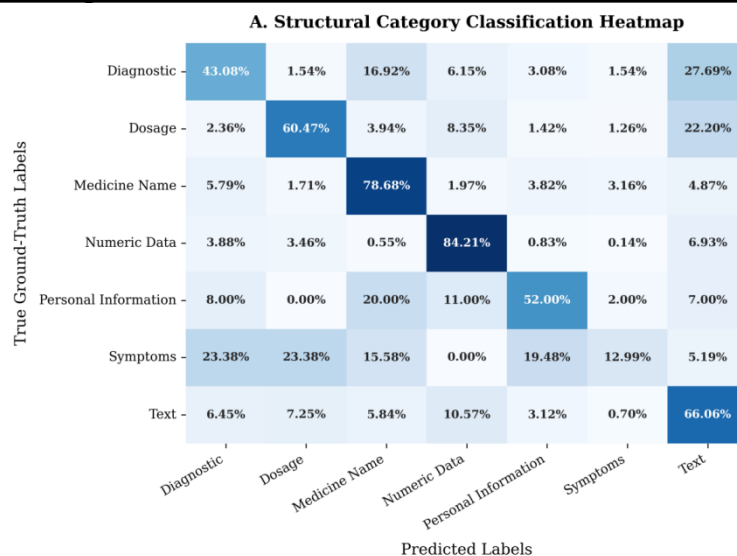


Figure 5: Heatmap of Category Classification

Task B achieved 87.77% test accuracy due to distinct visual script features (Latin alphanumeric, Nastaliq ligatures, Arabic numerals). English scored $F_1 = 0.88$ (96.55% recall). Urdu achieved 0.96 precision but

80.40% recall ($F_1 = 0.82$) due to 19.30% of Urdu tokens being misclassified as English. The heatmap of the confusion matrix shows the recall of the respective categories shown in figure below.

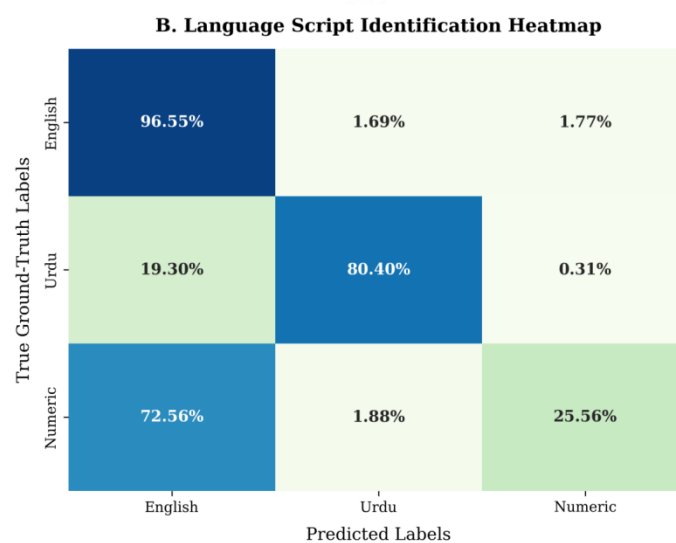


Figure 6: Heatmap of language class

A fine-tuned sequence-to-sequence TrOCR architecture was employed to transcribe the targeted text regions to produce character level recognition for the word segmentation loop without breaking. For fine-grained recognition without brittle text regions segmentation, a

fine-tuned sequence-to-sequence TrOCR architecture was used for transcription of the targeted text regions. The optimization path and error reduction path are plotted with epochs for the training in figure below.

