http://dx.doi.org/10.14500/aro.12820

Kurdish Handwritten Text Recognition: A DenseNet121-Transformer Architecture with Constrained Synthetic Line Generation

Karez A. Hamad[1,2†] and Shareef M. Shareef[1]

[1]Department of Software and Informatics Engineering, College of Engineering, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq.

[2]Department of Business Management, Mergasour Technical Institute, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq

***Abstract*—Based on the available peer-reviewed literature, Kurdish handwritten recognition remains at a nascent stage, with existing models limited to isolated character, digit, and word recognition. The demand for digitizing Kurdish handwritten documents has grown rapidly, particularly in regions undergoing government digitization, such as the Kurdistan Regional Government. The primary obstacle hindering progress beyond isolated characters or words is the absence of text-based handwritten datasets. To address this gap, a comprehensive Kurdish handwritten dataset encompassing paragraph-, line-, and word-level samples is first introduced. A recognition architecture combining DenseNet-121 as the CNN backbone with a Transformer-based encoder-decoder for sequence modeling is then proposed for Kurdish handwritten line recognition, representing one of the earliest reported efforts in this domain. To augment the limited line-level training data, a constrained recipe-based synthetic line generation framework is developed that concatenates real handwritten word images while enforcing text uniqueness, single-writer consistency, and leakage-free data partitioning. Additional training strategies are investigated, including cross-lingual transfer learning from Arabic handwritten data and fixed-content handwritten line integration for improving recognition accuracy. The best configuration achieved a character error rate (CER) of 0.0593 and a word error rate (WER) of 0.3083 without language model integration, further reduced to a CER of 0.0534 and a WER of 0.2746 with an 8-gram language model. The source code and trained models are publicly available at: https://huggingface.co/Karez/KHLR**

***Index Terms*—Arabic-script recognition, Data augmentation, DenseNet-Transformer, Kurdish handwriting recognition, Synthetic line generation.**

ARO-The Scientific Journal of Koya University
Vol. XIV, No.2 (2026), Article ID: ARO.12820. 12 pages
DOI: 10.14500/aro.12820
Received: 07 January 2026; Accepted: 09 May 2026
Regular research paper; Published: 23 June 2026
†Corresponding author's e-mail: karez.hamad@su.edu.krd

I. Introduction

Handwritten recognition is the process of converting handwritten text from scanned documents or images into machine-readable digital form (Hamad and Kaya, 2016). Although recognition systems for Latin scripts have advanced considerably, cursive scripts such as Arabic and Kurdish remain challenging due to character connections, confusing diacritical marks, and visually similar letters. While Arabic handwritten recognition has been extensively studied, Kurdish has received limited attention, with most existing research focusing only on isolated characters or digits. This limitation is mainly due to the lack of large Kurdish handwritten datasets at the line and paragraph levels. Fig. 1 highlights the main characteristics and challenges of Kurdish handwritten recognition.

This study addresses two major gaps in Kurdish handwritten recognition by introducing a large-scale dataset containing word-, line-, and paragraph-level samples, and by developing a recognition model that combines DenseNet-121 (Huang, et al., 2017) for feature extraction with a custom Transformer architecture (Vaswani, et al., 2017). The study also proposes a constrained recipe-based synthetic line generation framework that augments training data using concatenated real handwritten words. Based on the available peer-reviewed literature, no prior work has addressed Kurdish handwritten text recognition at the line level, and this work represents one of the earliest reported efforts in Kurdish line-based handwritten recognition.

The main contributions are summarized as follows:

- Introducing a comprehensive text-based Kurdish handwritten dataset comprising word-, line-, and paragraph-level annotations
- Developing a lightweight Transformer-based architecture for Kurdish handwritten line recognition
- Proposing a constrained recipe-based synthetic line generation framework that concatenates real handwritten word images while simultaneously enforcing text uniqueness, single-writer consistency, and data partition integrity
- Establishing baseline results for Kurdish handwritten line recognition based on the available peer-reviewed literature.

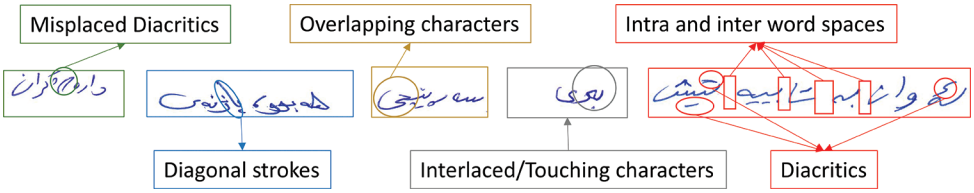


Fig. 1. Common characteristics and inherent challenges of the Kurdish script.

The remainder of this paper is organized as follows: Section II reviews related work, Section III presents the methodology, Section IV reports experimental results and analyses, Section V discusses findings and limitations, and Section VI concludes the paper.

II. Literature Review

This section reviews Kurdish handwritten recognition models reported in the literature. Based on the available peer-reviewed literature, and as further supported by the recent comprehensive review of (Hamad and Shareef, 2025), no Kurdish handwritten recognition system has been proposed beyond the word level, including line-level or paragraph-level recognition. Consequently, in addition to surveying existing Kurdish handwritten models, handwritten recognition approaches developed for script-related languages, namely, Arabic and Urdu, are also reviewed, with particular emphasis on Transformer-based architectures.

(Ahmed, et al., 2022) introduced one of the earliest CNN-based approaches for recognizing isolated Kurdish characters, reporting a test accuracy of 83%, although the model suffered from considerable overfitting. This limitation was later alleviated by (Ibrahim, Saman Nariman, and Majeed, 2023) through the use of improved preprocessing techniques. (Ali and Abdulrazzaq, 2024b) evaluated several pre-trained networks, including MobileNet, DenseNet121, and ResNet50, alongside custom CNN architectures using the KurdSet dataset, with DenseNet121 achieving the highest performance. However, the subsequent study by (Ali and Abdulrazzaq, 2024a) on character recognition was limited to 29 classes, which does not cover the full set of 33–35 Kurdish characters. At the word level, (Shareef and Ali, 2024) applied a Mask R-CNN framework with a ResNet-18 backbone to localize and recognize Kurdish strokes in printed word images. More recently, (Alsaqi and Fattah, 2025) introduced the Kurdish Handwritten Word Dataset and proposed a CNN-BiLSTM-CTC pipeline for Kurdish handwritten word recognition.

Recent advances in handwritten recognition for Arabic-like scripts have increasingly adopted Transformer and attention-based models. (Momeni and BabaAli, 2024) explored transformer and transformer transducer architectures on KHATT, reporting character error rate (CER) values of 12.11–19.76% using Data-efficient Image Transformer (DeiT) and Bidirectional Encoder Representations from Transformers (BERT) representations. Gader and Echi (2022) proposed a Convolutional Neural Network – Bidirectional Long Short-Term Memory – Connectionist Temporal Classification (CNN-BiLSTM-CTC) attention framework evaluated on Institut für Nachrichtentechnik / École Nationale d'Ingénieurs de Tunis (IFN/ENIT) and Institut für Angewandte Mathematik (IAM Handwriting Database) (IAM) datasets. For Urdu, (Anjum and Khan, 2023) introduced CALText with a contextual attention localizer and Densely Connected Convolutional Network – Gated Recurrent Unit - (DenseNet-GRU) decoder, while (Ganai and Khursheed, 2023) proposed a BERT-initialized Vision Transformer relying on explicit ligature segmentation, which may introduce error propagation. Dhiaf, et al. (2023) presented MSdocTr Lite, a lightweight 3.9M parameter model exploiting cross-script transfer learning, and (Abdo, et al., 2024) combined YOLOv5 with a Vision Transformer for Arabic bank check recognition, achieving 99.02% accuracy.

