JCRINN
Journal of Computing Research and Innovation
e-ISSN : 2600-8793

Available online at
https://jcrinn.com/
https://crinn.conferencehunter.com/

Journal of Computing Research and Innovation 10(2) 2025

Journal of Computing Research and Innovation

AI-Driven Forgery Detection in Offline Handwriting Signatures: Advances, Challenges, and the Role of Generative Adversarial Networks

Safura Adeela Sukiman[1*], Nor Azura Husin[2], Hazlina Hamdan[3], Masrah Azrifah Azmi Murad[4]

[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Johor Branch, Segamat Campus, 85000, Segamat, Malaysia.*
[2,3,4]*Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Malaysia.*

ARTICLE INFO

Article history:
Received 13 June 2025
Revised 19 July 2025
Accepted 22 July 2025
Published 1 September 2025

Keywords:
Offline Handwriting Signatures
Signature Forgery Detection
Generative Adversarial Networks
Siamese Networks
Autoencoders
Deep Learning

DOI:
10.24191/jcrinn.v10i2.532

ABSTRACT

Handwriting-based authentication continues to be a critical element in forensic analysis, particularly in the context of document fraud and signature forgery. Although deep learning (DL) techniques have shown promising results, there are still obstacles associated with the availability of limited datasets, the generalization of models, and their robustness. This review conducts a systematic examination of recent developments in DL methods for signature forgery detection. It employs the PRISMA protocol and retrieves literature from four well-established databases: Scopus, ACM Digital Library, Web of Science, and IEEE Xplore. Following a rigorous screening procedure, a total of 15 primary studies published between 2019 and 2025 were selected from an initial 115 records that were filtered by Computer Science subject area, English language, and original research articles. Five publicly accessible datasets: CEDAR, BHSig260, ICDAR 2011 SigComp, Kaggle signature verification dataset by RobinReni, and Kaggle handwritten signatures by Divyansh Rai were identified and analysed. The review indicates that Siamese networks dominate the DL architecture for signature forgery detection tasks, while alternative methods either employed fine-tuned pre-trained models (i.e., VGG16) or a hybrid of autoencoders and Convolutional Neural Networks (CNNs). An accuracy of 100% has been achieved through utilization of Siamese network leveraging the CEDAR dataset. This result is reasonable since CEDAR has the advantages of clean and balanced dataset. In response to the persisting limitations, this review emphasizes Generative Adversarial Networks (GANs) as the powerful data augmentation technique and a potential solution to enrich training datasets, simulate diverse forgery patterns, and enhance the robustness of models. Finally, a generative-aware conceptual framework is proposed at the end of the review to inform future research on the development of offline handwriting signature forgery detection system that is more resilient and forensic-ready.

[1*] Corresponding author. *E-mail address*: safur185@uitm.edu.my

https://doi.org/10.24191/jcrinn.v10i2.532

1. INTRODUCTION

Handwriting continues to play an important role in personal authentication, particularly in forensic and legal contexts where signature verification is central to identity validation and document security. Despite the proliferation of digital alternatives, handwritten signatures remain widely utilized in sectors such as such as banking, contracts, education, and healthcare, making them susceptible to forgery and fraud.

Offline handwriting signature forgeries can be broadly classified into three categories: random (or blind) forgeries, simple (or unskilled) forgeries, and skilled (or simulated) forgeries. Random forgeries occur when the forger has no access to the genuine signature and attempts to replicate it without any reference, resulting in signatures with little to no similarity to the original. Meanwhile, simple forgeries occur when the forger knows the person's name but does not have access to the actual signature, resulting in attempts that may mimic the general style but lack precision. Skilled forgeries, on the other hand, are produced by those who have access to actual signature samples and have practiced copying them, making them the most difficult to detect since they are so identical to original signatures (Hafemann et al., 2017).