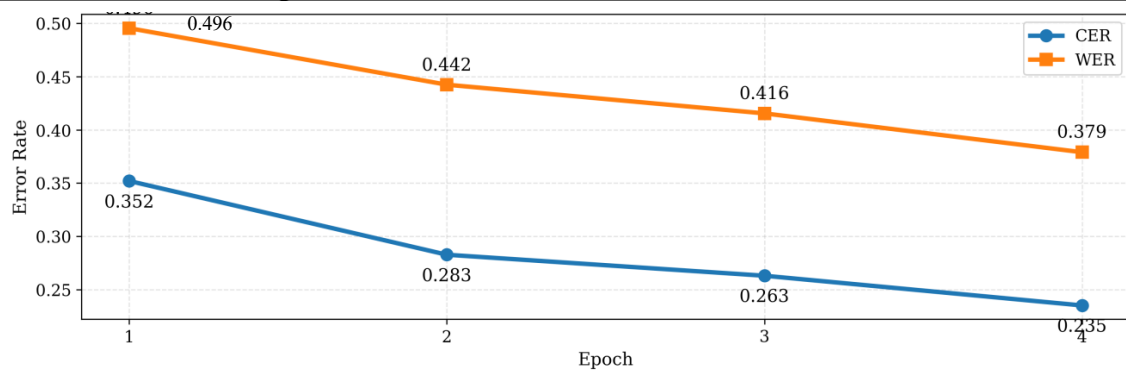


Figure 7: Convergence metrics for fine tuned-OCR

Fine-tuning reduced the Character Error Rate (CER) from 35.20% to 23.51% and the Word Error Rate (WER) from 49.55% to 37.91%. The persistent gap between WER and CER remains standard for handwritten document parsing, as any isolated character substitution invalidates the complete word token.

Conclusion:

Study proposed a two-stage pipeline for processing handwritten medical prescriptions, motivated by the safety risks posed by illegible, often bilingual (English/Urdu) prescription text and the lack of automated tools and dataset tailored to this domain. The experimental evaluation of the multi-task architecture demonstrates a robust capability in isolating highly structured textual entities, particularly within heavily populated classes like Numeric Data and Medicine Names. The vision encoder's high precision in identifying pharmaceutical nomenclature serves as a critical safety mechanism, drastically lowering false-positive rates where mis-classification could compromise data integrity in practical clinical workflows. However, a persistent bottleneck remains within the severely underrepresented semantic categories, such as Diagnostics and Symptoms.

Compared to prior handwritten prescription OCR work, our results reflect the added difficulty of an open-vocabulary, dictionary matched task. Limitation of this study include the moderate dataset size, the current independence of the classification and transcription stages and comparatively higher error rates observed for urdu script. Overall, this work demonstrates a practical end-to-end

approach toward digitizing handwritten prescriptions, contributing a step toward reducing medication errors caused by illegible handwriting in clinical settings.

REFERENCES:

- Ahmed, S. B., Naz, S., Swati, S., & Razzak, M. I. (2019). Handwritten Urdu character recognition using one-dimensional BLSTM classifier. *Neural Computing and Applications*, 31(4), 1143-1151.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Elliott, R. A., Camacho, E., Jankovic, D., Sculpher, M. J., & Faria, R. (2021). Economic analysis of the prevalence and clinical and economic burden of medication error in England. *BMJ Quality & Safety*, 30(2), 96-105.
- Fajardo, L. J., Sorillo, N. J., Garlit, J., Tomines, C. D., Abisado, M. B., Imperial, J. M. R., ... & Fabito, B. S. (2019, November). Doctor's cursive handwriting recognition system using deep learning. In *2019 IEEE 11th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management (HNICEM)* (pp. 1-6). IEEE.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. 2016. *Deep learning*. MIT press, Cambridge.

- Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. 2008. A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 31(5): 855-868.
- Kajla, N. I. BHCP-DOC: A Real-World Bilingual Dataset for Handwritten Clinical Prescription Detection and Recognition.
- Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., ... & Wei, F. (2023, June). Trocr: Transformer-based optical character recognition with pre-trained models. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 11, pp. 13094-13102).
- Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., ... & Wei, F. (2023). Trocr: Transformer-based optical character recognition with pre-trained models. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 11, pp. 13094-13102).
- Mulac, A., K. Taxis, E. Hagesaether, and A. G. Granas. 2021. Severe and fatal medication errors in hospitals: findings from the Norwegian Incident Reporting System. *Eur. J. Hosp. Pharm.* 28(e1): e56-e61.
- Raisi, Z., Naiel, M. A., Fieguth, P., Wardell, S., & Zelek, J. (2020). Handwritten recognition using transformer for medical prescription application. In *Virtual colloquium jointly hosted by University of Strathclyde and University of Waterloo*.
- Tabassum, S., Takahashi, R., Rahman, M. M., Imamura, Y., Sixian, L., Rahman, M. M., & Ahmed, A. (2021, May). Recognition of doctors' cursive handwritten medical words by using bidirectional LSTM and SRP data augmentation. In *2021 IEEE technology & engineering management conference-Europe (TEMSCON-EUR)* (pp. 1-6). IEEE.
- Waly, A., Tarek, B., Feteha, A., Yehia, R., Amr, G., & Fares, A. (2024, October). Arabic handwritten document OCR solution with binarization and adaptive scale fusion detection. In *2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (pp. 316-319). IEEE.
- World Health Organization. 2024. Medication without harm: policy brief. World Health Organization, Geneva, Switzerland.