TRISIG-430: A TRI-SCRIPT, MULTI-SESSION, MULTI-DEVICE DATASET TOWARD REAL-WORLD OFFLINE SIGNATURE VERIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Offline signature verification (OffSV) systems in real-world use must remain reliable when a query differs from the enrolled template in script, acquisition device, or time of writing. Existing OffSV datasets are mostly single-script, single-session, and single-device, so they cannot measure robustness to these shifts. We introduce **TriSig-430**, a writer-aligned tri-script offline signature dataset designed for this setting. It contains 76,374 real handwriting images from 430 writers, including 60,286 signature images and 16,088 content-controlled digit-string images. Each writer signs in three scripts: Han, Latin, and Arabic(Uyghur). A paired subset captures the same physical handwriting sheets with both a flatbed scanner and a smartphone, enabling acquisition-domain evaluation without confounding device changes with rewriting. A longitudinal cohort of 178 writers contributes samples from two sessions separated by approximately nine months, providing complementary sparse-enrollment and dense-sampling regimes. We define train-test protocols for matched and cross-script verification, cross-device transfer, few-shot enrollment, cross-session robustness, and content transfer between signatures and digit strings. TriSig-430 is intended as a unified dataset for learning and evaluating writer representations under realistic script, temporal, content, and device shifts.

1 INTRODUCTION

Handwritten signatures are one of the most widely accepted biometric traits. Signature verification decides whether a query signature is genuine by comparing it against a stored template. Depending on the data acquisition method, signature verification systems are categorized into two types: online (dynamic) (Zhang et al., 2025a) and offline (static) (Li et al., 2024). Online systems capture dynamic behavioral features such as pen pressure and velocity using specialized devices, while offline signature verification (OffSV) relies solely on scanned or photographed signature images. Due to the absence of dynamic characteristics, OffSV is inherently more challenging. Nevertheless, given the prevalence of paper-based documents, it remains indispensable across various domains, such as finance and law. This paper focuses on OffSV.

In practice, OffSV systems face substantial variation. For a given individual, a new query signature may differ from the stored template due to different capture devices or long time spans. More broadly, OffSV models should generalize across diverse writing scripts (*e.g.*, Latin and Chinese) used by different populations. Since no pen trajectory or timing information is available, a critical challenge is whether the systems still extract consistent signature representations across these variations (Singla & Mittal, 2025; Engin et al., 2020; Arrigoni, 2025).

Although many OffSV datasets have been published in recent decades, they fall short in diversity and real-world complexity, and therefore fail to evaluate the models' robustness under the aforementioned variations. (1) **Narrow script and linguistic scope.** Most datasets cover only Latin or Han script signatures. A few extend to other systems like the Brahmic script (Pal et al., 2016), but minority languages remain largely overlooked. This prevents signature verification systems from supporting minority languages and hinders automated forensic technologies in these regions. (2) **Absence of intra-writer script diversity.** Prior works restrict each writer to a single script, so models may overfit to specific structural traits rather than learning script-agnostic handwriting traits, limiting cross-script generalization. (3) **Omission of temporal shifts.** A person's signature natu-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

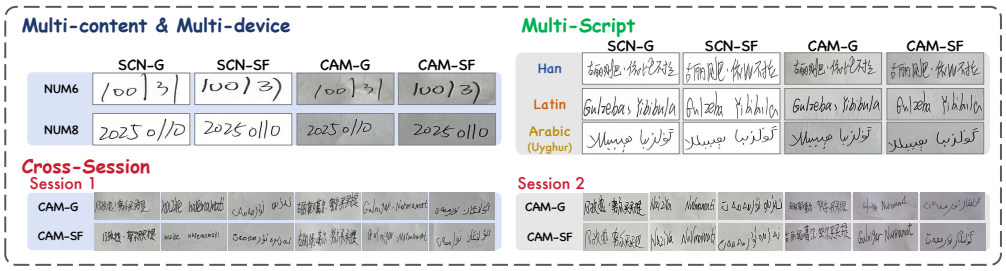


Figure 1: Overview of TriSig-430, a tri-script offline signature verification dataset of 76,374 images from 430 writers. SCN/CAM: scanner/camera; G/SF: genuine sample/skilled forgery. Every writer provides genuine and skilled-forgery samples in Han, Latin, and Arabic (Uyghur) scripts plus content-controlled digit strings. 178 writers are re-collected across two sessions with a nine-month gap. Samples are captured by both scanners and cameras for multi-device evaluation.

rally evolves over time through aging, motor skill changes, and habitual adaptation. Yet, existing datasets collect signatures only within a single session and thus miss intra-writer temporal variation. (4) **Lack of cross-device data.** In practice, signatures are digitized by various devices (*e.g.*, scanners or cameras), each introducing distinct artifacts. Current datasets lack paired cross-device samples, making it impossible to develop and evaluate systems that generalize across acquisition devices.

To address these limitations, we introduce **TriSig-430** (Fig. 1), a large-scale offline signature dataset designed to capture real-world diversity across scripts, time, and acquisition devices. Comprising 76,374 images from 430 writers (including pairwise genuine signatures and skilled forgeries), TriSig-430 introduces several unique properties:

- **Rich intra-writer script diversity.** TriSig-430 collects signatures in three distinct scripts (Han, Latin, and Arabic) from the same writer. Notably, it introduces Uyghur signatures (a right-to-left Arabic script), a language rarely covered by existing signature datasets, filling a critical gap for this minority language. This also enables the evaluation of cross-script robustness.
- **Multi-session acquisition.** We specifically recollect data from 178 writers after a nine-month gap, aiming to evaluate natural signature evolution and models' long-term stability.
- **Device control.** We digitize 19,953 physical signature sheets written using both a flatbed scanner and a smartphone, enabling the isolated investigation of device-induced variations.
- **Cross-domain identification & retrieval.** While standard datasets only support intra-script identification, our multi-script and multi-content design uniquely enables cross-domain writer identification (Hossain et al., 2025) and retrieval (Zhang & Jin, 2024).
- **Additional content-controlled digit string.** We further include same- and different-digit strings to evaluate models' generalization on out-of-distribution content from the same writers.

