

Article

Multi-Class Online Signature Verification Based on Hybrid Statistical Moments and UMAP-Based Nonlinear Dimensionality Reduction

Liyan Huang ¹, Yuanxiang Ruan ¹, Weijun Li ^{1,*}, Naisheng Xu ² and Pan Zheng ³

¹ Department of Forensic Science, Fujian Police College, Fuzhou 350007, China; huangliyan@fjpsc.edu.cn (L.H.); r2074497421@163.com (Y.R.)

² Criminal Technology Corps, Fujian Provincial Public Security Department, Fuzhou 350003, China; wenyeke9v@163.com

³ Criminal Investigation Detachment, Fuzhou Municipal Public Security Bureau, Fuzhou 350001, China; sjjy06@126.com

* Correspondence: xjlwj@fjpsc.edu.cn

Abstract

Online signature verification (OSV) is a challenging problem in behavioral biometrics, especially when skilled forgeries closely mimic genuine signatures in both appearance and dynamics. This study presents a multi-class OSV framework that combines hybrid statistical features and nonlinear dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP). A 40-dimensional feature set is created from statistical moments of dynamic writing parameters in both time and frequency (DCT) domains. Experimental results show that UMAP-based dimensionality reduction preserves category-related structures in a compact two-dimensional space. The proposed approach achieves an average classification accuracy of 0.989 ± 0.005 and a Cohen's kappa coefficient of 0.985 ± 0.006 , demonstrating robust performance across multiple classifiers. The results validate the effectiveness of combining multi-domain statistical feature fusion with UMAP for multi-class online signature verification, providing both high performance and interpretable visual representations.

Keywords: online signature verification; behavioral biometrics; multi-class classification; hybrid statistical moments; dimensionality reduction; UMAP; forgery detection

1. Introduction

Online Signature Verification (OSV) is a widely used behavioral biometric technique in applications such as financial transactions, legal document authentication, and access control systems [1]. Unlike static biometric traits, online signatures are acquired as time-series signals through digitizing devices, recording dynamic information including writing speed, pen pressure, stroke direction, and spatial trajectories [2]. These dynamic characteristics provide additional behavioral cues beyond static shape information and have been shown to improve robustness against imitation and counterfeiting [3,4].

A central challenge in OSV lies in distinguishing genuine signatures from skilled forgeries, which are produced by imitators who deliberately attempt to reproduce both the visual appearance and the dynamic execution patterns of the target signature [5,6]. Such

Academic Editor: Loris Nanni

Received: 8 January 2026

Revised: 29 January 2026

Accepted: 30 January 2026

Published: 1 February 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](https://creativecommons.org/licenses/by/4.0/).

forgeries often exhibit subtle differences that are difficult to capture using limited or single-domain feature representations. Many traditional approaches rely primarily on time-domain statistical descriptors or adopt linear dimensionality reduction techniques, which may have limitations in modeling the complex variability present in signature dynamics. Classical linear methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) assume linear separability of the underlying feature space [7,8]. However, online signature data are influenced by nonlinear biomechanical factors and individual writing habits, which often lead to complex manifold structures. As a result, linear projections may fail to preserve discriminative information after dimensionality reduction, especially when separating genuine signatures from high-similarity forgeries [9].

To address these limitations, this study explores a feature representation that integrates complementary information from multiple domains and adopts a nonlinear dimensionality reduction strategy. Statistical moment descriptors are extracted from core dynamic writing sequences in both the time domain and the Discrete Cosine Transform (DCT) domain. Time-domain moments characterize overall writing behavior and stability, whereas DCT-domain moments capture global frequency-related patterns that are known to be relatively robust to small temporal fluctuations and local distortions in signature signals [10,11].

The resulting high-dimensional feature vectors are subsequently projected into a low-dimensional space using Uniform Manifold Approximation and Projection (UMAP) [12]. Unlike linear methods, UMAP is designed to preserve neighborhood relationships while maintaining global structural consistency in nonlinear data manifolds [13]. Such properties motivate its use for modeling the intrinsic geometry of online signature features. The obtained low-dimensional representations facilitate efficient verification while retaining interpretable structural relationships among different signature categories.

Unlike traditional OSV studies that primarily address binary genuine forgery verification, this study focuses on multi-class forgery-type classification, which better reflects practical forensic and anti-fraud requirements. It should be clarified that they are complementary sub-directions in OSV: traditional OSV verifies identity ownership, while this study identifies forgery types and provides risk control/forensic assistance without replacing identity verification. The main contributions of this work can be summarized as follows:

- A hybrid statistical feature representation is developed by extracting moment descriptors from dynamic signature signals in both the time and DCT domains, enabling complementary characterization of temporal behavior and spectral patterns.
- A UMAP-based nonlinear dimensionality reduction strategy is incorporated to obtain compact embeddings that preserve category-related structural information, supporting both quantitative verification and visual interpretability.
- A systematic multi-class evaluation framework is established to analyze verification performance across genuine signatures and multiple forgery types, enabling a structured comparison across different classifiers and data splits.

The remainder of this paper is organized as follows. Section 2 reviews related work on feature extraction and dimensionality reduction in OSV. Section 3 describes the dataset and the proposed methodology. Experimental results and discussion are presented in Section 4, followed by conclusions in Section 5.

2. Related Work

The development of Online Signature Verification (OSV) systems has largely been driven by advancements in feature representation and the optimization of classification models. This section reviews existing literature focusing on multi-domain feature extraction strategies and dimensionality reduction techniques relevant to the proposed framework.

2.1. Feature Extraction Strategies in OSV

Feature extraction is the cornerstone of robust biometric systems. In the context of OSV, features are generally categorized into function-based (local) and parameter-based (global) features [14]. Early research predominantly focused on time-domain parameter-based features, such as total duration, average speed, and pen-up/pen-down ratios, due to their computational simplicity [15]. However, relying solely on these global statistics often fails to capture the intricate temporal variations inherent in skilled forgeries.

To address this, researchers turned to function-based features that represent the signature as a time-series signal. Vorugunti et al. proposed a compound feature extraction method, demonstrating that combining static images with dynamic time-series information significantly enhances verification performance [16]. Furthermore, recent studies have emphasized the importance of transforming time-domain signals into the frequency or transform domain to capture global energy distributions. Techniques such as the Fast Fourier Transform (FFT), the Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) have been widely adopted to extract compact and discriminative representations. Kahindo et al. [17] demonstrated that frequency-domain descriptors can provide a global characterization of online signatures and support effective verification using GMMs. Similarly, Yanikoglu et al. [18] proposed an FFT-based verification system, showing that frequency-domain representations enable compact feature descriptions with fixed-length coefficients and efficient matching. While FFT provides global frequency information, it lacks temporal resolution. To address this, Wavelet Transform (WT) and Continuous Wavelet Transform (CWT) have been introduced. These techniques provide a “scalogram” (time-scale representation), effectively capturing both the frequency content and when it occurs. Chavan et al. [19] utilized a Hybrid Wavelet Transform (HWT-2) constructed via Kronecker products of DCT, Haar, and Hadamard transforms. This approach captures detailed energy distribution across scales, demonstrating superior performance in distinguishing genuine from forged signatures. Liu et al. [20] proposed a method based on Discrete Cosine Transform (DCT) and sparse representation. The DCT was used to obtain a compact representation of an online signature using a fixed number of coefficients, leading to simple matching procedures while retaining essential dynamic features. The transition from the time-domain to the frequency or wavelet transform domain represents a significant advancement in online signature verification.

More recently, the concept of “Hybrid Statistical Moments”, proposed by Radhika et al. [21], has gained attention to further enhance robustness. This approach involves integrating global geometric invariants (e.g., Hu moments) with orthogonal descriptors (e.g., Zernike moments) or dynamic signal statistics. By combining these complementary descriptors, systems can capture both the holistic shape invariance and the local dynamic variations in the signature curves. However, many existing hybrid approaches rely on direct concatenation of high-dimensional moment vectors followed by conventional classifiers (e.g., SVMs), with limited consideration of the underlying geometric structure of the feature space. Such linear fusion strategies may fail to capture intrinsic nonlinear correlations between different moment orders, which can limit the ability to distinguish subtle distributional shifts caused by skilled forgeries.

2.2. Dimensionality Reduction and Manifold Learning

While high-dimensional feature fusion improves discriminative power, it inevitably introduces the “curse of dimensionality,” leading to computational inefficiency and potential overfitting. Consequently, dimensionality reduction (DR) is a critical step in modern OSV pipelines.

Traditional linear dimensionality reduction methods, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), have been widely applied in biometric systems. PCA focuses on maximizing data variance, whereas LDA emphasizes class separability. For instance, Vahab et al. [22] employed multilayer perceptrons (MLPs) on selected PCA features to analyze signature time-series signals, achieving an accuracy of 93.1%. Similarly, Kholmatov et al. [23] formulated online signature verification as a two-class classification problem and combined Linear Discriminant Analysis with dynamic time warping, reporting a low skilled forgery error rate of 2.8%. However, both methods are based on the assumption that data lie in a linear subspace, which is often violated by online signature data characterized by complex biomechanical variations and inherently nonlinear manifold structures. As a result, linear projections may fail to adequately unfold the intrinsic data geometry, leading to overlap between genuine and skilled forgery clusters in the reduced space [24].

To overcome the limitations of linear methods, researchers have explored nonlinear manifold learning algorithms such as t-SNE (t-Distributed Stochastic Neighbor Embedding) for preserving local neighborhood relationships of data [25]. However, t-SNE often distorts the global structure of data and incurs high computational costs, making it difficult to meet the dual requirements of performance and efficiency for OSV systems.

Notably, Uniform Manifold Approximation and Projection (UMAP), proposed by McInnes et al. [26], addresses this limitation appropriately. By prioritizing the preservation of local neighborhood relationships alongside the integrity of global topological structures, UMAP has emerged as a superior nonlinear dimensionality reduction framework. The empirical success of UMAP in high-dimensional bioinformatics [27] underscores its potential for modeling the intricate biomechanical trajectories inherent in online signatures. Unlike t-SNE, which frequently prioritizes local manifold clusters at the cost of global connectivity, UMAP maintains a robust equilibrium between local and global perspectives. In the context of OSV, this topological fidelity is vital for delineating the fine-grained boundaries between genuine specimens and high-similarity skilled forgeries. Nevertheless, the application of UMAP within the OSV domain remains nascent, particularly regarding its synergy with hybrid statistical-DCT feature sets. This study addresses this gap by integrating UMAP-based manifold learning with multi-domain features, aiming to enhance the discriminative robustness and class separability of verification systems.

3. Datasets and Methods

This chapter elaborates on the construction of the proposed online signature verification system. The study adopts a three-stage processing architecture comprising feature extraction, dimensionality reduction, and classification. The overall technical roadmap is illustrated in Figure 1.

