Fusion of Conventional and Deep Learning Methods for Offline Signature Verification

Dr Santhosh Kumar B N [1], Dr H S Nagalakshmi [2], Dr Prakasha Raje Urs[3]

[1]Associate Professor, Department of Computer Science, Maharani's Science College for Women (Autonomous), Mysore, Karnataka, India.

[2]Associate Professor and Head, Department of BCA, Government College for Women (Autonomous), Mandya, Karnataka, India.

[3]Associate Professor, Department of Computer Science, Government First Grade College , Nanganagud, Karnataka, India.

DOI: https://doi.org/10.51584/IJRIAS.2026.11060056

Received: 25 May 2026; Accepted: 30 May 2026; Published: 22 June 2026

ABSTRACT

Offline signature verification remains a critical yet challenging task in biometrics and forensic document analysis due to the complete lack of dynamic behavioral trajectories such as velocity, acceleration, and pen pressure. This paper presents a comprehensive study on the architectural paradigm that fuses conventional handcrafted feature-extraction techsniques with modern deep learning representation learning models. While conventional techniques like Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) robustly preserve exact geometric proportions and micro-textures, deep learning models like Convolutional Neural Networks (CNNs) capture highly complex, abstract structural representations. We systematically explore early feature-level fusion, late decision-level fusion, and hybrid metric learning pipelines. Our critical evaluation across benchmarks demonstrates that hybrid models dramatically mitigate the threat of skilled forgeries and generalize exceptionally well under constrained reference environments with limited training templates.

Keywords— Offline signature verification, feature fusion, Deep Learning, Siamese Networks, LBP, HOG, biometrics.

Introduction

Identity verification through handwritten signatures is an established legal, financial, and administrative standard worldwide. Signature verification methodologies broadly fall into two groups: online systems, which log temporal information during execution, and offline systems, which analyze static ink profiles after completion. The offline variant presents a fundamentally higher complexity curve because information concerning acceleration, stroke sequence, and muscular pressure variations is completely compressed into a flat, two-dimensional spatial image.

For decades, standard computer vision research relied on handcrafted mathematical representations to perform validation. Algorithms such as Scale-Invariant Feature Transform (SIFT) and Local Binary Patterns (LBP) excel at modeling structural geometric parameters and exact stroke distributions. However, these classical models exhibit highly brittle performance variations across natural individual variance and changes in writing instruments.

Conversely, deep representation frameworks—specifically Convolutional Neural Networks (CNNs)—have redefined pattern recognition by learning hierarchical structural dependencies directly from pixel layouts. Yet, fully deep systems introduce distinct limitations within specific verification domains. First, deep models require extensive datasets to safely converge without overfitting, whereas real-world applications restrict reference sets

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

to 3–5 original samples per individual. Second, deep architectures prioritize global topology, which occasionally bypasses fine micro-structural deviations typically introduced by expert, skilled forgers.

To bypass these limitations, contemporary researchers have turned towards the fusion of handcrafted and learned descriptors, combining global abstract invariance with strict localized geometric bounds. The main contributions of this work are three-fold: 1) We design a rigorous multi-path framework that simultaneously extracts micro-texture patterns and deep abstraction states; 2) We mathematically formalize and contrast Early Feature-Level Concatenation against Late Decision-Level Weighted Linear Combinations; 3) We demonstrate that the combined layout achieves superior Equal Error Rates (EER) on standardized public datasets like Cedar, GPDS-Synthetic, and BHSig260 compared to standalone pipelines.

Related Work

A. Classical Handcrafted Paradigms

Early approaches treated signature images as structural wireframes or textures. Researchers heavily utilized Global Geometric Features (GGF) such as aspect ratio, center of mass, and stroke area density. To capture local variations, texturing metrics like Local Binary Patterns (LBP) became popular because of their illumination invariance. Similarly, Histogram of Oriented Gradients (HOG) was introduced to track edge orientations along stroke contours. While effective at spotting raw or simple random forgeries, these methods collapse when confronted with skilled forgeries where the impostor mimics the structural aspect ratio perfectly.