The conventional methods of signature forgery detection, which frequently depend on human examiners or handcrafted features, are constrained in their scalability, objectivity, and accuracy. The automation of signature forgery detection has gained more prominence as a result of the emergence of artificial intelligence (AI) and DL. Sequential CNNs (Balaji et al., 2024), Siamese networks (Joe Harris & Anitha, 2023) and Autoencoder-based (Swamy et al., 2024) models have all shown significant potential in the detection of skilled and simulated forgeries. These models provide adaptability across writer-independent scenarios in addition to enhanced performance. Nevertheless, substantial challenges continue to exist, despite these advancements. The generalizability of the model is still constrained by dataset limitations, varying handwriting styles, and the complexity of skilled forgeries. Furthermore, most models trained on benchmark datasets often struggle to maintain robustness under real-world conditions that include noise, document degradation, and cross-domain variations (Engin et al., 2020).

Addressing these dataset limitations and generalizability issues, GANs have emerged as a powerful data augmentation component within the training pipeline. In contrast to discriminative models such as CNNs, which generally utilize conventional data augmentation methods like rotation, flipping, or scaling (Swamy et al., 2024), GANs is capable to generate entirely new and realistic samples that closely resemble the training distribution (Goodfellow et al., 2017; Wang & Jia, 2019). In the context of offline signature forgery detection, GANs can simulate high-fidelity synthetic forgeries, particularly skilled forgeries, which traditional augmentation methods often struggle to realistically re-create. This not only expands the dataset but also introduces controlled variability, allowing models to better generalize and resist overfitting. Consequently, GANs represent a significant paradigm shift in the approach to data scarcity, variability, and robustness in the context of AI-driven signature forgery detection. This review emphasizes GANs not merely as a supporting technique but as a transformative tool for enhancing the realism, diversity, and forensic-readiness of training datasets.

While several past reviews have examined handwriting and signature verification systems, many have predominantly concentrated on machine learning techniques (Soelistio et al., 2021) without incorporating the most recent advancements in deep learning and generative adversarial networks. Hafemann et al. (2017) conducted a survey of offline signature verification approaches but omitted current architectural innovations, like Transformer-based models and Siamese Networks. In contrast, this review incorporates recent peer-reviewed works (2019–2025) employing current deep learning architectures with multiple benchmark datasets. More importantly, it uniquely explores the emerging role of generative adversarial networks in addressing persistent challenges such as data scarcity, complexity, and model robustness, which remains underexplored in existing literature. The following shows the objectives of this review:

https://doi.org/10.24191/jcrinn.v10i2.532

a) To synthesize recent deep learning models in AI-driven offline handwriting signature forgery detection.
b) To analyse challenges especially the ones related to offline handwriting signature datasets.
c) To explore the emerging role of generative adversarial networks in addressing offline handwriting signature datasets challenges.
d) To propose a generative-aware conceptual framework that is recent, more resilient and forensic-ready for future development of offline handwriting signature forgery detection system.

2. METHODOLOGY

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol, a widely recognized framework for improving the rigor and reproducibility of systematic reviews in scientific domains, is invoked in this review to ensure a structured and transparent approach (Page et al., 2021). This protocol was modified to accommodate the scope of this review, which is focused on the application of deep learning in handwriting forensics, with a particular emphasis on the detection of offline signature forgery and identity-related manipulation. The PRISMA-based strategy facilitated a rigorous process of identification, screening, eligibility evaluation, and inclusion of relevant articles, thereby ensuring consistency and objectivity throughout the review process.

The review process consisted of four important stages: (a) the identification of relevant records from selected databases, (b) the removal of duplicates, (c) the assessment of titles and abstracts with respect to inclusion and exclusion criteria, and (d) the full-text evaluation of eligible studies.

2.1 Database selection and search strategy

Four established and high-impact databases were utilized for literature retrieval: Scopus, ACM Digital Library, IEEE Xplore, and Web of Science. The databases were chosen for their comprehensive coverage of peer-reviewed publications in the fields of computer science, artificial intelligence, and digital forensics. The following search queries were composed using meticulously structured Boolean expressions that combined domain-specific terms:

("handwriting forgery" OR "signature forgery" OR "document fraud") AND ("deep learning" OR "neural networks" OR "generative artificial intelligence" OR "generative adversarial networks")

The search was confined to articles published between 2019 and 2025, written in English, and classified under the Computer Science subject area. This period was chosen to indicate the time at which generative adversarial networks started to be practically utilized in handwriting forensics tasks including synthetic handwriting generation, anomaly detection, and counterfeit detection. Restricting the review to this time ensures that the included studies fit the present methodological setting and are pertinent to the state-of-the-art advancements in the domain. Reviews, editorials, and non-peer-reviewed works were excluded, and only original research articles and conference papers were considered.

