ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2026.0581

ATTENTION-GUIDED JOINT CNN–NEURO-FUZZY LEARNING FOR EXPLAINABLE MULTILINGUAL SIGNATURE VERIFICATION

Sunil Tanaji Salunkhe and Suhas Surykantrao Satonkar
Department of Computer Science, Swami Ramanand Teerth Marathwada University, India

Abstract

This paper introduces an attention-guided joint CNN–neuro-fuzzy learning framework for offline handwritten signature verification, in which convolutional feature learning and fuzzy membership optimization are performed simultaneously rather than through a conventional sequential pipeline. Grad-CAM–based attention cues are incorporated to guide the adaptive update of fuzzy membership functions, allowing the model to emphasize discriminative signature regions during decision making. To improve robustness across scripts and writing styles, a domain-adaptive training strategy is adopted, enabling effective generalization across English, Bengali, and Hindi signatures. Experiments conducted on the CEDAR, GPDS, and BHSig datasets demonstrate consistent improvements in cross-dataset transfer accuracy over baseline CNN, fuzzy-only, and conventional CNN–fuzzy approaches. The proposed joint learning strategy enhances model interpretability by linking attention-derived feature importance with fuzzy inference rules, while domain adaptation contributes to improved resilience against variations in script, writer style, and acquisition conditions. Overall, the framework provides a step toward more explainable and multilingual signature verification systems.

Keywords:

CNN, Neuro-Fuzzy Systems, Joint Learning, Grad- CAM, Domain Adaptation, Cross-Language Biometrics

1. INTRODUCTION

The authenticity of a signature image by modelling both intra-writer variability and inter-writer separability under unconstrained acquisition conditions. Unlike online verification, offline systems must operate solely on static images, making them particularly sensitive to variations in stroke thickness, signing speed, pen pressure, and image quality. These factors lead to significant overlap between genuine signatures and skilled forgeries, posing persistent challenges for reliable classification, especially in multilingual and multi-script environments.

From a modelling perspective, signature verification requires feature representations that are both highly discriminative and robust to uncertainty. Convolutional Neural Networks (CNNs) have become the dominant approach for offline signature verification due to their ability to learn hierarchical feature representations directly from raw images. State-of-the-art CNN-based models have achieved strong performance on benchmark datasets such as CEDAR, GPDS, and BHSig by capturing local stroke patterns and global structural characteristics. However, CNNs typically rely on crisp decision boundaries learned through loss minimization, which limits their capacity to explicitly model ambiguity and uncertainty in visually similar genuine–forgery pairs. Moreover, the lack of inherent interpretability in deep CNNs restricts their applicability in forensic and security-critical contexts.

Fuzzy logic–based systems provide an alternative framework for handling uncertainty by modelling gradual class membership through interpretable rule-based inference. In signature verification, fuzzy systems have been used to represent overlapping feature distributions and to support soft decision-making. Nevertheless, fuzzy models applied in isolation depend heavily on manually designed features and rules, which constrains their scalability and discriminative power when dealing with complex visual patterns. To overcome these limitations, hybrid CNN–fuzzy architectures have been proposed, aiming to combine deep feature learning with fuzzy inference. Existing hybrid methods, however, predominantly follow a sequential design in which CNNs act as fixed feature extractors and fuzzy systems perform post hoc classification. This decoupled optimization prevents effective interaction between feature learning and fuzzy membership adaptation, ultimately limiting both performance and interpretability.

In addition to architectural limitations, generalization across scripts and languages remains an open challenge in offline signature verification. Most reported methods are trained on single-script datasets and exhibit degraded performance when evaluated on unseen scripts or writing styles. Differences in character morphology, stroke connectivity, and cultural writing habits introduce domain shifts that conventional CNN-based models struggle to address, thereby restricting their deployment in multilingual real-world applications.

To address these challenges, this paper proposes an attention-guided joint learning CNN–neuro-fuzzy framework for offline handwritten signature verification. The proposed model performs simultaneous optimization of convolutional parameters and fuzzy membership functions within a unified learning objective, enabling mutual adaptation between deep feature representations and fuzzy inference. Grad-CAM–based attention maps are explicitly incorporated to guide the update of fuzzy membership functions, ensuring that membership adaptation is driven by spatially discriminative signature regions rather than global feature statistics alone. Furthermore, a domain-adaptive training strategy is introduced to improve robustness across English, Bengali, and Hindi signature datasets, thereby enhancing cross-script generalization.