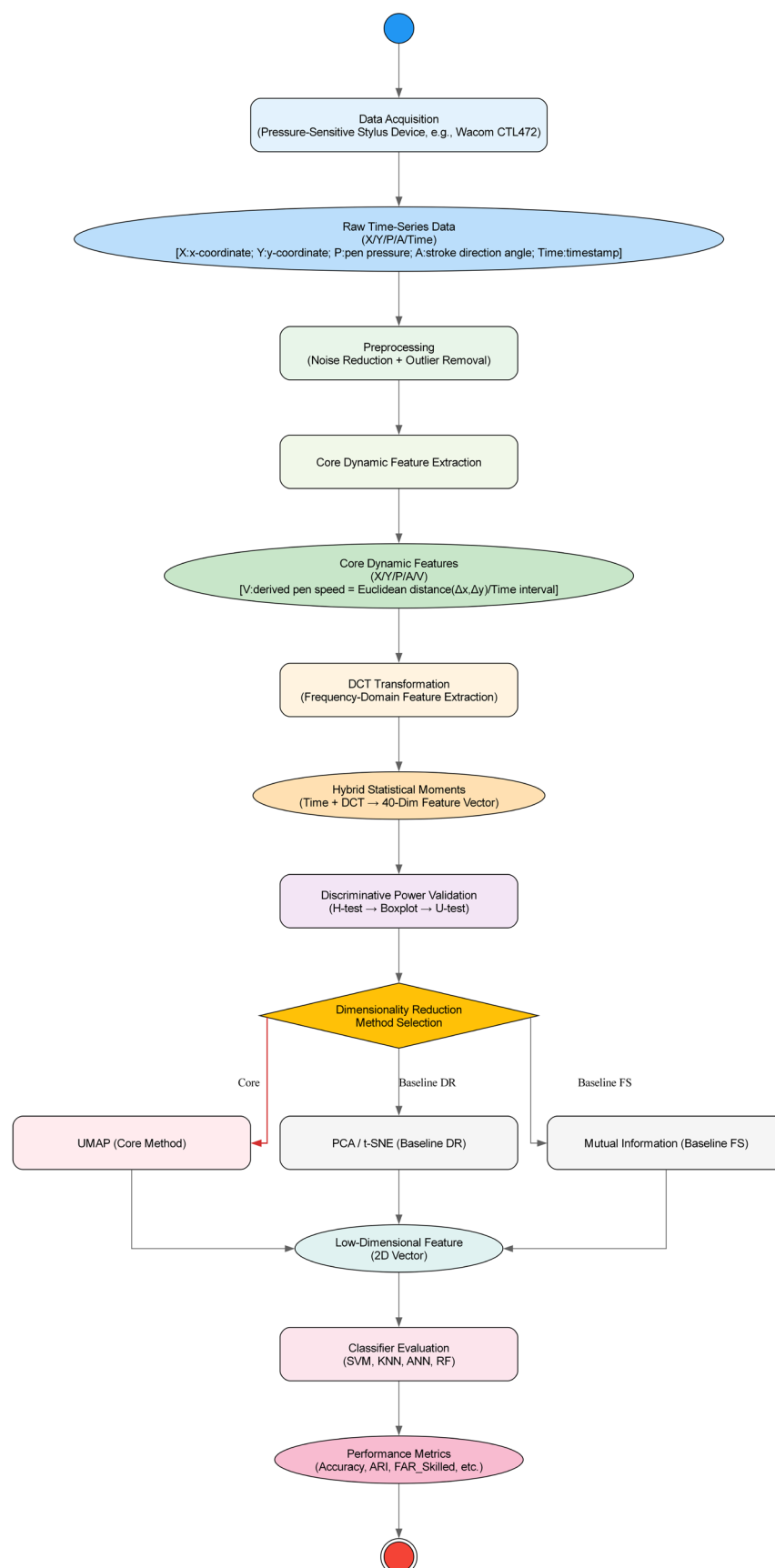


Figure 1. Overall workflow of the proposed online signature verification system: the blue circle denotes the start event of the workflow, the red circle denotes the end event, and the arrows represent the sequence flow indicating the direction of process execution.

3.1. Experimental Environment and Dataset Configuration

3.1.1. Experimental Configuration

All experiments were conducted in a workstation environment to ensure stable and reproducible execution. The system was equipped with an Intel® Core™ i9-14900HX CPU (2.20 GHz), 16 GB RAM, and an NVIDIA GeForce RTX 4060 GPU (8 GB GDDR6). The software environment was developed using Python 3.13.0, with all implementations executed under a unified environment. Key dependency libraries include NumPy 2.2.5 for numerical computation, Scikit-learn 1.7.0 for machine learning modeling, and UMAP-learn 0.5.9 for manifold learning dimensionality reduction.

3.1.2. Dataset

To ensure the targeted and practical value of the experimental results, a self-built online signature dataset was constructed in this study, as existing public datasets [28,29] lack sufficient samples for distinguishing multiple forgery types, particularly high-fidelity skilled forgeries produced through long-term imitation and repeated tracing. These forgery types are commonly encountered in real cases but are underrepresented in publicly available datasets. The dataset was collected in a controlled laboratory environment. Signature data were acquired using an online handwriting acquisition system based on a Wacom CTL-472 digitizing tablet (pressure sensitivity: 2048 levels) and an LP-190 digital pen, with a sampling rate of 100 Hz. The recorded parameters include x-coordinate, y-coordinate (resolution: 0.001 mm), pen pressure (0–1), stroke direction angle (-180° to $+180^\circ$), and time interval between consecutive sampling points (10 ms).

All participants were 6 adult volunteers who provided informed consent prior to data collection. Signature samples were anonymized and used exclusively for research purposes. No personally identifiable information was retained. According to institutional guidelines, this study did not require formal ethical approval. Participants sat at a standard office desk in a quiet laboratory setting, with the tablet angle adjusted to their natural writing habits to minimize postural variability during data acquisition.

The dataset comprises 800 valid signature samples, which were independently reviewed by two independent experts to exclude invalid cases (e.g., incomplete signatures, excessive pen breaks). Four categories (GEN, RF, TF, SF) [30] were defined based on authenticity and forgery methods, with a balanced distribution of 200 samples per category. Detailed characteristics of each category are summarized in Table 1, and typical examples of these categories are illustrated in Figure 2.

Table 1. Distribution and Generation Methods of Signature Samples.

Signature Category	Abbreviation	Generation Method	Sample Distribution
Genuine signature	GEN	Produced by 1 target participant (distinct from the 5 forgery writers, totaling 6 participants) across four independent sessions spaced one day apart, under natural conditions.	200 samples total (50 samples per session).
Random forgery signature	RF	Created by 5 forgery writers (distinct from the target participant) who observed the genuine signature and practiced briefly before execution.	200 samples total (40 samples per writer).
Tracing forgery signature	TF	Generated by the same 5 forgery writers (distinct from the target participant) who traced pre-printed genuine signatures placed under the digital tablet.	200 samples total (40 samples per writer).
Skilled forgery signature	SF	Produced by 3 out of the 5 forgery writers (distinct from the target participant, with prior handwriting imitation experience) after one week of intensive practice.	200 samples total (40, 80, 80 samples from three writers).

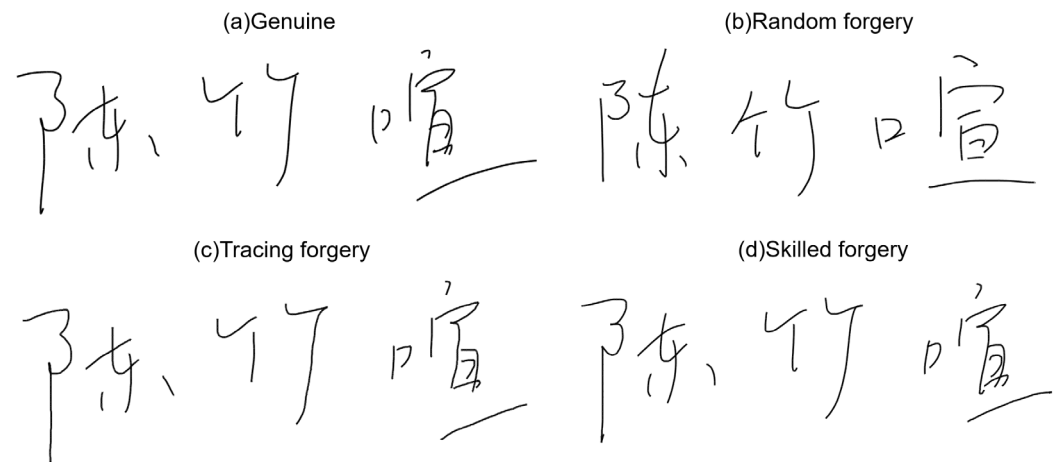


Figure 2. Examples of four signature categories based on authenticity and forgery methods.

The dataset was split into training (60%) and test (40%) sets via stratified sampling to maintain the same category distribution across sets. Fixed random seeds (1, 15, 70, 105, and 410) were used for the 5 independent data splits respectively to ensure the reproducibility of each split. Critical to avoiding data leakage, the training and test sets were kept fully independent throughout the experiment: dimensionality reduction models, clustering algorithms, and classifiers were exclusively trained on the training set, with no access to test set data during training. The test set was only used for final performance evaluation after all models were trained. Final reported metrics represent the average of results across the 5 independent splits to mitigate sampling bias.

Under this sample-level stratified splitting strategy, although the training and test sets remain independent at the model level, forged signatures produced by the same impersonator may appear in both sets. To assess the potential influence of this factor, we further considered a more restrictive partitioning scheme in which all forged samples generated by a given impersonator were assigned exclusively to either the training or the test set. This stricter setting increases the difficulty of the classification task, while the relative performance trends across feature configurations and classifiers remain consistent. Therefore, to ensure methodological consistency with commonly adopted evaluation practices in the literature, the main experimental results reported in this paper are based on the standard splitting protocol.

3.2. Dual-Domain Feature Extraction and Discriminative Power Verification

Feature extraction is a core component of signature verification, as it directly influences the discriminative capability of the system. This study focuses on dynamic behavioral characteristics of signatures and adopts an “extraction–verification” workflow: dual-domain features are first extracted, followed by statistical validation of their discriminative potential, to support reliable subsequent classification.

3.2.1. Dual-Domain Feature Extraction

Five core dynamic features were selected: X (x coordinates), Y (y coordinates), P (pen pressure), A (stroke direction angle), and V (derived pen speed, which is a critical behavioral marker not directly recorded by the acquisition device). Pen speed was calculated as the Euclidean distance between consecutive sampling points divided by the corresponding time interval:

$$V_i = \frac{\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}}{t_i - t_{i-1}} \quad (i \geq 2), \quad (1)$$

where V_i denotes the pen speed at the i -th sampling point, X_i and Y_i are coordinate values at the i -th point, and t_i is the timestamp of the i -th point.

The first four statistical moments were calculated for each feature sequence [31]. The choice of the first four moments ensures a comprehensive description of the feature probability distribution: mean and variance define the baseline style and its stability, while skewness and kurtosis are sensitive to non-linear irregularities. Higher-order moments are sensitive to asymmetry and peakedness in feature distributions and may reflect subtle irregularities introduced during the signing process.

First Moment (μ_1 , Mean): Central tendency.

$$\mu_1 = \frac{1}{n} \sum_{i=1}^n f_i, \quad (2)$$

where f_i denotes the feature value of the i -th sampling point; n is the total number of sampling points for the feature.

Second Moment (μ_2 , Variance): Dispersion.

$$\mu_2 = \frac{1}{n} \sum_{i=1}^n (f_i - \mu_1)^2, \quad (3)$$

Third Moment (μ_3 , Skewness): Degree of asymmetry.

$$\mu_3 = \frac{\frac{1}{n} \sum_{i=1}^n (f_i - \mu_1)^3}{(\mu_2)^{3/2}}, \quad (4)$$

Fourth Moment (μ_4 , Kurtosis): Degree of peakedness.

$$\mu_4 = \frac{\frac{1}{n} \sum_{i=1}^n (f_i - \mu_1)^4}{(\mu_2)^2} - 3 \quad (5)$$

Two complementary feature branches were constructed and fused into a 40-dimensional (40D) feature set:

Time-domain features (20D): 5 features $\times$ 4 moments. For each dynamic feature (e.g., pen pressure), μ_1 – μ_4 were calculated directly from its time-series data.

DCT-domain features (20D): Each time-series feature was transformed using the Discrete Cosine Transform (DCT) to extract frequency-domain structural information [32]. The DCT transformation formula is defined as [33]:

$$F(u) = \sqrt{\frac{2}{N}} c(u) \sum_{i=0}^{N-1} f(i) \cos \left[\frac{\pi(2i+1)u}{2N} \right] \quad (u=0,1,\dots,N-1), \quad (6)$$

where $f(i)$ denotes the original time-series feature value of the i -th sampling point, N denotes the total number of sampling points, $c(u)$ denotes the normalization factor ($c(u) = 1/\sqrt{2}$ if $u = 0$, and 1 otherwise), $F(u)$ denotes the DCT-domain coefficient ($u = 0$: DC component; larger u : high-frequency components), and $\cos \left[\frac{\pi(2i+1)u}{2N} \right]$ is the DCT basis function weighting $f(i)$'s contribution to $F(u)$.

As illustrated in Figure 3, the temporal trajectories of the online signature features (X , Y , P , V , A) and the amplitude distributions of their top 20% DCT coefficients are presented for one representative sample. Table 2 further quantifies the cumulative energy proportions of the top 10% (41 coefficients) and top 20% (82 coefficients) DCT coefficients for this sample. The results demonstrate that the top 20% DCT coefficients capture over 85% of the cumulative energy across all features: 100% for X and Y , 96.1% for P , 88.5% for V , and 87.6% for A . Notable differences in energy concentration are also observed among the features. These analyses demonstrate the high energy concentration characteristics of the signature's DCT-domain features, suggesting that retaining all DCT coefficients for subsequent statistical moment computation is a reasonable design choice.