2.2 Study identification and screening process

A preliminary collection of 115 records was obtained: Scopus (n = 47), ACM (n = 18), IEEE Xplore (n = 37), and Web of Science (n = 13). Following the elimination of 59 duplicate entries, 56 distinct records underwent a multi-phase screening process that included title screening, abstract screening, and full-text assessment (before any screening process n = 56).

The relevance of the articles was initially determined by looking at their titles. A total of eleven records were disregarded because they were deemed irrelevant. These records included studies that concentrated on hardware implementation or emotion detection rather than forgery analysis. After filtering the titles, the total number of articles is 45 (title screened n=45).

https://doi.org/10.24191/jcrinn.v10i2.532

Following that, the relevance of articles was screened by reading their abstracts. Nine studies were disqualified because they presented models that were simply conceptual in nature, frameworks that were end-to-end but did not have empirical validation, or content that was not in English. The total number of articles after the screening of abstracts (abstract screened n = 36).

Finally, the inclusion criteria were met by 15 articles after the full texts were evaluated. The remaining 21 articles were excluded due to the following reasons: the use of privately collected or real-time data, the absence of performance evaluation, or the presentation of conceptual frameworks that were not reproducible. The total number of articles after full-text screening is 15 (full-text screened n = 15). Fig. 1. depicted the flow diagram of PRISMA protocol for both identification and screening process of this review.

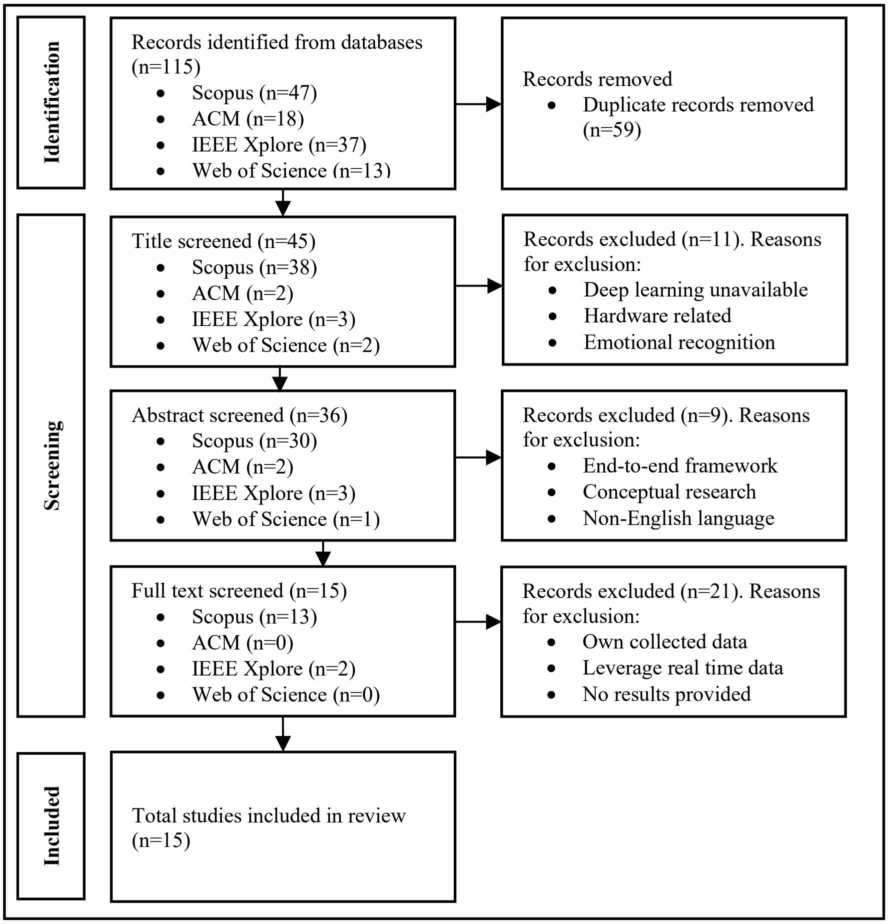

Fig. 1. PRISMA protocol flow diagram for filtering 15 relevant articles

2.3 Inclusion and exclusion criteria

This study employed inclusion and exclusion criteria to maintain the relevance and quality of the literature examined during the article selection process. Studies were selected if published in peer-reviewed journals or conference proceedings and focused on offline handwriting signature forgery detection. Moreover, qualifying studies must have utilized artificial intelligence (AI) or DL methodologies, encompassing, but not restricted to CNNs, Transformer-based architecture, hybrid models, or generative models. A fundamental requirement was the reporting of evaluation metrics pertinent to forgery detection,