We define three complementary training splits and five evaluation aspects: within-domain, cross-dataset, cross-device, and cross-session verification, as well as cross-script and cross-content writer identification and retrieval. All settings are evaluated on a single held-out test set, which is writer-disjoint from all training splits. Four key findings emerge: (1) Relative to existing Latin-dominated benchmarks, verification lags well behind, most notably on Uyghur, the most difficult script for most methods; (2) models trained on existing public datasets transfer poorly to TriSig-430; (3) Training on both devices yields the best results in most cross-device scenarios and the best or near-best performance in intra-device settings; and (4) verification performance degrades markedly over nine months. Together, these results highlight the critical gap between existing datasets and real-world complexities, establishing TriSig-430 as a challenging new frontier for offline signature analysis.

2 Related Work

2.1 Offline Signature Dataset

Existing offline signature (OffSV) datasets mostly contain only one writing script. Early datasets, such as CEDAR (Kalera et al., 2004), MCYT (Ortega-Garcia et al., 2003), and GPDS (Vargas et al., 2007), cover Latin signatures. Later datasets add Persian, Arabic, and Han scripts, including UT-Sig (Soleimani et al., 2017), AS-VAR (Balat et al., 2026), ChiSig (Yan et al., 2022), and Han-

Sig (Huang & Lu, 2023). A few datasets cover more than one script. BHSig260 (Pal et al., 2016) spans Bengali and Hindi, but the two scripts come from disjoint signer populations rather than the same writer. MLSig (Cairang et al., 2022) collects three scripts from the same writer but is single-session, with limited samples per writer. Hossain et al. (Hossain et al., 2024) study cross-script writer identification on a Bangla–Roman corpus, extending an earlier Roman–Devanagari dataset in which 80 writers wrote only Roman and a separate, non-overlapping 46 wrote only Devanagari. The authors note that a true same-writer test was not possible because no dataset collected multiple scripts from the same writer (Obaidullah et al., 2021).

Multi-session acquisition has been studied for online handwriting datasets (Tolosana et al., 2020; Zhang et al., 2022; 2025b). For example, MobileTouchDB (Tolosana et al., 2020) collects online digits across six sessions with intervals exceeding two days. MSDS (Zhang et al., 2022) captures data across two sessions with a gap of at least 21 days, though it is restricted to Han script signatures. However, multi-session acquisition remains unexplored for offline signature datasets. Therefore, we construct TriSig-430, which is the first OffSV dataset to combine tri-script coverage, paired devices, and a nine-month multi-session acquisition.

2.2 Offline Signature Verification Method

OffSV methods have moved from hand-crafted descriptors to learned representations (Pandey et al., 2025). CNNs extract stroke and texture cues directly from images (Hafemann et al., 2017a), and Siamese or metric-learning objectives build shared embedding spaces for genuine and forged signatures (Dey et al., 2017; Wei et al., 2019). More recent work explores meta-learning (Hafemann et al., 2019), self-supervised representation learning that avoids skilled forgeries during pretraining (Manna et al., 2022), and prototype-based training data that reduces reliance on random forgeries (de Moura et al., 2026).

A growing line of work adopts hybrid CNN–transformer or attention-based architectures. DetailSemNet preserves fine stroke detail through detail-semantic matching (Shih et al., 2024), HTC-SigNet combines convolution and transformer branches to capture local and global features (Zheng et al., 2025), PAST applies pairwise attention between reference and query signatures (Xiong et al., 2026), and CNN–CrossViT fuses multi-scale convolutional and cross-attention features (Li et al., 2025). These architectural advances target within-script discriminative power rather than cross-script, cross-device, or longitudinal robustness, and reported accuracy also depends on writer partitioning, enrollment size, and impostor construction (Hafemann et al., 2017b; Diaz et al., 2019; Musgrave et al., 2020). Hence, we introduce TriSig-430, a comprehensive dataset designed to evaluate whether architectural advances translate into robustness across script, device, and session shifts, rather than just in-domain accuracy.

3 TriSig-430

TriSig-430 is an offline signature verification (OffSV) dataset that comprises 76,374 real handwriting images from 430 writers, including 60,286 signature images and 16,088 content-controlled digit-string images. Figure. 2 illustrates the data construction pipeline, with details as follows.

3.1 Data Acquisition

We recruit a group of participants from a multilingual region, who are literate in Han, Latin, and Arabic (Uyghur). To guarantee privacy, each writer is assigned with an anonymized writer identifier, and is required to write full names (abbreviated marks not permitted) in three scripts. Samples were digitized with a flatbed scanner (HP LaserJet M227fdw, 600 dpi, yielding 7020×4964-pixel grayscale scans) and a smartphone camera (iPhone 13, 12-megapixel ultra-wide lens, 3412×2407-pixel color images); for a paired subset, the same physical handwriting sheet was digitized by both devices, yielding 19,953 instance-level scanner–camera pairs in which any difference reflects acquisition rather than rewriting. Each target writer was also assigned four or five classmates as forgers, who imitated the target writer's signature and digit strings while viewing reference samples, producing skilled forgeries against a consistent target-writer population.