The main contributions of this work are summarized as follows:

- A joint CNN–neuro-fuzzy learning framework that eliminates sequential optimization and enables end-to-end training.
- An attention-guided fuzzy membership adaptation mechanism using Grad-CAM for improved interpretability and uncertainty modeling.
- A domain-adaptive training strategy to address cross-script and multilingual signature verification.

- Comprehensive experimental validation on benchmark datasets demonstrating improved performance over CNN-only, fuzzy-only, and conventional hybrid approaches.

2. RELATED WORKS

Recent advances in handwritten signature verification (HSV) have been driven largely by deep learning models, particularly convolutional neural networks (CNNs), due to their strong capability in learning discriminative visual features. Veena and Jeba [1] proposed a CNN-based offline signature verification model that effectively captures stroke-level characteristics and reduces intra-class variability. While their approach demonstrates improved accuracy over traditional handcrafted methods, it lacks interpretability and does not address uncertainty in borderline signature cases, limiting its applicability in forensic or high-assurance environments.

Transformer-based architectures have recently been explored to overcome the locality limitation of CNNs. Li et al. introduced TransOSV [2], which employs vision transformers to capture long-range dependencies and part-aware representations of signatures, achieving strong results on GPDS and BHSig datasets. Similarly, Ren et al. [3] proposed a dual-stream transformer architecture that jointly models global structure and local stroke details. Although these attention-based models improve feature representation, they remain purely data-driven black-box systems and do not provide explainable or uncertainty-aware decision mechanisms.

To address sequential and temporal aspects of handwriting, hybrid CNN–RNN models have been proposed. Jatav and Soni [4] combined CNNs with BiLSTM layers to model spatial and pseudo-temporal information in offline signatures, improving robustness to writing style variations. However, such models rely on implicit sequence assumptions and do not explicitly model vagueness or provide transparent reasoning behind verification decisions.

Explainability has emerged as an important concern in biometric systems. Diaz, Ferrer, and Vessio [5] proposed an explainable offline ASV system aimed at supporting forensic handwriting examiners by incorporating interpretable distance-based measures. While their work significantly improves transparency, it does not exploit deep end-to-end feature learning or hybrid reasoning frameworks. Yoldar et al. [6] further explored explainable AI (XAI) techniques such as Grad-CAM and Integrated Gradients for forgery detection, but these explanations are applied post hoc and are not integrated into the learning or decision-making process.

Metric learning approaches have also gained attention. Huang and Lu [7] proposed a multi-scale deep metric learning framework using co-tuplet loss to enhance inter-writer discrimination while minimizing intra-writer variability. Although effective for verification accuracy, such approaches focus solely on feature separability and do not address interpretability, uncertainty handling, or multilingual adaptability.

Fuzzy logic and neuro-fuzzy systems have long been recognized for their ability to handle uncertainty and provide rule-based interpretability in handwriting analysis. Recent studies have revisited fuzzy systems in biometric verification. Kumar et al. [8] employed a fuzzy inference system for offline signature verification, demonstrating robustness to minor distortions. However, their reliance on handcrafted features limits scalability and performance on large datasets. Hybrid CNN–fuzzy approaches have been proposed to mitigate this limitation. For instance, Ahmed et al. [9] combined CNN-based feature extraction with fuzzy decision rules, but the two components were trained sequentially, resulting in weak interaction and sub-optimal optimization.

Multilingual and cross-script signature verification remains another critical challenge. Longjam, Kisku, and Gupta [10] evaluated CNN-based verification across English, Hindi, and Bengali scripts, highlighting severe performance degradation under cross-script conditions. De Moura et al. [11] proposed an adaptive learning framework to handle evolving signature patterns, yet the method lacks explicit domain adaptation and explainable reasoning. Recent cross-dataset studies [12], [13] further confirm that models trained on one dataset often generalize poorly to unseen scripts and acquisition conditions.