The reviewed studies exhibit several limitations, including reliance on large-scale synthetic printed data or external language models, complex multi-stage pipelines with error-prone explicit segmentation, and the absence of targeted data augmentation strategies that enhance the visual discrimination of similar dot-differentiated characters without relying on generative models or additional learnable parameters. The methodologies and results of comparable Transformer-based studies for Arabic-like scripts are summarized in Table XI of Section IV for direct comparison with the proposed architecture.

III. Proposed Methodology

This section presents the proposed DASTNUS Kurdish handwritten dataset, the DenseNet121-Transformer architecture for recognizing Kurdish handwritten lines, the constrained recipe-based synthetic line generation framework, and the fixed-content handwritten line integration strategy.

A. Proposed Dataset

The progress of Central Kurdish handwritten text recognition has been significantly constrained by the lack of publicly available datasets beyond isolated characters and digits, as indicated by the available peer-reviewed literature and further supported by the recent review of (Hamad and Shareef, 2025). To address this limitation, the DASTNUS (a Kurdish term meaning handwriting) dataset is introduced, a large-scale Kurdish handwritten dataset comprising word-, line-, and paragraph-level samples. A total of 1,134 forms were distributed to writers with diverse demographic backgrounds, of which 1,000 completed forms were retained after excluding forms with extremely unreadable handwriting and unreturned forms, as reported in Table I.

http://dx.doi.org/10.14500/aro.12820

DASTNUS was collected using a five-page A4 form. The first page records writer metadata and instructions. The second page contains a fixed paragraph designed to cover all Kurdish characters in their positional forms. The third page collects unique handwritten texts, with each writer assigned a distinct passage selected from 85 Central Kurdish books across 15 subject fields, with management-related content forming the largest share (≈15%), reflecting digitization needs in public sectors in the Kurdistan regional government. Fig. 2 presents the distribution of content topic categories across the 15 subject fields covered by the dataset. The division of the unique handwritten paragraph samples into approximately 70% for training, 15% for validation, and 15% for testing considered factors such as subject field diversity and a balanced distribution of Kurdish characters to ensure proper representation across all subsets. This writer-to-split assignment served as the baseline distribution for all other subsets of the proposed DASTNUS dataset for preventing writer leakage across subsets.

Handwritten text lines were extracted from both the fixed and unique paragraph images using a standard text line segmentation technique (Rabaev and Litvak, 2025), namely, projection profile analysis, followed by connected component analysis for boundary refinement, as illustrated in Fig. 3.

The fourth and fifth pages collect handwritten words and characters/digits, including personal and place names, month names, and 2,750 unique Kurdish words selected from a standard dictionary introduced by (Rustay, et al., 2012) and distributed across 1,000 forms. All forms were scanned at a resolution of 300 dpi using a Canon DR-C240 scanner and saved as TIF format files with the CaptureOnTouch V5 Pro software. Full dataset statistics are reported in Table II.

Dataset verification was performed in two stages. First, forms with missing content, poor handwriting quality, or scanning errors were removed. Second, ground truth annotations were created at the paragraph, line, and word levels in plain text, Excel, and customized PAGE XML formats containing writer metadata. The transcription process strictly preserved the original handwriting, including writing mistakes, omitted dots, extra words, and scratched corrections encoded with special symbols. Fig. 4 presents the complete dataset development pipeline and sample collection forms.

TABLE I
DEMOGRAPHIC DISTRIBUTION OF DASTNUS DATASET PARTICIPANTS (N=1,000)

Attribute	Distribution
Age	Below 15: 8.8%, 15–25: 53.4%, 26–50: 29.8%, Above 50: 8.0%
Gender	Male: 43.3%, Female: 56.7%
Handedness	Right: 88.6%, Left: 11.4%
Occupation	Student: 60.8%, Teacher: 17.4%, Employee: 9.6%, Freelancing: 5.4%, Others: 6.8%

B. Proposed Baseline Architecture

The baseline model first extracts features using the well-known pre-trained CNN DenseNet-121 (Huang, et al., 2017), which has proven effective for Kurdish handwritten recognition by (Ali and Abdulrazzaq, 2024a). These features are then processed by a custom transformer network (Vaswani, et al., 2017) designed to model sequential dependencies and decode the entire handwritten lines. Fig. 5 illustrates the complete structure of the proposed architecture for Kurdish handwritten line recognition.

As illustrated in Fig. 5, the proposed architecture first feeds the handwritten line image X into DenseNet-121 for feature extraction. DenseNet-121 establishes dense connectivity by linking each layer to all preceding layers, enabling effective feature reuse and alleviating the vanishing gradient problem. Accordingly, the output of the l_{th} layer is given by:

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \quad (1)$$

where $[x_0, \ldots, x_{l-1}]$ is the concatenation of all preceding feature maps, and $H_l()$ represents a composite function of

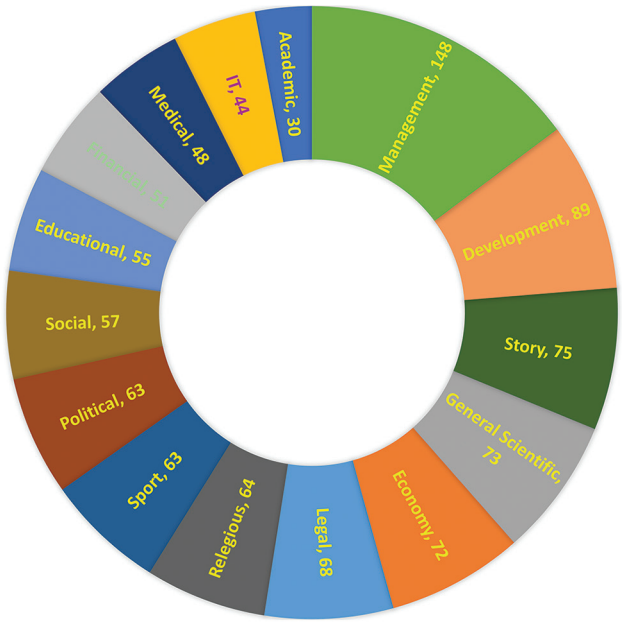


Fig. 2. Distribution of unique handwritten text content across the 15 subject fields in the DASTNUS.

TABLE II
STATISTICS OF ALL THE SUBSETS OF THE PROPOSED DASTNUS DATASET ACROSS SPLITS

Splits	Unique paragraphs	Fixed paragraphs	Unique lines	Fixed lines	Character and digits	Person names	Place names	Month names	Unique words	Total word counts	Total character counts	Bi-grams	Tri-grams
Training	710	712	3,575	7,433	17,771	2,983	2,985	8,527	37,954	255,978	1,428,682	177,858	173,198
Validation	144	144	655	1,521	3,594	1,000	1,000	1,723	8,201				
Testing	144	144	649	1,518	3,589	998	1,000	1,726	8,036				
Total	998	1,000	4,879	10,472	24,954	4,981	4,985	11,976	54,191				