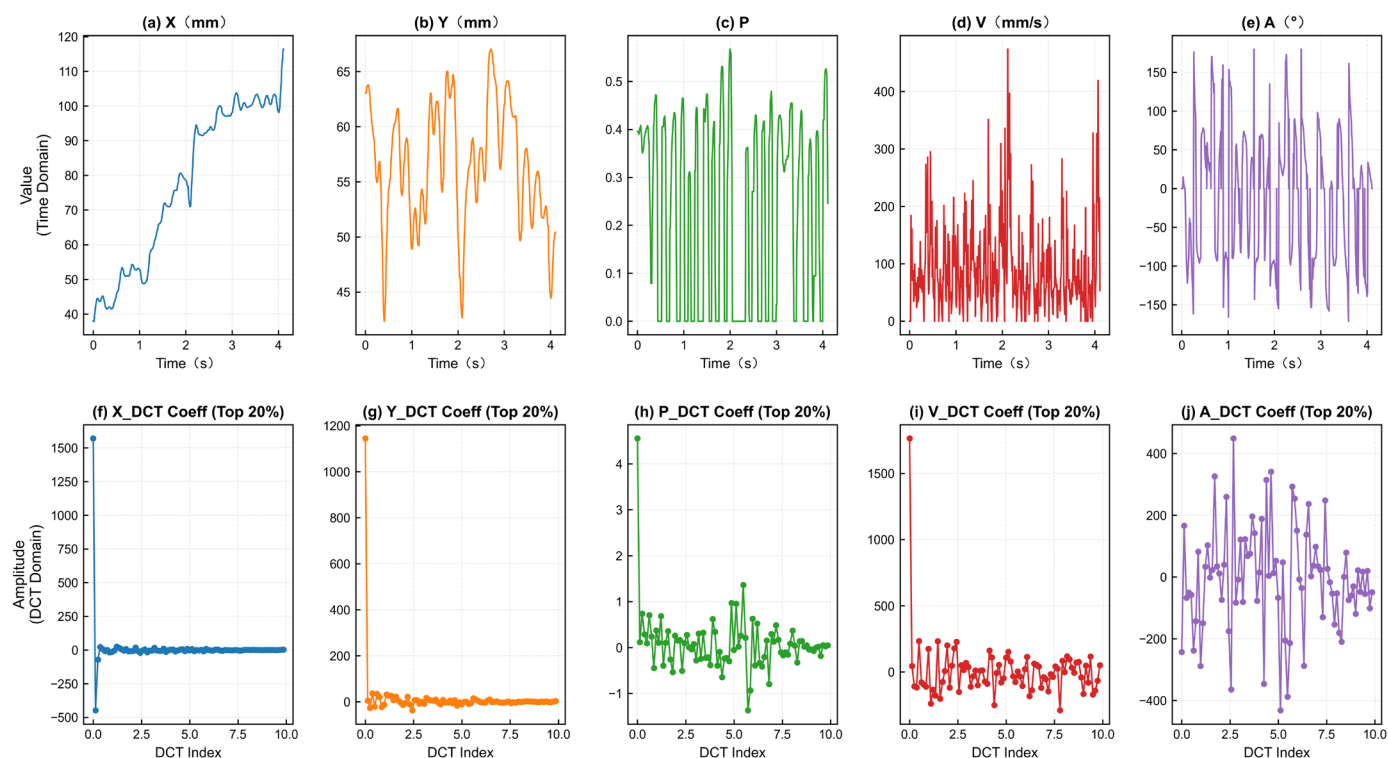


Figure 3. Representative time-domain trajectories and corresponding DCT-domain coefficient distributions of an online signature: (a–e) Time-domain signals of the five dynamic parameters (X, Y, pressure, velocity, and stroke angle); (f–j) Corresponding DCT-domain coefficient magnitude distributions.

Table 2. Cumulative energy proportions of leading DCT coefficients for different dynamic signature parameters.

Feature	Top 10% (41 pts)	Top 20% (82 pts)
X_DCT	100.0	100.0
Y_DCT	99.9	100.0
P_DCT	73.5	96.1
V_DCT	79.9	88.5
A_DCT	49.8	87.6

Note: “Top 10%/20%” corresponds to the first 10% (41) and 20% (82) DCT coefficients (ordered by frequency index). Values represent the cumulative energy proportion (%) for each signature parameter.

We did not filter the DCT coefficients and directly utilized all of them to calculate four statistical moments (μ_1 – μ_4), thereby constructing the final DCT-domain feature set for subsequent analysis.

3.2.2. Statistical Protocol for Feature Discriminability Validation

To examine the discriminative potential of the constructed 40D dual-domain features, a three-step validation framework integrating statistical testing and visualization was established. Prior to validation, the Shapiro–Wilk test confirmed that all features violated the normality assumption ($p < 0.05$), necessitating the adoption of non-parametric tests. The specific workflow is detailed as follows.

Kruskal–Wallis H-test (overall discriminability screening) [34]: This test was utilized to assess whether each feature exhibits significant distribution differences across the four

signature categories (GEN, RF, TF, SF). The judgment criterion was set as $p < 0.05$ ($\alpha = 0.05$), ensuring that features meet the fundamental requirement for inter-category differentiation.

Boxplot visualization (intuitive distribution verification): For features passing the Kruskal–Wallis H-test, boxplot visualization was employed to intuitively characterize their distribution differences across categories. Key observation indicators included the overlap degree of boxes (less overlap indicating stronger discriminability) and the separation of medians (larger median differences reflecting more distinct distribution patterns).

Mann–Whitney U-test (pairwise discriminability refinement) [35]: This test was applied to all six category pairs (GEN-RF, GEN-TF, GEN-SF, RF-TF, RF-SF, TF-SF) to verify pairwise discriminative efficacy. A feature was considered to exhibit consistent pairwise discriminability if statistical significance ($p < 0.05$) was observed in at least three category pairs.

This validation framework is designed to screen effective features and avoid information loss caused by premature pruning, thereby providing a reliable foundation for subsequent dimensionality reduction and classification tasks.

3.3. Dimensionality Reduction Method Selection and UMAP Parameter Configuration

Dimensionality reduction is a key step to mitigate the curse of dimensionality and provide compact representations for subsequent clustering and classification. Four representative techniques were considered for comparison: three projection-based methods (UMAP [26], t-SNE [25], PCA [22]) and one feature selection method (Mutual Information Feature Selection, MIFS). The selection was based on two core criteria aligned with experimental practice: (1) Preservation of category-related structural information, evaluated via qualitative visual inspection of embedded clusters and quantitative clustering metrics (ARI, DB, SC, CH); (2) Compatibility with downstream classification performance, assessed via the test accuracy of subsequent classifiers (to ensure reduced-dimensional features retain discriminative power for signature verification).

3.3.1. Dimensionality Reduction Evaluation Criteria

To evaluate the preservation of category-related structural information after dimensionality reduction, K-means clustering ($k = 4$, matching the number of signature categories) was adopted as a reference method. Four complementary clustering metrics were employed to quantitatively assess structural preservation efficacy: the Adjusted Rand Index (ARI) quantifies the consistency between clustering results and ground-truth labels, with higher values indicating better alignment (range: $[-1, 1]$); the Davies-Bouldin Index (DB) evaluates the balance between cluster compactness and inter-cluster separation, where lower values denote more well-separated and compact clusters (range: $[0, +\infty)$); the Silhouette Coefficient (SC) measures the trade-off between intra-cluster cohesion and inter-cluster separation, with higher values reflecting superior cluster structure (range: $[-1, 1]$); and the Calinski-Harabasz Index (CH) reflects the ratio of between-cluster variance to within-cluster variance, where larger values indicate more distinct cluster separation [36]. These metrics collectively provide a comprehensive benchmark for structural preservation, directly supporting UMAP parameter optimization (Section 3.3.2) and the comparative evaluation of dimensionality reduction methods (Section 4.2). Detailed calculation formulas for these metrics are provided in Section 3.5.1 to ensure reproducibility.

3.3.2. UMAP Parameter Configuration

To ensure reliable low-dimensional embedding of the extracted dual-domain features, UMAP hyperparameters were determined through systematic search on the training set only, avoiding any use of test data. Three key parameters were considered: $n_neighbors$, min_dist , and $n_components$, which respectively control the balance between local and global structure preservation, cluster compactness, and embedding dimensionality.

The search ranges were defined based on commonly adopted practices in prior studies [26], with $n_neighbors \in [1, 30]$, $min_dist \in [0.01, 0.80]$, and $n_components \in [2, 10]$. To prevent data leakage and simulate a realistic deployment, the UMAP manifold was learned exclusively from the training partition of each stratified split. This learned projection was then frozen and applied to map unseen test samples into the embedding space, ensuring that the feature transformation remained strictly out-of-sample and independent of the testing data.

Considering the emphasis on interpretability and computational efficiency in downstream classification, two-dimensional embeddings were adopted for all subsequent experiments. A detailed comparison between UMAP-2D and other dimensionality reduction methods is provided in Section 4.2.

3.4. Classification Model Construction

Four representative classifiers—KNN, SVM, ANN, and RFC—were constructed and evaluated based on the 2D UMAP-reduced features to examine the robustness of the proposed feature representation across different classification paradigms. The hyperparameters of all classifiers were optimized using grid search combined with 5-fold cross-validation on the training set, with cross-validation accuracy used as the optimization criterion. Core parameters, search ranges, and selected values are summarized in Table 3, together with brief configuration rationales. The parameter settings reported in Table 3 were applied consistently across all experiments to ensure fair comparison among classifiers.

Table 3. Hyperparameter settings of the evaluated classifiers.

Classifier Type	Core Hyperparameters	Selected Values	Configuration Details
KNN	Neighbors (K), Metric	$K = 7$, Euclidean	$K = 7$ was selected to effectively balance noise smoothing with the capture of local structural features.
SVM	Kernel, Penalty (C), Gamma (γ)	RBF, $C = 10$, $\gamma = 0.1$	The Radial Basis Function (RBF) kernel was chosen to handle non-linear decision boundaries. C and γ were optimized to enhance generalization.
ANN	Architecture, Activation, Learning Rate	Input = d , Hidden = $128/64$, Output = 4	Fully connected feedforward network with two hidden layers (128 neurons in the first, 64 in the second). The number of input nodes d scales dynamically with the reduced dimension. Hidden layers use ReLU activation to mitigate gradient vanishing, with a learning rate of 0.001.
RFC	Estimators, Max Depth	100 Trees, MaxDepth = 15	An ensemble of 100 decision trees. The maximum depth was limited to 15, and the feature sampling rate was set to 2/3 of total features to ensure diversity among base learners.

3.5. Evaluation Metrics and Protocols

Two groups of metrics comprehensively assess the framework's performance: clustering metrics (for dimensionality reduction) and classification/verification metrics (for signature verification).

3.5.1. Clustering Metrics (Consistent with Section 3.3.1)

Clustering metrics are used to evaluate the structural preservation of dimensionality-reduced features. Supplementary calculation formulas are provided below for reproducibility:

The Adjusted Rand Index (*ARI*) is calculated as:

$$ARI = \frac{2(TP \times TN - FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (7)$$

where *TP* denotes correctly clustered sample pairs, *TN* denotes correctly separated sample pairs, *FP* denotes falsely clustered pairs, and *FN* denotes falsely separated pairs.

The Davies–Bouldin Index (*DB*) is calculated as:

$$DB = \frac{1}{k} \sum_{i=1}^k \max \left(\frac{S_i + S_j}{d_{ij}} \right), \quad (8)$$

where *k* denotes number of clusters, *S_i* denotes average distance from samples in cluster *i* to cluster center, *S_j* denotes average distance from samples in cluster *j* to cluster center, *d_{ij}* denotes Euclidean distance between centers of cluster *i* and *j*.

The Silhouette Coefficient (*SC*) is calculated as:

$$SC = \frac{1}{N} \sum_{i=1}^N \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (9)$$

where *N* denotes total number of samples, *a_i* denotes average distance from sample *i* to other samples in the same cluster, *b_i* = minimum average distance from sample *i* to samples in other clusters.

The Calinski–Harabasz Index (*CH*) is calculated as:

$$CH = \frac{\frac{SSB}{k-1}}{\frac{SSW}{N-k}}, \quad (10)$$

where *SSB* denotes sum of squared distances between cluster centers and global center, *SSW* denotes sum of squared distances within each cluster, *k* denotes number of clusters, *N* denotes total number of samples.

3.5.2. Classification and Verification Metrics

These metrics assess signature verification performance, with macro-averaging adopted to ensure equal contribution from each class. *FAR_{Skilled}* and *FRR_{Genuine}* are calculated exclusively on the test set, and confusion matrices are used to visualize misclassification patterns.

Test Set Accuracy (*Acc*) quantifies overall classification correctness, calculated as:

$$Acc = \frac{\sum_{i=1}^4 TP_i}{N}, \quad (11)$$

where *TP_i* denotes true positives of category *i* in test set, *N* denotes total number of samples in test set.

Macro-Precision is the arithmetic mean of precision across all four categories, with the formula:

$$Pre_{macro} = \frac{1}{4} \sum_{i=1}^4 \frac{TP_i}{TP_i + FP_i}, \quad (12)$$

where *FP_i* denotes false positives of category *i*; average precision across all categories.

Macro-Recall represents the average recall across all categories, with recall for each class calculated as:

$$Rec_{macro} = \frac{1}{4} \sum_{i=1}^4 \frac{TP_i}{TP_i + FN_i}, \quad (13)$$

where FN_i denotes false negatives of category i ; average recall across all categories.