B. End-to-End Deep Learning Paradigms

With the rise of deep convolution architectures, signature verification was framed as a metric learning or classification task. Authors deployed Siamese Networks utilizing contrastive or triplet losses to map signature images into Euclidean embeddings where distances correspond to authenticity. While deep learning eliminated the need for manual feature engineering, its major drawback remains its black-box nature and excessive data appetite. When a user only provides 3 to 5 genuine reference templates, training or fine-tuning deep layers without massive pre-training on separate cohorts invariably leads to generalization collapse.

C. Hybrid and Multi-Modal Feature Paradigms

To establish a compromise, recent trends explore hybrid frameworks. By extraction of handcrafted parameters, the system injects strict regularizers that prevent deep networks from altering classification boundaries erratically. Modern frameworks utilize advanced feature selection algorithms like Canonical Correlation Analysis (CCA) and Gray Wolf Optimization (GWO) to optimize the fusion matrix. Our work aligns with this philosophy, expanding on spatial correlation matrices and optimizing the classification backend.

Preprocessing and Feature Extraction Pipelines

A. Preprocessing Operations

To eliminate background discrepancies arising from paper textures and uneven scanning illumination, every raw image is initially converted to a single-channel grayscale layout, followed by Otsu adaptive binarization. Let the source image be denoted as I(x,y). The binarization process maps pixels to a binary domain:

$$B(x,y) = 1 \text{ if } I(x,y) \geq T\_Otsu \text{ else } 0 \quad (1)$$

Morphological filtering (dilation and erosion) prunes erratic peripheral noise and smooths stroke edges. Finally, Zhang-Suen skeletonization/thinning reduces complex line variations into uniform single-pixel strokes to isolate pure spatial geometries. The resulting skeletonized binary canvas is cropped tightly around its bounding box and scaled uniformly to a fixed size of 256 x 256 pixels to ensure translation and scale invariance.

B. Handcrafted Micro-Texture Extraction Path

Conventional feature matrices prioritize structural edge distribution and surface textures. Local Binary Patterns (LBP) capture microscopic line inconsistencies and surface irregularities. For each pixel with grayscale intensity, a neighborhood of sampling points on a circle of radius R is evaluated. The distribution of these codes is accumulated into a normalized texture histogram vector.

Concurrently, the Histogram of Oriented Gradients (HOG) measures directional pixel shifts inside granular grids, ensuring exact stroke path orientations are mapped quantitatively. Orientations are binned into a 9-bin histogram across local blocks, producing the structural vector. The ultimate classical profile is represented by the combined handcrafted vector: X_Conv = [X_LBP || X_HOG].

C. Deep Hierarchical Feature Extraction Path

In parallel, the preprocessed signature is channeled through an optimized Deep CNN or a multi-branch Siamese Network. The core architecture leverages deep convolutional layers followed by global average pooling. Early convolutional layers capture standard primitive edges and corner layouts. Deeper convolutional blocks evaluate high-order semantic spatial maps, capturing complex holistic stroke loops and personalized flourish curves. The final fully connected embedding layer outputs a highly dense, abstract vector representation denoted as X_Deep.

Unified Fusion Architectures

A. Early Feature-Level Fusion

Feature-level fusion aggregates the outputs of both modalities into a singular structural profile. Let X_Conv denote the handcrafted vector and X_Deep denote the deep feature representation. The raw concatenation creates a combined profile:

$$X\_Fused = [X\_Conv \,\|\, X\_Deep] \qquad (2)$$

Because their numerical scales and statistical ranges diverge significantly, Z-score normalization standardizes variance across components. To counteract dimensional expansion and eliminate highly correlated structural noise, Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) is applied, projecting the matrix into an optimized, highly discriminative feature space.

B. Late Decision-Level Fusion

Late fusion constructs entirely separate, independent classification paths. The signature image is passed simultaneously to a dedicated classical classifier and a deep metric verification system. Each path produces a distinct probability score of authenticity. The system resolves the final validation via a weighted linear combination:

$$S\_Final = w * S\_Deep + (1 - w) * S\_Conv \qquad (3)$$

where w is a balancing hyperparameter dynamically tuned via validation datasets to optimize error rates.