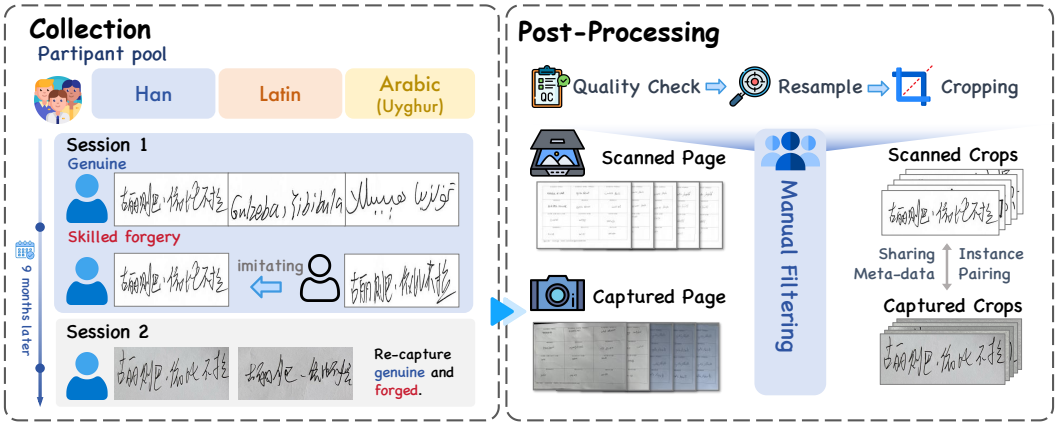


Figure 2: Data construction pipeline of TriSig-430. Left: Participants write genuine signatures in three scripts in Session 1. Skilled forgeries are produced by imitating a reference sample. Partial writers are re-collected about nine months later for Session 2. Right: Sheets are digitized by both scanner smartphone camera, followed by quality checking, resampling, and manual cropping. Crops from the same physical instance share meta-data and are paired across devices.

In addition to natural signatures, participants wrote two types of content-controlled digit-string: an eight-digit date string that was identical for every writer (same-token), and a six-digit string encoding class and seat numbers that varied across writers (different-token). Digit strings were collected alongside matched genuine signatures and skilled forgeries using the same target–forger assignments, which separates writer-specific style from lexical content.

When producing forgery data, forger identities were recorded alongside each skilled-forgery sample, which allows forgery detection to be evaluated against a known target–forger assignment rather than an anonymous impostor pool. To keep signatures comparable within and across sessions, participants were instructed to keep each signature's content, spelling, and letter case consistent every time it was written (the Latin portion in lowercase throughout), to center it horizontally within its grid cell without splitting it across lines, and to stay inside the cell boundary; samples that extended beyond the boundary were discarded.

Before data acquisition, participants were informed that their handwriting would be anonymized and released publicly for signature-verification research, and could withdraw their data before anonymization was finalized. They gave written informed consent. Forgers separately consented to imitate another participant's signature. Only anonymized writer identifiers are released, and the mapping back to participant identities is retained under restricted access.

3.2 Handwriting Annotation and Quality Control

Each signature or digit-string sample on a sheet was localized with an OCR-assisted annotation tool that records its four corner coordinates, and samples captured at an angle were deskewed and padded with a white mask according to their orientation before being cropped into an individual image.

The cropped samples then underwent a three-stage quality-control process. Automated rules first flagged empty or near-duplicate crops and images below a sharpness threshold for review. Two annotators then independently checked writer identity, script, session, device, authenticity, and content metadata against the original collection records. A senior annotator resolved all disagreements using the source sheet and acquisition log. A sample was removed only if severe blur, exact duplication, or a blank region made it unusable; ordinary variation in stroke thickness, illumination, perspective, and writing size was retained as realistic acquisition noise. Table 1 shows a comparison between TriSig-430 and existing OffSV datasets.

3.3 Statistics

Data collection was carried out in two sessions, denoted as Session 1 (S1) and Session 2 (S2). S1 prioritized writer breadth, collecting 3–5 genuine signatures and 3–5 skilled forgeries per script from

Table 1: Comparison of offline handwritten signature datasets. TriSig-430 is a publicly available dataset containing Uyghur signatures (a right-to-left Arabic script). "Script/writer" is the number of writing systems provided by each writer. "OOD Cont." indicates the inclusion of out-of-distribution handwritten content (*e.g.*, content-controlled digit strings in TriSig-430) for generalization evaluation. "SF" denotes skilled forgery. † indicates that 178 writers participated in two acquisition sessions separated by approximately nine months.

Dataset	Writers	# Script/writer	Script Families	# Session	Multi-Device	SF	OOD Cont.	# Image
Single-script, real handwriting								
CEDAR (Kalera et al., 2004)	55	1	Latin	1	✗	✓	✗	2,640
MCYT-75 (Ortega-Garcia et al., 2003)	75	1	Latin	1	✗	✓	✗	2,250
GPDS-960 (Vargas et al., 2007)	881	1	Latin	1	✗	✓	✗	47,574
SigComp2011 (Liwicki et al., 2011)	10	1	Latin	1	✗	✓	✗	1,177
BHSig260 (Pal et al., 2016)	260	1	Brahmic	1	✗	✓	✗	14,040
ChiSig (Yan et al., 2022)	102	1	Han	1	✗	✓	✗	10,242
HanSig (Huang & Lu, 2023)	238	1	Han	1	✗	✓	✗	35,400
AS-VAR (Balat et al., 2026)	70	1	Arabic	1	✗	✓	✗	3,471
Multi-script, script-parallel								
BRS-ID (Hossain et al., 2025)	37	2	Brahmic, Latin	1	✗	✗	✗	3,744
MLSig (Cairang et al., 2022)	200	3	Han, Latin, Tibetan	1	✗	✓	✗	18,000
TriSig-430 (ours)	430	**3**	Han, Latin, Arabic (Uyghur)	**2**†	✓	✓	✓	**76,374**

406 writers (23,926 signature images), together with all 16,088 digit-string images. S2 prioritized per-writer depth, collecting 30 genuine and 30 skilled-forgery signatures per script from 202 writers (36,360 signature images), using the smartphone camera only. The 178 writers who took part in both sessions form a multi-session cohort, with an interval of approximately nine months between S1 and S2; 228 writers appear only in S1, and 24 appear only in S2.

