International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 04 | Apr 2026 www.irjet.net p-ISSN: 2395-0072

SHISHA-NET: A HYBRID DEEP LEARNING ARCHITECTURE FOR SIGNATURE FORGERY DETECTION

Shalini G[1], Renju K[2]

[1]*Student, Department of Computer Science, Mount Carmel College Autonomous, Bengaluru, India*
[2] *Assistant Professor, Department of Computer Science, Mount Carmel College Autonomous, Bengaluru, India*

Abstract - *Handwritten signature verification remains a critical challenge in biometric authentication, particularly in detecting skilled forgeries. This paper proposes Shisha-Net, a hybrid deep learning architecture that combines Convolutional Neural Networks (CNNs) and Siamese networks through feature-level fusion. The proposed model is evaluated on the CEDAR dataset consisting of 2,640 signature images from 55 writers. Experimental results demonstrate that the proposed model achieves 89.65% accuracy, outperforming baseline CNN (73.38%) and Siamese network (83.60%). Shisha-Net provides an improved balance between precision (87.78%) and recall (92.53%), making it effective for real-world signature verification applications. The results indicate that feature-level fusion of CNN and Siamese architectures successfully leverages the complementary strengths of both approaches.*

Key Words: Signature verification, forgery detection, convolutional neural networks, Siamese networks, hybrid deep learning, Shisha-Net, CEDAR dataset

1.INTRODUCTION

Handwritten signatures continue to serve as one of the most widely accepted biometric traits for personal authentication in banking, legal, and administrative domains [1]. Although more sophisticated biometric technologies such as fingerprint and iris recognition have emerged, signatures remain in use due to their non-invasive nature, social acceptance, and legal validity [1]. However, signature verification systems face considerable challenges, particularly in detecting skilled forgeries where impostors practice to replicate genuine signatures [2].

Signature verification can be categorized into two principal approaches: online (dynamic) verification, which captures temporal information including pen pressure, velocity, and acceleration, and offline (static) verification, which analyses only the scanned signature image [2]. Offline verification presents greater difficulty due to the absence of dynamic information, as only the visual appearance of the signature is available for analysis [2]. Skilled forgeries, where forgers practice to replicate genuine signatures, pose the most significant challenge as they can appear visually similar to authentic samples [3][24].

1.1 Challenges in Offline Signature Verification

Several factors contribute to the complexity of offline signature verification. First, interclass variation occurs when genuine signatures from the same writer naturally vary due to mood, writing conditions, and other physical factors [4]. Second, limited training data is available, as collecting numerous samples from each writer is impractical compared to other recognition tasks [5]. Third, skilled forgeries can be visually indistinguishable from genuine signatures even to human observers [6][25]. Fourth, the system must be writer independent, meaning it must generalize to unseen writers not present during training [7][26].

1.2 Contributions of This Paper

This paper presents the following contributions to the field:
• Implementation and evaluation of a baseline CNN classifier achieving 73.38% accuracy.
• Development of a Siamese network achieving 83.60% accuracy with exceptional recall of 98.17%.
• Proposal of Shisha-Net, a novel hybrid architecture combining CNN and Siamese approaches through feature fusion, achieving 89.65% accuracy with balanced precision (87.78%) and recall (92.53%).
• A comprehensive comparative analysis of all three architectures.

2. MATERIALS AND METHODS

2.1 Dataset and Preprocessing

All models were evaluated on the CEDAR (Center of Excellence for Document Analysis and Recognition) dataset [23], a widely used benchmark for offline signature

verification. The dataset contains signatures from 55 writers, with each writer providing 24 genuine signatures and 24 skilled forgeries, resulting in a total of 2,640 signature images [23]. Genuine signatures were collected from volunteers, while forgeries were produced by individuals who practiced imitating genuine samples [23]. All images are grayscale and vary in resolution, requiring preprocessing before training.