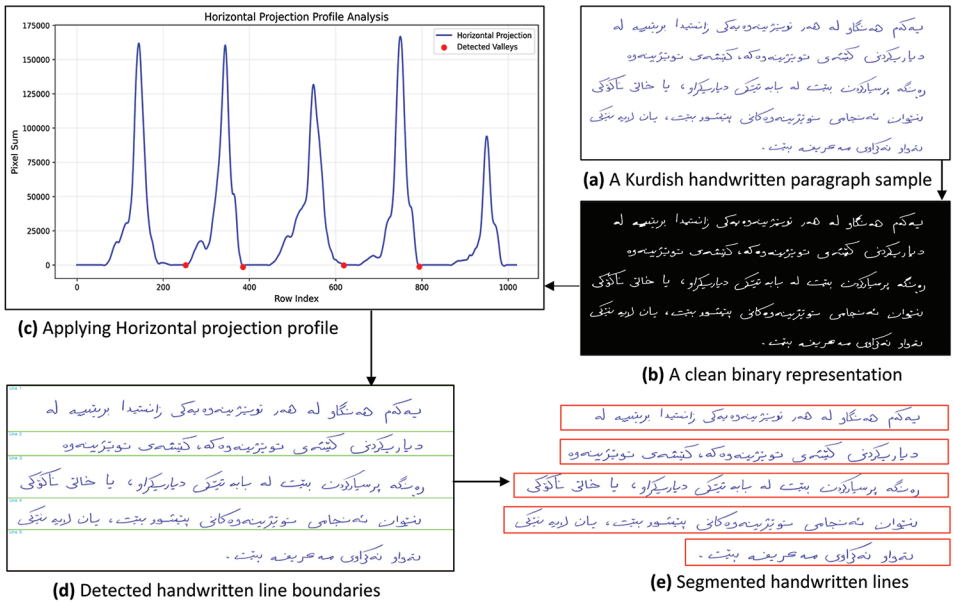


Fig. 3. Procedure for segmenting Kurdish handwritten paragraphs into individual handwritten lines.

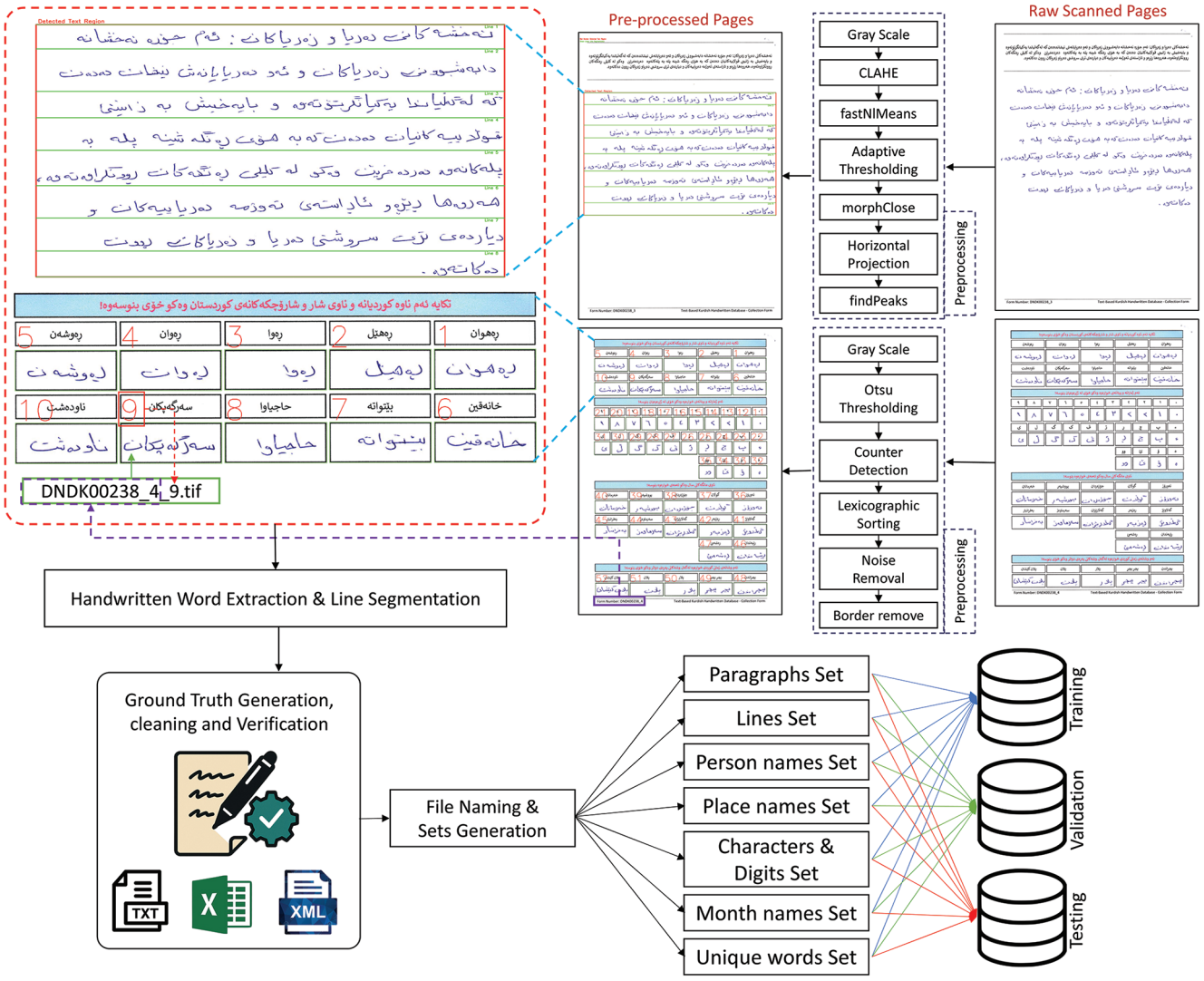


Fig. 4. Complete overview of the development procedure of the proposed Kurdish handwritten dataset, including two representative samples.

http://dx.doi.org/10.14500/aro.12820

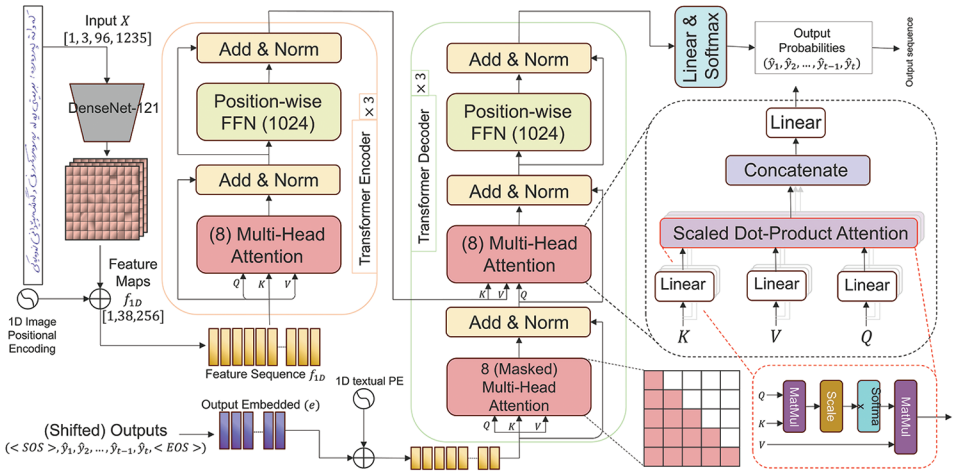


Fig. 5. Proposed DenseNet121-transformer architecture for recognizing Kurdish Handwritten lines.

BatchNorm, ReLU, and convolution. This dense connectivity improves gradient flow and parameter efficiency.

The extracted feature sequence $f \in R^{B\times L\times 256}$, where L is the sequence length determined by the spatial width after feature extraction, is then augmented with sinusoidal positional encoding to obtain sequential information from the input feature vectors. The resulting feature sequence is subsequently fed into the transformer architecture. The proposed custom transformer consists of three encoder layers and three decoder layers. Each encoder layer comprises multi-head self-attention with eight scaled dot-product attention heads, followed by a two-layer position-wise feed-forward network (FFN) with an inner dimension of 1024. Both submodules are equipped with residual (shortcut) connections and followed by normalization layers to ensure stable training, as follow:

$$X_A = NormalizationLayer(G_f[X]) + X \quad (2)$$

where X is the input, X_A is the output of the formulation, and G_f denotes the output of the previous Transformer sublayer, which is either the position-wise FFN or the multi-head attention module.