The *Macro-F1* Score balances precision and recall for each category, computed as the arithmetic mean of class-specific *F1* values:

$$F1_{macro} = \frac{1}{4} \sum_{i=1}^4 2 \times \frac{Pre_i \times Rec_i}{Pre_i + Rec_i}, \quad (14)$$

where Pre_i denotes precision of category i , Rec_i denotes recall of category i ; balances precision and recall for each category.

The Area Under ROC Curve (*AUC*) measures discriminative capability across decision thresholds, calculated via the trapezoidal rule with the One-vs-Rest strategy for multi-classification:

$$AUC = \frac{1}{2} \sum_{k=1}^{m-1} (x_{k+1} - x_k)(y_k + y_{k+1}), \quad (15)$$

where x_k denotes false positive rate (FPR) at the k -th threshold, y_k denotes true positive rate (TPR) at the k -th threshold, m denotes number of thresholds; calculated via trapezoidal rule, adopts One-vs-Rest for multi-classification.

$FAR_{Skilled}$ quantifies the probability of skilled forgeries being falsely judged as genuine, with the formula:

$$FAR_{Skilled} = \frac{FP_{SF}}{N_{SF}}, \quad (16)$$

where FP_{SF} denotes skilled forgery (*SF*) samples falsely judged as genuine (*GEN*), N_{SF} denotes total *SF* samples in test set.

$FRR_{Genuine}$ represents the probability of genuine signatures being falsely judged as forgery, calculated as:

$$FRR_{Genuine} = \frac{FN_{GEN}}{N_{GEN}}, \quad (17)$$

where FN_{GEN} denotes genuine (*GEN*) samples falsely judged as forgery, N_{GEN} denotes total *GEN* samples in test set.

4. Results and Discussion

4.1. Discriminability and Distribution of Dual-Domain Features

This section examines the discriminative behavior of the constructed 40-dimensional hybrid feature set (20 time-domain and 20 DCT-domain features) across the four signature categories (GEN, TF, RF, and SF), following the statistical validation protocol described in Section 3.2.2.

4.1.1. Global Significance and Rationale for Feature Inclusion

The Kruskal–Wallis H-test was independently applied to each of the 40 extracted features to evaluate whether their distributions differed significantly across the four signature categories. For all features, the asymptotic p -values were below 0.001, leading to rejection of the null hypothesis that the feature distributions are identical among classes. These results indicate that each extracted feature exhibits statistically distinguishable behavior across at least part of the signature categories under the adopted non-parametric criterion. Consequently, all 40 features are retained for subsequent distributional and pairwise analyses.

4.1.2. Time-Domain Feature Characteristics and Discriminative Efficacy

Time-domain features describe the spatiotemporal properties of the signature generation process, with boxplot distributions in Figure 4a–e revealing recurring class-dependent differences. Genuine signatures (GEN) are generally characterized by compact interquartile ranges and relatively concentrated median values, reflecting stable writing behavior. In contrast, random forgeries (RF) display wider interquartile ranges accompanied by a larger number of outliers, while tracing forgeries (TF) and skilled forgeries (SF) typically occupy intermediate positions in terms of dispersion and central tendency.

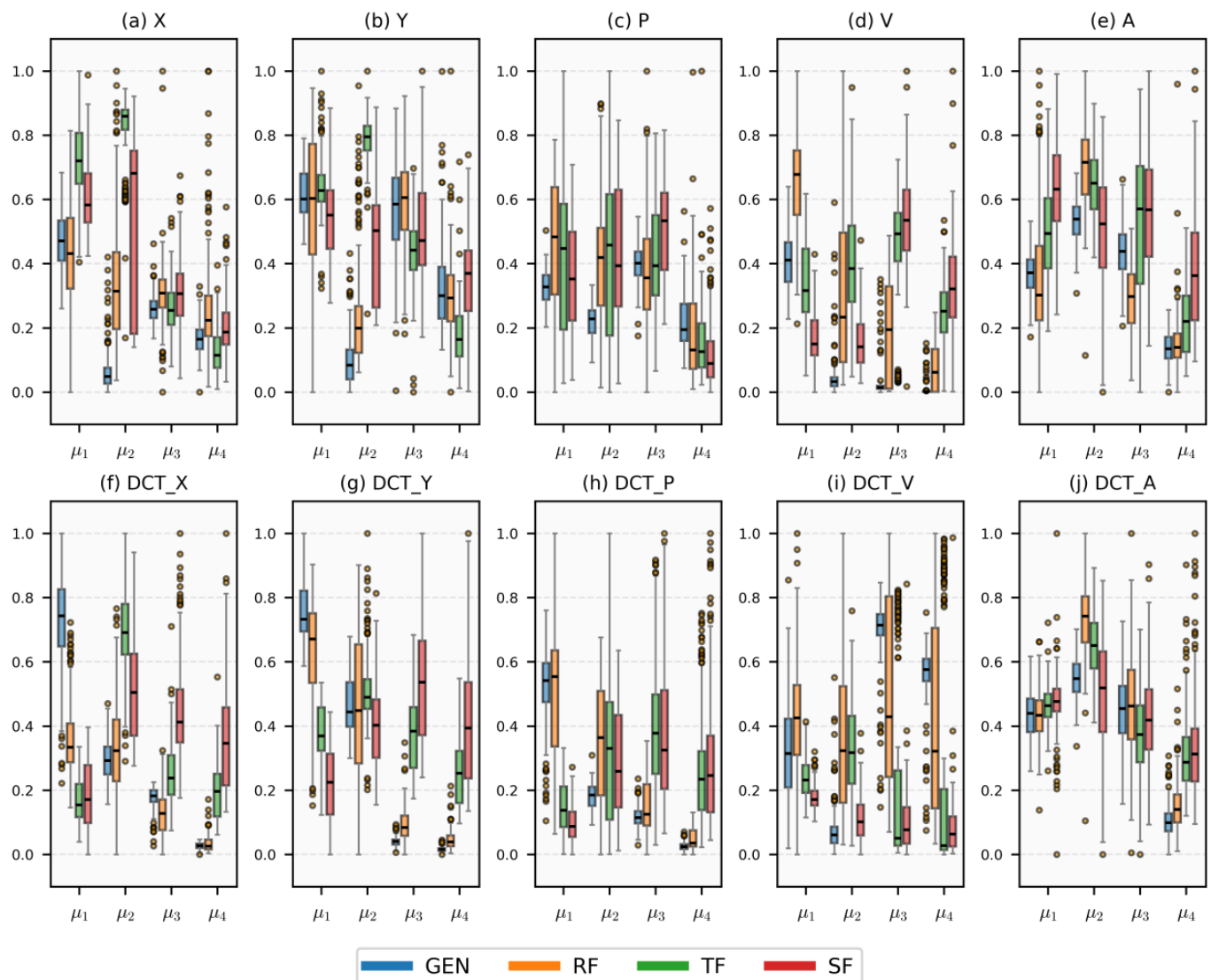


Figure 4. Boxplot distributions of the first four statistical moments for five dynamic signature parameters. Panels (a–e) show time-domain features, and panels (f–j) show the corresponding DCT-domain features. All features are normalized to [0, 1] for visualization.

For coordinate-related features, X- and Y-based moments show observable distribution shifts among categories. In particular, the second-order moment X_{μ_2} exhibits a gradual increase in median values from GEN to forged signatures, with TF and SF tending to occupy higher median ranges than RF and GEN. Although overlaps remain across classes, the interquartile ranges of X_{μ_2} and Y_{μ_2} differ sufficiently to provide useful separation cues, especially for distinguishing tracing forgeries from genuine signatures.

Pressure-related features demonstrate category-dependent distribution characteristics. As shown in Figure 4c, the first- and second-order moments ($P_{\mu 1}$ and $P_{\mu 2}$) of GENs are associated with lower medians and more compact boxplots, whereas forged signatures exhibit increased dispersion. Higher-order moments such as $P_{\mu 3}$ and $P_{\mu 4}$ further emphasize differences between RF and SF through shifts in median values and spread. However, substantial overlap across interquartile ranges persists among several classes.

Among the time-domain descriptors, velocity-related features exhibit relatively stable distributional structures across different moment orders. As illustrated in Figure 4d, GENs cluster around lower median values with narrow interquartile ranges for $V_{\mu 2}$ – $V_{\mu 4}$, whereas forged categories show broader distributions and higher variability. Although partial overlap persists, the reduced dispersion observed for GEN suggests that velocity-related moments reflect differences in writing rhythm stability across signature types.

Stroke direction angle features provide complementary information. As shown in Figure 4e, SF signatures generally exhibit higher median values than GEN and RF for several angular moments, while the fourth-order moment $A_{\mu 4}$ further differentiates GEN from forged categories through its lower median and reduced spread. However, the overlapping whiskers across categories indicate that angle-based features are most informative when combined with other dynamic descriptors rather than used independently.

4.1.3. DCT-Domain Feature Characteristics and Complementary Value

DCT-domain features describe the spectral characteristics of the signature signals and provide a complementary representation to time-domain statistics. As shown in Figure 4f–j, DCT-domain features exhibit class-dependent distribution patterns, with separation varying by feature type and moment order.

For coordinate-related DCT-domain features (Figure 4f,g), the first-order moments ($DCT_X_{\mu 1}$ and $DCT_Y_{\mu 1}$) show relatively higher median values for GENs, while forged categories occupy lower and partially overlapping ranges. Higher-order coordinate-related DCT moments ($\mu 3$ and $\mu 4$) exhibit compact interquartile ranges for GENs, with RF showing similarly concentrated distributions. In contrast, TF and SF display noticeably increased dispersion, indicating that these moments are more associated with increased spatial variability observed in tracing and skilled forgeries.

Pressure-related DCT-domain features (Figure 4h) display clearer differences in dispersion than in median levels. GENs tend to maintain more compact interquartile ranges across moments, while TF and SF show increased variability, especially in higher-order moments ($\mu 3$ and $\mu 4$). This behavior suggests that DCT-domain pressure moments emphasize variability patterns associated with forged writing behaviors.

As shown in Figure 4i, velocity-related DCT features exhibit order-dependent distribution patterns across signature categories. For the first-order moment ($\mu 1$), GEN and RF display comparable interquartile ranges, whereas TF and SF are more compact. In contrast, higher-order moments ($\mu 2$ – $\mu 4$) reveal a clear dispersion hierarchy, with RF showing the widest interquartile ranges, followed by TF and SF, while GEN remains consistently concentrated.

Angle-related DCT-domain features (Figure 4j) exhibit comparatively limited separability. Although slight median shifts can be observed for certain moments, substantial overlap persists across interquartile ranges for all four categories, indicating that angular information in the DCT domain mainly serves a supplementary role.

Overall, Figure 4f–j shows that DCT-domain features exhibit distributional tendencies distinct from their time-domain counterparts. Coordinate and velocity-related moments emphasize variability among forged signatures, while pressure and angle moments provide complementary information, highlighting their complementary statistical behavior relative to time-domain features.

4.1.4. Pairwise Discriminability: Mann–Whitney U-Test Results

To further refine the pairwise discriminative efficacy of the 40D dual-domain feature set, we conducted 240 independent Mann–Whitney U-tests, covering all six category pairs (GEN-RF, GEN-TF, GEN-SF, RF-TF, RF-SF, TF-SF) across each feature dimension. Figure 5 presents the key results of these tests, with detailed statistics for non-significant cases ($p \geq 0.05$) tabulated in Table 4. Overall, 201 tests (83.75%) reached extreme significance ($p < 0.01$), and an additional 9 tests (3.75%) achieved statistical significance ($0.01 \leq p < 0.05$), resulting in an aggregate significance rate of 87.5% (Figure 5a). Notably, each individual feature demonstrated statistically significant differences in no fewer than three category pairs, indicating that no feature exhibits uniform redundancy across comparison scenarios.