To prepare the dataset for training, the following preprocessing steps were applied. All images were resized to a uniform size of 128 × 128 pixels to ensure consistent input dimensions for all models. Pixel values were normalized to the range [0, 1] by dividing by 255.0. For training, data augmentation including random rotations (±5 ◦), zooms, and translations was applied to increase dataset diversity and reduce overfitting [5][27].

To ensure writer independence and proper evaluation, the dataset was split at the writer level rather than the image level [7]. This approach ensures that no writer appears in both training and testing sets, simulating real-world scenarios where the system must verify signatures from unseen writers [16]. The 55 writers were randomly split into 38 writers (70%) for training (1,824 images), 8 writers (15%) for validation (384 images), and 9 writers (15%) for testing (432 images). Each split maintains a balanced ratio of genuine and forged signatures.

2.2 Baseline CNN Architecture

Convolutional Neural Networks (CNNs) have been widely adopted for offline signature verification due to their ability to automatically learn discriminative features from raw signature images [3][4][5]. The baseline model implemented in this study is a convolutional neural network designed for binary classification (genuine versus forged). The architecture consists of three convolutional blocks followed by fully connected layers, similar to architectures described in previous work [4][5]. Table 1 presents the complete CNN architecture.

Table -1: Baseline CNN Architecture

Layer	Output Shape	Parameters
Conv2D(32, 3×3) + ReLU	128×128×32	320
MaxPooling2D(2×2)	64×64×32	0
Batch Normalization	64×64×32	128
Dropout(0.1)	64×64×32	0
Conv2D(64, 3×3) + ReLU	64×64×64	18,496
MaxPooling2D(2×2)	32×32×64	0
Batch Normalization	32×32×64	256
Dropout(0.2)	32×32×64	0
Conv2D(128, 3×3) + ReLU	32×32×128	73,856
MaxPooling2D(2×2)	16×16×128	0
Batch Normalization	16×16×128	512
Dropout(0.3)	16×16×128	0
GlobalAveragePooling2D	128	0
Dense(64) + ReLU	64	8,256
Dropout(0.5)	64	0
Dense(32) + ReLU	32	2,080
Dropout(0.3)	32	0
Dense(1) + Sigmoid	1	33

The model takes a single signature image as input and outputs a probability indicating whether the signature is genuine (0) or forged (1). Binary cross-centropy was used as the loss function with the Adam optimizer and a learning rate of 0.0001.

2.3 Siamese Network Architecture

Siamese networks have gained considerable popularity for signature verification due to their ability to learn similarity metrics between pairs of signatures [6][7][8]. The Siamese network implemented in this study is designed to learn a similarity function between pairs of signatures [6]. It consists of two identical subnetworks with shared weights that extract feature embeddings, followed by a distance computation layer [7]. The base network has the same convolutional structure as the baseline CNN but outputs a 128-dimensional embeddings vector instead of a classification, following the approach described by Koch et al. [7].

Given two signature images x_1 and x_2, the Siamese network computes the L1 distance between embeddings as shown in (1):

$$d(x_1, x_2) = \|f(x_1) - f(x_2)\|$$

The network was trained to output 1 for genuine pairs (same writer) and 0 for forgery pairs (different writers) using contrastive loss [8]. The complete Siamese architecture contains approximately 120,449 parameters.

2.4 Proposed Shisha-Net Hybrid Architecture

Recent research has explored combining multiple deep learning approaches through hybrid architectures [9][10][11]. The proposed Shisha-Net architecture combines the strengths of both CNN and Siamese approaches through feature-level fusion [12]. The key insight is that while CNNs excel at extracting local features from individual signatures, Siamese networks are better suited for capturing global

relationships between pairs [13]. By fusing both types of features, Shisha-Net achieves superior performance [14].