The scaled dot-product attention used in each head of the multi-head attention mechanism can be expressed as a unified function:

$$\text{ScaledDotProductAttention}(Q,K,V) = \text{softmax}(\alpha QK^T)V \quad (3)$$

where Q, K, and V are the query, key, and value matrices with $Q \in R^{N\times Dk}$, $K \in R^{M\times Dk}$, and $V \in R^{M\times Dv}$. Here, N and M denote the lengths of the query and key/value sequences, respectively, and $D_k = D_v = 32$ are the dimensions of each attention head (corresponding to 256/8 heads). The matrices Q, K and V are obtained through linear transformations using the weight matrices W_q, W_k, and W_v. The product QK^T is scaled by a factor $\alpha = 1/\sqrt{D_k}$ to prevent the gradient vanishing problem.

Finally, the decoder generates the output sequence autoregressively using causal masking, where each token prediction depends on previously generated tokens and the encoder memory:

$$Y_t = \arg\max(\text{softmax}(Wo \cdot \text{Decoder}(M, E([y_0,\ldots,y_{t-1}])))) \quad (4)$$

Here, M is the encoder memory, E() is the token embedding with positional encoding, $Wo \in R^{256\times|V|}$ is the output projection matrix, and |V| is the vocabulary size. Generation begins with the SOS token and stops once the EOS token is predicted.

C. Constrained Recipe-Based Synthetic Line Generation

Augmenting HTR datasets by concatenating real handwritten image samples has attracted attention as a practical alternative to generative synthesis. Shen and Messina (2016) synthesized text lines from isolated Chinese character images, Ingle, et al. (2019) concatenated online ink strokes into longer lines within a multilingual pipeline, Cascianelli, et al. (2021) composed synthetic text from real character instances under limited data conditions, and (Coquenet, Chatelain, and Paquet, 2023) synthesized document-level training data by concatenating real word and line crops. However, all existing approaches target left-to-right scripts, and none jointly enforces text uniqueness, single-writer consistency, and leakage-free data partitioning. Based on the available peer-reviewed literature, the framework presented here represents one of the earliest reported applications of constrained real-world concatenation to a right-to-left Arabic-script language.

The proposed framework operates over two handwritten word sources from the DASTNUS dataset, namely, unique words and Kurdish person names. The unique vocabulary subset contains 2,750 words distributed across 1,000 collection forms, where every 20 consecutive writers share the same set of words, forming big groups G_b of 20 writers and yielding 54,191 word samples in total. The person names subset contains 1,000 Kurdish male and female names distributed across 1,000 forms, where every 5 consecutive writers share the same set of names, forming nested sub-groups g_s of 5 writers within each big group and yielding 4,981 samples in total. This dual-level grouping arises naturally from the DASTNUS collection design and is exploited by the proposed framework to maximize recipe diversity. Let $W = \{w_1,\ldots,w_N\}$ denote the writer set, where

each writer *w* possesses samples from one or both sources. Every sample receives a source tag τ_i = (*type, id*) with type ∈ {*uniqueWord, personName*} to distinguish unique words from person names and prevent the system from treating two different words as identical when they share the same numerical identifier across the two sources. A recipe $r = \{\tau_1,\ldots,\tau_k\}$ is defined as a unique combination of *k* tagged word identifiers that specifies the content of one synthetic line, where *k* follows:

$$P(k) = \begin{cases} 0.50 & k = 8 \\ 0.25 & k = 7 \\ 0.25/3 & k \in \{4,5,6\} \end{cases} \quad (5)$$

The signature function σ(r) = sort (r) maps each recipe to its canonical form, and acceptance requires $\sigma(r) \notin \Sigma_{used}$, guaranteeing zero duplicate text sequences across the entire synthetic corpus. Recipes are classified as pure when containing only unique word tags, making them eligible for all 20 writers in G_b, or mixed when containing person name tags, restricting eligibility to the 5 writers in the corresponding G_s. Mixed recipes receive assignment priority due to their narrower eligibility constraints. For each recipe, the first eligible writer *w* is selected such that all required samples exist and no tagged word appears in the consumed set U_w.

On assignment, ink regions from real handwritten word samples of DASTNUS dataset are extracted using adaptive Otsu thresholding:

$$\theta^* = \text{Otsu}(I_g)+20 \quad (6)$$

where the +20 adjustment preserves thin diacritical marks critical in Kurdish script. Each word is scaled to height $h_t = 0.88 \times H$ and aligned to baseline $y_b = 0.75 \times H$ with ±1 pixel vertical jitter. Words are composed right-to-left with spacing sampled from Uniform (10,30) pixels. Algorithm 1 formalizes this process.

This procedure produced 3,762 training, 614 validation, and 596 testing synthetic lines, each containing an entirely unique text sequence. Authentic handwriting style is preserved by enforcing single-writer consistency within every generated line, ensuring that all words composing a given line originate from the same writer. Data leakage between partitions is prevented by restricting each split to the same set of writers assigned to the corresponding split in the other two data sources (unique handwritten lines and fixed handwritten lines) of the proposed DASTNUS dataset.

D. Fixed-Content Handwritten Line Integration

The fixed handwritten line subset of the DASTNUS dataset consists of 7,380 training line images collected from 1,000 writers, each of whom transcribed the same paragraph designed to cover all Kurdish characters in their isolated, initial, medial, and final forms. To introduce writer diversity without encouraging vocabulary memorization, 50 writers from the training split are randomly selected, and their corresponding line samples are integrated with the unique and synthetic handwritten lines. This strategy increases exposure to diverse handwriting styles and improves the representation of rare Kurdish characters that may be underrepresented in the unique or synthetic handwritten lines.

ALGORITHM 1: Constrained Recipe-Based Synthetic Line Generation

INPUT: Word pools $S_{uniqeWord}$, $S_{personName}$; writer set $W=\{w_1,\ldots, w_N\}$; Writer group structures $\{G_b, G_s\}$

OUTPUT: Synthetic line set L with ground truth labels

```
Begin
  For each big group G_b of 20 writers that share the same 55 unique words:
    Assign source tag τ_i=(type, id) to every word, type ∈ {uniqueWord, personName}
    Initialize Σ_used ←∅; U_w←∅ ∀w∈G_b
    For i=1 to R_attempts:
      Sample line length k from P(k); select k tagged words
      Classify recipe r as pure (only uniqueWord tags) or mixed (uniqueWord+personName tags)
      σ(r)←sort (r)
      If σ(r) ∈ Σ_used: Reject and continue⊳Uniqueness guarantee
      Σ_used ←Σ_used ∪{σ(r)}
    Assign mixed recipes first, then pure recipes
    For each recipe r:
      For each eligible w where all tags exist and no tag of r∈U_w:
        Extract ink from each word using θ*=Otsu (I_g)+20⊳diacritical preservation
        Scale to h_t=0.88×H; align to y_b=0.75×H with±1 px jitter
        Compose words right-to-left on canvas; spacing ~Uniform (10,30) ⊳ RTL layout
        U_w←U_w ∪ {tags of r}; save to L; break⊳Single-use guarantee
End
```

TABLE III
Training Hyperparameters

Hyperparameter	Value
Image size	96×1235
Optimizer	AdamW
Learning rate	5×10^{-4}
Weight decay	1×10^{-4}
Optimizer momentum	β_1=0.9, β_2=0.999
Batch size	64
Training epochs	80
LR scheduler	OneCycleLR+ReduceLROnPlateau
Gradient clipping	5.0
Early stopping patience	10 epochs
Dropout rate	0.4
Total trainable parameters	~12.8M

IV. Experimental Results

This section presents the experimental evaluation of the proposed model, covering datasets, metrics, training configurations, strategy analysis, baseline comparisons, ablation studies, and exploratory experiments including language model integration.