4 Experiments

4.1 Experimental Protocol

Unless otherwise stated, all experiments are conducted on TriSig-430. To assess cross-dataset generalization, we additionally employ two public offline signature datasets, namely CEDAR (Kalera et al., 2004) and BHSig260 (Pal et al., 2016), the latter comprising a Bengali subset and a Hindi subset. Throughout this paper, HAN, LAT, and ARA denote signatures written in the Han, Latin, and Arabic(Uyghur) scripts, respectively, whereas NUM8 denotes the 8-digit date string shared by all writers (same-token) and NUM6 denotes the 6-digit code specific to each writer (different-token).

Dataset splitting. TriSig-430 is contributed by 430 writers, of whom 406 participated in the first session (S1) and 202 in the second session (S2), with 178 writers present in both sessions. We partition the data at the writer level into one held-out test set and three training sets. The test set is writer-disjoint from all training sets, whereas the training sets may overlap with one another.

- **EVAL40** is the only test set and is shared by all evaluation settings within TriSig-430. Among the 178 two-session writers, 86 provide at least five genuine signatures and five skilled forgeries per capture device in S1. EVAL40 consists of 40 of these 86 writers, drawn by stratified sampling with respect to gender and age group.
- **CORE-v1** is the main training set. It consists of the 162 S2 writers that remain after excluding EVAL40, each providing 30 genuine signatures and 30 skilled forgeries per script in S2.
- **DEV-MULTI** is the cross-device training set. It consists of the 366 S1 writers that remain after excluding EVAL40, each providing 3–5 specimens for every combination of script and content type, captured by both a smartphone camera and a scanner.
- **DEV-LONG** is the cross-session training set. It consists of the 138 two-session writers that remain after excluding EVAL40.

Training and testing strategies. All models trained on TriSig-430 are tested on EVAL40, whereas the training set and the reference–query configuration vary with the evaluation setting.

- **Within-domain verification.** Models are trained on CORE-v1 and tested on S2 samples of EVAL40, with references and queries captured in the same session and by the same device.

Table 2: Comparison with different baselines on the TriSig-430 dataset in terms of EER. Models are trained on CORE-v1 and evaluated on EVAL40 Session 2. For each setting and script, the best result is shown in **bold**, and the second-best result is underlined. Dashes denote unavailable results.

Method	Venue	WD			WI		
		HAN	LAT	ARA	HAN	LAT	ARA
General-purpose methods							
ResNet-18 (He et al., 2016)	CVPR'16	10.75	13.39	12.85	20.94	19.87	20.66
MobileNet-V3 (Howard et al., 2019)	ICCV'19	40.50	39.72	40.92	38.83	41.33	42.45
EfficientNet (Tan & Le, 2019)	ICML'19	28.79	33.43	32.75	33.43	38.94	39.07
ViT (Dosovitskiy et al., 2021)	ICLR'21	10.79	15.19	15.50	27.90	38.49	36.72
HorNet (Rao et al., 2022)	NeurIPS'22	30.91	29.96	30.48	31.95	31.10	31.65
ConvNeXt (Liu et al., 2022)	CVPR'22	11.73	12.14	13.61	22.53	27.26	24.60
FasterNet (Chen et al., 2023)	CVPR'23	17.63	17.46	19.65	25.68	27.78	28.07
Offline signature verification methods							
SigNet-F (Hafemann et al., 2017a)	PR'17	8.50	10.57	12.08	–	–	–
Meta-Learning (Hafemann et al., 2019)	T-IFS'20	24.06	25.07	22.77	–	–	–
Stream-HSV (de Moura et al., 2024)	ICPR'24	11.96	13.84	15.13	–	–	–
DCCCAE (Zhao et al., 2025)	PR'25	7.94	10.12	11.62	–	–	–
SigNet (Dey et al., 2017)	arXiv'17	–	–	–	30.12	36.83	39.57
DeepHSV (Li et al., 2019)	ICDAR'19	–	–	–	29.81	36.02	38.74
IDN (Wei et al., 2019)	CVPR'19	–	–	–	29.46	35.45	38.19
DetailSemNet (Shih et al., 2024)	ECCV'24	–	–	–	16.17	19.50	19.75
HTCSigNet (Zhao et al., 2025)	PR'25	**7.51**	**9.54**	**11.05**	**11.33**	**12.89**	**15.87**

- **Writer identification and retrieval.** Models are trained on CORE-v1, where the signatures of a participant in all three scripts are assigned to a single writer identity. Testing uses the S1 samples of all five content types in EVAL40. In cross-script and cross-content trials, references and queries are drawn from different scripts and different content types, respectively.
- **Cross-device verification.** Models are trained on DEV-MULTI and tested on the S1 samples of EVAL40 in two directions: scanner-to-smartphone (scanned references and smartphone-captured queries) and the reverse smartphone-to-scanner direction.
- **Cross-session verification.** Models are trained on DEV-LONG. To isolate inter-session variation from device variation, testing is restricted to EVAL40 samples captured by the same smartphone in both sessions. All queries are drawn from S2: S1-to-S2 trials measure cross-session performance, whereas S2-to-S2 trials provide the within-session reference.
- **Cross-dataset validation.** Models are trained on one of TriSig-430, CEDAR, and BHSig260 and tested on another.

Each shifted setting is reported alongside its matched counterpart, which serves as the reference. Combined domain shifts are evaluated only for list of combinations, since the remaining combinations lack sufficient samples under the above constraints.

Template selection. For each writer, we construct an enrollment template from K genuine reference signatures and evaluate the resulting K-versus-1 setting for $K \in \{K_1, \ldots, K_m\}$. To ensure reproducibility, the references are selected as the first K genuine samples under a fixed ordering.

Metrics. Verification performance is measured by the equal error rate (EER, %). In the writer-dependent (WD) setting, we report the user-specific EER (EER_{user}). An EER is computed separately for each test writer using a writer-specific decision threshold, and the per-writer values are then averaged. In the writer-independent (WI) setting, we report the global EER ($\text{EER}_{\text{global}}$). Writer identification is evaluated using Top-1 and Top-10 accuracy, and writer retrieval is evaluated using Top-1 accuracy and mean average precision (mAP).