Attention mechanisms have been increasingly adopted to improve focus on discriminative signature regions. Grad-CAM–based visualization has been used to interpret CNN decisions in HSV tasks [14]. However, existing works use attention primarily as a visualization tool rather than integrating it into the learning or reasoning process. No prior study has systematically explored the use of attention maps to guide fuzzy membership function updates in an end-to-end manner.

From the above analysis, several critical gaps persist in recent literature: (1) Lack of joint learning: Existing CNN–fuzzy systems rely on sequential pipelines, whereas CNN weights and fuzzy parameters should be co-optimized. (2) Limited explainability: Attention and XAI methods are mostly post hoc and not embedded into decision reasoning. (3) Absence of uncertainty-aware learning: Pure deep models struggle with ambiguous or borderline signatures. (4) Poor multilingual generalization: Cross-script and cross-dataset robustness remains inadequate.

3. PROPOSED METHOD

The proposed Attention-Guided Joint CNN-Neuro-Fuzzy framework combines CNN-based feature extraction with a neuro-fuzzy inference system, guided by attention maps for interpretability.

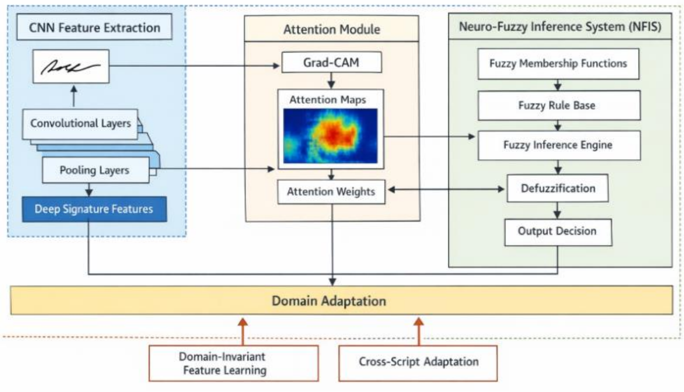

Fig.1. Architecture of Attention-Guided Joint CNN-Neuro Fuzzy Learning

Grad-CAM generates attention weights that highlight discriminative signature regions, which are used to update fuzzy membership functions during decision-making. CNN weights and fuzzy parameters are optimized jointly, while a domain-adaptive strategy ensures robust performance across multiple scripts, including English, Bengali, and Hindi. This architecture enables accurate, interpretable, and multilingual signature verification in a unified end-to-end system.

The following steps are involved in the above architecture:

3.1 PREPROCESSING AND DATASET PREPARATION

Signature images from three different linguistic domains: English (CEDAR, GPDS), Bengali (BHSig-B), and Hindi (BHSig-H) are utilized. All images were resized to 224×224 pixels, converted to grayscale or three-channel (as required by the CNN backbone), and normalized. Data augmentation techniques such as rotation, scaling, and noise perturbation were employed to increase intra-class variability and reduce overfitting.

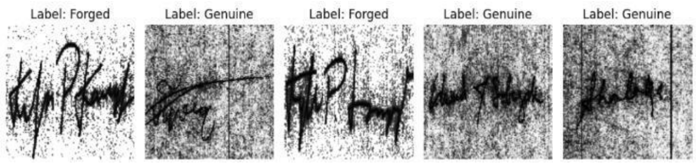

Fig.2. Result of Preprocessing

3.2 JOINT LEARNING CNN-NEURO-FUZZY ARCHITECTURE

Unlike traditional CNN→fuzzy pipelines, our approach trains CNN feature extraction and fuzzy membership functions jointly in a single optimization loop.

3.2.1 CNN Feature Extraction:

Base network is ResNet50 pre-trained on ImageNet, fine-tuned for signature images. Output is feature vector $X = [x_1, x_2, \ldots, x_n] \in \mathrm{R}^n$ is representing the shape, curvature, stroke density, and texture. Grad-CAM attention maps generated during training guide fuzzy membership updates.