A. Experimental Setup

Three datasets were used for evaluation. The proposed DASTNUS dataset served as the primary benchmark for all

training and evaluation experiments. The KHATT Arabic handwritten dataset introduced by (Mahmoud, et al., 2014), containing 4,582 training, 569 validation, and 589 testing line samples, and the PUCIT Urdu handwritten dataset introduced by (Anjum and Kha, 2023), containing 5,554 training, 935 validation, and 912 testing samples, were additionally employed for cross-script transfer learning, zero-shot and few-shot evaluation, domain adaptation, and cross-dataset generalization experiments.

Recognition performance was measured using CER and word error rate (WER), both computed based on Levenshtein distance as below:

$$CER=\frac{(S+D+I)}{N} \qquad WER=\frac{(S_w+D_w+I_w)}{N_w} \tag{7}$$

In these formulas, *S*, *D*, and *I* denote substitutions, deletions, and insertions at the character level, with *N* representing the total character count. The subscript *w* refers to word-level operations, and N_w denotes the total word count. Character Recognition Rate, calculated as 1-CER, is also reported as a percentage-based accuracy measure. In addition, inference time and frames per second are reported for language model integration experiments.

The model was implemented using the PyTorch framework. Table III lists the training hyperparameters used across all experiments. All training was conducted on a workstation equipped with an Intel Core i9-14900K processor (24 cores, 3.20 GHz), 128 GB RAM (6000 MT/s), and an NVIDIA GeForce RTX 5090 GPU with 32 GB VRAM.

B. Results

Table IV presents the progressive improvement in recognition performance as each training strategy described in Section III is applied incrementally. All experiments used a fixed random seed of 42 to ensure reproducibility. The baseline model, trained without augmentation, exhibited substantial overfitting and memorization of training data. Each subsequent strategy reduced error rates and improved generalization, with the best configuration +AA+SKHL+FHL-50: Adaptive Augmentation + Synthetic Kurdish Handwritten Lines + Fixed-content Handwritten Lines from 50 writers (+AA+SKHL+FHL-50) achieving a CER of 0.0593 and a CRR of 94.07%, representing a relative CER reduction of 27.6% compared to the baseline.

TABLE IV
Progressive Improvement Through Proposed Training Strategies on Dastnus Test Set

No.	Model configuration	CER	WER	CRR (%)
1	Baseline model (no augmentation)	0.0819	0.3812	91.81
2	+ Standard augmentation (SA)	0.0666	0.3323	93.34
3	+ Adaptive augmentation (AA)	0.0664	0.3251	93.36
4	+ AA+Synthetic Kurdish handwritten lines (SKHL)	0.0613	0.3139	93.87
5	+ AA+Fixed-content handwritten lines from 50 writers (FHL-50)	0.0649	0.3232	93.51
6	+ AA+SKHL+FHL-50	0.0593	0.3083	94.07

CER: Character error rate, WER: Word error rate

To ensure that the observed improvements were not caused by random variation, paired statistical significance testing was performed using the paired bootstrap test (B = 100,000) and the Wilcoxon signed-rank test, and the results are presented in Table V. Results showed that both standard and adaptive augmentation significantly improved performance over the baseline, while combining synthetic handwritten lines with adaptive augmentation also significantly reduced CER. In contrast, adding fixed-content lines alone did not achieve statistical significance, suggesting that writer diversity without new text content offers limited benefit. However, combining synthetic and fixed-content lines produced a statistically significant improvement with a 10.7% relative CER reduction, demonstrating that both augmentation approaches complement each other effectively. Fig. 6 illustrates the bootstrap distribution, paired difference histogram, per-line CER distribution, and scatter plot for the comparison between configurations 3 and 6.

To further evaluate model robustness, multi-seed validation was conducted for the +AA and +AA+SKHL+FHL-50 configurations across five random seeds. Table VI reports the results, confirming consistent improvement of the proposed augmentation strategy across different initializations.

Fig. 7 illustrates the training and validation loss curves together with CER progression across epochs for the best-performing model, demonstrating stable convergence with the validation loss closely tracking the training loss, indicative of effective regularization and minimal overfitting.

Fig. 8 presents confusion matrices for the 15 most frequently misrecognized characters, comparing the +AA model against the +AA+SKHL+FHL-50 model. Fig. 9 demonstrates the deployment of the best-performing model through a web-based recognition system that accepts Kurdish handwritten paragraph images and outputs the recognized text.

C. Additional Exploratory Experiments

To evaluate the effect of language model integration on recognition performance, beam search decoding with character-level n-gram language models trained on the AsoSoft Kurdish text corpus (Veisi, MohammadAmini, and Hosseini, 2020) and a Central Kurdish RoBERTa model (Abdullah, et al. 2024) based on XLM-RoBERTa-large were evaluated. Table VII reports the best-performing

TABLE V
Statistical Significance Testing for Model Comparisons on Dastnus Test Set (Model Numbers Refer to Table IV)

Comparison	ΔCER	Relative Δ (%)	Bootstrap p-value	95% Confidence interval	Wilcoxon p-value
1 versus 2	0.0143	18.7	<0.001	(0.0107, 0.0181)	<0.001
1 versus 3	0.0151	18.9	<0.001	(0.0112, 0.0191)	<0.001
3 versus 4	0.0042	7.8	0.008	(0.0008, 0.0076)	0.002
3 versus 5	0.0019	2.3	0.090	(−0.0009, 0.0048)	0.192
3 versus 6	0.0069	10.7	<0.001	(0.0037, 0.0101)	<0.001

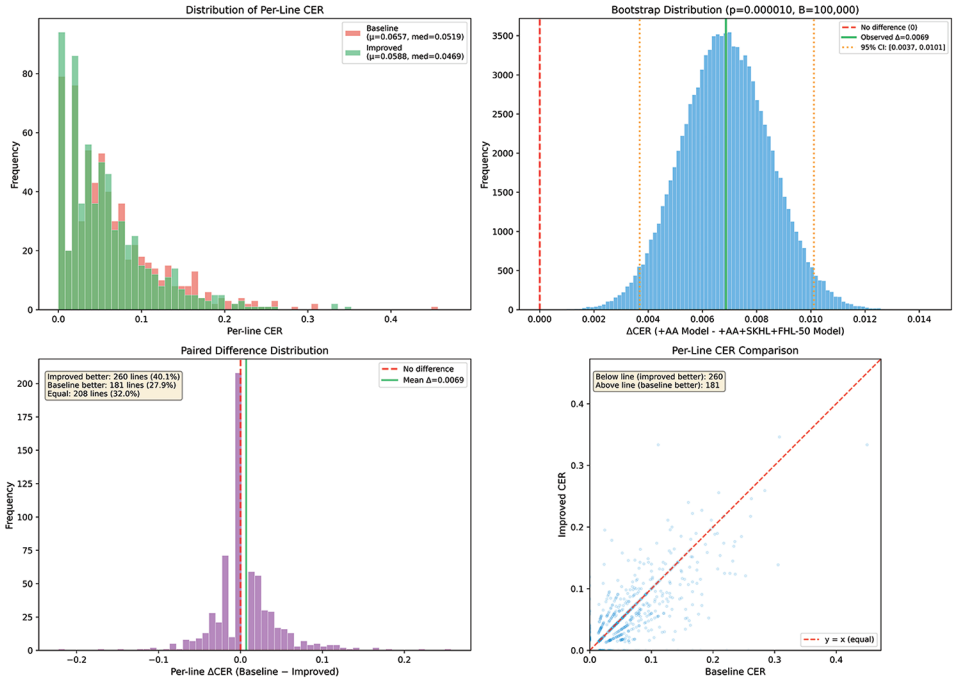
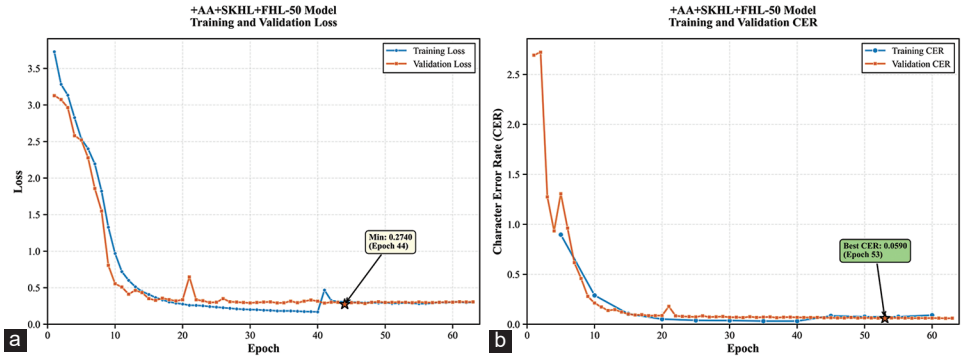


