applied sciences

MDPI

Article

Automated Writer and Acquisition-Condition Classification of Digitally Captured Handwriting Using Statistical Dynamic Features and Support Vector Machines

Long-Huang Tsai [1], Hsiang-Ju Lai [2], Wen-Chao Yang [3], Jiajun Jiang [4,*] and Chung-Hao Chen [4,*]

[1] Forensic Science Section, Keelung City Police Bureau, Keelung City 201201, Taiwan; fs821213@klg.gov.tw
[2] Forensic Science Division, Ministry of Justice Investigation Bureau, New Taipei City 231209, Taiwan; raylin@mjib.gov.tw
[3] Department of Forensic Science, Central Police University, Taoyuan City 333322, Taiwan; una135@mail.cpu.edu.tw
[4] Electrical and Computer Engineering Department, Old Dominion University, Norfolk, VA 23529, USA
* Correspondence: jjian001@odu.edu (J.J.); cxchen@odu.edu (C.-H.C.)

Abstract

Digitally captured handwriting preserves pen trajectories and dynamic signals, but it also records hardware- and input-dependent properties that can confound forensic interpretation. This study revises a support vector machine (SVM) screening framework using 16,500 samples from 30 writers, 11 writing-content categories, and five acquisition conditions spanning three tablets and stylus or finger input. Twenty-four raw and derived time-series variables were summarized by maximum, minimum, mean, median, and standard deviation, yielding 120 features; the mode statistic was removed. Writing direction and angular velocity were recalculated with atan2-based vector formulas. Unavailable device/API channels were encoded as zero, and Z-score parameters were estimated only from training folds. Writer and content evaluations used rotating pooled-"other" categories as rejection-class proxies, whereas acquisition-condition classification remained closed-set. Every outer five-fold split contained an inner five-fold forward-selection loop; RBF-SVM hyperparameters were fixed a priori (C = 1.0, gamma = scale, balanced class weights, and random seed 42). Writer classification achieved 94.85% accuracy (descriptive 95% CI: 94.69–95.01%) and 86.44% pooled-other recall. Content classification achieved 96.64% accuracy (95% CI: 95.74–97.53%) and 98.74% pooled-other recall. Acquisition-condition classification achieved 99.99% accuracy (99.98–100.00%), with one error among 16,500 outer-test predictions. The acquisition result is interpreted primarily as evidence that channel availability and device-specific measurement scales are strongly encoded in the feature space. Because the folds were sample-level, the samples were collected contemporaneously, and pooled-other writers were represented during training, these results do not establish session-disjoint, cross-device writer, or strict open-set generalization. The proposed workflow should therefore be regarded as an experimental triage aid that supports, rather than replaces, examiner-led comparison.

Keywords: digitally captured handwriting; forensic document examination; support vector machine; dynamic handwriting features; feature selection; pooled-other rejection; acquisition condition; writer classification

Check for updates

Academic Editor: Christos Bouras

Received: 14 July 2026
Revised: 29 August 2026
Accepted: 30 August 2026
Published: 1 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Electronic signing and handwriting workflows increasingly occur on tablets, mobile devices, and signature pads. A digitally captured signature or handwriting (DCS) sample is not merely a static image: it is a time-ordered record that may include X–Y coordinates, pressure, timestamps, stroke events, and stylus orientation. These signals can represent aspects of writing production that are unavailable in scanned handwriting and can therefore provide useful comparison information [1,2].

Forensic guidance nevertheless requires special treatment of digitally captured material. The ENFSI best-practice manual and ISO/IEC 19794-7 emphasize capture-process documentation, data preservation, and interpretation of time-series channels [1,2]. ANSI/ASB Standard 070 retains the questioned–known comparison framework for handwritten items [3]. Recent forensic guidance further emphasizes that dynamic measurements are useful only when the limitations of non-original and digitally captured writing are explicitly considered [4]. The previously cited ASTM E2290 [5] guide is therefore not used as a current authority in this revision.

DCS measurements are shaped by both the writer and the acquisition system. Device-specific sampling frequency, coordinate resolution, pressure range, latency, screen friction, and stylus sensitivity can alter observed trajectories [6–8]. Cross-device signature experiments have reported degraded genuine–impostor separation when enrollment and testing devices differ [9]. Finger and stylus input can also produce different dynamic profiles; sensor-fusion studies have shown strong input-modality effects and have used ablation analysis to quantify the contribution of individual channels [10]. Accordingly, a high acquisition-condition classification rate may reflect hardware scales or channel availability rather than neuromotor behavior.

Automatic online signature and handwriting research spans handcrafted statistical features, distance-based matching, and deep representation learning. Earlier surveys and DeepSign established the importance of temporal modeling [11,12]. Writer-independent Siamese networks, transformers, graph-attention models, and hybrid CNN–recurrent architectures can learn rich sequence representations [13–16]. Statistical and handcrafted-feature approaches remain relevant. Shape-context and function-feature combinations have been used effectively for online signature verification [17], while nonparametric time-series quantifiers combined with SVM or gradient boosting also perform competitively [18]. Pressure can be highly discriminative [19], and feature-vector splitting has shown that coordinate-pressure fusion is often beneficial, while stylus-angle channels have mixed effects [20]. Rank-based aggregation methods such as reciprocal rank fusion can further provide a simple mechanism for combining feature-selection evidence across repeated evaluations [21]. Mobile-device CNN systems [22], augmentation studies [23], and multimodal consistency learning [24] further illustrate the breadth of the current verification research. Offline writer-independent and explainable systems provide complementary evidence about generalization and interpretability but are not directly comparable to the present online multiclass task [25–28].

2. Materials and Methods

2.1. Dataset and Acquisition Conditions

The study used the DCS dataset established by Lai et al. [29]. Thirty adult volunteers provided samples in a contemporaneous acquisition period to limit short-term changes in writing condition. This design reduces temporal variability but prevents session-disjoint evaluation; handwriting can vary with time, fatigue, illness, emotional state, and other conditions [1,30]. The institutional review board number was FUJ-IRB NO: C113187. Each writer produced 11 content categories, repeated 10 times under each of five acquisition

conditions, yielding $30 \times 11 \times 10 \times 5 = 16{,}500$ samples. The writing area was approximately 100×50 mm, and custom iOS 26 and Android 14 applications recorded the time series. The raw DCS channels and their measurement units are summarized in Table 1.

Table 1. Raw DCS channels and measurement units.

Raw Channel	Unit	Definition
Timestamp	s	Elapsed time from the first sampled point
LocationX	pixel	Horizontal pen-tip or contact position
LocationY	pixel	Vertical pen-tip or contact position
Pressure	device/API-specific unit	Pressure value when exposed by the device API
Stroke	count/index	Stroke order from contact-down to contact-up events
Azimuth angle	radian	Projection angle of the stylus on the tablet surface when available
Altitude angle	radian	Angle between the stylus shaft and tablet surface when available

The overall experimental design and acquisition conditions are summarized in Table 2.

Table 2. Experimental design and acquisition conditions.

Component	Description
Participants	30 volunteer writers
Writing contents	11 categories: Chinese uppercase-number (BCN); two Chinese organization names (CG1–CG2); three Chinese names (CN1–CN3); three English names (EN1–EN3); alphanumeric ID (NID); and numeric string (NUM)
Acquisition conditions	iPad mini + Apple Pencil (IPMS); iPad Pro + Apple Pencil Pro (IPPS); iPad Pro + finger (IPPF); Samsung Galaxy Tab S10 Ultra + S-Pen (SPPS); Samsung Galaxy Tab S10 Ultra + finger (SPPF)
Repetitions	10 per writer, content category, and acquisition condition
Total samples	16,500
Software	Python 3.12.8; scikit-learn 1.8.0; NumPy 1.26.4; pandas 2.3.1

2.2. Preprocessing, Derived Variables, and Statistical Summaries

For each sample, seven raw channels were combined with 17 derived time-series variables for a total of 24 variables. Let t_n, x_n, y_n, and p_n denote timestamp, horizontal coordinate, vertical coordinate, and pressure at point n. The position radius is $r_n = \sqrt{x_n^2 + y_n^2}$. Adjacent changes are $\Delta x_n = x_{n+1} - x_n$, $\Delta y_n = y_{n+1} - y_n$, $\Delta r_n = \sqrt{\Delta x_n^2 + \Delta y_n^2}$, and $\Delta p_n = p_{n+1} - p_n$. Velocity and acceleration are calculated from the recorded time increments. No temporal resampling or smoothing stage was included in the documented pipeline; derivative-based findings are therefore interpreted as sensitive to sampling noise and timing irregularity.