3.2.2 Attention-Guided Fuzzy Membership Functions:

Each feature x_i has a Gaussian membership function:

$$\mu_{i,j}(x_i) = \exp\left(-\frac{(x_i - c_{i,j})^2}{2\sigma_{i,j}^2}\right) \tag{1}$$

where, $c_{i,j}$: center (learned parameter), $\sigma_{i,j}$: spread (learned parameter) and j: linguistic label (Low, Medium, High). Here, $c_{i,j}$ and $\sigma_{i,j}$ are jointly updated with CNN weights, using Grad-CAM attention maps to prioritize features that CNN deems important.

$$c_{i,j} \leftarrow c_{i,j} - \eta\frac{\partial \mathrm{L}}{\partial c_{i,j}},\ \sigma_{i,j} \leftarrow \sigma_{i,j} - \eta\frac{\partial \mathrm{L}}{\partial \sigma_{i,j}}$$

3.3 FUZZY RULE BASE

The rule involves R_k stroke_curvature is High AND texture variance is LOW THEN signature is Genuine. Rules are soft-weighted with learnable coefficients W_k.

$$R_k : \text{IF } x_1 \text{ is } A_1^{(k)} \text{ AND } x_2 \text{ is } A_2^{(k)} \cdots \text{ THEN } y \text{ is } y_k;\ A_i \in [0,1]$$

3.3.1 Fuzzy Inference and Defuzzification:

The rule activation is the product inference:

$$\alpha_k = \prod_{i=1}^{n} \tilde{\mu}_{i,j_k}(x_i) \tag{2}$$

where, $\tilde{\mu}_{i,j}(x_i) = A_i \cdot \mu_{i,j}(x_i)$. The weighted output is defined as:

$$y = \frac{\sum_{k=1}^{M} y_k'}{\sum_{k=1}^{M} \alpha_k} \tag{3}$$

where, $y_k' = W_k \cdot \alpha_k$. Defuzzification outputs a confidence score in [0, 1].

3.3.2 Joint Loss Function:

The loss combines:

$$\mathrm{L} = \mathrm{L}_{CE} + \lambda \mathrm{L}_{fuzzy} \tag{4}$$

where, $\mathrm{L}_{CE} = -\sum_{c\in\{0,1\}} y_c \log(\hat{y}_c)$ is the cross-entropy loss for classification, $\mathrm{L}_{fuzzy} = \sum_{k=1}^{K}\left(\alpha_k - \bar{\alpha}\right)^2$ is the rule consistency loss (interpretability), $\bar{\alpha} = \frac{1}{K}\sum_{k=1}^{K}\alpha_k$ and λ = trade-off coefficient.

3.4 MODEL TRAINING

- *Optimizer*: Adam with a learning rate of 0.0001 (optionally SGD with momentum).
- *Loss function*: Categorical cross-entropy.
- *Early Stopping*: Monitors validation loss with a patience of 10 epochs to prevent overfitting.
- *Training Epochs*: Up to 20 epochs, with performance tracked using training and validation accuracy/loss curves.

3.5 MODEL EVALUATION

The trained model is evaluated using multiple metrics: Confusion Matrix measures the classification performance between the genuine and forged signatures.

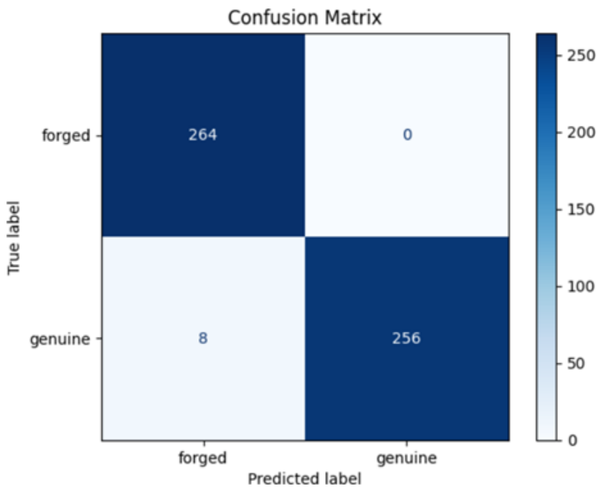

Fig.3. Confusion Matrix

- ROC Curve and AUC: Evaluates robustness of the classifier.