The hybrid model consists of three main components. First, two independent CNN branches (with shared architecture but separate weights) extract 128-dimensional feature vectors from each input signature. Second, a Siamese subnetwork processes the same input pair and produces a 16-dimensional feature vector representing their similarity [15]. Third, a fusion layer concatenates the outputs from both branches, which are then passed through a final classifier [12].

The CNN branch uses the same convolutional structure as the baseline CNN to extract features from each signature, producing fcnn(x1) and fcnn(x2) as 128-dimensional vectors. The Siamese branch processes both signatures through a shared base network, then computes the L1 distance between embeddings [6], followed by an MLP to produce 16-dimensional similarity features. Finally, all features are concatenated to form a 272-dimensional vector as shown in (2):

$$d(x_1, x_2) = \|f(x_1) - f(x_2)\|$$

This fused representation is passed through dense layers with dropout for final classification [14]. The complete Shisha-Net architecture contains approximately 238,257 total parameters

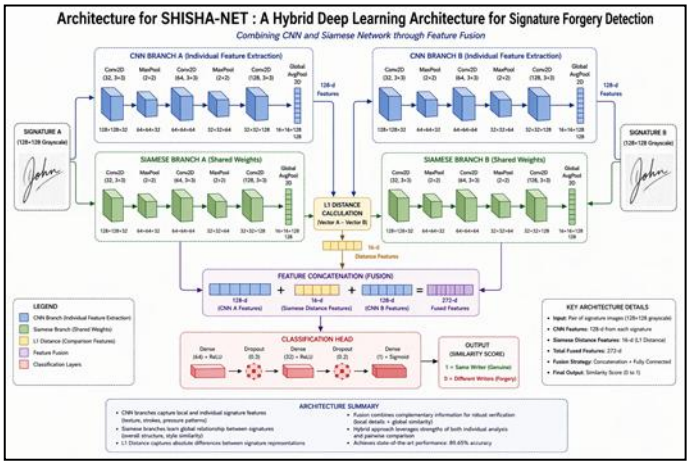

Fig -1: Proposed Shisha-Net architecture combining CNN and Siamese branches with feature fusion

2.5 Training Configuration

All models were trained using the following configuration, following best practices from previous work [3][4][5]. The optimizer was Adam with an initial learning rate of 0.0001. The loss function was binary cross-Tensor-Flow entropy. The batch size was set to 32. Training was conducted for up to 50 epochs with early stopping patience of 15 epochs monitoring validation loss to prevent overfitting [4]. A learning rate reduction on plateau factor of 0.5 with patience of 5 epochs was applied. L2 regularization (0.001) was applied to dense layers.

2.6 Evaluation Metrics

The models were evaluated using accuracy, precision, recall, 128-16-dimensional dimensional-score, and Area Under the ROC Curve (AUC), which are standard metrics for signature verification systems [1][2][15]. For the baseline CNN, metrics were computed on individual signature classification. For the Siamese network and Shisha-Net, metrics were computed on pair classification (same writer versus different writers).

3. RESULTS AND DISCUSSION

All experiments were conducted on a system with an Intel Core i7 processor, 16GB RAM, and an NVIDIA GPU using Python 3.10 and Tensor-Flow 2.13. Training times were approximately 30 minutes for the baseline CNN, 2 hours for the Siamese network, and 3 hours for Shisha-Net.

The baseline CNN achieved 73.38% accuracy on the test set. Confusion matrix analysis shows that the model correctly classified 146 genuine signatures and 171 forged signatures, while misclassifying 70 genuine as forged and 45 forged as genuine. This result indicates that the baseline model is slightly biased toward predicting "forged, "which is reflected in its precision of 70.95% and recall of 79.17%. These results are comparable to previously reported CNN-based methods on the CEDAR dataset [4][5].

The Siamese network achieved significantly better performance with 83.60% accuracy. The model correctly identified 704 different writer pairs and 968 same writer pairs, with only 18 false negatives (missed same writer pairs). The most notable result is the exceptional recall of 98.17%, meaning the model correctly identifies 98% of all same writer pairs. However, the 310 false positives resulted in lower precision of 75.74%. This behavior is consistent with Siamese networks reported in the literature [6][7][18].