Fig. 6. Statistical significance analysis comparing the +AA and +AA+SKHL+FHL-50 models on DASTNUS test set.



Training and validation loss of best performing model Training and validation CER of best performing model

Fig. 7. (a and b) Training and validation loss and character error rate curves of the best-performing model.

TABLE VI
MULTI-SEED VALIDATION RESULTS (TEST CER) FOR THE+AA AND+AA+SKHL+FHL-50 MODELS

Model	Seed (7)	Seed (42)	Seed (123)	Seed (456)	Seed (789)	Mean±standard deviation
+AA	0.0661	0.0664	0.0653	0.0652	0.0672	0.0660±0.0008
+AA+SKHL+FHL-50	0.0602	0.0593	0.0586	0.0575	0.0611	**0.0593±0.0014**

Note: Bold values indicate the best-performing result in the corresponding comparison.

TABLE VII
LANGUAGE MODEL INTEGRATION RESULTS ON DASTNUS TEST SET USING THE+AA+SKHL+FHL-50 MODEL

Decoding strategy	CER	WER	Inference time (ms)	Frames per second
Greedy	0.0593	0.3083	**176.3**	**5.67**
Beam-3+7-g (w=0.7)	0.0550	0.2852	847.6	1.18
Beam-5+8-g (w=0.7)	**0.0534**	**0.2746**	2026.2	0.49
Beam-10+8-g (w=0.7)	0.0547	0.2774	5881.4	0.17
Beam-5+RoBERTa (w=0.2)	0.0550	0.2876	2026.2	0.49

CER: Character error rate, WER: Word error rate. Note: Bold values indicate the best-performing result in the corresponding comparison.

configuration for each decoding strategy. The best overall result was achieved by beam-5 with an 8-gram language model at weight 0.7, yielding a CER of 0.0534 and a WER of 0.2746. Wider beam widths and RoBERTa-based rescoring did not surpass this configuration.

Beyond the proposed training strategies, additional experiments were conducted on the best-performing +AA+SKHL+FHL-50 configuration (CER = 0.0593) to evaluate whether alternative architectural modifications and training objectives could further improve performance. As summarized in Table VIII, none of the tested techniques

http://dx.doi.org/10.14500/aro.12820

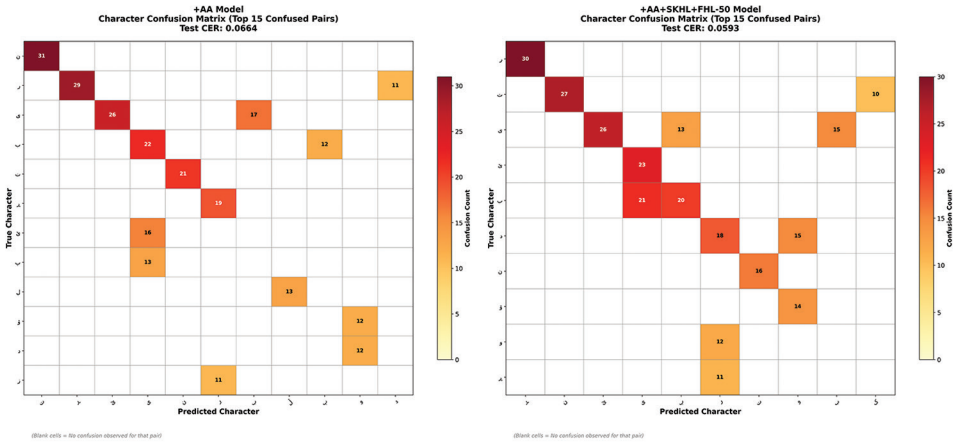


Fig. 8. Confusion matrices of the 15 most frequently misrecognized characters for the +AA and +AA+SKHL+FHL-50 models.

Fig. 9. (a-d) Screenshots of the deployed web-based interface for Kurdish handwritten text recognition.

TABLE VIII
ADDITIONAL EXPLORATORY EXPERIMENTS ON DASTNUS TEST SET

No.	Technique	Description	CER	WER
1	BCA-KSFM	Bidirectional context-aware attention modeling RTL/LTR directions and positional letter forms	0.1510	0.4783
2	Printed Pre-training	Pre-training on synthetic printed Kurdish lines followed by fine-tuning	0.0634	**0.3051**
3	K-STAR	Script-aware transformer with auxiliary position prediction and confusion-aware loss	0.0592	0.3095
4	BV-Decoder	Factorizes characters into base shape and variant heads for separate prediction	0.0590	0.3133
5	HW-Decoder-MT	Multi-task decoder jointly predicting character and word sequences	0.0675	0.3384
6	HW-Decoder-AG	Extends HW-Decoder-MT with agreement loss enforcing character-word consistency	0.0670	0.3354
7	AFA	Adaptive focus attention module for multi-resolution feature extraction in ambiguous regions	0.0602	0.3126
8	DualPath-Net	Parallel decoder paths separately predicting base characters and diacritical marks	0.3240	0.8266
9	MCRL	Multi-level contrastive learning on unigram, bigram, and trigram representations	0.0587	0.3112
10	SCL	Script-specific contrastive loss for similar character separation	**0.0579**	0.3079

Note: Bold values indicate the best-performing result in the corresponding comparison. BCA-KSFM: Bidirectional Context-aware Attention with Kurdish Script-Form Modeling. K-STAR: Kurdish Script-aware Transformer with Auxiliary Refinement, BV-Decoder: Base–Variant Multi-head Decoder, HW-Decoder-MT: Hierarchical Word–Character Multi-Task Decoder, HW-Decoder-AG: Hierarchical Word–Character Agreement Decoder, AFA: Adaptive Focus Attention

achieved a statistically significant improvement over the reference model. The lowest CER values were obtained by Script-specific Contrastive Learning (SCL) (0.0579) and Multi-level Contrastive Representation Learning (MCRL) (0.0587), but paired bootstrap and Wilcoxon signed-rank tests confirmed that these improvements were not statistically significant at $\alpha = 0.05$. These findings are reported to guide future research directions.