https://doi.org/10.24191/jcrinn.v10i2.532

including accuracy, precision, recall, and F1-score. Only research published in English were included to ensure consistency and interpretative accessibility.

On the other hand, studies were excluded if they were based on non-peer-reviewed sources, such as blog entries, technical reports, or preprints that did not undergo rigorous peer review. Articles that were exclusively theoretical, which lacking experimental validation or implementation results were also excluded. In addition, this review did not include research that exclusively focused on handwriting recognition (i.e., character or word transcription) without a direct emphasis on forgery or forensic detection applications.

Consequently, the following key attributes were extracted from each study: the specific forgery scenario addressed, the year of publication, the dataset(s) used, the learning model architecture, the use of generative techniques, performance results, and limitations noted by authors. This data was collected to facilitate comparative synthesis, trend analysis, and gap identification in the subsequent sections.

3. RESULTS

The results section is organized to initially provide an overview of the datasets, followed by the performance comparative analysis of all 15 deep learning models from the PRISMA protocol in offline handwriting signature forgery detection.

3.1 Dataset overview

This review incorporates five (5) publicly available offline signature datasets, which are extensively used in offline handwriting signature forgery detection studies. These datasets contain a balanced mix of genuine and forged samples from a wide range of signers and serve as important benchmarks for training and evaluating deep learning models. The datasets chosen include CEDAR (Srihari & Leedham, 2001), BHSig260 (Kathuria, 2022), ICDAR 2011 SigComp (Liwicki et al., 2011), Kaggle signature verification dataset by RobinReni (Reni, 2019), and Kaggle handwritten signatures by Divyansh Rai (Rai, 2020).

The CEDAR dataset (Srihari & Leedham, 2001) is widely regarded as a fundamental benchmark in offline signature verification. The dataset comprises 2,640 grayscale signature images from 55 signers, with each signer providing 24 genuine and 24 forged samples. A primary advantage is its balanced architecture and regulated acquisition settings, rendering it optimal for training and assessing models in a noise-free setting. Nonetheless, a significant disadvantage of the dataset is its comparatively small participant pool, which constrains its capacity to generalize across varied writing styles. The dataset consists solely of English-language signatures, hence constraining its relevance in multilingual or cross-script research.

Meanwhile, BHSig260 dataset (Kathuria, 2022) is a large-scale signature dataset with 260 signers: 100 in Bengali and 160 in Hindi. It comprises 24 authentic and 30 counterfeit signatures per user, amounting to almost 14,000 samples. Its strength is its wide range of languages and big sample size, which make it a useful benchmark for evaluating signature verification systems in scripts other than Latin. However, its constraints include script specificity, making it less relevant for Latin-based handwriting models as well as diversity in the quality and style of forgeries, which can generate errors during training or evaluation.

On the other hand, the ICDAR 2011 SigComp dataset (Liwicki et al., 2011) was initially developed for a competition. The signature images of 106 individuals are included in this dataset, with 64 of them being used for training and 42 for testing. The test set contains a minimum of 12 genuine and 12 forged samples per signer, while the training set contains a minimum of 24 genuine and 8 forged signatures per user. The dataset's primary advantage is its design for benchmarking, which allows for a rigorous writer-independent evaluation due to appropriate separation between training and testing users. Nevertheless, its constraints include the fact that it is relatively older and has some image quality constraints, as well as limited cultural or linguistic diversity, as it exclusively studies Dutch signatures.

https://doi.org/10.24191/jcrinn.v10i2.532

Reni (2019) created the Kaggle signature verification dataset, which is a preprocessed version of Dutch signature samples from the ICDAR 2011 SigComp. It contains 2,149 labeled images of genuine and forged signatures. Its key advantages are accessibility and usability. This dataset is curated in a format that simplifies loading and training, making it ideal for rapid prototyping and model development. On the downside, it does not adhere to a rigorous train/test split and has a moderate dataset size, which may limit its appropriateness for models that require larger-scale evaluations.