The direction angle is a signed two-argument inverse tangent. Angle change is computed directly from the cross and dot products of consecutive displacement vectors, which avoids the quadrant ambiguity of a one-argument arctangent and removes the need to subtract wrapped absolute angles:

$$\theta_n = \text{atan2}(\Delta y_n, \Delta x_n), \tag{1}$$

$$\Delta\theta_n = \text{atan2}(\Delta x_n \Delta y_{n+1} - \Delta y_n \Delta x_{n+1},\ \Delta x_n \Delta x_{n+1} + \Delta y_n \Delta y_{n+1}), \tag{2}$$

$$\omega_n = \frac{\Delta\theta_n}{t_{n+1} - t_n}. \tag{3}$$

These definitions follow the function-feature framework in online signature analysis [17] while explicitly preserving signed angular change. Each of the 24 variables was summarized by five statistics: maximum, minimum, arithmetic mean, median, and standard deviation. The mode was removed because its value for continuous measurements depends on an unreported binning or discretization rule. The resulting vector therefore contained $24 \times 5 = 120$ features. The complete set of 24 raw and derived time-series variables is summarized in Table 3.

Table 3. Complete list of the 24 raw and derived time-series variables. Ratios are evaluated where denominators are nonzero; undefined or unavailable values are represented as zero in the implemented feature matrix.

Group	Variable	Definition	Interpretation
Raw	timestamp	t_n	Elapsed acquisition time
Raw	locationX	x_n	Horizontal position
Raw	locationY	y_n	Vertical position
Position	displacement	$r_n = \sqrt{x_n^2 + y_n^2}$	Distance from the coordinate origin
Position	change_locationX	$\Delta x_n = x_{n+1} - x_n$	Horizontal increment
Position	change_locationY	$\Delta y_n = y_{n+1} - y_n$	Vertical increment
Position	change_displacement	$\Delta r_n = \sqrt{\Delta x_n^2 + \Delta y_n^2}$	Step length
Pressure	pressure	p_n	Recorded pressure or zero when unavailable
Pressure	change_pressure	$\Delta p_n = p_{n+1} - p_n$	Pressure increment
Velocity	x_velocity	$v_{x,n} = \Delta x_n / \Delta t_n$	Horizontal velocity
Velocity	y_velocity	$v_{y,n} = \Delta y_n / \Delta t_n$	Vertical velocity
Velocity	total_velocity	$v_n = \sqrt{v_{x,n}^2 + v_{y,n}^2}$	Speed magnitude
Acceleration	x_acceleration	$a_{x,n} = (v_{x,n+1} - v_{x,n}) / \Delta t_n$	Horizontal acceleration
Acceleration	y_acceleration	$a_{y,n} = (v_{y,n+1} - v_{y,n}) / \Delta t_n$	Vertical acceleration
Acceleration	total_acceleration	$a_n = \sqrt{a_{x,n}^2 + a_{y,n}^2}$	Acceleration magnitude
Acceleration	centripetal_acceleration	$a_{c,n} = (v_{x,n} a_{y,n} - v_{y,n} a_{x,n}) / v_n$	Signed normal component
Angle	cos_alpha	$\Delta x_n / \Delta r_n$	Cosine of local path direction
Angle	sin_alpha	$\Delta y_n / \Delta r_n$	Sine of local path direction
Angle	cos_beta	$v_{x,n} / v_n$	Cosine between x-velocity and speed vector
Angle	theta	$\theta_n = \text{atan2}(\Delta y_n, \Delta x_n)$	Signed local direction
Angle	angle_velocity	$\omega_n = \Delta\theta_n / \Delta t_n$	Signed angular velocity
Raw	stroke	s_n	Stroke count/index
Raw	azimuthAngle	az_n	Stylus azimuth when available
Raw	altitudeAngle	alt_n	Stylus altitude when available

Unavailable channels were encoded as zero. This includes stylus-orientation channels for finger input and any pressure-related channels not exposed by a device/API condition. Coordinates remained in pixels and pressure remained in device-specific units. This choice intentionally preserves acquisition-condition information but creates a potential confound for writer and content analysis. No common-channel-only ablation or physical

unit calibration was available, so the acquisition classifier is interpreted as a detector of the complete capture signature, including missingness patterns and hardware scales.

Z-score standardization was performed inside the validation hierarchy. For every training fold and feature k, the mean μ_k and standard deviation σ_k were estimated only from the training data, and both the training and corresponding validation/test data were transformed as

$$z_{ik} = \frac{x_{ik} - \mu_k}{\sigma_k}. \quad (4)$$

The final outer-test transformation used parameters estimated from the complete outer-training fold. This prevents validation or test data from influencing scaling.

2.3. Classification Tasks and Validation Design

The objective of this study is to characterize the discriminative information contained in DCS rather than to construct a device-invariant writer-identification system. Specifically, we investigate whether statistical dynamic representations of DCS contain sufficient information to distinguish writer identity, writing content, and acquisition condition. Because acquisition hardware and input modality constitute part of the DCS-generation process, device-related characteristics are intentionally retained in the present analysis. Consequently, the writer-classification experiment should be interpreted as within-dataset discrimination under the observed acquisition conditions rather than as an estimate of cross-device or unseen-session writer generalization.

Three supervised tasks were evaluated. The writer task used 30 writers arranged as 10 groups of three. In each round, one group was collapsed into a pooled "other" class, and the remaining 27 writers retained their individual labels for a total of 10 rounds. The content task grouped the 11 contents into Chinese (CN; six categories), English (EN; three categories), and numeric/alphanumeric (NUM; two categories). In each of three rounds, one macro-group was collapsed into "other" while the remaining categories retained their original labels. The acquisition-condition task used the five closed-set labels IPMS, IPPS, IPPF, SPPS, and SPPF.

Interpretation of the pooled-other protocol. The pooled-other writers or content categories contributed training examples for the "other" label within each outer fold. This protocol evaluates rejection-class discrimination, not strict open-set recognition in which unknown identities are entirely absent from training. It is therefore reported as a pooled-other or open-category proxy, not as writer-disjoint open-set validation.

For each round, the data were partitioned using five-fold stratified outer cross-validation, yielding an 80% outer-training fold and a 20% outer-test fold. Stratification preserved the current task-label distribution. Within each outer-training fold, a second five-fold stratified cross-validation loop performed scaling and forward feature selection. A candidate feature was added when it maximized mean inner-fold accuracy. When the current subset achieved the same mean inner-fold accuracy as the historical best, the tied subset could be retained according to the implemented selection rule. Selection stopped after six consecutive candidate additions failed to improve the historical best mean inner-fold accuracy; the best subset was retained. This nested procedure follows established cross-validation and early-stopping principles [31,32].

Because timestamp was defined relative to the first sampled point, `Min_timestamp` was identically zero for all the DCS samples and therefore contained no discriminative information. Under the implemented forward-selection procedure and its patience/tie-handling mechanism, however, a zero-variance candidate could be retained when its inclusion did not change the inner-fold validation accuracy. Consequently, the occasional appearance of `Min_timestamp` in the fold-specific selected subsets reflects the behavior of the feature-selection procedure rather than a predictive contribution of this feature.

The writer, content, and acquisition tasks therefore produced 50, 15, and 5 outer-test evaluations, respectively. All the splitters used shuffling with `random_state = 42`. Because splitting was performed at the sample level, repetitions from the same combination of writer, content, and acquisition condition could occur in both the training and test folds. The overall feature-construction and nested evaluation procedure is summarized in Figure 1.

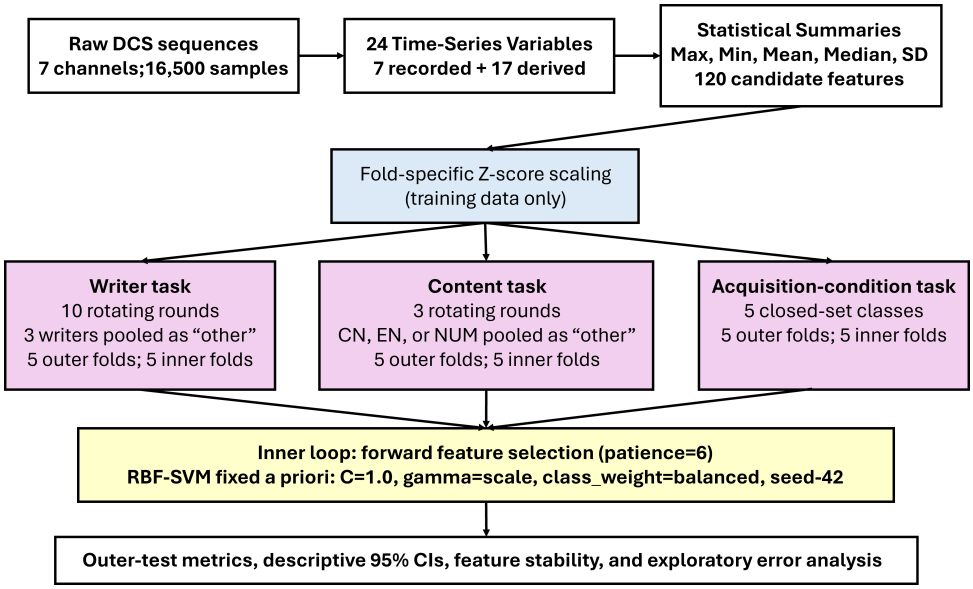


Figure 1. Feature-construction and nested evaluation workflow. The inner loop performs scaling and forward feature selection; the outer loop reports the test metrics.