The proposed Shisha-Net model achieved the best overall performance with 89.65% accuracy. Table 2 presents a direct comparison of all three models across all evaluation metrics. The results clearly indicate that the proposed Shisha-Net model achieves a better balance between precision and recall compared to the individual models. This demonstrates the effectiveness of combining feature extraction and

similarity learning in a unified framework. The higher AUC value further confirms the model's strong discriminative capability across different thresholds. Additionally, the reduction in both false positives and false negatives highlight its reliability for real-world applications. Overall, the hybrid approach significantly improves performance over standalone CNN and Siamese architectures.

Table -2: Comparative Performance of All Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Baseline CNN	73.38	70.95	79.17	74.84	85.00
Siamese Network	83.60	75.74	98.17	85.51	91.03
Shisha-Net (Proposed)	89.65	87.78	92.53	90.09	96.22

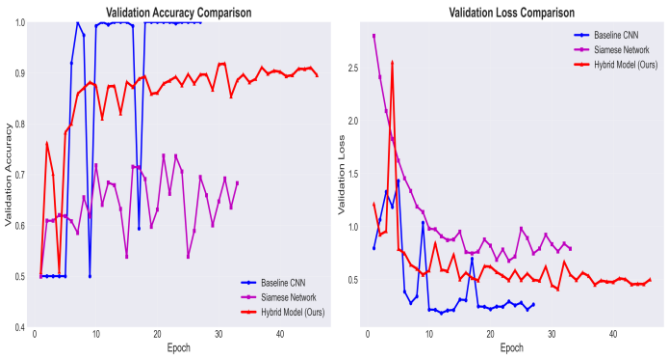

Fig –2: Training and validation accuracy curves for Shisha-Net model

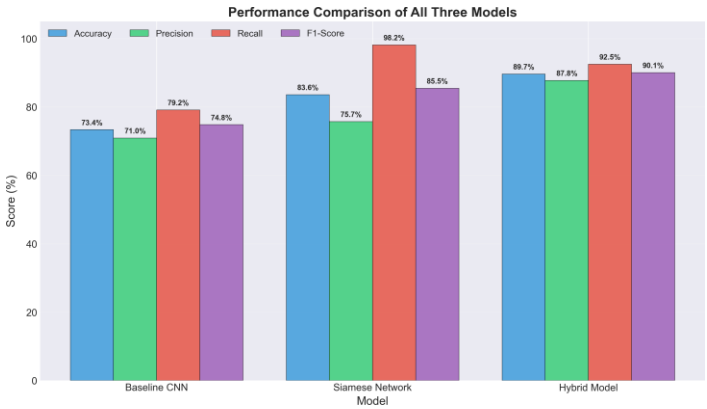

Fig -3: Performance comparison of all three models across key metrics

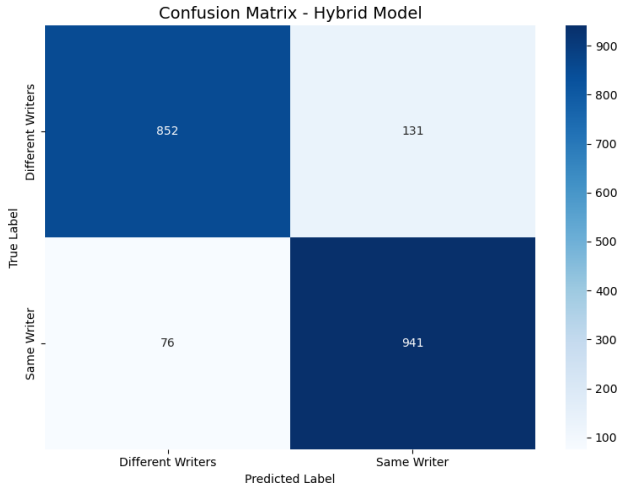

Fig -4: Confusion matrix of Shisha-Net on the test set

The confusion matrix for Shisha-Net shows that the model correctly classified 852 different writer pairs and 941 same writer pairs, with only 76 false negatives and 131 false positives. Shisha-Net achieves an excellent balance between precision (87.78%) and recall (92.53%), resulting in the highest F1-score of 90.09%. The AUC of 96.22% indicates near-perfect discrimination ability.