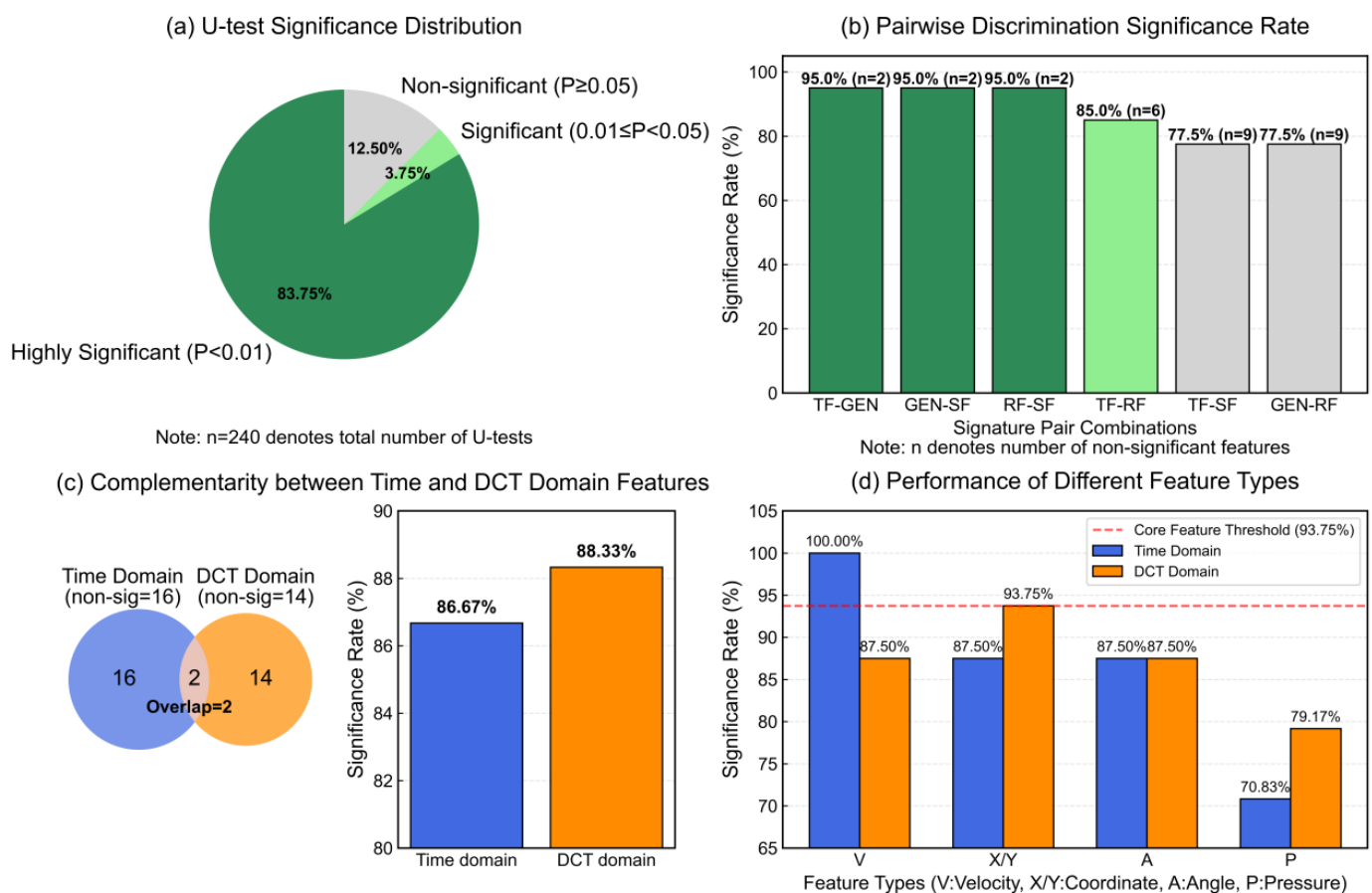


Figure 5. Results of pairwise Mann–Whitney U-tests for the 40-dimensional feature set. The analysis is based on 240 pairwise tests (6 signature pairs $\times$ 40 features): (a) Distribution of significance levels across all tests; (b) Relative discriminative difficulty of different signature pairs; (c) Complementarity between time-domain and DCT-domain features; (d) Contribution of different feature types across comparison scenarios.

Table 4. Statistics of non-significant features in the Mann–Whitney U-test for signature pairs with different discriminative difficulties ($p \geq 0.05$).

Pairing Difficulty Level	Signature Pairing Type	Non-Significant Features (Time/DCT Domain)	p-Value
Low Difficulty (95%)	TF-GEN	$P_{\mu 3}, X_{\mu 3}$	0.79, 1.00
	GEN-SF	$A_{\mu 2}, P_{\mu 1}$	0.25, 0.25
	RF-SF	$X_{\mu 3}, P_{\mu 2}$	0.81, 0.95

Medium Difficulty (85%)	TF-RF	$P_{\mu_1}, P_{\mu_2}, P_{\mu_4},$ $P_DCT_{\mu_2}, V_DCT_{\mu_2},$ Y_{μ_1}	0.12, 0.87, 0.3, 0.12, 0.31, 0.05
	TF-SF	$P_{\mu_2}, P_DCT_{\mu_2-\mu_4},$ $A_{\mu_3}, A_DCT_{\mu_3},$ $V_DCT_{\mu_3-\mu_4},$ $X_DCT_{\mu_1}$	0.77, 0.27, 0.16, 0.16, 0.2, 0.08, 0.57, 0.34, 0.45
High Difficulty (77.5%)	GEN-RF	$Y_{\mu_1}, Y_{\mu_3-\mu_4},$ $Y_DCT_{\mu_2}, A_{\mu_4},$ $A_DCT_{\mu_1}, A_DCT_{\mu_3},$ $P_DCT_{\mu_1}, X_DCT_{\mu_4}$	0.96, 0.21, 0.08, 0.91, 0.13, 0.86, 0.25, 0.85, 0.73

Based on the proportion of significant features, the six category pairs can be grouped into three levels of discriminative difficulty (Figure 5b and Table 4). Low-difficulty pairs, including TF–GEN, GEN–SF, and RF–SF, each achieved a 95% significance rate, with only two non-significant features per pair, reflecting limited distributional overlap between the corresponding classes. The TF–RF pair represents a medium-difficulty scenario, exhibiting an 85% significance rate with six non-significant features, suggesting partial overlap in certain dynamic attributes. High-difficulty pairs, namely TF–SF and GEN–RF, show comparatively lower but still substantial significance rates of 77.5%, corresponding to nine non-significant features per pair. Importantly, even in these cases, more than 30 feature dimensions remain statistically effective, ensuring robust pairwise discriminability.

The comparative analysis further reveals a complementary relationship between time-domain and DCT-domain features (Figure 5c). Time-domain features achieve an overall significance rate of 86.67%, including 101 extremely significant tests, while DCT-domain features reach 88.33%, with 100 extremely significant outcomes. Only two features are simultaneously non-significant across both domains, indicating limited redundancy. For example, in the challenging GEN–RF pairing, the DCT-domain angular first-order moment ($A_DCT_{\mu_1}$) is non-significant ($p = 0.86$), whereas its time-domain counterparts (A_{μ_1} to μ_3) exhibit extreme significance ($p < 0.001$), illustrating effective cross-domain compensation.

A category-wise examination of feature significance further highlights differentiated yet complementary roles among the five feature groups (Figure 5d). Velocity-related features demonstrate consistently high discriminative power, achieving full significance in the time domain and 87.50% significance in the DCT domain, thereby providing broad coverage across most comparison scenarios. Coordinate-based features (X/Y) yield overall significance rates of 87.5% and 83.33%, respectively, with non-significant cases primarily concentrated in high-difficulty pairings, yet still contributing stable spatial discrimination. Angle-related features maintain balanced significance rates of 87.50% across both domains, partially compensating for pressure-related limitations in medium-difficulty scenarios such as TF–RF. Pen pressure features exhibit comparatively lower overall significance (70.83% in the time domain and 79.17% in the DCT domain); however, they retain multiple extremely significant dimensions in the most challenging GEN–RF pairing, where other feature categories show reduced effectiveness.

Taken together, these results indicate that the discriminative contributions of individual feature groups vary across comparison scenarios, while their combined use mitigates pair-specific weaknesses. These results provide a statistical context for interpreting the behavior of different feature groups in subsequent analyses.

4.2. Assessment of Structural Preservation by Dimensionality Reduction

4.2.1. Visual Assessment of Embedded Structures

To intuitively compare the ability of different dimensionality reduction methods to preserve category-related structural information, Figure 6 presents two-dimensional embeddings of the 40-dimensional hybrid feature set generated by PCA, t-SNE, and UMAP. The displayed results correspond to the run whose clustering performance is closest to the five-run average. These embeddings provide a qualitative visualization of how different dimensionality reduction methods organize samples in a low-dimensional space.

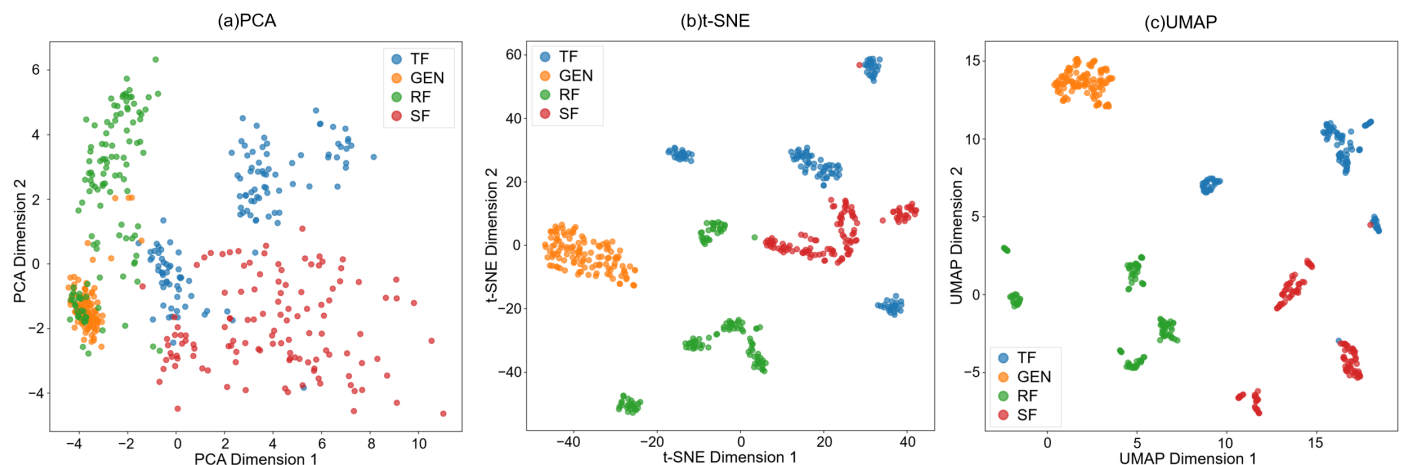


Figure 6. Two-dimensional embeddings of the hybrid feature set obtained using PCA, t-SNE, and UMAP.

The PCA embedding exhibits relatively broad and partially overlapping sample regions, in which genuine and forged signatures are not consistently separated, indicating limited class separability in the two-dimensional projection.

The t-SNE embedding exhibits several locally compact regions, suggesting improved local grouping for certain categories. However, noticeable class fragmentation and inconsistencies in the global arrangement are also observed, reducing overall structural coherence across classes.

In contrast, the UMAP embedding forms a compact cluster for genuine signatures and exhibits clearer spatial separation among different forgery types. Compared with PCA and t-SNE, the relative positions of clusters in the UMAP embedding appear more stable and less fragmented, indicating improved preservation of category-related structures.

It should be noted that these visualizations are intended for qualitative illustration only, as a single two-dimensional projection cannot fully represent the underlying high-dimensional topology or reflect the stability of a dimensionality reduction method across different runs. Accordingly, the visual assessment is interpreted in conjunction with the quantitative clustering results reported in the following section, which provide the primary evidence for evaluating structural preservation.

4.2.2. Quantitative Evaluation of Clustering Performance

To quantitatively assess the ability of different dimensionality reduction methods to preserve category-related structural information, clustering performance was evaluated in the two-dimensional embedding space. The results, summarized in Table 5, are reported as mean $\pm$ standard deviation over five independent runs under identical experimental settings.

Table 5. Clustering performance of PCA, t-SNE, and UMAP in the two-dimensional embedding space on the training set (mean $\pm$ SD over five runs).

Dimensionality Reduction Method	Output Dimensionality	ARI $\uparrow$	SC $\uparrow$	DB $\downarrow$	CH $\uparrow$
PCA	2	0.441 ± 0.009	0.523 ± 0.006	0.742 ± 0.018	589.5 ± 27.2
t-SNE	2	0.651 ± 0.065	0.498 ± 0.031	0.748 ± 0.019	568.4 ± 95.6
UMAP	2	0.988 ± 0.005	0.637 ± 0.012	0.582 ± 0.017	1068.1 ± 67.5

Note: $\uparrow$ indicates higher values are better; $\downarrow$ indicates lower values are better.

UMAP achieves the highest clustering performance with the lowest variability across runs, yielding an Adjusted Rand Index (ARI) of 0.988 ± 0.005 , a Silhouette Coefficient (SC) of 0.637 ± 0.012 , a Davies–Bouldin Index (DB) of 0.582 ± 0.017 , and a Calinski–Harabasz Index (CH) of 1068.1 ± 67.5 . The high ARI indicates strong agreement between cluster assignments in the reduced space and ground-truth labels, while the SC and DB values reflect compact intra-class structures and effective inter-class separation, respectively, and the high CH index confirms well-defined cluster organization.

t-SNE demonstrates intermediate clustering performance, with an ARI of 0.651 ± 0.065 . Although its average clustering metrics exceed those of PCA, the comparatively larger standard deviations indicate greater variability across runs, suggesting sensitivity to data partitioning and initialization.