2.4. SVM Specification and Reproducibility

All the models used scikit-learn SVC with a radial basis function kernel and one-vs-one multiclass decision strategy [33–35]. Hyperparameters were fixed a priori rather than optimized: `kernel = rbf`, $C = 1.0$, $\gamma =$ scale, `class_weight = balanced`, and `random_state = 42`. Fixing C and gamma avoids adding another adaptive search to the inner loop; the nested procedure therefore estimates the performance of the pre-specified SVM with data-dependent feature selection rather than the best result from a broad hyperparameter grid. Balanced class weights were used to reduce the influence of class-frequency differences [36].

The evaluation workflow is specified by the round-specific target definitions, shuffled stratified outer and inner five-fold partitions, the fixed random seed (`random_state = 42`), fold-specific training-only scaling, forward feature selection with `patience = 6`, refitting on the complete outer-training fold, and the error-aggregation rule described in Section 2.7. These procedural details, together with Table 4, provide an explicit algorithmic description of the model-development and evaluation workflow.

Table 4. Reproducible training and evaluation procedure.

Step	Operation
1	Define the round-specific target labels, including the pooled-other label for writer/content tasks.
2	Create shuffled five-fold stratified outer partitions with `random_state = 42`.
3	Within each outer training fold, create a shuffled stratified five-fold inner partition with the same seed.
4	For every candidate feature subset, fit the Z-score scaler on the inner-training fold only, transform inner-training/validation data, fit the fixed RBF-SVM, and calculate validation accuracy.

Table 4. *Cont.*

Step	Operation
5	Apply forward selection with patience = 6 and retain the historically best subset.
6	Refit the scaler and fixed RBF-SVM on the full outer-training fold using the selected subset; evaluate once on the outer-test fold.
7	Aggregate outer-test metrics across all folds/rounds and compute feature-selection stability.
8	For the descriptive error-concentration analysis, concatenate the five outer-test prediction sets within each round and then apply the cross-round union rule defined in Section 2.7 so that each unique DCS sample contributes at most one aggregated error.

2.5. Evaluation Metrics and Descriptive Confidence Intervals

For each outer-test fold, the multiclass confusion matrix was converted to a one-vs-rest table for each class k, with TP_k, TN_k, FP_k, and FN_k. The reported task-level metrics were:

- Accuracy = number of correctly classified samples/N.
- Macro precision = $$\frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k+FP_k}.$$
- Macro recall = $$\frac{1}{K}\sum_{k=1}^{K}\frac{TP_k}{TP_k+FN_k}.$$
- Macro F1 = $$\frac{1}{K}\sum_{k=1}^{K}\frac{2P_kR_k}{P_k+R_k}.$$
- Macro FPR = $$\frac{1}{K}\sum_{k=1}^{K}\frac{FP_k}{FP_k+TN_k}.$$
- Macro FRR = $$\frac{1}{K}\sum_{k=1}^{K}\frac{FN_k}{TP_k+FN_k};$$ this is the macro false-negative rate.
- Known-class accuracy = correctly classified known-label samples/number of known-label samples.
- Other-class recall = correctly predicted pooled-other samples/number of pooled-other samples.

For a metric m measured over M outer-test evaluations, the descriptive interval was calculated as

$$\bar{m} \pm t_{0.975,M-1}\frac{s_m}{\sqrt{M}}. \quad (5)$$

Bounds were truncated to the admissible range $[0, 1]$. Because outer folds and rounds reuse samples, these intervals summarize cross-split variability and should not be interpreted as independent-sample population confidence intervals.

The task-level accuracy metrics reported in Table 5 are calculated from the individual outer-test evaluations and are distinct from the union-based post hoc error-concentration metric defined in Section 2.7.

Table 5. Mean outer-test performance in percent, with descriptive 95% t intervals in parentheses.

Task	Accuracy	Macro Precision	Macro Recall	Macro F1	Macro FPR	Macro FRR	Known-Class Accuracy	Other Recall
Writer	94.85 (94.69–95.01)	94.82 (94.67–94.98)	95.45 (95.31–95.60)	95.09 (94.94–95.24)	0.19 (0.18–0.20)	4.55 (4.40–4.69)	95.79 (95.64–95.94)	86.44 (85.43–87.44)
Content	96.64 (95.74–97.53)	95.94 (95.20–96.68)	96.04 (95.30–96.77)	95.98 (95.23–96.72)	0.42 (0.35–0.50)	3.96 (3.23–4.70)	95.66 (94.93–96.38)	98.74 (98.17–99.32)
Acquisition condition	99.99 (99.98–100.00)	99.99 (99.98–100.00)	99.99 (99.98–100.00)	99.99 (99.98–100.00)	0.00 (0.00–0.01)	0.01 (0.00–0.02)	–	–

2.6. Feature-Selection Stability

Forward selection produced a different subset in each outer fold. Consequently, there is no single task-wide selected feature set. Appendix A reports the complete cross-fold union for each task, together with selection rate, mean/minimum/maximum selection rank, and reciprocal rank fusion (RRF) scores.

Following the reciprocal rank fusion formulation [21], for feature j, with selected-fold set F_j and rank r_{jf} in fold f, the stability score was

$$\mathrm{RRF}_j = \sum_{f \in F_j} \frac{1}{60 + r_{jf}}. \tag{6}$$

The constant 60 is the conventional RRF offset. A direct approximation, $n_j/(60 + \text{mean rank}_j)$, is shown only as a consistency check.

2.7. Descriptive Error-Concentration Analysis

For each classification round, the predictions from the five mutually exclusive outer-test folds were first concatenated to obtain one out-of-fold prediction for each of the 16,500 DCS samples. For the subsequent descriptive error-concentration analysis, errors were then aggregated across rotating rounds at the level of the unique DCS sample using a union rule. Specifically, a DCS sample was counted as an error if it was misclassified in one or more rounds regardless of the number of rounds in which the error occurred. Thus, each unique DCS sample contributed at most one error to the aggregated error analysis.

For sample i, the cross-round error indicator was defined as

$$E_i = \begin{cases} 1, & \text{if } \exists r : \hat{y}_{i,r} \neq y_{i,r}, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $\hat{y}_{i,r}$ and $y_{i,r}$ denote the predicted and true labels, respectively, for sample i in round r.

For a subgroup G, the union-based error-occurrence rate was calculated as

$$\mathrm{ErrorRate}_G = \frac{\sum_{i \in G} E_i}{|G|}. \tag{8}$$

Accordingly, the denominator represents the number of unique DCS samples rather than the total number of round-level prediction events. For the writer task, the 10 rotating rounds generated $16{,}500 \times 10 = 165{,}000$ round-level prediction events, whereas the descriptive union-based error analysis contained 16,500 unique DCS samples, including 550 unique samples per writer. Similarly, the three content rounds generated $16{,}500 \times 3 = 49{,}500$ round-level prediction events, whereas the union-based content error analysis contained 16,500 unique DCS samples, including 1500 samples per writing-content category.

The resulting union-based error-occurrence rate is methodologically distinct from the mean outer-test accuracy reported in Table 5 and should not be interpreted as 1 − Accuracy. Because repeated observations from the same writer are not statistically independent, no inferential chi-square testing, standardized-residual analysis, mixed-effects modeling, generalized estimating equations, or multiple-testing procedure was applied. These summaries are therefore used only to describe the observed concentration of classification errors.

3. Results

3.1. Outer-Test Classification Performance

Table 5 summarizes the performance across the outer-test evaluations. Writer classification achieved a mean accuracy of 94.85% and macro F1 of 95.09%. The known-writer accuracy was 95.79%, whereas the pooled-other recall was 86.44%, indicating that the heterogeneous rejection class was more difficult than the known labels. Content classification achieved 96.64% accuracy, 95.98% macro F1, and 98.74% pooled-other recall. Acquisition-condition classification reached 99.99% accuracy and macro F1, with one outer-test prediction incorrectly classified among the 16,500 predictions. The mean accuracy and macro F1 values across the three classification tasks are visualized in Figure 2, together with their descriptive 95% intervals.