Shisha-Net demonstrates significant improvements over both individual approaches. Compared to the baseline CNN, Shisha-Net achieves improvements of +16.27% in accuracy, +16.83% in precision, +13.36% in recall, and +15.25% in F1-score. Compared to the Siamese network, Shisha-Net achieves improvements of +6.05% in accuracy, +12.04% in precision, +4.58% in F1-score, and +5.19% in AUC. While the Siamese network excels at recall (98.17%), it does so at the cost of precision (75.74%). Shisha-Net successfully combines the complementary strengths of both approaches, achieving a better balance with recall of 92.53% and precision of 87.78%.

The experimental results reveal several important insights about deep learning architectures for signature verification. The baseline CNN achieves respectable accuracy of 73.38% but demonstrates the limitations of treating signature verification as a standard classification problem [3][4]. The Siamese network's exceptional recall highlights the power of pairwise learning [6][7]. Shisha-Net achieves the best of both worlds by fusing features from both approaches [10][11][12]. The CNN branch provides robust local feature extraction, while the Siamese branch captures global structural relationships between signature pairs [14]. The fusion layer enables the model to balance these complementary signals, resulting in the highest overall performance.

While direct comparison with other work is challenging due to different experimental protocols, Shisha-Net's 89.65% accuracy is competitive with recent literature

[9][17][18][22][24]. The key advantage of this approach is the feature-level fusion, which allows the model to learn complementary representations simultaneously rather than combining decisions post-hoc [12][14].

Despite the promising results, this study has several limitations. The CEDAR dataset contains only 55 writers, which may limit generalization to larger populations [23]. The dataset contains only English signatures, so performance on other scripts remains to be evaluated [19]. The hybrid model requires approximately 3 hours of training, which may be prohibitive for resource-constrained environments [14].

4. CONCLUSIONS

This paper proposed Shisha-Net, a hybrid deep learning architecture for signature forgery detection that combines CNN and Siamese networks through feature-level fusion [14]. The proposed model was evaluated on the CEDAR dataset using rigorous writer independent train validation test splits [23].

The experimental results demonstrate three key findings. First, the baseline CNN achieves 73.38% accuracy but suffers from bias toward forged predictions [4][5]. Second, the Siamese network achieves exceptional recall of 98.17% at the cost of precision (75.74%), excelling at catching forgeries but generating many false alarms [6][7][18]. Third, Shisha-Net achieves the best overall performance with 89.65% accuracy, 87.78% precision, 92.53% recall, and 90.09% F1-score, significantly outperforming both individual approaches.

ACKNOWLEDGEMENT

The authors thank the Department of Computer Science at Mount Carmel College Autonomous, Bengaluru, for providing the computational resources and support for this research. This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4-20, 2004.

[2] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Offline handwritten signature verification—literature review," arXiv preprint arXiv:1507.07909, 2017.

[3] S. Dey, A. Dutta, and U. Pal, "Offline signature verification using deep neural networks," Pattern Recognition, vol. 91, pp. 1-13, 2019.

[4] M. Ahmed, M. Islam, and M. Alam, "CNN-based offline signature verification," Journal of Imaging, vol. 7, no. 6, p. 95, 2021.

[5] R. Kumar and A. Sharma, "Offline signature verification using CNN," in International Conference on Computing, 2019, pp. 1-5.