D. Ablation Study and Comparison

Ablation experiments were conducted to validate the architectural design choices of the proposed model. Three alternative pre-trained CNN backbones were evaluated against DenseNet-121 for feature extraction, and three encoder-decoder depth configurations were tested. All ablation experiments were performed using the best-performing +AA+SKHL+FHL-50 configuration. Table IX presents the results. DenseNet-121 with 3 encoders and 3 decoder layers achieved the lowest CER among all configurations, confirming the suitability of the selected architecture.

To establish baseline comparisons on the DASTNUS dataset, four well-known recognition architectures were trained and evaluated under identical data conditions. Table X reports the results. Both CRNN and ABINet produced high error rates on DASTNUS, indicating limited suitability for Kurdish cursive script without script-specific adaptation. TrOCR, despite having over 333 million parameters, exhibited severe overfitting with a training CER of 0.0011 against a validation CER of 0.1290 at the final epoch. The gMLP-based model achieved reasonable performance but remained inferior to the proposed architecture.

To assess cross-script transferability, the proposed architecture was pre-trained on DASTNUS and fine-tuned on two external datasets. Table XI presents a comparative analysis with recent Transformer-based handwritten line recognition studies alongside the cross-dataset generalization results of the proposed model.

To evaluate domain adaptation capability, the proposed model was pre-trained on the KHATT Arabic dataset and evaluated on DASTNUS under zero-shot and few-shot conditions with varying proportions of Kurdish training data. Table XII reports the results. Zero-shot transfer from Arabic to Kurdish yielded a CER of 0.6199, reflecting the substantial script-level differences despite shared cursive characteristics. Performance improved progressively with increasing amounts of Kurdish fine-tuning data, reaching a CER of 0.0581 when the full training set was utilized.

TABLE IX
Ablation Study Results on Dastnus Test Set (+AA+SKHL+FHL-50 Configuration)

Configuration	Parameters	Character error rate	Word error rate
CNN backbone			
DenseNet-121 (Proposed)	~12.8M	**0.0593**	**0.3083**
ResNet-18 (He, et al., 2016)	~16.9M	0.0670	0.3335
EfficientNet-B0 (Tan and Le, 2019)	~9.9M	0.0662	0.3247
MobileNetV3-Large (Howard, et al., 2019)	~8.8M	0.0771	0.3579
Encoder-decoder depth			
3E-3D (proposed)	~12.8M	**0.0593**	**0.3083**
2E-2D	~11.0M	0.0595	0.3111
3E-2D	~11.8M	0.0664	0.3328
4E-4D	~14.6M	0.0674	0.3410

Note: Bold values indicate the best-performing result in the corresponding comparison.

V. Discussion and Limitations

The experimental results show that the proposed constrained recipe-based synthetic line generation framework significantly improves Kurdish handwritten line recognition. Starting from a baseline CER of 0.0819, adaptive augmentation reduced the error by 18.9%, while combining synthetic and fixed-content handwritten lines (+AA+SKHL+FHL-50) achieved a total relative CER reduction of 27.6%. Statistical testing further confirmed that combining synthetic and fixed-content handwritten lines produced a significant 10.7% relative CER reduction over adaptive augmentation alone ($p < 0.001$), demonstrating that text diversity and handwriting style diversity play complementary roles in enriching training data.

Baseline comparisons on DASTNUS showed that the proposed DenseNet121-Transformer architecture outperformed CRNN, ABINet, TrOCR, and gMLP under identical conditions while using only 12.8 million parameters. Cross-dataset evaluation demonstrated strong generalization across Arabic-script languages, achieving CRR values of 88.65% on KHATT and 90.68% on PUCIT. In domain adaptation experiments, Arabic pre-training with only 25% DASTNUS fine-tuning data achieved a CER of 0.0892, confirming the effectiveness of cross-script transfer learning in low-resource settings. In addition, the synthetic printed Kurdish pre-training provided limited benefit compared to Arabic handwritten transfer learning, indicating that handwriting-specific representations transfer more effectively than printed text features.

Despite these findings, several limitations remain. First, the proposed system operates on pre-segmented and annotated handwritten lines, whereas real-world applications require end-to-end paragraph recognition that would demand substantially larger volumes of paragraph-level data than currently represented in the DASTNUS dataset. Second, the architecture assumes clean, well-preprocessed images

TABLE X
Baseline Architecture Comparisons on Dastnus Test Set

References	Methodology	Parameters	Character error rate	Word error rate	CRR (%)
(Shi, Bai and Yao, 2016)	CNN-BiLSTM-CTC (CRNN)	~14.3M	0.6375	0.9530	36.25
(Fang, et al. 2021)	Bidirectional iterative LM (ABINet)	~43.2M	0.6076	0.9690	39.24
(Li, et al., 2023)	Pre-trained Transformer (TrOCR)	~334.0M	0.1270	0.4482	87.30
(Bensouilah, Taffar, and Zennir, 2024)	Gated MLP with CNN-BiLSTM	~15.5M	0.0679	0.3486	93.21
Proposed	DenseNet121-Transformer	**~12.8M**	**0.0593**	**0.3083**	**94.07**

Note: Bold values indicate the best-performing result in the corresponding comparison.

TABLE XI
COMPARATIVE ANALYSIS WITH TRANSFORMER-BASED STUDIES AND CROSS-DATASET GENERALIZATION RESULTS

References	Methodology	Dataset	CRR
(Dhiaf, et al., 2023)	MSdocTr-lite transformer	KHATT	86.52
(Bensouilah, Taffar, and Zennir, 2024)	Gated MLP with CNN-BiLSTM	KHATT	87.89
(Momeni and BabaAli, 2024)	Transducer transformer	KHATT	80.24
	Vanilla transformer		81.55
(Anjum and Khan, 2020)	Contextual attention with GRU-DenseNet	PUCIT	77.05
(Anjum and Khan, 2023)	CALText contextual attention	PUCIT	82.06
Proposed	DenseNet121-Transformer	DASTNUS	**94.07**
		KHATT*	**88.65**
		PUCIT*	**90.68**

*The proposed model was pre-trained on DASTNUS before fine-tuning on the respective target dataset. Note: Bold values indicate the best-performing result in the corresponding comparison.

TABLE XII
ZERO-SHOT AND FEW-SHOT DOMAIN ADAPTATION RESULTS (PRE-TRAINED ON KHATT, FINE-TUNED ON DASTNUS)

Condition	Training data (%)	Character error rate	Word error rate
Zero-shot	0	0.6199	1.2474
Few-shot	5	0.2142	0.6526
Few-shot	10	0.1307	0.5034
Few-shot	25	0.0892	0.4149
Few-shot	50	0.0710	0.3568
Full fine-tuning	100	0.0581	0.3098

with standard layouts. Documents exhibiting heavy noise, degradation, or complex structures such as multi-column text or mixed orientations would require additional preprocessing or specialized layout analysis before reliable recognition could be achieved.

VI. Conclusion and Future Works

This study addresses a fundamental research gap in Kurdish handwritten text recognition, a domain of considerable importance for government and public service digitization. A lightweight DenseNet121-Transformer architecture with 12.8 million parameters is proposed for Kurdish handwritten line recognition. A constrained recipe-based synthetic line generation framework is developed to augment the limited training data by concatenating real handwritten word images. Combined with adaptive augmentation and fixed-content handwritten line integration, the proposed training strategy achieved a statistically significant 27.6% relative CER reduction over the baseline, reaching a best CER of 0.0593 without language model assistance. Language model integration through beam search with character-level n-gram models further improved performance to a CER of 0.0534. Baseline comparisons confirmed the superiority of the proposed architecture over CRNN, ABINet, TrOCR, and gMLP on DASTNUS, while cross-dataset evaluation demonstrated effective generalization to Arabic and Urdu handwritten datasets.