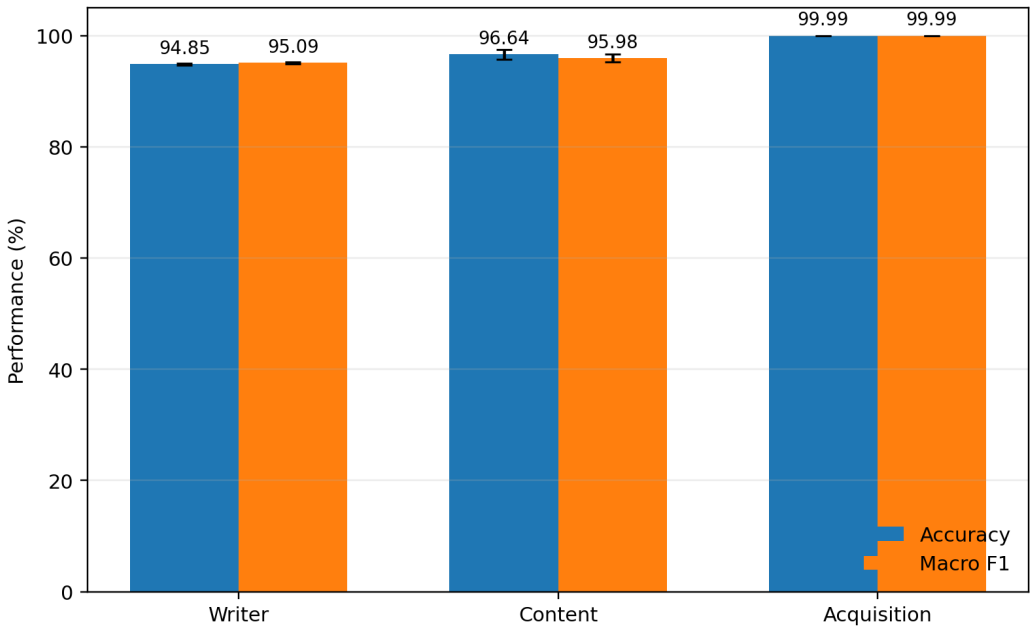


Figure 2. Mean accuracy and macro F1 across outer-test evaluations. The y-axis begins at zero; error bars show the descriptive 95% intervals.

3.2. Selected-Feature Counts and Stability

The writer task selected a mean of 47.2 features per outer fold (range 36–65), the content task selected 29.27 features (range 25–38), and the acquisition-condition task selected exactly four features per fold. The full cross-fold unions were larger because the selected subset changed across folds. The numbers of features selected per outer fold are summarized in Table 6.

Table 6. Number of features selected per outer fold.

Task	Mean Count	Minimum	Maximum	SD
Writer	47.2	36	65	7.21
Content	29.27	25	38	3.59
Acquisition condition	4	4	4	0

The writer features with the largest stability scores included minimum displacement, standard deviation of total velocity, minimum pressure change, displacement variability, mean centripetal acceleration, and mean x-velocity. The content features emphasized stroke count, horizontal position, centripetal acceleration, displacement variability, and vertical velocity. The acquisition-condition union contained pressure-change variability, displacement change, acceleration, velocity, and altitude-angle summaries. The most stable selected features are summarized in Table 7.

Table 7. Most stable selected features. Appendix A gives the complete cross-fold unions.

Task	Feature	Selection Rate	Mean Rank	RRF Score
Writer	Min_displacement	100%	3.74	0.786373
Writer	Std_total_velocity	98%	5.49	0.750061
Writer	Min_change_pressure	94%	3.98	0.735286
Writer	Std_displacement	100%	9.26	0.725763
Writer	Mean_centripetal_acceleration	100%	10.36	0.711776
Writer	Mean_x_velocity	100%	13.82	0.683915
Writer	Mean_sin_alpha	100%	13.84	0.679574
Writer	Median_locationX	100%	15.04	0.667990
Writer	Max_stroke	94%	11.32	0.666942
Writer	Min_locationY	100%	15.58	0.662643
Content	Max_stroke	100%	2.73	0.241002
Content	Median_locationX	100%	2.67	0.239415
Content	Mean_centripetal_acceleration	100%	4.93	0.231155
Content	Max_locationX	100%	6.87	0.225105
Content	Mean_y_velocity	100%	7.93	0.221157
Content	Min_locationX	100%	8.07	0.220641
Content	Min_locationY	100%	10.13	0.216073
Content	Std_displacement	100%	10.53	0.212794
Content	Mean_x_velocity	100%	12.20	0.209220
Content	Max_change_locationX	100%	12.93	0.205810
Acquisition condition	Std_change_pressure	100%	2	0.080645
Acquisition condition	Mean_change_displacement	80%	3	0.063492
Acquisition condition	Mean_total_acceleration	80%	4	0.062500
Acquisition condition	Median_altitudeAngle	60%	1	0.049180
Acquisition condition	Mean_altitudeAngle	40%	1	0.032787
Acquisition condition	Median_change_displacement	20%	3	0.015873
Acquisition condition	Mean_total_velocity	20%	4	0.015625

3.3. Writer-Task Error Patterns

For the descriptive writer-level error analysis, the five outer-test folds within each round were first combined to obtain predictions for all the DCS samples, after which the errors were aggregated across the 10 rotating writer rounds using the union rule defined in Section 2.7. Each unique DCS sample therefore contributed at most one error. Accordingly, each writer contributed 550 unique DCS samples, corresponding to 11 writing-content categories × 10 repetitions × 5 acquisition conditions rather than 5500 round-level prediction events.

The writer-task error rates were the highest for writers A3 (208/550, 37.82%), D2 (184/550, 33.45%), C3 (174/550, 31.64%), F2 (138/550, 25.09%), and B3 (128/550, 23.27%). These counts represent union-based error occurrences among unique DCS samples and should not be interpreted as independent round-level prediction events.

The writer error concentration varied substantially across writing-content categories: NUM produced 466 errors among 1500 unique DCS samples (31.07%) and NID produced 379 (25.27%), compared with 6.27% for EN1. This pattern is consistent with the hypothesis that short constrained strings contain fewer individualized movement patterns.

The distribution of the writer-task error rates across the five acquisition conditions is shown in Figure 3.

The writer error concentration also varied across acquisition conditions. IPPF produced 701 errors among 3300 unique DCS samples (21.24%) and SPPF produced 575 (17.42%). The stylus conditions IPMS, IPPS, and SPPS produced error rates of 9.91%, 9.30%, and 5.76%, respectively. These results agree with the prior DTW analysis of the same dataset in which stylus input produced more stable same-writer/different-writer separation [29,37,38].

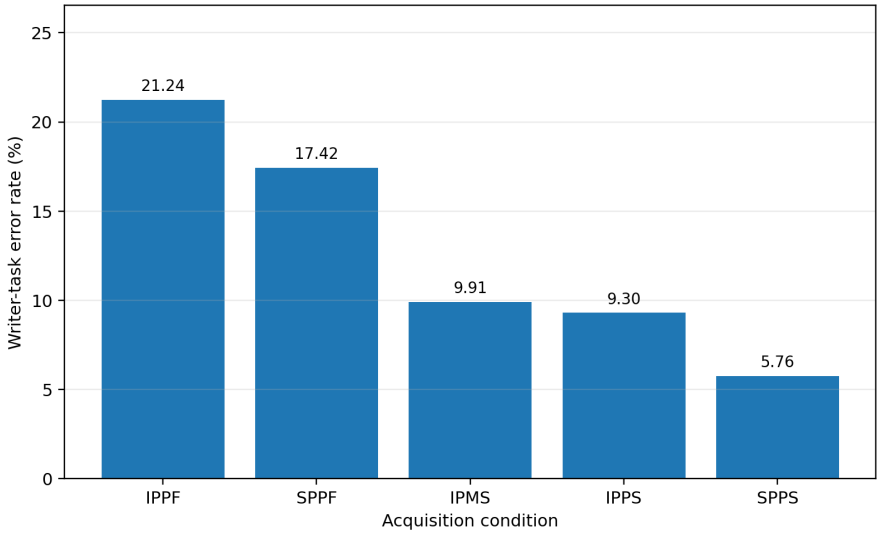


Figure 3. Union-based writer-task error-occurrence rate by acquisition condition, calculated over unique DCS samples across the 10 rotating writer rounds. A sample was counted once if it was misclassified in one or more rounds. These rates are descriptive and are not complements of the task-level accuracy reported in Table 5.

3.4. Content-Task Error Patterns

For the descriptive content-level error analysis, errors were aggregated across the three rotating content rounds using the same union rule defined in Section 2.7. Each unique DCS sample therefore contributed at most one error. Accordingly, each writing-content category contributed 1500 unique DCS samples, corresponding to 30 writers × 10 repetitions × 5 acquisition conditions, rather than 4500 round-level prediction events.

The content-task errors were concentrated in CN2 (165/1500, 11.00%), EN2 (147/1500, 9.80%), CN1 (115/1500, 7.67%), and CN3 (111/1500, 7.40%). NUM was the easiest category (18/1500, 1.20%). At the writer level, F3, A3, and B3 had the highest content-error rates. IPPF again had the highest union-based acquisition-condition error-occurrence rate (227/3300, 6.88%), whereas SPPS had the lowest (117/3300, 3.55%).