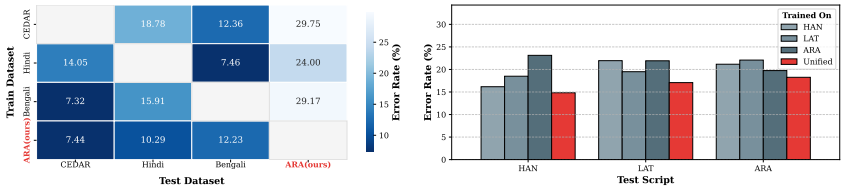
Implementation details. No validation split is used. EVAL40 serves solely for reporting and is not involved in hyperparameter tuning, model selection, or the selection of any operating threshold. Accordingly, all results are obtained with the checkpoint from the final training epoch.

4.2 Data Preprocessing

TriSig-430 images are captured with different devices and thus vary in background intensity, skew, position, and scale. We standardize each image as follows. After converting it to 8-bit grayscale I,

Table 3: Writer identification accuracy under script and content mismatch on TriSig-430. Rows denote the reference content, columns the query content.

Ref / Query	HAN		LAT		ARA		NUM6		NUM8	
	Top-1	Top-10	Top-1	Top-10	Top-1	Top-10	Top-1	Top-10	Top-1	Top-10
HAN	**96.67**	**100.00**	9.25	48.75	6.00	39.75	2.50	30.25	2.50	33.00
LAT	13.25	52.25	**94.72**	**98.89**	3.25	39.75	5.25	28.75	2.75	26.75
ARA	4.00	47.75	1.75	40.75	**90.28**	**96.67**	4.50	29.00	2.50	28.50
NUM6	5.50	33.50	3.75	33.50	3.50	32.75	**72.22**	**94.17**	8.50	43.25
NUM8	3.50	34.50	3.25	32.75	5.00	29.25	8.00	38.25	**56.94**	**86.67**



(a) Cross-Dataset Validation Results (b) Cross-Script Validation Results

Figure 3: Cross-dataset (a) and cross-script (b) validation results (EER, %) of DetailSemNet.

we remove the background using a capped Otsu threshold (Otsu, 1979):

$$t = \min(100,\, t_{\text{Otsu}}), \qquad \tilde{I}(x,y) = \begin{cases} I(x,y), & I(x,y) \le t, \\ 255, & \text{otherwise}, \end{cases} \tag{1}$$

where the cap suppresses light artifacts such as paper texture and shading. We then estimate the skew angle θ with a script-specific estimator and rotate by $-\theta$, padding to a $\lceil \sqrt{H^2 + W^2} \rceil$ square to avoid clipping strokes. Finally, we crop to the foreground bounding box with a 15-pixel margin, invert the intensities so the background is zero, and resize to $H_0 \times W_0$ preserving the aspect ratio.

4.3 Verification Result and Analysis

Table 2 reports the single-script verification results on the S2 signatures of EVAL40 for seven general-purpose backbones and nine OffSV methods.

Writer-dependent and writer-independent verification. Except for MobileNet-V3 on HAN, most models report a higher EER under WI. On LAT, the rise is modest for HTCSigNet (9.54% to 12.89%, 3.35 points) but severe for ViT (15.19% to 38.49%, 23.30 points). Strong within-writer accuracy therefore does not imply writer-independent generalization, and TriSig-430 exposes how this deployment-critical gap varies across scripts.

Script-wise performance. Most OffSV methods reach their lowest EER on HAN and their highest on ARA, so ARA is the hardest of the three scripts. The EER of SigNet, DeepHSV, and IDN grows from 29.46–30.12% on HAN to 38.19–39.57% on ARA. HTCSigNet keeps this cross-script gap within 3.54 points under WD and 4.54 points under WI, yet it never matches its HAN accuracy on ARA. The consistently higher error on Uyghur signatures confirms that this rarely studied right-to-left script is genuinely difficult, underscoring the value of a publicly available Uyghur corpus for OffSV research.

General-purpose versus OffSV methods. General-purpose backbones can match or exceed earlier OffSV methods, but their behavior varies widely across architectures. ResNet-18 surpasses SigNet, DeepHSV, and IDN under WI, whereas MobileNet-V3, EfficientNet, and HorNet stay above 30% EER on most scripts. HTCSigNet obtains the best result on every script and both protocols. This consistent advantage in cross-writer and cross-script robustness is why we adopt it as the primary baseline in the following experiments.

Table 4: Writer retrieval under script and content mismatch on TriSig-430. Rows denote the reference content, columns the query content.

Ref / Query	HAN		LAT		ARA		NUM6		NUM8	
	Top-1	mAP	Top-1	mAP	Top-1	mAP	Top-1	mAP	Top-1	mAP
HAN	**100.00**	**96.47**	9.00	13.67	8.25	10.49	2.75	7.37	3.50	7.33
LAT	11.75	12.74	**100.00**	**94.78**	2.50	7.31	5.00	8.92	3.25	8.03
ARA	3.75	8.87	3.25	7.61	**99.50**	**90.78**	2.75	7.63	3.00	6.33
NUM6	5.00	6.80	6.50	7.12	5.00	7.51	**99.25**	**74.58**	8.75	10.71
NUM8	4.50	8.42	3.50	6.79	4.50	7.24	11.25	10.31	**98.25**	**60.60**

Table 5: Cross-device evaluation across three scripts. CAM: camera; SCN: scanner; Ref: reference. "Both" denotes training with camera and scanner images.