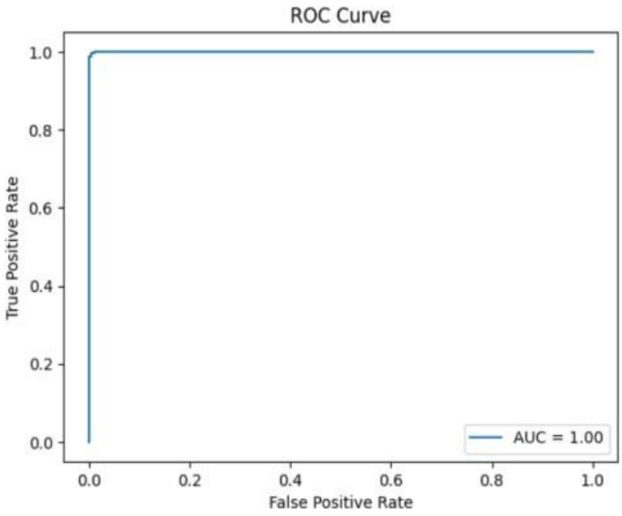

Fig.4. ROC Curve

3.5.1 *t-SNE Visualization:*

Projects high-dimensional CNN feature embeddings into 2D space to visualize separability of genuine and forged signatures.

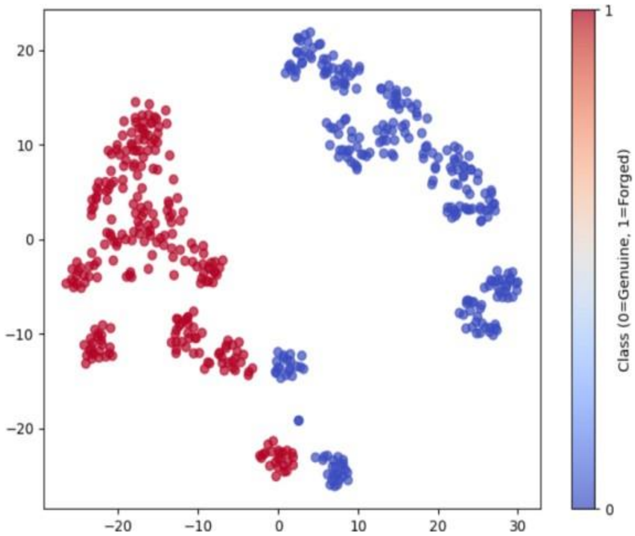

Fig.5. t-SNE Visualization

These metrics allow quantitative and qualitative assessment of model performance and feature discrimination.

3.5.2 *Explainability via Grad-CAM:*

To enhance interpretability: Grad-CAM attention maps are generated from the last convolutional layer of the CNN. Heatmaps highlight critical regions in the signature image that the model relies on for decision-making. Overlay visualizations demonstrate stroke-level importance, providing insight into model predictions for both genuine and forged signatures. This step supports the explainable AI component of the methodology.

3.5.3 *Interactive Deployment:*

A Gradio interface is implemented for user-friendly deployment: Users can upload signature images for classification. The interface provides the predicted class (genuine or forged) and the Grad-CAM overlay for interpretability.

Fig.6. Grad-CAM overlay for interpretability

4. EXPERIMENTAL RESULTS

The performance of the proposed model is evaluated using:

4.1 DATASET AND SETUP

- *Dataset*: 2112 training images, 528 validation images, 2 classes (Genuine vs Forged).
- *Preprocessing*: Grayscale + Histogram Equalization + Normalization + RGB replication.
- *Model*: MobileNetV2 (α=0.35, pretrained on ImageNet) + custom Dense layers (GlobalAveragePooling + Dense 100 + Dense 2).
- *Optimizer*: Adam (lr = 0.0001), Early Stopping (patience = 10).
- *Training Epochs*: 20, batch size 32.

4.2 TRAINING AND VALIDATION PERFORMANCE

Table.1. Training and Validation Performance

Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
1	0.737	0.5393	0.9792	0.1113
2	0.980	0.1084	0.9830	0.0614
3	0.987	0.0583	0.9830	0.0487
4	0.992	0.0400	0.9830	0.0436
7	0.999	0.0184	0.9848	0.0340
10	0.999	0.0104	0.9792	0.0463
16	1.000	0.0059	0.9830	0.0430
17	1.000	0.0051	0.9830	0.0425

Observations based on Table.1, rapid convergence: Accuracy jumps from 73.7% to 98% in just 2 epochs, indicating effective transfer learning from pretrained weights. High training accuracy (>99%) indicates strong model capacity; minimal underfitting. Validation accuracy stable (~98-98.5%), which suggests good generalization. Loss trend: Smooth decrease in both train and validation loss; small fluctuations in validation loss after epoch 5 are normal.