As shown in Figure 4, the highest content-task error rates occurred for CN2 and EN2, whereas NUM had the lowest error rate.

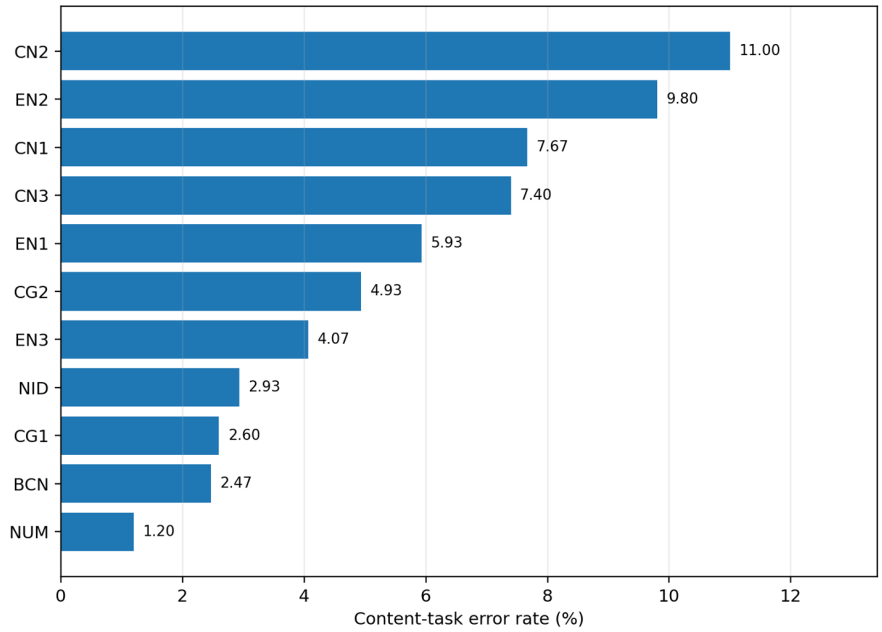


Figure 4. Union-based content-task error-occurrence rate by writing-content category, calculated over unique DCS samples across the three rotating content rounds. A sample was counted once if it was misclassified in one or more rounds. These rates are descriptive and are not complements of the task-level accuracy reported in Table 5.

4. Discussion

4.1. Information Preserved by Statistical Dynamic Features

Five summary statistics of 24 DCS variables retained substantial task information despite discarding detailed temporal alignment. The writer and content results indicate that global summaries of displacement, pressure change, velocity, acceleration, direction, and stroke structure can support efficient multiclass discrimination. The selected feature patterns are also interpretable: writer classification emphasized motion variability and pressure changes, whereas content classification emphasized spatial extent and stroke count. These findings align with prior work showing that pressure, coordinate dynamics, derivatives, and statistical descriptors can be discriminative [18–20].

The compact representation is computationally simpler than pairwise DTW or end-to-end sequence learning. This property is relevant for triage, where large collections may need to be screened before detailed comparison. However, summary features cannot preserve all the local temporal dependencies. Transformer, graph-attention, Siamese, and hybrid recurrent models can model long-range or pairwise relationships more directly [13–16], and multimodal consistency methods can address image–sequence–video agreement and attack scenarios [24]. These approaches are promising comparators for future external validation.

4.2. Acquisition-Condition Encoding Is a Confound, Not a Forensic Identification Result

Importantly, the term ”confound” should be interpreted relative to the classification objective. For acquisition-condition classification, hardware scales, channel availability, and input modality are legitimate characteristics of the DCS acquisition process and are therefore part of the phenomenon being characterized. They become confounding factors only when the objective is to infer writer identity independently of the acquisition system. The present study addresses the former question—the characterization of information encoded in a DCS sample—whereas device-normalized or leave-one-device-out writer

identification addresses the latter. Distinguishing these two research questions is essential when interpreting the reported results.

The near-perfect acquisition-condition result is the most important methodological finding. Coordinates were retained in pixels, pressure in device/API-specific units, and unavailable channels were zero-coded. Moreover, altitude-angle and pressure-change statistics were repeatedly selected. The classifier therefore had access to measurement scales and missingness patterns that are intrinsic to the capture system. The result shows that the acquisition condition is easily recoverable from the feature matrix; it does not demonstrate that handwriting dynamics alone identify a device.

This interpretation is consistent with cross-device studies reporting greater distances and degraded ROC performance under device mismatch [9] and with mobile-device research emphasizing sensor precision and interoperability [22]. Forensic comparison should therefore use comparable capture conditions whenever possible and should not treat a cross-device difference as writer evidence without validated calibration or domain adaptation. A common-channel-only analysis, per-device physical calibration, coordinate normalization, and leave-one-device-out writer evaluation are necessary to determine how much writer information survives removal of hardware cues.

4.3. Why SVM Was Retained

SVM was selected because the input is a moderate-dimensional fixed-length vector, the sample size is moderate, nonlinear decision boundaries are plausible, and forward selection yields a traceable subset [33,34,39]. The fixed RBF-SVM also provides a reproducible baseline with substantially lower modeling complexity than deep sequence networks. This is consistent with statistical learning studies in which SVM remains competitive with engineered online signature features [18].

The limitation is that SVM does not directly learn temporal representations. Deep models can exploit local geometry, long-range dependencies, data augmentation, and writer-independent metric learning [13–16,22,23]. Offline studies likewise show strong generalization from CNN, CNN–BiLSTM, deformable attention, and manifold metric learning [26–28], although their images and decision protocols differ from the present online multiclass setting. A fair future comparison should use the same outer partitions, common preprocessing, calibrated operating points, and writer-disjoint evaluation.

4.4. Pooled-Other Rejection and Generalization

The rotating pooled-other experiments add a rejection-class stress test but do not solve the open-set problem raised by forensic deployment. Because the pooled other identities were used during training as examples of the other class, the model learned a supervised rejection category. Strict open-set writer recognition requires that unknown writers be entirely absent from training, typically using writer-disjoint development/test partitions and a calibrated rejection threshold. Recent writer-independent systems explicitly use such partitions or pairwise verification protocols [13,16,27,28].

Generalization is also limited by the contemporaneous collection. The absence of independent sessions prevents evaluation of natural change over time even though contemporaneous samples are appropriate for controlling short-term variability in forensic comparison [1,30]. The current sample-level folds can include repetitions of the same writer, content, and acquisition condition in both the training and test data, so the reported performance may be optimistic. The reported accuracies should not be interpreted as operational identification or verification rates for an unknown case.

4.5. Forensic Role and Explainability

In forensic contexts, automated output should support examiner reasoning rather than substitute for it. Explainable signature-verification research emphasizes transparent features, simple distance or probability scores, and communication that a non-expert can understand [25]. The present feature-selection stability tables help to identify which physical or kinematic summaries drive the SVM, but class predictions are not evidential likelihood ratios. No claim is made that the scores quantify the probability of common authorship.

The most defensible use is preliminary triage: flagging obviously inconsistent acquisition conditions, prioritizing samples for detailed examination, and identifying conditions that are associated with reduced stability. Any casework application would require validated operating thresholds, known error rates under relevant propositions, strict separation of training and case data, and a procedure for examiner review. The model should not be used to issue a source conclusion, authorship opinion, or legal determination.

4.6. Limitations and Future Work

The main limitations of the present study are as follows:

- The outer partitions are sample-level rather than repetition-, session-, writer-, or device-disjoint. Therefore, the reported writer and content accuracies quantify discrimination among samples collected within the observed experimental domain and should not be interpreted as estimates of generalization to independent sessions, unseen writers, or unseen devices.
- All the samples were collected in a contemporaneous period, so longitudinal stability is unknown.
- The pooled-other class is supervised and is not strict open-set recognition.
- Coordinates and pressure retain device-specific units, and missing channels are zero-coded; no common-channel ablation or physical calibration was performed.
- Derivative features were calculated from adjacent recorded timestamps without a documented resampling, smoothing, or sensitivity analysis.
- No repeated-measures inferential model, such as a mixed-effects logistic regression or generalized estimating equation, was applied to the descriptive subgroup error analysis.
- The dataset contains 30 natural writers and no skilled forgery, disguise, cross-dataset, or cross-session evaluation.
- Exact computational replication additionally depends on preservation of the fold-index files and fold-specific selected subsets; the present study therefore reports the splitting algorithm, software versions, fixed random seed, and step-by-step evaluation procedure.

Future work should prioritize strict writer-disjoint verification, longitudinal sessions, leave-one-device-out testing, common-channel and per-device-normalization ablations, and external datasets. Probabilistic calibration and likelihood-ratio research should be considered only after stable cross-domain performance is demonstrated. Multimodal fusion and attack-detection frameworks may also improve resilience when online sequences, static images, or videos are available [24].