Script	Train	Ref: CAM		Ref: SCN	
		Query: CAM	Query: SCN	Query: CAM	Query: SCN
HAN	CAM	34.50	36.50	**33.00**	30.50
	SCN	**30.00**	35.00	35.00	29.50
	Both	32.00	**31.50**	33.50	**29.00**
LAT	CAM	**29.50**	35.50	31.50	34.50
	SCN	30.00	37.00	35.00	31.50
	Both	32.50	**34.50**	**30.00**	**28.50**
ARA	CAM	40.50	38.00	37.50	36.50
	SCN	37.00	39.00	**34.50**	34.00
	Both	**35.50**	**35.50**	36.50	**33.00**

4.4 Cross-Dataset and Cross-Script Generalization

Figure 3 studies the cross-dataset (a) and cross-script (b) generalization of DetailSemNet (Shih et al., 2024). For clarity, we take the cross-dataset scores between CEDAR and BHSig260 from the original DetailSemNet paper, whereas we produce all results involving the ARA subset of TriSig-430 and all cross-script results.

Public benchmarks transfer poorly to Uyghur signatures. The left heatmap (Fig. 3a) reveals a large, asymmetric gap on the ARA (Uyghur) subset. Models trained on CEDAR, BHSig260-B (Bengali), and BHSig260-H (Hindi) reach EERs of 29.75%, 24.00%, and 29.17% on ARA, whereas a model trained on ARA scores 14 to 22 points lower on the same benchmarks (7.44%, 10.29%, and 12.23%). Existing Latin- and Brahmic-centric corpora thus fail to cover our Uyghur signatures, marking the ARA subset as a harder testbed and supporting our finding that public datasets do not transfer to the new scripts in TriSig-430.

Writer-aligned scripts expose a large cross-script gap. The right chart (Fig. 3b) shows the same pattern within TriSig-430: single-script models are accurate on their matched script (the diagonal) but degrade sharply on unseen scripts, such as HAN or LAT to ARA. Because every writer contributes all three scripts, this comparison isolates script shift from writer identity, which is impossible when each writer uses only one script. This confirms our motivation that single-script training overfits to script-specific structure rather than script-agnostic style.

Multi-script training improves cross-script robustness. Unified training on all three scripts gives more balanced results than any single-script model, shrinking the cross-script gap without hurting matched-script accuracy. The writer-aligned tri-script design is therefore not just broader coverage but a training signal for transferable writer representations, validating intra-writer script diversity as a central contribution.

4.5 Cross-content Writer Identification and Retrieval

TriSig-430 contains signatures in tri-scripts and content types from the same writers, which makes it possible to evaluate writer recognition when the reference and query have different content (Tables 3 and 4). When the content matches, performance is high: Top-1 identification accuracy reaches 90.28–96.67% on the three scripts, and Top-1 retrieval accuracy reaches 98.25–100.00%. Notably,

Table 6: Cross- and within-session performance under different training configurations and values of K (Sec. 4.1). Ref denotes the reference session. S1 and S2 denote Session 1 and Session 2, respectively. "Both" denotes training with data from both sessions.

Ref → Query	Train	HAN			LAT			ARA		
		$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$	$K=1$	$K=3$	$K=5$
S1 → S2	S1	38.22	37.14	34.97	44.51	45.51	39.47	**40.10**	**39.02**	35.72
	S2	36.22	34.47	34.05	**40.10**	**39.93**	**35.22**	40.93	41.26	36.64
	Both	**34.55**	**33.64**	**32.14**	46.76	45.01	39.55	40.77	39.85	**35.47**
S2 → S2	S1	36.89	29.89	26.81	36.69	33.11	30.37	32.95	30.20	26.79
	S2	30.14	25.65	24.15	**30.62**	**24.88**	**23.13**	33.19	28.70	26.37
	Both	**28.56**	**22.06**	**20.32**	38.69	34.11	33.61	**27.37**	**25.46**	**23.79**

NUM8 is harder than NUM6 even when the content matches (56.94% vs. 72.22% Top-1 identification accuracy, 60.60% vs. 74.58% mAP). Because NUM8 is content-controlled, all writers write the same digit string, so the model cannot tell writers apart by what they wrote and must rely on writing style alone. When the content differs, performance falls sharply. Top-1 identification accuracy is at most 13.25%, and retrieval mAP is at most 13.67%. Even between NUM6 and NUM8, Top-1 identification accuracy stays below 9%. This suggests that current representations still mix up a writer's identity with what is written. These results show that cross-content writer recognition is an open and challenging problem, and that TriSig-430 provides a dataset for it.

4.6 Cross-Device Evaluation

Table 5 presents cross-device evaluation results across three scripts under various device configurations. A noticeable domain gap exists: models trained on a single device (*e.g.*, Train: CAM) suffer performance drops when evaluated on cross-device pairs (*e.g.*, Ref: CAM / Query: SCN). However, joint training mitigates this discrepancy. Training on both devices ("Both") mostly yields the best results in cross-device scenarios and achieves optimal or near-optimal performance in intra-device settings (*e.g.*, Ref: SCN / Query: SCN). This proves that increased device diversity during training significantly enhances model generalizability. These trends hold across all scripts, though ARA exhibits higher overall error rates, indicating it remains challenging in cross-device conditions.

4.7 Cross-Session Evaluation

Table 6 presents the cross- and within-session performance under different training configurations and varying values of K. First, increasing K consistently improves results across all scripts and configurations, aligning with the common observation that more templates enhance verification performance. Second, we observe a clear session gap: training on the target session (S2) significantly outperforms cross-session training (S1) in S2 → S2 evaluations. Finally, joint training ("Both") proves highly robust, achieving optimal performance in most HAN and ARA scenarios. Although LAT slightly favors target-session training, combining session data generally mitigates session-related discrepancies and enhances model generalizability.

5 Conclusion

We introduce **TriSig-430**, an offline signature verification (OffSV) dataset of 430 writers, each writing in Han, Latin, and Arabic(Uyghur), a script rarely covered by existing OffSV datasets. It comprises 76,374 images (60,286 signatures and 16,088 content-controlled digit strings) and uniquely combines intra-writer tri-script diversity, multi-session acquisition (178 writers re-collected after nine months), and paired scanner–camera samples (19,953 pairs). Using three complementary training splits and five evaluation aspects on a writer-disjoint test set, our benchmark reveals four findings: verification performance is weakest on Uyghur, the most challenging script for most methods; models trained on existing public datasets transfer poorly to TriSig-430; training on both devices yields the best results in most cross-device scenarios while remaining best or near-best within each device; and verification performance degrades markedly over nine months. These results highlight the gap between existing datasets and real-world complexity, establishing TriSig-430 as a challenging new dataset for robust OffSV across scripts, devices, and time.