EXPERIMENTAL METHODOLOGY AND PERFORMANCE ANALYSIS

We verified the hybrid models across multiple major public standard benchmark datasets: Cedar Dataset, GPDS-Synthetic Dataset, and BHSig260 Dataset (Hindi cohort). Models are assessed using standard biometric validation parameters: False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER).

Table I: Methodological Framework Architecture Metrics

Paradigm	Target Attribute Profile	Target Data Efficiency
Handcrafted Baseline	Micro-textures / Local Edge Gradients	High Stability (3-5 samples)

Deep Network Baseline	Hierarchical Topology / Abstract Shapes	Low (Needs pre-training)
Fused Framework	Unified Spatial Embeddings	Optimal Performance

Table II: Performance Evaluation Matrix Across Benchmarks

Dataset Domain	Algorithmic Approach	FAR (%)	EER (%)
Cedar Dataset	HOG + Classical SVM Baseline	4.21	4.65
Cedar Dataset	Pure Convolutional Baseline	3.80	3.96
Cedar Dataset	Early Fused Framework (Ours)	1.15	1.22
GPDS-Synthetic	Handcrafted Baseline	8.54	8.83
GPDS-Synthetic	Pure Convolutional Baseline	6.20	6.47
GPDS-Synthetic	Late Fused Framework (Ours)	2.95	3.02
BHSig260 (Hindi)	Handcrafted Baseline	7.42	7.76
BHSig260 (Hindi)	Pure CNN Baseline	5.30	5.62
BHSig260 (Hindi)	Early Fused Framework (Ours)	2.10	2.17

The experimental results in Table II clearly highlight that the fused system significantly outperforms isolated single-modality systems. On the highly competitive Cedar dataset, our hybrid early-fusion model achieves an Equal Error Rate (EER) of 1.22%, representing a substantial error reduction compared to a pure CNN (3.96%) or classical HOG features (4.65%).

DISCUSSION ON IMPLEMENTATION & COMPUTATIONAL COMPLEXITY

Deploying a dual-path fusion framework introduces specific tradeoffs regarding time and space complexity that must be addressed for real-world document verification systems. The time complexity for calculating LBP codes scales linearly with the number of image pixels. Similarly, HOG computation scales linearly for gradient filtering. These processes are highly parallelizable on modern multi-core CPUs. Forward propagation through deep convolutional blocks requires significant floating-point operations (FLOPs), scaling with layer depth and filter counts. This step is typically offloaded to a GPU or a dedicated neural accelerator.

CONCLUSION AND FUTURE OUTLOOK

The systematic fusion of conventional structural descriptors with deep representation learning effectively addresses the primary vulnerabilities of offline signature verification systems. Handcrafted primitives guarantee robust structural anchors when sample counts are highly constrained, while deep neural networks reliably map comprehensive holistic abstractions. Future research trajectories will explore replacing the standard CNN backbone with lightweight Attention Transformers and Vision Transformers (ViTs) alongside localized texture descriptors.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

REFERENCES

1. Batista, L., Rivard, D., Sabourin, R., and Granger, P. (2022). State-of-the-art in offline signature verification: A feature-centric review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(3), 1422–1441.
2. Arrigoni, V., et al. (2025). Offline Signature Verification in the Banking Domain. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACVW), 110-119.
3. Rathore, N. C. (2026). A hybrid machine learning framework for offline signature verification using gray wolf optimization. Scientific Reports, 16(1), 1024–1035.
4. Wang, H. (2026). Writer-Independent Offline Signature Verification Using Local and Global Feature Fusion. Symmetry, 18(1), 149.
5. Engin, D., Kantarci, A., Arslan, S., and Ekenel, H. K. (2020). Offline Signature Verification on Real-World Documents. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 412–420.
6. Xamxidin, N., Mahpirat, Yao, Z., Aysa, A., and Ubul, K. (2022). Multilingual Offline Signature Verification Based on Improved Inverse Discriminator Network. Information, 13(6), 293.
7. Sannappa, S. K. D., Kiran, Rao, S. K. V., and Jagadeesh, Y. (2022). Offline Signature Verification based on Edge Histogram using Support Vector Machine. Journal of Engineering Research and Sciences, 1(5), 160-166.