Future work will focus on end-to-end paragraph recognition to eliminate the dependency on pre-segmented lines, expanding the dataset with historical manuscripts and degraded document samples, and further exploration of cross-script transfer learning for low-resource Arabic-script languages.

VII. Acknowledgment

This work was supported by Salahaddin University-Erbil, Kurdistan Region, Iraq. The authors thank the university for providing the workstation used in the experiments and express gratitude to all contributors involved in creating the DASTNUS dataset, including the writers and data collection assistants.

VIII. Data Availability

The data used in this study will be made available on request for non-commercial scientific research purposes only.

References

Abdo, H.A., Abdu, A., Al-Antari, M.A., Manza, R.R., Talo, M., Gu, Y.H., and Bawiskar, S., 2024. End-to-end deep learning framework for Arabic handwritten legal amount recognition and digital courtesy conversion. *Mathematics,* 12(14), p.2256.

Abdullah, A.A., Abdulla, S.H., Toufiq, D.M., Maghdid, H.S., Rashid, T.A., Farho, P.F., Sabr, S.S., Taher, A,H., Hamad, D.S., Veisi, H., and Asaad, A.T., 2024. *NER- RoBERTa: Fine-Tuning Roberta for Named Entity Recognition (Ner) Within Low-Resource Languages*. Available from: https://arxiv.org/abs/2412.15252

Ahmed, R.M., Rashid, T.A., Fattah, P., Alsadoon, A., Bacanin, N., Mirjalili, S., Vimal, S., and Chhabra, A., 2022. Kurdish handwritten character recognition using deep learning techniques. *Gene Expression Patterns,* 46, p.119278.

Ali, S.H., and Abdulrazzaq, M.B., 2024a. KurdSet: A Kurdish Handwritten characters recognition dataset using convolutional neural network. *Computers, Materials and Continua,* 79(1), pp.429-448.

Ali, S.H., and Abdulrazzaq, M.B., 2024b. KurdSet handwritten digits recognition based on different convolutional neural networks models. *TEM Journal,* 13(1), pp.221-233.

Alsaqi, I.M., and Fattah, P., 2025. Kurdish Handwritten Word Recognition withCRNN Using Multi-Backbone Evaluation. *Academic Journal of International University of Erbil,* 2(4), pp.439-449.

Anjum, T., and Khan, N., 2020. An Attention Based Method for Offline Handwritten Urdu Text Recognition. In: *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. pp.169-174.

Anjum, T., and Khan, N., 2023. CALText: Contextual attention localization for offline handwritten text. *Neural Processing Letters,* 55(6), pp.7227-7257.

Bensouilah, M., Taffar, M., and Zennir, M.N., 2024. gMLP guided deep networks model for character-based handwritten text transcription. *Multimedia Tools and Applications,* 83(5), pp.13557-13575.

Cascianelli, S., Cornia, M., Baraldi, L., Piazzi, M.L., Schiuma, R., and Cucchiara, R., 2021. Learning to read L'Infinito: Handwritten text recognition with synthetic training data. In: Tsapatsoulis, N., Panayides, A., Theocharides, T., Lanitis, A., Pattichis, C., and Vento, M., Eds. *Computer Analysis of Images and Patterns*. International Publishing, Springer, Cham, pp.340-350.

Coquenet, D., Chatelain, C., and Paquet, T., 2023. DAN: A segmentation-free document attention network for handwritten document recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(7), pp.8227-8243.

Dhiaf, M., Rouhou, A.C., Kessentini, Y., and Salem, S.B., 2023. MSdocTr-Lite:

A lite transformer for full page multi-script handwriting recognition. *Pattern Recognition Letters,* 169, pp.28-34.

Fang, S., Xie, H., Wang, Y., Mao, Z., and Zhang, Y., 2021. Read Like Humans: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp.7098-7107.

Gader, T.B.A., and Echi, A.K., 2022. Attention-based deep learning model for Arabic handwritten text recognition. *Machine Graphics and Vision,* 31(1/4), pp.49-73.

Ganai, A.F., and Khursheed, F., 2023. Computationally efficient recognition of unconstrained handwritten Urdu script using BERT with vision transformers. *Neural Computing and Applications,* 35(34), pp.24161-24177.

Hamad, K.A., and Shareef, S.M., 2025. Towards the digitization of Kurdish Handwritten script using deep learning: A comprehensive analysis. *International Journal on Document Analysis and Recognition (IJDAR)*, 29(2), pp.302-330.

He, K., Zhang, X., Ren, S., and Sun, J., 2016. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770-778.

Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., and Adam, H., 2019. Searching for MobileNetV3. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q., 2017. Densely Connected Convolutional Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp.4700-4708.

Ibrahim, R.M., Saman Nariman, G., and Majeed, H.D., 2023. offline kurdish character handwritten recognition (okchr) using cnn with various preprocessing techniques. *Journal of Engineering Science and Technology,* 18, pp.3113-3127.

Ingle, R.R., Fujii, Y., Deselaers, T., Baccash, J., and Popat, A.C., 2019. A Scalable Handwritten Text Recognition System. In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. pp.17-24.

Hamad, K., and Kaya, M., 2016. A detailed analysis of optical character recognition technology. *International Journal of Applied Mathematics Electronics and Computers,* 4(1), pp.244-249.

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., and Wei, F., 2023. TrOCR: Transformer-based optical character recognition with pre-trained models. *Proceedings of the AAAI Conference on Artificial Intelligence,* 37(11), pp.13094-13102.

Mahmoud, S.A., Ahmad, I., Al-Khatib, W.G., Alshayeb, M., Tanvir Parvez, M., Märgner, V., and Fink, G.A., 2014. KHATT: An open Arabic offline handwritten text database. *Pattern Recognition,* 47(3), pp.1096-1112.

Momeni, S., and BabaAli, B., 2024. A transformer-based approach for Arabic offline handwritten text recognition. *Signal, Image and Video Processing,* 18(4), pp.3053-3062.

Rabaev, I., and Litvak, M., 2025. Recent advances in text line segmentation and baseline detection in historical document images: A systematic review. *International Journal on Document Analysis and Recognition (IJDAR)*, 29, pp.3-39.

Rustay, S., Mohammed, J., Rasool, R., Taha, B., and Khdir, H., 2012. فەرهەنگی هەرزە. First. Available from: https://kurdipedia.org/default.aspx?q=2016022118152294085 [Last accessed on 2024 Aug 19].

Shareef, S.M., and Ali, A.M., 2024. Deep learning-based digitization of Kurdish text handwritten in the e-government system. *Indonesian Journal of Electrical Engineering and Computer Science,* 35(3), p.1865.

Shen, X., and Messina, R., 2016. A Method of Synthesizing Handwritten Chinese Images for Data Augmentation. In: *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. pp.114-119.

Shi, B., Bai, X., and Yao, C., 2016. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(11), pp.2298-2304.

Tan, M., and Le, Q., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Chaudhuri, K., and Salakhutdinov, R., Eds. *Proceedings of the 36th International Conference on Machine Learning*. Proceedings of Machine Learning Research. PMLR, pp.6105-6114. Available from: https://proceedings.mlr.press/v97/tan19a.html. [Last accessed on 2025 Sep 10].

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I., 2017. *Attention is All you Need*. *Advances in Neural Information Processing Systems 30*. Neural Information Processing Systems Foundation, Long Beach.

Veisi, H., MohammadAmini, M., and Hosseini, H., 2020. Toward Kurdish language processing: Experiments in collecting and processing the AsoSoft text corpus. *Digital Scholarship in the Humanities,* 35(1), pp.176-193.