Ethics Statement

TriSig-430 was collected from volunteer participants who gave written informed consent after being told that their handwriting would be anonymized and released publicly for signature-verification research. Participants who acted as forgers separately consented to imitate another participant's signature, and any participant could withdraw their data before anonymization was finalized. We release only anonymized writer identifiers and no personally identifiable information; the mapping to participant identities is kept under restricted access. The dataset is distributed under a non-commercial research license (CC BY-NC 4.0). Because TriSig-430 contains skilled forgeries, it is intended solely for research on signature verification and its robustness, and must not be used to produce fraudulent signatures or for any other unlawful purpose. The Uyghur portion is included to improve script coverage for an underrepresented language, and all scripts were collected under the same consent and anonymization procedure.

AI use statement

In this work, we used generative AI tools to assist with language editing, environment configuration, and data processing during the reproduction of baseline methods. The authors reviewed and verified all AI-assisted outputs. We take full responsibility for the final content of this work.

Reproducibility Statement

The main text describes TriSig-430 and its experimental setup, including data acquisition and quality-control procedures, writer-disjoint data splits, the preprocessing pipeline, and evaluation protocols and metrics. All writer-level partitions, template selections, and trial lists are fixed and released with the dataset. TriSig-430 will be publicly released for research use, distributing only anonymized writer identifiers, consistent with the informed consent obtained from participants. We will publicly release the implementation code, experiment configurations, and instructions for reproducing the reported results.

References

Valentina Arrigoni. Offline signature verification in the banking domain. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, pp. 1264–1273. IEEE, 2025.

Mazen Balat, Esraa Mamdouh, Rewaa Awaad, Mohamed Elhoseny, and Ahmed M Anter. Signatureguard: hybrid cnn–transformer model for signature verification and identification across arabic and english datasets. *Scientific Reports*, 2026.

Xianmu Cairang, Duoji Zhaxi, Xiaolong Yang, Yan Hou, Qijun Zhao, Dingguo Gao, Danzeng Pubu, and Dorji Gesang. Learning generalisable representations for offline signature verification. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7. IEEE, 2022.

Jierun Chen, Shiu-hong Kao, Hao He, Weipeng Zhuo, Song Wen, Chul-Ho Lee, and S-H Gary Chan. Run, don't walk: chasing higher flops for faster neural networks. In *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 12021–12031. IEEE, 2023.

Kecia G de Moura, Robert Sabourin, and Rafael MO Cruz. Protosig: Enhancing training data for offline handwritten signature verification using prototypical signatures. *Pattern Recognition*, pp. 114258, 2026.

Kecia Gomes de Moura, Rafael Menelau O Cruz, and Robert Sabourin. Offline handwritten signature verification using a stream-based approach. In *International Conference on Pattern Recognition*, pp. 271–286. Springer, 2024.

Sounak Dey, Anjan Dutta, J Ignacio Toledo, Suman K Ghosh, Josep Lladós, and Umapada Pal. Signet: Convolutional siamese network for writer independent offline signature verification. *arXiv preprint arXiv:1707.02131*, 2017.

Moises Diaz, Miguel A Ferrer, Donato Impedovo, Muhammad Imran Malik, Giuseppe Pirlo, and Réjean Plamondon. A perspective analysis of handwritten signature technology. *Acm Computing Surveys (Csur)*, 51(6):1–39, 2019.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Deniz Engin, Alperen Kantarcı, Seçil Arslan, and Hazım Kemal Ekenel. Offline signature verification on real-world documents. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3518–3526. IEEE, 2020.

Luiz G Hafemann, Robert Sabourin, and Luiz S Oliveira. Learning features for offline handwritten signature verification using deep convolutional neural networks. *Pattern Recognition*, 70:163–176, 2017a.

Luiz G Hafemann, Robert Sabourin, and Luiz S Oliveira. Offline handwritten signature verification—literature review. In *2017 seventh international conference on image processing theory, tools and applications (IPTA)*, pp. 1–8. IEEE, 2017b.

Luiz G Hafemann, Robert Sabourin, and Luiz S Oliveira. Meta-learning for fast classifier adaptation to new users of signature verification systems. *IEEE Transactions on Information Forensics and Security*, 15:1735–1745, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Sk Golam Sarowar Hossain, Mridul Ghosh, Sk Md Obaidullah, and Kaushik Roy. Writer identification using cross-script signature images. In *Congress on Intelligent Systems*, pp. 45–55. Springer, 2024.

Sk Golam Sarowar Hossain, Mridul Ghosh, Tonmoy Mete, Maria Zdimalova, Kaushik Roy, and Sk Md Obaidullah. Brs-id: A new bangla-roman signature dataset for benchmarking writer identification with adt-net framework in cross-script scenarios. In *2025 International Conference on Emerging Technologies in Electronics and Green Energy (ICETEG)*, pp. 1–9. IEEE, 2025.

Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314–1324, 2019.

Fu-Hsien Huang and Hsin-Min Lu. Multiscale feature learning using co-tuplet loss for offline handwritten signature verification. *arXiv preprint arXiv:2308.00428*, 2023.

Meenakshi K Kalera, Sargur Srihari, and Aihua Xu. Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(07):1339–1360, 2004.

Chuang Li, Feng Lin, Zhiyong Wang, Gang Yu, Liou Yuan, and Haiqiang Wang. Deephsv: User-independent offline signature verification using two-channel cnn. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 166–171. IEEE, 2019.