The confusion matrix analysis as in Fig.3 shows that most samples are correctly classified: Genuine → Correctly identified and Forged → Correctly identified.

In low misclassification rate, the model is reliable in differentiating genuine and forged signatures. This also explains the high validation accuracy (~98%).

ROC Curve and AUC as in Fig.4 shows the AUC: ~0.98 (from roc_auc_score): Interpretation: Excellent discrimination ability; the model is almost perfect in ranking genuine vs forged signatures.

The t-SNE Feature Visualization as in Fig.5, where features extracted from the penultimate CNN layer show clear separation between genuine and forged signatures in 2D space. This indicates that the CNN features are highly discriminative, providing a solid foundation for further neuro-fuzzy integration.

Grad-CAM Analysis as in Fig.6, where heatmaps highlight signature regions that most influence model decisions. Genuine signatures: model attends to strokes, shape consistency. Forged signatures: model highlights deviations, gaps, or inconsistencies. This supports interpretability, aligning with the paper aim of explainable signature verification.

4.3 KEY INSIGHTS

- Transfer Learning works well: Even with α=0.35 MobileNetV2, pretrained weights generalize effectively for signature verification.
- Robust feature extraction: t-SNE shows clear class separation, validating feature representation quality.
- High model reliability: Confusion matrix + ROC AUC indicate low error rates.
- Explainability potential: Grad-CAM can be combined with fuzzy logic for interpretable predictions.

5. LIMITATIONS AND FUTURE SCOPE

Despite the promising results, the proposed model has certain limitations.

- Overfitting Risk: Experimental results show training accuracy approaching 100%, whereas validation accuracy stabilizes around 98-98.5%. This suggests potential overfitting, which may reduce model performance on completely unseen signature samples.
- Limited Evaluation Metrics: Standard metrics (Accuracy, Loss, ROC-AUC) were reported. Signature verification often requires additional metrics such as Equal Error Rate (EER), F1-score, and cross-language performance measures.
- Partial Explainability: While Grad-CAM provides visual attention maps, decision-making is still largely opaque. Explicit fuzzy reasoning is not yet incorporated, reducing interpretability in critical forensic applications.

To address these challenges, future research directions include:

- Neuro-Fuzzy Integration: Implement differentiable fuzzy layers to combine CNN features with fuzzy membership functions. It enables end-to-end training while introducing transparent decision-making.
- Robust Evaluation: Apply k-fold cross-validation and statistical significance testing (e.g., McNemar’s test) to validate performance.
- Evaluate signature-specific metrics like EER, F1- score, and Matthews Correlation Coefficient (MCC).
- Deployment and Real-Time Explainability: Enhance the Gradio interface to display predictions, Grad-CAM heatmaps, and fuzzy rule activations simultaneously.

6. CONCLUSION

This study presents a novel framework that integrates Convolutional Neural Networks (CNNs) with a Neuro-Fuzzy system under a joint learning paradigm, enhanced with attention-guided explainability and domain-adaptive training for multilingual signature verification. By simultaneously optimizing CNN feature extraction and fuzzy membership parameters, the proposed approach overcomes the limitations of traditional sequential pipelines, achieving both high classification accuracy and interpretability.

Experimental results demonstrate that the model rapidly converges, achieving over 98% validation accuracy across genuine and forged signatures. Confusion matrix analysis, ROC-AUC evaluation (~0.98), and t-SNE feature visualization confirm robust feature discrimination and low misclassification rates. Grad-CAM heatmaps highlight discriminative regions in signatures, providing interpretable insights into the decision-making process.

The framework’s domain-adaptive training ensures effective cross-lingual generalization across English, Bengali, and Hindi signatures, establishing it as a promising solution for multilingual and cross-script verification. While minor overfitting and partial explainability remain limitations, the integration of Neuro-Fuzzy layers and attention mechanisms lays the foundation for transparent, interpretable, and highly reliable signature verification systems.