5. Conclusions

This revision presents a transparent nested SVM evaluation of 120 statistical DCS features for writer, content, and acquisition-condition classification. The writer and content accuracies were 94.85% and 96.64%, and the pooled-other recalls were 86.44% and 98.74%, respectively. The acquisition-condition accuracy was 99.99%. The corrected angular formulas, removal of the mode, fold-specific scaling, fixed hyperparameters, explicit metrics, confidence intervals, and feature-stability tables improve methodological clarity and reproducibility.

The results also show why forensic claims must remain limited. More specifically, the present results demonstrate that a DCS sample contains simultaneously discriminative writer-, content-, and acquisition-related information; they do not establish that writer-specific information is invariant to acquisition hardware. The acquisition condition is strongly encoded by device scales and missing-channel patterns; the splits are at the sample level; no independent sessions or strict unknown writers are available; and the error analysis is exploratory. The workflow is therefore best understood as an experimental screening aid for prioritizing comparison and identifying acquisition mismatch, not as a standalone writer-identification or signature-verification system.

Author Contributions: Conceptualization, L.-H.T. and W.-C.Y.; methodology, L.-H.T. and W.-C.Y.; software, L.-H.T.; validation, J.J. and L.-H.T.; formal analysis, L.-H.T.; resources, H.-J.L.; writing—original draft preparation, W.-C.Y. and J.J.; writing—review and editing, W.-C.Y.; supervision, W.-C.Y. and C.-H.C.; funding acquisition, H.-J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Justice Investigation Bureau, China (Taiwan), grant number 115-1301-10-28-01.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Fu Jen Catholic University (protocol code C113187; date of approval: 20 August 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from Wen-Chao Yang upon reasonable request.

Acknowledgments: The authors thank the volunteers who participated in the handwriting-acquisition experiment and colleagues who provided technical comments during the development of the DCS capture and analysis workflow.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

DCS	Digitally captured signature or handwriting
DTW	Dynamic time warping
FPR	Macro one-vs-rest false-positive rate
FRR	Macro one-vs-rest false-rejection/false-negative rate
GEE	Generalized estimating equation
RBF	Radial basis function
RRF	Reciprocal rank fusion
SVM	Support vector machine

Appendix A. Complete Cross-Fold Selected-Feature Unions

The following tables report every feature selected in at least one outer fold. Selection rate is the proportion of outer folds in which the feature was retained. RRF score is the sum of $1/(60 + \text{rank})$ over selected folds. Because feature selection was nested, these unions are descriptive stability summaries rather than a single fixed model signature.

Because timestamp is measured relative to the first sampled point, `Min_timestamp` is identically zero and has no discriminative value. Its nonzero selection rate in Tables A1 and A2 results from the patience/tie-handling behavior of the implemented forward-selection procedure, as described in Section 2.3, and should not be interpreted as evidence that this feature contributes predictive information.

Table A1. Writer-task complete selected-feature union across 50 outer-test evaluations.

Feature	Selection Rate	Mean Rank	Min Rank	Max Rank	RRF Score	Direct Approximation
Min_displacement	100.0%	3.74	1	14	0.786373	0.784437
Std_total_velocity	98.0%	5.49	1	16	0.750061	0.748208
Min_change_pressure	94.0%	3.98	3	9	0.735286	0.734619
Std_displacement	100.0%	9.26	5	28	0.725763	0.721917
Mean_centripetal_acceleration	100.0%	10.36	3	17	0.711776	0.710631
Mean_x_velocity	100.0%	13.82	6	35	0.683915	0.677323
Mean_sin_alpha	100.0%	13.84	9	26	0.679574	0.677140
Median_locationX	100.0%	15.04	4	22	0.667990	0.666311
Max_stroke	94.0%	11.32	3	49	0.666942	0.659010
Min_locationY	100.0%	15.58	9	27	0.662643	0.661551
Mean_x_acceleration	100.0%	15.92	6	28	0.661890	0.658588
Std_centripetal_acceleration	82.0%	2.66	1	10	0.655540	0.654340
Max_locationX	100.0%	17.86	1	31	0.645443	0.642178
Min_locationX	100.0%	19.66	8	27	0.629376	0.627668
Max_change_locationY	100.0%	20.04	12	40	0.627783	0.624688
Max_timestamp	100.0%	21.68	1	46	0.620237	0.612145
Mean_change_locationY	100.0%	21.04	13	30	0.618082	0.616979
Std_locationX	94.0%	17.74	9	44	0.613817	0.604543
Mean_y_velocity	98.0%	21.06	10	44	0.609497	0.604481
Mean_theta	100.0%	22.58	10	34	0.607812	0.605473
Max_change_locationX	100.0%	24.34	17	40	0.593984	0.592839
Median_locationY	100.0%	25.86	17	42	0.583788	0.582343
Max_pressure	94.0%	21.43	12	42	0.582706	0.577215
Max_locationY	96.0%	28.29	19	41	0.545867	0.543653
Std_cos_alpha	74.0%	8.49	2	45	0.545281	0.540253
Std_sin_alpha	98.0%	33.47	8	60	0.527193	0.524236
Std_change_pressure	96.0%	34.56	12	61	0.511992	0.507601
Mean_angle_velocity	96.0%	34.44	26	47	0.509883	0.508273
Mean_cos_alpha	80.0%	21.65	7	58	0.505681	0.489896
Min_change_locationY	92.0%	31.76	23	50	0.503157	0.501303
Mean_change_locationX	92.0%	32.91	17	50	0.498168	0.495087
Min_pressure	84.0%	30.79	14	52	0.469188	0.462628

Table A1. *Cont.*

Feature	Selection Rate	Mean Rank	Min Rank	Max Rank	RRF Score	Direct Approximation
Max_y_velocity	76.0%	36.66	26	55	0.394861	0.393139
Std_x_velocity	72.0%	34.81	24	54	0.382216	0.379725
Median_stroke	72.0%	37.47	28	63	0.370963	0.369336
Min_change_locationX	66.0%	37.64	28	53	0.339108	0.337989
Min_y_velocity	62.0%	37.45	26	59	0.320033	0.318107
Median_total_velocity	56.0%	31.21	1	56	0.312049	0.306969
Mean_y_acceleration	60.0%	37.50	26	60	0.309713	0.307692
Std_cos_beta	48.0%	22.67	5	61	0.305928	0.290323
Median_change_displacement	56.0%	35.54	16	58	0.296723	0.293084
Median_total_acceleration	52.0%	37.15	2	65	0.273567	0.267617
Median_angle_velocity	54.0%	41.04	28	54	0.268268	0.267229
Median_cos_alpha	52.0%	39.12	28	53	0.263148	0.262321
Mean_change_displacement	38.0%	18.74	4	49	0.250973	0.241310
Max_x_velocity	46.0%	38.00	24	64	0.236799	0.234694
Mean_cos_beta	36.0%	19.56	7	57	0.234098	0.226257
Std_theta	40.0%	31.70	12	60	0.225855	0.218103
Std_total_acceleration	44.0%	39.45	6	59	0.224807	0.221207
Mean_total_acceleration	32.0%	26.44	4	43	0.192119	0.185105
Median_y_velocity	36.0%	40.94	32	56	0.179035	0.178316
Mean_stroke	32.0%	34.62	4	64	0.174326	0.169089
Std_stroke	32.0%	35.19	2	55	0.174126	0.168089
Std_angle_velocity	32.0%	36.81	25	62	0.167111	0.165268
Min_change_displacement	22.0%	9.82	1	51	0.165631	0.157552
Max_displacement	24.0%	18.67	1	55	0.160891	0.152542
Min_stroke	20.0%	6.90	2	50	0.153996	0.149477
Min_timestamp	20.0%	7.90	3	51	0.151618	0.147275
Min_total_acceleration	20.0%	8.90	4	52	0.149313	0.145138
Min_total_velocity	20.0%	9.90	5	53	0.147078	0.143062
Std_pressure	28.0%	39.21	20	57	0.143032	0.141109
Median_change_locationY	28.0%	42.21	31	53	0.137428	0.136967
Std_locationY	28.0%	46.43	37	59	0.132007	0.131544
Std_y_velocity	26.0%	40.92	30	54	0.129558	0.128811
Max_altitudeAngle	26.0%	43.92	34	63	0.125817	0.125093
Mean_timestamp	22.0%	43.18	29	51	0.107082	0.106608