PCA yields the weakest clustering performance, with an ARI of 0.441 ± 0.009 and comparatively lower SC (0.523 ± 0.006) and CH (589.5 ± 27.2) values, indicating limited effectiveness in preserving category-related structures in the two-dimensional space.

Overall, the quantitative results indicate that UMAP provides more effective preservation of both local continuity and global class separation than PCA and t-SNE for the proposed hybrid feature representation. These properties are achieved within a compact two-dimensional embedding, which motivates its use in the subsequent classification and verification experiments.

4.3. Comprehensive Evaluation of Core Components and Classifier Performance

This section presents a comprehensive evaluation of the interaction between different core components (including dimensionality reduction/feature selection strategies, transform domain selection schemes) and four classifiers (KNN, SVM, ANN, and RFC) in the multi-class online signature verification task. All reported results are obtained from five independent stratified train–test splits, in which models are trained exclusively on the training set and performance is averaged over the corresponding held-out test sets.

4.3.1. Sensitivity of Feature Dimension and Test Accuracy Trends

Figure 7 compares the average test accuracy of four classifiers under four dimensionality reduction or feature selection methods (PCA, t-SNE, MIFS, and UMAP) as the feature dimension varies from 2 to 10. The results reveal distinct performance trends and differing sensitivity to feature dimensionality across methods.

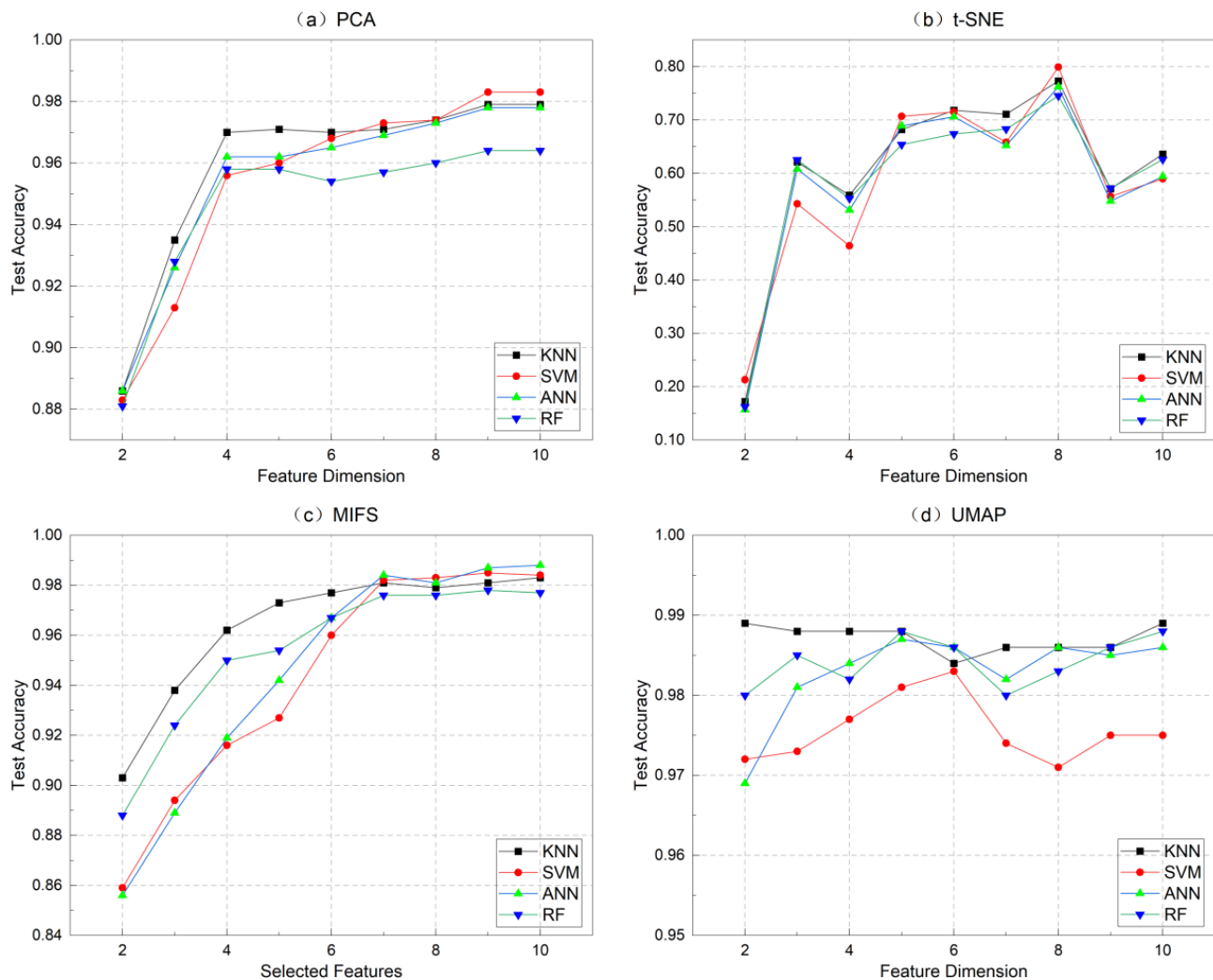


Figure 7. Comparison of average test accuracy as a function of feature dimension (2–10) for different feature selection and dimensionality reduction methods: (a) PCA (linear dimensionality reduction), (b) t-SNE (nonlinear dimensionality reduction), (c) MIFS (mutual information–based feature selection), and (d) UMAP (nonlinear dimensionality reduction).

Under PCA, classification accuracy improves progressively as the feature dimension increases. At dimension 2, the accuracy of representative classifiers (e.g., KNN) remains relatively limited (0.886), but increases steadily with dimensionality, reaching 0.979 for KNN at dimension 9 before gradually saturating. This trend suggests that PCA requires a sufficiently large subspace to retain discriminative information relevant for classification.

t-SNE exhibits pronounced instability across different embedding dimensions. At dimension 2, classifier accuracy drops to very low levels (e.g., KNN at 0.172), increases sharply by dimension 6 (KNN at 0.718), and then fluctuates at higher dimensions, decreasing to 0.636 for KNN at dimension 10. The resulting non-monotonic trend, together with increased variability across classifiers, limits the reliability of t-SNE for quantitative classification analysis in this setting.

MIFS demonstrates a comparatively smooth and stable performance trend. Classification accuracy starts at 0.903 for KNN at dimension 2, increases consistently as additional features are selected, and reaches a plateau by dimension 10 (e.g., KNN at 0.983). In this plateau regime, performance differences among classifiers narrow to marginal levels (within 0.01 across all four models), indicating that mutual information–based feature selection preserves discriminative characteristics effectively.

In contrast, UMAP exhibits a distinctly different behavior. Even at dimension 2, all classifiers achieve high and consistent test accuracy (e.g., KNN at 0.989 and RF at 0.980). Increasing the dimensionality beyond two results in only negligible performance changes—for example, KNN maintains an accuracy of 0.989 at dimension 10—with no systematic improvement observed across classifiers. This behavior indicates that UMAP captures the essential discriminative structure of the data within a compact embedding.

Based on these observed trends, the 2D UMAP embedding is selected for subsequent classifier performance analysis, as it provides a balance between representational compactness and classification effectiveness.

4.3.2. Detailed Classifier Performance in UMAP-2D Space

Based on the dimensionality sensitivity analysis in Section 4.3.1, the UMAP 2D embedding is fixed, and the performance of four classifiers is further examined using a comprehensive set of evaluation metrics. In addition to standard multi-class indicators, two security-oriented metrics— FAR_{Skilled} and FRR_{Genuine} —are introduced to explicitly assess the system’s robustness against high-skilled forgery attacks and its usability for legitimate users.

As summarized in Table 6, KNN consistently achieves the highest overall performance among the evaluated classifiers, attaining a test accuracy of 0.989 ± 0.005 together with high Macro-Precision (0.992 ± 0.003), Macro-Recall (0.988 ± 0.004), and Macro-F1 Score (0.990 ± 0.003). This indicates balanced classification behavior across all four signature categories, while the corresponding Cohen’s Kappa coefficient reaches 0.985 ± 0.006 , confirming a high level of agreement beyond chance.

Table 6. Classification performance metrics of four classifiers evaluated on the test set in the selected UMAP two-dimensional embedding space (mean $\pm$ SD over five runs).

Classifier	Dimension	Test Accuracy	Macro-Precision	Macro-Recall	Macro-F1 Score	Cohen’s Kappa	FAR_{Skilled}	FRR_{Genuine}
ANN	2	0.969 ± 0.023	0.970 ± 0.022	0.969 ± 0.023	0.969 ± 0.023	0.958 ± 0.031	0.010 ± 0.008	0.031 ± 0.023
RFC	2	0.980 ± 0.011	0.981 ± 0.011	0.980 ± 0.011	0.980 ± 0.011	0.973 ± 0.015	0.007 ± 0.004	0.020 ± 0.011
SVM	2	0.972 ± 0.032	0.975 ± 0.024	0.972 ± 0.032	0.971 ± 0.033	0.963 ± 0.042	0.009 ± 0.011	0.028 ± 0.032
KNN	2	0.989 ± 0.005	0.989 ± 0.005	0.989 ± 0.005	0.989 ± 0.005	0.985 ± 0.006	0.004 ± 0.002	0.011 ± 0.005

The RFC classifier achieves the second-best overall performance with strong accuracy and stability across splits, though it remains slightly inferior to KNN in all evaluated metrics. In comparison, ANN and SVM exhibit lower average performance and larger standard deviations, indicating reduced robustness under the same low-dimensional representation.

The advantage of KNN becomes more pronounced when security-oriented metrics are considered. Specifically, KNN achieves the lowest FAR_{Skilled} value of 0.004 ± 0.002 , indicating a substantially reduced likelihood of incorrectly accepting high-skilled forgeries as genuine signatures. At the same time, its FRR_{Genuine} value of 0.011 ± 0.005 is also the lowest among all classifiers, reflecting improved system usability by minimizing the rejection of legitimate signatures.

These results can be interpreted in relation to the characteristics of the UMAP embedding. By preserving local neighborhood relationships in the low-dimensional space, UMAP forms compact clusters for samples belonging to the same class while maintaining clear separation between different classes. In such a representation, the local, distance-based decision mechanism of KNN can directly exploit the geometric structure of the data, enabling more reliable discrimination between genuine signatures and visually similar high-skilled forgeries.

Overall, the experimental results indicate that the UMAP 2D representation provides an effective and compact embedding for multi-class online signature verification. Within this fixed low-dimensional space, KNN exhibits the most favorable balance between classification accuracy, robustness, and security-related performance, supporting its adoption as the final classifier in this study.

4.3.3. Comparative Verification of Transform Domain Selection (DCT vs. DST vs. DWT)

To verify the rationality of DCT domain selection, only the transform domain was replaced (other experimental conditions remained consistent) based on the optimal UMAP 2D + KNN model, comparing the performance of DCT, DST, and DWT. As shown in Table 7, DCT is significantly optimal across core metrics; DWT is completely impractical due to its extremely high false rejection rate of genuine signatures (35.8%). DCT's advantage stems from its energy concentration characteristic (consistent with the analysis in Section 3.2.1), which effectively preserves the global discriminative information of signatures. In contrast, DST is sensitive to local noise in signature trajectories, and DWT focuses on local time-frequency details—both fail to capture the global nonlinear behavioral differences between genuine signatures and forgeries. Based on these experimental results, DCT is confirmed as the most suitable choice for feature transformation in this study.

Table 7. Performance Comparison of Different Transform Domains (Based on UMAP 2D + KNN Model).

Transform Domain	Test Accuracy	Macro-Precision	Macro-Recall	Macro-F1 Score	Cohen's Kappa	FAR _{Skilled}	FRR _{Genuine}
DST	0.917 ± 0.013	0.917 ± 0.013	0.911 ± 0.013	0.911 ± 0.013	0.881 ± 0.017	0.032 ± 0.010	0.028 ± 0.009
DWT	0.770 ± 0.025	0.773 ± 0.027	0.770 ± 0.025	0.769 ± 0.026	0.693 ± 0.033	0.098 ± 0.020	0.358 ± 0.090
DCT	0.989 ± 0.005	0.989 ± 0.005	0.989 ± 0.005	0.989 ± 0.005	0.985 ± 0.006	0.004 ± 0.002	0.011 ± 0.005

4.4. Comparison with Baseline Methods

To verify whether the proposed method outperforms conventional techniques in the OSV field, we selected a suite of representative baseline models for comparative assessment. These baselines span three major categories: traditional statistical models, deep learning models, and metric-learning-based approaches. We implemented all baseline methods with standard Python open-source libraries (Scikit-learn and PyTorch), using the default hyperparameter configurations provided therein. No additional parameter tuning was performed, which ensured consistency across models and minimized potential biases in the comparison.