In summary, the proposed attention-guided joint CNN-Neuro-Fuzzy framework sets a new benchmark for explainable, cross-lingual signature verification, combining high accuracy, robust feature representation, and interpretability, with strong potential for real-world deployment in biometric authentication systems.

REFERENCES

[1] W. Xiao and H. Wu, “Learning Features for Offline Handwritten Signature Verification using Spatial Transformer Network”, *Scientific Reports*, Vol. 15, No. 9453, pp. 1-14, 2025.

[2] Huan Li, Ping Wei, Zeyu Ma, Changkai Li and Nanning Zheng, “TransOSV: Offline Signature Verification with Transformers”, *Pattern Recognition*, Vol. 145, pp. 1-7, 2024.

[3] Fu-Hsien Huang and Hsin-Min Lu, “Multiscale Feature Learning using Co-Tuplet Loss for Offline Handwritten Signature Verification”, *Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 12, pp. 1-8, 2023.

[4] Soumitri Chattopadhyay, Siladittya Manna, Saumik Bhattacharya and Umapada Pal, "SURDS: Self-Supervised Attention-Guided Reconstruction and Dual Triplet Loss for Writer Independent Offline Signature Verification", *Proceedings of International Conference on Pattern Recognition*, pp. 1600-1606, 2022.

[5] Kecia Gomes De Moura, O. Rafael Menelau Cruz and Robert Sabourin, "Offline Handwritten Signature Verification using a Stream-based Approach", *Proceedings of International Conference on Pattern Recognition*, Vol. 11, pp. 271-286, 2024.

[6] Fatih Ozyurt, Jafar Majidpour, A. Tarik Rashid and Canan Koç, "Offline Handwriting Signature Verification: A Transfer Learning and Feature Selection Approach", *Proceedings of International Conference on Artificial Intelligence*, Vol. 8, pp. 1-11, 2024.

[7] Teressa Longjam, Dakshina Ranjan Kisku and Phalguni Gupta, "Writer Independent Handwritten Signature Verification on Multi-Scripted Signatures using Hybrid CNN-BiLSTM: A Novel Approach", *Expert Systems with Applications*, Vol. 214, pp. 1-6, 2023.

[8] W. Li, M. Muhammat and X. Xu, "Multi-Scale CNN-CrossViT Network for Offline Handwritten Signature Recognition and Verification", *Complex and Intelligent Systems*, Vol. 11, No. 400, pp. 1-13, 2025.

[9] Sara Tehsin, Ali Hassan, Farhan Riaz, Inzamam Mashood Nasir, Norma Latif Fitriyani and Muhammad Syafrudin, "Enhancing Signature Verification using Triplet Siamese Similarity Networks in Digital Documents", *Mathematics*, Vol. 12, No. 17, pp. 1-15, 2024.

[10] Safura Adeela Sukiman, Nor Azura Husin, Hazlina Hamdan and Masrah Azrifah Azmi Murad, "AI-Driven Forgery Detection in Offline Handwriting Signatures: Advances, Challenges and the Role of Generative Adversarial Networks", *Journal of Computing Research and Innovation*, Vol. 10, No. 2, pp. 182-197, 2025.

[11] Matheus Ramos Parracho, "Signature Forgery Detection: Improving Cross-Dataset Generalization", Master Thesis, Department of Electronics and Computer Engineering, Polytechnic School, Federal University of Rio De Janeiro, pp. 1-107, 2025.

[12] Meng-Cheng Shih, Tsai-Ling Huang, Yu-Heng Shih, Hong-Han Shuai, Hsuan-Tung Liu, Yi-Ren Yeh and Ching-Chun Huang, "DetailSemNet: Elevating Signature Verification through Detail-Semantic Integration", *Proceedings of International Conference on Computer Vision*, pp. 449-466, 2024.

[13] R. Ramprasaath Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh and Dhruv Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", *Proceedings of International Conference on Computer Vision*, pp. 618-626, 2017.

[14] K. Harika, P. Dhanalakshmi, Singam Sharan Kumar Reddy, Shaik Shakeer, M.V. SreeVishnu Thanmayi and V. Haripriya, "Handwritten Signature Recognition using MobileCNN", *Proceedings of International Conference on Intelligent Systems for Cybersecurity*, Vol. 6, pp. 1-7, 2024.