Huan Li, Ping Wei, Zeyu Ma, Changkai Li, and Nanning Zheng. Transosv: Offline signature verification with transformers. *Pattern Recognition*, 145:109882, 2024.

W. Li, M. Muhammat, X. Xu, A. Aysa, and K. Ubul. Multi-scale cnn-crossvit network for offline handwritten signature recognition and verification. *Complex & Intelligent Systems*, 11:400, 2025.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11966–11976. IEEE, 2022.

Marcus Liwicki, Muhammad Imran Malik, C Elisa Van Den Heuvel, Xiaohong Chen, Charles Berger, Reinoud Stoel, Michael Blumenstein, and Bryan Found. Signature verification competition for online and offline skilled forgeries (sigcomp2011). In *2011 International conference on document analysis and recognition*, pp. 1480–1484. IEEE, 2011.

Siladittya Manna, Soumitri Chattopadhyay, Saumik Bhattacharya, and Umapada Pal. Swis: Self-supervised representation learning for writer independent offline signature verification. In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 1411–1415. IEEE, 2022.

Kevin Musgrave, Serge Belongie, and Ser-Nam Lim. A metric learning reality check. In *European Conference on Computer Vision*, pp. 681–699. Springer, 2020.

Sk Md Obaidullah, Mridul Ghosh, Himadri Mukherjee, Kaushik Roy, and Umapada Pal. Automatic signature-based writer identification in mixed-script scenarios. In *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 364–377. Springer, 2021.

Javier Ortega-Garcia, Julian Fierrez-Aguilar, Danilo Simon, Joaquin Gonzalez, Marcos Faundez-Zanuy, Vicente Espinosa, Anna Satue, I Hernaez, J-J Igarza, Carles Vivaracho, et al. Mcyt baseline corpus: a bimodal biometric database. *IEE Proceedings-Vision, Image and Signal Processing*, 150(6):395–401, 2003.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

Srikanta Pal, Alireza Alaei, Umapada Pal, and Michael Blumenstein. Performance of an off-line signature verification method based on texture features on a large indic-script signature dataset. In *2016 12th IAPR workshop on document analysis systems (DAS)*, pp. 72–77. IEEE, 2016.

Gaurav Kumar Pandey, V. Raj, A. Agarwal, M. Dixit, S. Chauhan, and Sumit Srivastava. Offline signature verification: An extensive survey of deep learning methods. In *International Conference on Sentiment Analysis and Deep Learning (ICSADL)*. IEEE, 2025.

Yongming Rao, Wenliang Zhao, Yansong Tang, Jie Zhou, Ser Nam Lim, and Jiwen Lu. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *Advances in Neural Information Processing Systems*, 35:10353–10366, 2022.

Meng-Cheng Shih, Tsai-Ling Huang, Yu-Heng Shih, Hong-Han Shuai, Hsuan-Tung Liu, Yi-Ren Yeh, and Ching-Chun Huang. Detailsemnet: elevating signature verification through detail-semantic integration. In *European Conference on Computer Vision*, pp. 449–466. Springer, 2024.

Aman Singla and Ajay Mittal. Exploring offline signature verification techniques: A survey based on methods and future directions. *Multimedia Tools and Applications*, 84(6):2835–2875, 2025.

Amir Soleimani, Kazim Fouladi, and Babak N Araabi. Utsig: A persian offline signature dataset. *IET Biometrics*, 6(1):1–8, 2017.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PmLR, 2019.

Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, and Javier Ortega-Garcia. Biotouchpass2: Touchscreen password biometrics using time-aligned recurrent neural networks. *IEEE Transactions on Information Forensics and Security*, 15:2616–2628, 2020.

Francisco Vargas, M Ferrer, Carlos Travieso, and J Alonso. Off-line handwritten signature gpds-960 corpus. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pp. 764–768. IEEE, 2007.

Ping Wei, Huan Li, and Ping Hu. Inverse discriminative networks for handwritten signature verification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5757–5765. IEEE, 2019.

Yu-Jie Xiong, Jian-Xin Ren, Dong-Hai Zhu, Xi-Jiong Xie, and Xi-He Qiu. Past: Pairwise attention swin transformer for offline signature verification: Y.-j. xiong et al. *International Journal on Document Analysis and Recognition (IJDAR)*, 29(1):137–149, 2026.

Kaihong Yan, Ying Zhang, Haoran Tang, Chengkai Ren, Jian Zhang, Gaoang Wang, and Hongwei Wang. Signature detection, restoration, and verification: A novel chinese document signature forgery detection benchmark. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 5159–5168. IEEE, 2022.

Peirong Zhang and Lianwen Jin. Online Writer Retrieval with Chinese Handwritten Phrases: A Synergistic Temporal-Frequency Representation Learning Approach. *IEEE Transactions on Information Forensics and Security*, 19:10387–10399, 2024.

Peirong Zhang, Jiajia Jiang, Yuliang Liu, and Lianwen Jin. MSDS: A Large-Scale Chinese Signature and Token Digit String Dataset for Handwriting Verification. In *Advances in Neural Information Processing Systems*, volume 35, pp. 36507–36519, 2022.

Peirong Zhang, Kai Ding, and Lianwen Jin. Capturing More: Learning Multi-Domain Representations for Robust Online Handwriting Verification. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 1471–1479, 2025a.

Peirong Zhang, Yuliang Liu, Songxuan Lai, Hongliang Li, and Lianwen Jin. Privacy-Preserving Biometric Verification with Handwritten Random Digit String. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):3049–3066, 2025b.

Xingbiao Zhao, Lidong Zheng, Panli Yuan, and Yuchen Zheng. Fusing deep and hand-crafted features by deep canonically correlated contractive autoencoder for offline signature verification. *Pattern Recognition*, 168:111834, 2025.

Lidong Zheng, Da Wu, Shengjie Xu, and Yuchen Zheng. Htcsignet: A hybrid transformer and convolution signature network for offline signature verification. *Pattern Recognition*, 159:111146, 2025.