The selected baseline methods are listed as follows:

- Gaussian Mixture Model (GMM) with 5 mixture components, a default configuration for OSV-related tasks in the corresponding library;
- Hidden Markov Model (HMM) with 4 hidden states, the standard setup for sequence modeling tasks;
- Convolutional Neural Network (CNN) consisting of two hidden layers (64 and 32 neurons respectively), with ReLU as the activation function;
- Siamese LSTM, configured with 64 hidden units and optimized by the Adam optimizer;
- Triplet LSTM, equipped with 64 hidden units and a margin parameter set to 1.0.

We evaluated all models—including the proposed UMAP + KNN framework—under strictly identical experimental conditions. These conditions cover three key aspects: consistent dataset partitioning (60% for training and 40% for testing) and a unified set of evaluation metrics. A summary of the comparative results is presented in Table 8.

Table 8. Performance Comparison with Baseline Models.

Model	Test Accuracy	Macro-Precision	Macro-Recall	Macro-F1 Score	Cohen's Kappa	FAR _{Skilled}	FRR _{Genuine}
GMM	0.983 ± 0.013	0.983 ± 0.012	0.983 ± 0.013	0.983 ± 0.013	0.965 ± 0.026	0.010 ± 0.012	0.033 ± 0.084
HMM	0.981 ± 0.009	0.981 ± 0.009	0.981 ± 0.009	0.981 ± 0.009	0.963 ± 0.018	0.015 ± 0.020	0.032 ± 0.100
CNN	0.988 ± 0.003	0.988 ± 0.003	0.988 ± 0.003	0.988 ± 0.003	0.973 ± 0.006	0.006 ± 0.010	0.016 ± 0.002
Siamese LSTM	0.655 ± 0.054	0.655 ± 0.054	0.655 ± 0.054	0.655 ± 0.054	0.310 ± 0.108	0.345 ± 0.054	0.315 ± 0.042
Triplet LSTM	0.981 ± 0.014	0.981 ± 0.014	0.981 ± 0.014	0.981 ± 0.014	0.963 ± 0.027	0.0188 ± 0.014	0.022 ± 0.011
Proposed UMAP + KNN	0.989 ± 0.005	0.989 ± 0.005	0.989 ± 0.005	0.989 ± 0.005	0.985 ± 0.006	0.004 ± 0.002	0.011 ± 0.005

Results in Table 6 demonstrate that the proposed UMAP + KNN method yields the optimal overall performance across all evaluation metrics. Statistical baseline models (GMM and HMM) attain relatively high accuracy, yet they exhibit elevated false acceptance rates for skilled forgeries, revealing their inherent limitations in fine-grained forgery type discrimination. Deep learning and metric-learning baselines show uneven performance under the multi-class classification setting, while several of these models present notable confusion when distinguishing between visually similar forgery types. In contrast, the proposed method consistently maintains low false acceptance rates for skilled forgeries and low false rejection rates for genuine signatures, along with reduced performance variance in repeated evaluations.

These findings indicate that the performance advantage of the proposed method stems primarily from two core designs: the effective fusion of dual-domain statistical features and the capacity of UMAP to capture nonlinear structures within the feature space.

4.5. Security Validation and Error Analysis of the Optimal Model

Building upon the performance comparison presented in Section 4.3, this section conducts a security-oriented evaluation and error analysis of the selected model, namely the KNN classifier operating in the UMAP-2D space. It should be emphasized that the multi-class classification in this study assists identity verification through risk stratification of SF/TF/RF, enhancing the practical value of the OSV system. While averaged performance metrics indicate strong overall effectiveness, class-level analysis is further required to assess robustness against high-similarity forgeries and to identify representative error patterns.

4.5.1. Multi-Class ROC Analysis

To evaluate the discriminative behavior of the proposed model across different decision thresholds, a multi-class one-vs-rest ROC analysis was conducted on the test set. Figure 8 illustrates the ROC curves averaged over five independent test splits for the four signature categories, with corresponding AUC values of 0.995 for GEN vs. Rest, 0.991 for SF vs. Rest, 0.990 for TF vs. Rest, and 0.989 for RF vs. Rest.

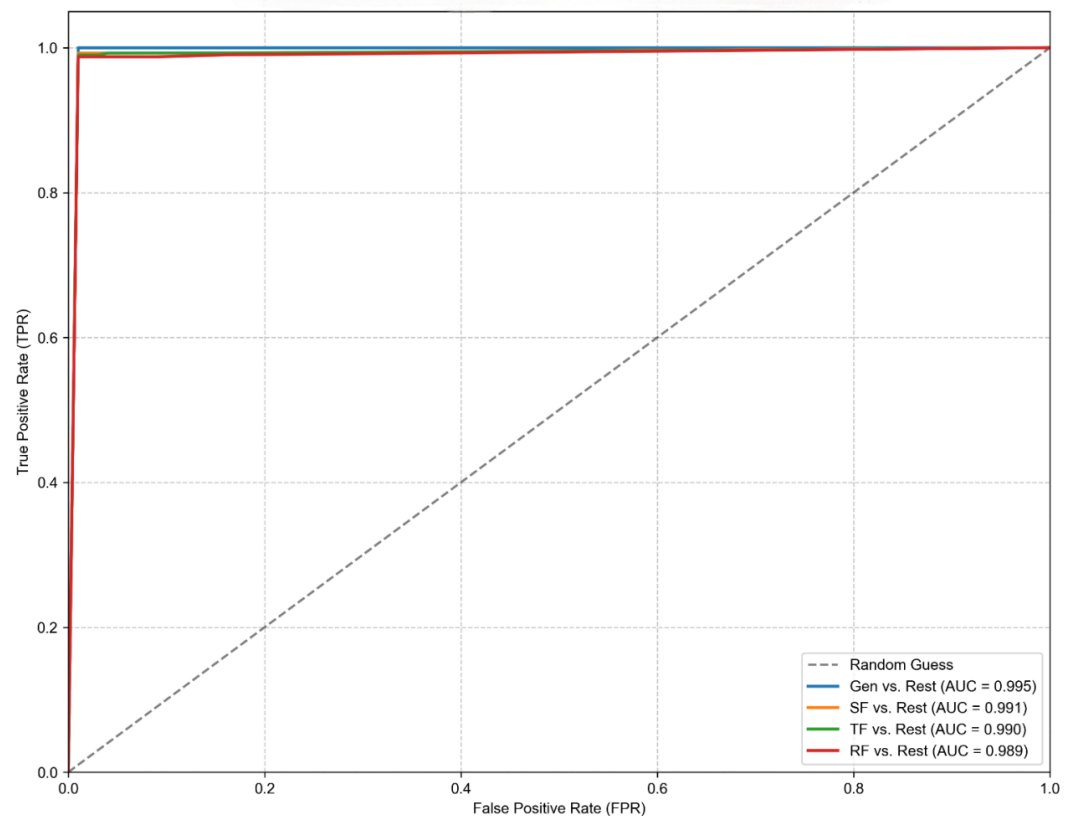


Figure 8. Mean one-vs-rest ROC curves of the KNN classifier evaluated in the UMAP two-dimensional embedding space.

Consistently high AUC values across all classes indicate strong separability in the learned low-dimensional space. In particular, the near-perfect AUC value (0.995) observed for the GEN class indicates a clear separation between genuine signatures and forgery samples across a wide range of decision thresholds, which is relevant for reducing false acceptance risk.

The similarly high AUC values observed for RF, SF, and TF suggest that the UMAP-based embedding retains discriminative structure among forgery subclasses that are visually and dynamically similar. Notably, low false acceptance rates for skilled forgeries are consistently observed across different classifier architectures. This observation suggests that the discriminative structure captured by the UMAP-based low-dimensional representation is not strongly dependent on a specific classifier.

4.5.2. Confusion Matrix and Robustness Against High-Similarity Forgeries

To complement the threshold-independent ROC analysis with an interpretable instance-level inspection, a confusion matrix from a representative test split is analyzed, as shown in Figure 9. This split was selected because its overall accuracy (0.988) is closest to the five-run average accuracy reported in Section 4.3 (0.989), thereby reflecting typical model behavior rather than an extreme or outlier case.

Each class in the test split contains 80 samples. All Gen samples in this split are correctly classified, resulting in a recall of 1.000 for the Gen class. This indicates that no false rejections of genuine signatures occur in this representative evaluation. Likewise, no forgery samples are misclassified as Gen in this split, indicating the absence of false acceptance events in this specific evaluation. Misclassifications are limited to interactions among forgery subclasses, where three TF samples are incorrectly predicted as RF and one as SF.

Based on this confusion matrix, the class-wise recall values are 1.000 for Gen, RF, and SF, and 0.950 for TF, resulting in a macro-averaged recall of 0.9875 for this split. The corresponding precision values are 1.000 (Gen), 0.9639 (RF), 0.9877 (SF), and 1.000 (TF). These values yield F1-scores of 1.000, 0.982, 0.994, and 0.974, respectively. It should be emphasized that these class-level metrics are derived from a single representative test split and are reported to illustrate typical error patterns rather than to replace the averaged performance statistics presented earlier.

Across all five independent test splits (400 Gen samples in total), no Gen signature is misclassified as any forgery type. Only a single RF sample is misclassified as Gen in one split. This observation indicates strong resistance to false acceptance while maintaining high usability for legitimate users across repeated evaluations.

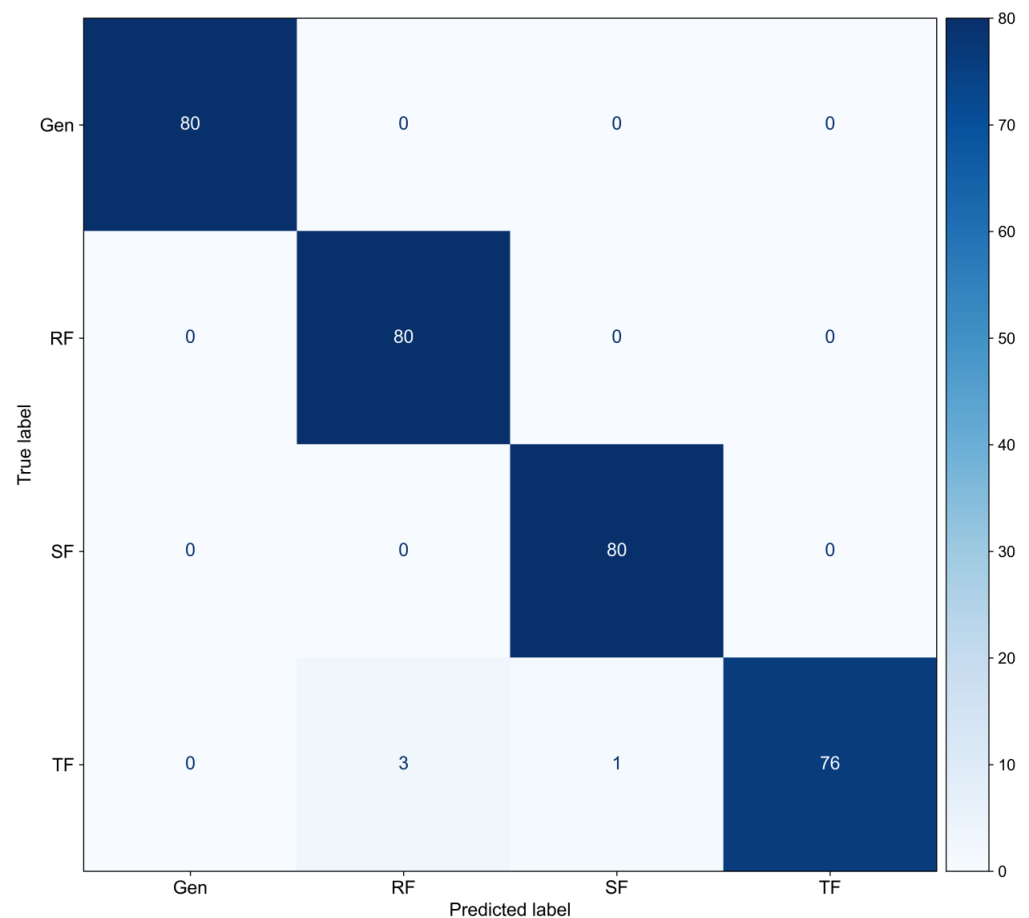


Figure 9. Confusion matrix of a representative test split for the KNN classifier in the UMAP two-dimensional embedding space.