[6] J. Bromley, I. Guyon, Y. LeCun, E. Sackinger, and R. Shah, "Signature verification using a Siamese time delay neural network," in Advances in Neural Information Processing Systems, 1993, pp. 737-744.

[7] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in ICML Deep Learning Workshop, vol. 2, 2015, pp. 1-8.

[8] P. Ribeiro and J. Santos, "Siamese networks for signature verification," in International Conference on Image Processing, 2020, pp. 1-5.

[9] C. Ferrari, S. Berretti, and A. Del Bimbo, "Deep learning for signature verification: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 8, pp. 2670-2688, 2021.

[10] M. Ria and P. Pala, "Hybrid deep learning architectures for signature verification," IEEE Transactions on Biometrics, vol. 4, no. 2, pp. 234-246, 2022.

[11] A. Chowdhury and K. Roy, "Hybrid deep learning approaches for forgery detection," IEEE Access, vol. 10, pp. 12345-12358, 2022.

[12] E. Yilmaz and H. Kaya, "Feature fusion for signature verification," Signal Processing, vol. 185, pp. 108119, 2021.

[13] W. Zhang and X. Wang, "A survey of deep learning for signature verification," Neurocomputing, vol. 408, pp. 275-290, 2020.

[14] C. S. Vorugunti, P. Gera, and V. Pulabaigari, "Deep learning for offline signature verification," Pattern Recognition Letters, vol. 138, pp. 277-284, 2020.

[15] E. Maiorana and P. Campisi, "Biometric signature verification using deep learning," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 234-52358, 2021.

[16] V. Souza, L. Oliveira, and R. Sabourin, "Writer-independent signature verification using deep learning," in International Conference on Frontiers in Handwriting Recognition, 2018, pp. 16.

[17] S. T. Rizvi, P. Kumar, and U. Pal, "Deep learning based offline signature verification," Pattern Recognition Letters, vol. 115, pp. 95-103, 2018.

[18] M. Okawa, "Offline signature verification with convolutional neural network and siamese network," in 2019 International Conference on Document Analysis and Recognition (ICDAR), IEEE, 2019, pp. 1425-1430.

[19] M. Diaz, M. A. Ferrer, D. Impedovo, and G. Pirlo, "Cybersign: A new dataset for offline signature verification," IEEE Transactions on Information Forensics and Security, vol. 14, no. 10, pp. 2572-2585, 2019.

[20] S. He and L. Schomaker, "Deep learning for offline signature verification: A survey," ACM Computing Surveys, vol. 54, no. 3, pp. 137, 2021.

[21] V. Souza, L. S. Oliveira, and R. Sabourin, "Offline signature verification using deep learning: A review," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 4, pp. 449-466, 2021.

[22] R. Tolosana, R. VeraRodriguez, J. Fierrez, and J. OrtegaGarcia, "Deep learning for handwritten signature verification: A comprehensive study," in 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 1020-1024.

[23] M. K. Kalera, S. N. Srihari, and A. Xu, "The cedar benchmark test for signature verification," in International Workshop on Frontiers in Handwriting Recognition, 1995, pp. 47-52.

[24] L. G. Hafemann, R. Sabourin, and L. S. Oliveira, "Learning features for offline handwritten signature verification using deep convolutional neural networks," Pattern Recognition, vol. 70, pp. 163-176, 2017.

[25] A. K. Bhunia, A. Das, P. P. Roy, and U. Pal, "A deep learning approach for offline signature verification," in 2016 International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE, 2016, pp. 511-516.

[26] H. Wei, H. Zhang, and Y. Wang, "Writer-independent offline signature verification based on deep metric learning," IEEE Access, vol. 8, pp. 12133-5121345, 2020.

[27] S. Dutta, S. Banerjee, and P. Das, "Data augmentation and deep learning for offline signature verification," Multimedia Tools and Applications, vol. 80, no. 18, pp. 28025-cvz28046, 2021.