Table A1. *Cont.*

Feature	Selection Rate	Mean Rank	Min Rank	Max Rank	RRF Score	Direct Approximation
Std_change_locationX	22.0%	44.91	24	57	0.105939	0.104853
Std_change_locationY	20.0%	37.70	12	62	0.105460	0.102354
Max_change_displacement	18.0%	46.00	22	58	0.086500	0.084906
Median_centripetal_acceleration	16.0%	42.50	29	54	0.078571	0.078049
Median_displacement	16.0%	46.25	36	54	0.075579	0.075294
Std_y_acceleration	14.0%	35.14	9	43	0.074879	0.073574
Median_cos_beta	14.0%	39.71	30	52	0.070485	0.070201
Max_total_velocity	12.0%	39.00	33	47	0.060725	0.060606
Median_sin_alpha	12.0%	40.50	30	47	0.059871	0.059701
Mean_change_pressure	12.0%	45.67	34	54	0.057011	0.056782
Median_change_pressure	12.0%	47.00	43	55	0.056149	0.056075
Max_change_pressure	8.0%	29.50	13	40	0.045351	0.044693
Median_x_velocity	8.0%	37.25	31	46	0.041256	0.041131
Mean_azimuthAngle	8.0%	41.50	34	54	0.039614	0.039409
Mean_altitudeAngle	8.0%	41.25	37	45	0.039549	0.039506
Max_azimuthAngle	8.0%	46.25	41	60	0.037844	0.037647
Mean_displacement	8.0%	47.75	43	53	0.037196	0.037123
Std_x_acceleration	6.0%	36.33	34	41	0.031178	0.031142
Min_azimuthAngle	6.0%	42.00	35	46	0.029484	0.029412
Std_altitudeAngle	6.0%	44.00	34	51	0.028993	0.028846
Median_pressure	6.0%	43.67	40	47	0.028961	0.028939
Mean_total_velocity	4.0%	16.50	2	31	0.027118	0.026144
Min_x_velocity	6.0%	52.00	48	57	0.026815	0.026786
Median_theta	4.0%	43.00	41	45	0.019425	0.019417
Min_altitudeAngle	4.0%	46.00	44	48	0.018875	0.018868
Median_change_locationX	4.0%	46.50	42	51	0.018813	0.018779
Std_timestamp	4.0%	48.00	46	50	0.018525	0.018519
Median_timestamp	4.0%	49.50	49	50	0.018265	0.018265
Min_sin_alpha	2.0%	1.00	1	1	0.016393	0.016393
Median_azimuthAngle	2.0%	38.00	38	38	0.010204	0.010204
Min_theta	2.0%	49.00	49	49	0.009174	0.009174
Median_y_acceleration	2.0%	54.00	54	54	0.008772	0.008772
Median_x_acceleration	2.0%	58.00	58	58	0.008475	0.008475

Table A2. Content-task complete selected-feature union across 15 outer-test evaluations.

Feature	Selection Rate	Mean Rank	Min Rank	Max Rank	RRF Score	Direct Approximation
Max_stroke	100.0%	2.73	1	27	0.241002	0.239107
Median_locationX	100.0%	2.67	2	4	0.239415	0.239362
Mean_centripetal_acceleration	100.0%	4.93	3	7	0.231155	0.231006
Max_locationX	100.0%	6.87	4	15	0.225105	0.224327
Mean_y_velocity	100.0%	7.93	6	18	0.221157	0.220805
Min_locationX	100.0%	8.07	6	13	0.220641	0.220372
Min_locationY	100.0%	10.13	5	22	0.216073	0.213878
Std_displacement	100.0%	10.53	8	15	0.212794	0.212665
Mean_x_velocity	100.0%	12.20	6	21	0.209220	0.207756
Max_change_locationX	100.0%	12.93	10	16	0.205810	0.205667
Median_stroke	100.0%	13.07	10	20	0.205561	0.205292
Mean_change_locationY	100.0%	14.67	12	18	0.201026	0.200893
Median_locationY	100.0%	15.13	9	22	0.200314	0.199645
Max_locationY	100.0%	15.60	9	25	0.199214	0.198413
Std_stroke	100.0%	20.60	1	26	0.187296	0.186104
Max_change_displacement	80.0%	14.00	7	29	0.163695	0.162162
Mean_cos_alpha	80.0%	14.50	7	36	0.162766	0.161074
Max_change_locationY	80.0%	18.75	7	31	0.154525	0.152381
Mean_sin_alpha	73.3%	18.45	8	27	0.140855	0.140209
Std_cos_alpha	53.3%	2.62	2	3	0.127752	0.127745
Mean_theta	60.0%	19.56	15	25	0.113374	0.113128
Std_locationX	60.0%	20.33	13	24	0.112192	0.112033
Std_cos_beta	46.7%	2.71	2	3	0.111623	0.111617
Std_change_pressure	60.0%	23.56	18	30	0.108021	0.107713
Mean_x_acceleration	60.0%	26.78	21	37	0.103977	0.103713
Max_y_velocity	46.7%	20.29	13	27	0.087582	0.087189
Std_theta	40.0%	18.17	14	25	0.077017	0.076759
Median_total_acceleration	40.0%	21.00	18	28	0.074207	0.074074
Min_change_locationY	40.0%	23.00	20	26	0.072346	0.072289
Mean_y_acceleration	40.0%	26.17	23	29	0.069675	0.069632
Min_change_locationX	33.3%	19.60	15	23	0.062933	0.062814
Max_timestamp	33.3%	22.80	19	27	0.060465	0.060386
Mean_angle_velocity	33.3%	28.20	23	34	0.056789	0.056689
Std_centripetal_acceleration	26.7%	18.75	12	26	0.050996	0.050794
Mean_change_locationX	26.7%	22.00	18	30	0.048936	0.048780

Table A2. *Cont.*

Feature	Selection Rate	Mean Rank	Min Rank	Max Rank	RRF Score	Direct Approximation
Min_displacement	26.7%	23.75	21	25	0.047780	0.047761
Mean_altitudeAngle	26.7%	25.75	20	35	0.046849	0.046647
Max_x_velocity	26.7%	28.75	26	32	0.045097	0.045070
Mean_total_velocity	20.0%	19.00	14	24	0.038077	0.037975
Min_pressure	20.0%	21.33	17	29	0.037043	0.036885
Max_pressure	20.0%	22.00	20	24	0.036600	0.036585
Mean_total_acceleration	20.0%	22.33	19	28	0.036522	0.036437
Median_cos_alpha	20.0%	24.33	23	27	0.035591	0.035573
Std_pressure	20.0%	24.67	20	28	0.035492	0.035433
Mean_timestamp	13.3%	20.50	16	25	0.024923	0.024845
Max_change_pressure	13.3%	20.50	20	21	0.024846	0.024845
Mean_cos_beta	13.3%	21.00	18	24	0.024725	0.024691
Std_y_velocity	13.3%	21.50	16	27	0.024652	0.024540
Std_change_locationY	13.3%	22.50	21	24	0.024250	0.024242
Std_change_locationX	13.3%	30.50	29	32	0.022106	0.022099
Mean_stroke	13.3%	31.00	29	33	0.021989	0.021978
Median_timestamp	6.7%	15.00	15	15	0.013333	0.013333
Mean_change_displacement	6.7%	17.00	17	17	0.012987	0.012987
Mean_displacement	6.7%	23.00	23	23	0.012048	0.012048
Min_change_displacement	6.7%	25.00	25	25	0.011765	0.011765
Std_timestamp	6.7%	25.00	25	25	0.011765	0.011765
Min_y_velocity	6.7%	26.00	26	26	0.011628	0.011628
Min_x_velocity	6.7%	27.00	27	27	0.011494	0.011494
Std_altitudeAngle	6.7%	27.00	27	27	0.011494	0.011494
Max_azimuthAngle	6.7%	28.00	28	28	0.011364	0.011364
Min_change_pressure	6.7%	28.00	28	28	0.011364	0.011364
Std_change_displacement	6.7%	29.00	29	29	0.011236	0.011236
Median_y_velocity	6.7%	31.00	31	31	0.010989	0.010989
Min_stroke	6.7%	31.00	31	31	0.010989	0.010989
Min_timestamp	6.7%	32.00	32	32	0.010870	0.010870
Min_total_acceleration	6.7%	33.00	33	33	0.010753	0.010753
Min_total_velocity	6.7%	34.00	34	34	0.010638	0.010638
Median_total_velocity	6.7%	35.00	35	35	0.010526	0.010526
Median_change_locationY	6.7%	38.00	38	38	0.010204	0.010204

Table A3. Acquisition-condition complete selected-feature union across five outer-test evaluations. Each fold selected four features; the union contains seven.