5. Conclusions

This study examined a multi-class online signature verification framework that integrates multi-domain statistical feature representation with nonlinear dimensionality reduction. By jointly exploiting statistical moments derived from dynamic signature signals and their DCT-domain counterparts, the proposed feature representation captures complementary temporal and spectral characteristics relevant to distinguishing genuine signatures from different types of forgeries. This characteristic supports the applicability of the proposed framework to practical OSV scenarios where reliable and fine-grained forgery-type identification is required, particularly under conditions involving skilled and high-fidelity forgeries.

Experimental results confirm that UMAP effectively preserves category-related structural relationships in a compact low-dimensional space, enabling stable verification performance and consistent class separation across multiple classifiers. This structure-preserving property is particularly beneficial in skilled forgery scenarios, where high inter-class similarity often undermines the effectiveness of conventional linear or single-domain verification approaches.

Beyond quantitative performance, the resulting low-dimensional representations also provide interpretable visual structures that facilitate the analysis of inter-class relationships and representative error patterns. Future work will focus on validation using larger and publicly available datasets, to examine the transferability of the proposed framework to broader acquisition conditions and populations, as well as assessing the generalizability of compact embeddings under varying acquisition conditions and application settings.

Author Contributions: Conceptualization, L.H.; methodology, L.H. and Y.R.; software, L.H. and Y.R.; validation, L.H., Y.R. and W.L.; formal analysis, L.H.; investigation, L.H., N.X. and P.Z.; resources, Y.R.; data curation, Y.R.; writing—original draft preparation, L.H. and Y.R.; writing—review and editing, N.X., P.Z. and W.L.; visualization, L.H. and Y.R.; supervision, L.H.; project administration, L.H. and W.L.; funding acquisition, L.H. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Funds for the Innovation of Policing Science and Technology, Fujian Province, Grant number: 2024Y0069&2024Y0072, and the Science and Technology Special Project for Provincial Units in 2023 from the Department of Finance of Fujian Province, Grant number: Mincai Zhi [2023] 891.

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

Data Availability Statement: The dataset supporting the conclusions of this article is publicly available at the Gitee repository: <https://gitee.com/bai-minan/FJPC-OSig-MCF-DB> (accessed on 29 January 2026).

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

Abbreviations

ANN	Artificial Neural Network
ARI	Adjusted Rand Index
AUC	Area Under ROC Curve
CH	Calinski–Harabasz Index
CNN	Convolutional Neural Network
DCT	Discrete Cosine Transform
DB	Davies–Bouldin Index
DST	Discrete Sine Transform
DWT	Discrete Wavelet Transform
FAR _{Skilled}	False Acceptance Rate for Skilled Forgeries
FN	False Negative
FRR _{Genuine}	False Rejection Rate for Genuine Signatures
GEN	Genuine Signature
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
KNN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LSTM	Long Short-Term Memory
MIFS	Mutual Information Feature Selection

OSV	Online Signature Verification
PCA	Principal Component Analysis
RFC	Random Forest Classifier
RF	Random Forgery Signature
SC	Silhouette Coefficient
SF	Skilled Forgery Signature
SSB	Sum of Squared Distances Between Cluster Centers and Global Center
SSW	Sum of Squared Distances Within Each Cluster
SVM	Support Vector Machine
TF	Tracing Forgery Signature
TN	True Negative
TP	True Positive
t-SNE	t-Distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection

References

1. Liu, R.; Xin, Y. Online Handwritten Signature Verification Method Based on Uni-Feature Correlation Coefficient between Signatures. *Sensors* **2023**, *23*, 9341. <https://doi.org/10.3390/s23239341>.
2. Cantó Navarro, E.; Ramos Lara, R.; López García, M. Online Signature Verification Systems on a Low-Cost FPGA. *Appl. Sci.* **2022**, *12*, 378. <https://doi.org/10.3390/app12010378>.
3. Chen, M.; Lin, J.; Zou, Y.; Wu, K. Acoustic Sensing Based on Online Handwritten Signature Verification. *Sensors* **2022**, *22*, 9343. <https://doi.org/10.3390/s22239343>.
4. Kurowski, M.; Sroczynski, A.; Bogdanis, G.; Czyżewski, A. An Automated Method for Biometric Handwritten Signature Authentication Employing Neural Networks. *Electronics* **2021**, *10*, 456. <https://doi.org/10.3390/electronics10040456>.
5. Zhou, Y.; Zheng, J.; Hu, H.; Wang, Y. Handwritten Signature Verification Method Based on Improved Combined Features. *Appl. Sci.* **2021**, *11*, 5867. <https://doi.org/10.3390/app11135867>.
6. Chavan, M. Handwritten Online Signature Verification and Forgery Detection Using Hybrid Wavelet Transform-2 with HMM Classifier. *Journal of Electrical Systems* **2024**, *20*(4s), 2463–2470. <https://doi.org/10.52783/jes.2799>.
7. Yang, Y.; Han, X.; Qin, D. Dynamic Feature Analysis of Handwritten Electronic Signatures Based on Fourier Transform. *J. Forensic Sci.* **2024**, *1*, 69. <https://doi.org/10.1111/1556-4029.15386>.
8. Okawa, M. Online Signature Verification Using Locally Weighted Dynamic Time Warping via Multiple Fusion Strategies. *IEEE Access* **2022**, *10*, 40806–40817. <https://doi.org/10.1109/ACCESS.2022.3167413>.
9. Singhal, M.; Shinghal, K. An Efficient, Fast and Accurate Online Signature Verification Using Blended Feature Vector and Deep Learning. *IETE J. Res.* **2024**, *70*, 7354–7364. <https://doi.org/10.1080/03772063.2024.2351567>.
10. Luan, F.; Cao, W.; Yuan, S. Relative Position Matrix and Multi-Scale Feature Fusion for Writer-Independent Online Signature Verification. *Heliyon* **2024**, *10*, e37655. <https://doi.org/10.1016/j.heliyon.2024.e37655>.
11. Rashidi, S.; Fallah, A.; Towhidkhah, F. Feature Extraction Based DCT on Dynamic Signature Verification. *Sci. Iran.* **2012**, *19*, 1810–1819. <https://doi.org/10.1016/j.scient.2012.05.007>.
12. Trozzi, F.; Wang, X.; Tao, P. UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: A comparison study. *J. Phys. Chem. B* **2021**, *125*(19), 5022–5034. <https://doi.org/10.1021/acs.jpcc.1c02081>.
13. Kobak, D.; Linderman, G. Initialization Is Critical for Preserving Global Data Structure in Both t-SNE and UMAP. *Nat. Biotechnol.* **2021**, *39*, 156–157. <https://doi.org/10.1038/s41587-020-00809-z>.
14. Parodi, M.; Gómez, J.C. Online Signature Verification: Improving Performance through Pre-Classification Based on Global Features. In *International Conference on Image Analysis and Processing*; Springer: Berlin/Heidelberg, Germany, 2013. https://doi.org/10.1007/978-3-642-41190-8_8.
15. Fayyaz, M.; Saffar, M.H.; Sabokrou, M.; Hoseini, M.; Fathy, M. Online Signature Verification Based on Feature Representation. In *Proceedings of the 2015 the International Symposium on Artificial Intelligence and Signal Processing (AISP)*, Mashhad, Iran, 3–5 March 2015. <https://doi.org/10.1109/AISP.2015.7123528>.
16. Vorugunti, C.S.; Pulabaigari, V.; Mukherjee, P.; Gautam, A. COMPOSV: Compound Feature Extraction and Depthwise Separable Convolution-Based Online Signature Verification. *Neural Comput. Appl.* **2022**, *34*, 10901–10928. <https://doi.org/10.1007/s00521-022-07018-6>.

17. Kahindo, C.; Garcia-Salicetti, S.; Houmani, N. A Signature Complexity Measure to Select Reference Signatures for Online Signature Verification. In Proceedings of the 2015 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 9–11 September 2015. <https://doi.org/10.1109/BIOSIG.2015.7314595>.
18. Yanikoglu, B.; Kholmatov, A. Online Signature Verification Using Fourier Descriptors. *EURASIP J. Adv. Signal Process.* **2009**, *2009*, 260516. <https://doi.org/10.1155/2009/260516>.
19. Kumar, R.; Ghosh, R. Person Verification and Recognition by Combining Voice Signal and Online Handwritten Signature Using Hyperbolic Function Based Transformer Neural Network. *Neurocomputing* **2025**, *632*, 129751. <https://doi.org/10.1016/j.neucom.2025.129751>.
20. Liu, Y.; Zhang, Z.; Yang, L. Online Signature Verification Based on DCT and Sparse Representation. *IEEE Transactions on Cybernetics* **2015**. <https://doi.org/10.1109/TCYB.2014.2375959>.
21. Radhika, K.R.; Venkatesha, M.K.; Shekar, G.N. On-Line Signature Authentication Using Zernike Moments. In Proceedings of the 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, Washington, DC, USA, 28–30 September 2009; pp. 1–4. <https://doi.org/10.1109/BTAS.2009.5339022>.
22. Iranmanesh, V.; Ahmad, S.M.S.; Adnan, W.A.W.; Yussof, S.; Arigbabu, O.A.; Malallah, F.L. Online Handwritten Signature Verification Using Neural Network Classifier Based on Principal Component Analysis. *Sci. World J.* **2014**, *2014*, 381469. <https://doi.org/10.1155/2014/381469>.
23. Kholmatov, A.; Yanikoglu, B. Identity Authentication Using Improved Online Signature Verification Method. *Pattern Recognit. Lett.* **2005**, *26*, 2400–2408. <https://doi.org/10.1016/j.patrec.2005.04.017>.
24. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Ortega-Garcia, J. DeepSign: Deep Learning for On-Line Signature Verification. *IEEE Trans. Biom. Behav. Identity Sci.* **2021**, *3*, 229–239. <https://doi.org/10.1109/TBIOM.2021.3054533>.
25. Shen, Q.; Luan, F.; Yuan, S. Multi-Scale Residual Based Siamese Neural Network for Writer-Independent Online Signature Verification. *Appl. Intell.* **2022**, *52*, 14571–14589. <https://doi.org/10.1007/s10489-022-03318-5>.
26. McInnes, L.; Healy, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *J. Open Source Softw.* **2018**, *3*, 861. <https://doi.org/10.21105/joss.00861>.
27. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.A.; Kwok, I.W.; Ng, L.G.; Ginhoux, F.; Newell, E.W. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat. Biotechnol.* **2019**, *37*, 38–44. <https://doi.org/10.1038/nbt.4314>.
28. Lai, S.; Jin, L.; Lin, L.; Zhu, Y.; Mao, H. SynSig2Vec: Learning Representations from Synthetic Dynamic Signatures for Real-World Verification. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 735–742. <https://doi.org/10.1609/AAAI.V34I01.5416>.
29. Parziale, A.; Anatrella, A.; Marcelli, A. Stability, Speed and Accuracy for Online Signature Verification. *J. Am. Soc. Quest. Doc. Exam.* **2015**, *18*, 39–49. <https://doi.org/10.69525/jasqde.221>.
30. Impedovo, D.; Pirlo, G. Automatic Signature Verification: The State of the Art. *IEEE Trans. Syst. Man Cybern. Part C* **2008**, *38*, 609–635. <https://doi.org/10.1109/TSMCC.2008.923866>.
31. Rao, G.A.; Syama, K.; Suribabu, K. Statistical Features Based Selfie Sign Language Recognition System (SSLRS) with ANN Classifier. *AIP Conf. Proc.* **2023**, *2958*, 020005. <https://doi.org/10.1063/5.0165675>.
32. Zois, E.N.; Tsourounis, D.; Theodorakopoulos, I.; Kesidis, A.L.; Economou, G. A Comprehensive Study of Sparse Representation Techniques for Offline Signature Verification. *IEEE Trans. Biom. Behav. Identity Sci.* **2019**, *1*, 68–81. <https://doi.org/10.1109/TBIOM.2019.2897802>.
33. Khatua, P.; Ray, K.C. An Efficient DCT-II Based Harmonic Wavelet Transform for Time-Frequency Analysis. *J. Signal Process. Syst.* **2022**, *94*, 1381–1394. <https://doi.org/10.1007/s11265-022-01805-z>.
34. Chetry, B.P.; Kar, B. Kruskal Wallis and mRMR Feature Selection Based Online Signature Verification System Using Multiple SVM and KNN. *Int. J. Exp. Res. Rev.* **2024**, *42*, 298–311. <https://doi.org/10.52756/ijerr.2024.v42.026>.
35. Jafseer, K.T.; Shailesh, S.; Sreekumar, A. Feature Drift Detection Using Overlapping Window and Mann-Whitney U Test. In Proceedings of the 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 11–12 February 2023; pp. 1–5. <https://doi.org/10.1109/ICITIIT57246.2023.10068710>.
36. Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. On Clustering Validation Techniques. *J. Intell. Inf. Syst.* **2001**, *17*, 107–145. <https://doi.org/10.1023/A:1012801612483>.

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.