Feature	Selection Rate	Mean Rank	Min Rank	Max Rank	RRF Score	Direct Approximation
Std_change_pressure	100.0%	2	2	2	0.080645	0.080645
Mean_change_displacement	80.0%	3	3	3	0.063492	0.063492
Mean_total_acceleration	80.0%	4	4	4	0.062500	0.062500
Median_altitudeAngle	60.0%	1	1	1	0.049180	0.049180
Mean_altitudeAngle	40.0%	1	1	1	0.032787	0.032787
Median_change_displacement	20.0%	3	3	3	0.015873	0.015873
Mean_total_velocity	20.0%	4	4	4	0.015625	0.015625

References

1. European Network of Forensic Science Institutes (ENFSI). *Best Practice Manual for the Forensic Handwriting Examination*, 4th ed.; ENFSI: Wiesbaden, Germany, 2022. Available online: https://enfsi.eu/wp-content/uploads/2023/02/BPM-Handwriting-Ed.-4.pdf (accessed on 22 August 2026).
2. *ISO/IEC 19794-7:2021*; Information Technology—Biometric Data Interchange Formats—Part 7: Signature/Sign Time Series Data. ISO: Geneva, Switzerland, 2021. Available online: https://www.iso.org/standard/77910.html (accessed on 22 August 2026).
3. Academy Standards Board (ASB). *ANSI/ASB Standard 070-22: Standard for Examination of Handwritten Items*, 1st ed.; Academy Standards Board: Colorado Springs, CO, USA, 2022. Available online: https://www.nist.gov/osac/standards-library/ansiasb-standard-070-22 (accessed on 22 August 2026).
4. Jones, K.; Bird, C.L. Non-Original and Digitally Captured Handwriting: Considerations for Forensic Handwriting Examinations. *WIREs Forensic Sci.* **2025**, *7*, e1540. [CrossRef]
5. *ASTM E2290*; Standard Guide for Examination of Handwritten Items. ASTM: West Conshohocken, PA, USA, 2017.
6. Linden, J.; Marquis, R.; Bozza, S.; Taroni, F. Dynamic signatures: A review of dynamic feature variation and forensic methodology. *Forensic Sci. Int.* **2018**, *291*, 216–229. [CrossRef] [PubMed]
7. Heckeroth, J.; Kupferschmid, E.; Dziedzic, T.; Kalantzis, N.; Geistova Cakovska, B.; Fernandes, C.; Branco, M.J.; Axelsson Spjuth, K.; Kerkhoff, A.; Vaccarone, P.; et al. Features of digitally captured signatures vs. pen-and-paper signatures: Similar or completely different? *Forensic Sci. Int.* **2021**, *318*, 110587. [CrossRef] [PubMed]
8. Zimmer, J.; Kalantzis, N.; Dziedzic, T.; Heckeroth, J.; Kupferschmid, E.; Fernandes, C.; Geistova Cakovska, B.; Branco, M.J.; Axelsson Spjuth, K.; Vaccarone, P.; et al. The challenge of comparing digitally captured signatures registered with different software and hardware. *Forensic Sci. Int.* **2021**, *327*, 110945. [CrossRef] [PubMed]
9. Reyes-Garcia, G.; Garcia-Barrientos, A.; Zambrano-Serrano, E.; Algredo-Badillo, I. S-Pen Technology and Online Signatures: Cross-Device Variability and Its Implications for Mobile Biometric Authentication. *Sensors* **2026**, *26*, 1451. [CrossRef] [PubMed]
10. Salkanovic, A.; Sušanj, D.; Batistić, L.; Ljubic, S. Beyond Signatures: Leveraging Sensor Fusion for Contextual Handwriting Recognition. *Sensors* **2025**, *25*, 2290. [CrossRef] [PubMed]
11. Impedovo, D.; Pirlo, G. Automatic signature verification: The state of the art. *IEEE Trans. Syst. Man Cybern. C* **2008**, *38*, 609–635. [CrossRef]
12. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Ortega-Garcia, J. DeepSign: Deep on-line signature verification. *IEEE Trans. Biom. Behav. Identity Sci.* **2021**, *3*, 229–239. [CrossRef]
13. Ahrabian, K.; BabaAli, B. Usage of autoencoders and Siamese networks for online handwritten signature verification. *Neural Comput. Appl.* **2019**, *31*, 9321–9334. [CrossRef]
14. Melzi, P.; Tolosana, R.; Vera-Rodriguez, R.; Delgado-Santos, P.; Stragapede, G.; Fierrez, J.; Ortega-Garcia, J. Exploring Transformers for On-Line Handwritten Signature Verification. *arXiv* **2023**, arXiv:2307.01663. [CrossRef]
15. Yuan, H.-J.; Zhang, H.; Yin, F. Online Handwritten Signature Verification Based on Temporal-Spatial Graph Attention Transformer. *arXiv* **2025**, arXiv:2510.19321. [CrossRef]
16. Ngin, K.; Valy, D.; Phauk, S.; Pin, V.; Yin, B.; Ang, S. KhmerWriterID: Toward Robust Khmer Writer Verification Using Deep Learning: CNN, RNN, and Hybrid Approaches. *IEEE Access* **2026**, *14*, 30339–30354. [CrossRef]
17. Jia, Y.; Huang, L.; Chen, H. A two-stage method for on-line signature verification using shape contexts and function features. *Sensors* **2019**, *19*, 1808. [CrossRef] [PubMed]
18. Ospina, R.; Costa, R.D.; Rêgo, L.C.; Marmolejo-Ramos, F. Application of nonparametric quantifiers for online handwritten signature verification: A statistical learning approach. *Stat. Anal. Data Min. ASA Data Sci. J.* **2024**, *17*, e11673. [CrossRef]
19. Liu, R.; Xin, Y. Online Handwritten Signature Verification Method Based on Uni-Feature Correlation Coefficient between Signatures. *Sensors* **2023**, *23*, 9341. [CrossRef] [PubMed]
20. Faundez-Zanuy, M.; Diaz, M.; Ferrer, M.A. Online Signature Recognition: A Biologically Inspired Feature Vector Splitting Approach. *Cogn. Comput.* **2024**, *16*, 265–277. [CrossRef]
21. Cormack, G.V.; Clarke, C.L.A.; Buettcher, S. Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Boston, MA, USA, 19–23 July 2009*; Association for Computing Machinery: New York, NY, USA, 2009; pp. 758–759. [CrossRef]
22. Roszczewska, K.; Niewiadomska-Szynkiewicz, E. Online Signature Biometrics for Mobile Devices. *Sensors* **2024**, *24*, 3524. [CrossRef] [PubMed]
23. Najda, D.; Saeed, K. Impact of augmentation methods in online signature verification. *Innov. Syst. Softw. Eng.* **2024**, *20*, 477–483. [CrossRef]
24. Shi, Z.; Li, F.; Hao, D.; Sun, Q. Handwritten Signature Verification via Multimodal Consistency Learning. *IEEE Trans. Inf. Forensics Secur.* **2025**, *20*, 3995–4007. [CrossRef]
25. Diaz, M.; Ferrer, M.A.; Vessio, G. Explainable offline automatic signature verifier to support forensic handwriting examiners. *Neural Comput. Appl.* **2024**, *36*, 2411–2427. [CrossRef]

26. Okawa, M. Multistage Convolutional Neural Network With Deformable Attention for Word-Level Offline Text-Independent Writer Identification. *IEEE Access* **2025**, *13*, 118504–118514. [CrossRef]
27. Zois, E.N.; Tsourounis, D.; Kalivas, D. Similarity Distance Learning on SPD Manifold for Writer Independent Offline Signature Verification. *IEEE Trans. Inf. Forensics Secur.* **2024**, *19*, 1342–1356. [CrossRef]
28. Longjam, T.; Kisku, D.R.; Gupta, P. Writer independent handwritten signature verification on multi-scripted signatures using hybrid CNN-BiLSTM: A novel approach. *Expert Syst. Appl.* **2023**, *214*, 119111. [CrossRef]
29. Lai, H.-J.; Tsai, L.-H.; Hsu, K.-Y.; Yang, W.-C. The research on the handwriting stability in different devices and conditions. *Sensors* **2026**, *26*, 538. [CrossRef] [PubMed]
30. Rogers, J.H. The Identification of Handwriting. *Aust. J. Forensic Sci.* **1968**, *1*, 23–33. [CrossRef]
31. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95), Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1145.
32. Prechelt, L. Early stopping—But when? In *Neural Networks: Tricks of the Trade*; Orr, G.B., Müller, K.-R., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69. [CrossRef] [PubMed]
33. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
34. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992*; Association for Computing Machinery: New York, NY, USA, 1992; pp. 144–152. [CrossRef]
35. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. Available online: https://jmlr.org/papers/v12/pedregosa11a.html (accessed on 22 August 2026).
36. Batuwita, R.; Palade, V. Class imbalance learning methods for support vector machines. In *Imbalanced Learning: Foundations, Algorithms, and Applications*; He, H., Ma, Y., Eds.; Wiley: Hoboken, NJ, USA, 2013; pp. 83–99. [CrossRef]
37. Salvador, S.; Chan, P. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **2007**, *11*, 561–580. [CrossRef]
38. Parziale, A.; Diaz, M.; Ferrer, M.A.; Marcelli, A. SM-DTW: Stability modulated dynamic time warping for signature verification. *Pattern Recognit. Lett.* **2019**, *121*, 113–122. [CrossRef]
39. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.