Lastly, the Kaggle handwritten signatures of Divyansh Rai (Rai, 2020). This small collection has 300 signature images from 30 signers, with each providing five genuine and five forged samples. Its tiny size makes it ideal for educational applications, rapid model prototyping, and resource-constrained scenarios. However, its scale has severe limitations in terms of unpredictability, generalizability, and real-world use. It might not be appropriate for training deep models or conducting large-scale studies.

Collectively, these datasets encompass a wide variety of forgery types, writing scripts, and signer diversity, rendering them appropriate for assessing the robustness and efficacy of deep learning models in the detection of offline signature forgeries. Their public availability guarantees the reproducibility and accessibility of academic and applied research. Each dataset is further synthesized in Table 1, which includes its key features, advantages, as well as disadvantages.

Table 1. Key features, advantages, and disadvantages of selected dataset

Dataset	Key features	Advantage(s)	Disadvantage(s)
CEDAR (Srihari & Leedham, 2001)	55 signers, 24 genuine 24 forged samples each, total of 2640 grayscale images	Balanced and clean dataset, advisable for benchmarking	Limited signer diversity, English only, less suitable for cross-script research
BHSig260 (Kathuria, 2022)	260 signers (100 Bengali, 160 Hindi), 24 genuine 30 forged samples each	Large-scale, multilingual script support, high variation	Limited to Indic script, variability in forgery quality
ICDAR 2011 SigComp (Liwicki et al., 2011)	106 signers, 64 signers for training, 42 signers for testing,	Supports writer-independent evaluation	Dutch only, some outdated samples, limited scalability
Kaggle signature verification dataset (Reni, 2019)	2149 pre-processed Dutch signature images	Easy to use, curated format, ideal for fast prototyping	Medium sized, no strict train/test split, limited cultural diversity
Kaggle handwritten signatures of Divyansh Rai (Rai, 2020)	30 signers, 5 genuine 5 forged samples each, total of 300 offline handwriting samples	Lightweight, useful for small-scale or educational projects	Quite small for deep learning models, limited variability and generalization

3.2 Data augmentation techniques

Data augmentation is often crucial for improving the robustness and generalizability of deep learning models, particularly when working with limited datasets such as those common in offline signature forgery detection. Nevertheless, data augmentation approaches were predominantly absent from the examined studies. Traditional augmentation methods were reported in a restricted number of publications (Swamy et al., 2024), specifically applying geometric transformations such as image flipping, rotation, cropping, and contrast or brightness adjustments to artificially increase training diversity. Despite their simplicity, these methods may offer only marginal improvements in simulating skilled forgeries, which frequently encompass nuanced, high-fidelity handwriting characteristics that such techniques are incapable of replicating.

The lack of complete augmentation in most studies reveals a fundamental gap in the existing research landscape. This constraint has obvious implications for model overfitting and poor generalization, especially when evaluated against unknown or degraded samples. In response, this review proposes the integration of Generative Adversarial Networks (GANs) specifically Signature GANs, as a more advanced

https://doi.org/10.24191/jcrinn.v10i2.532

data augmentation technique. GANs provide a unique benefit by generating realistic synthetic forgeries that closely resemble variances in human handwriting (Goodfellow et al., 2017; Wang & Jia, 2019). In contrast to traditional augmentations, GANs learn from authentic data distributions and generate intricate and realistic forgeries, which introduce controlled variability and improve the training set diversity. These properties allow GANs to surpass the constraints of traditional data augmentation techniques, which frequently fail to capture the intricate, skilled-level features of handwritten signatures. As a result, GAN-based augmentation is essential for mitigating dataset scarcity and improving model robustness in the offline signature forgery detection process. This augmentation strategy not only strengthens model learning but also enhances resilience against adversarial inputs and real-world noise, supporting the forensic readiness of future signature forgery detection systems.

3.3 Comparative analysis

This section provides a synthesis of the accuracy findings obtained from the total studies included in review (n=15) which are shown in Table 2. All studies employed the deep learning methods to the forgery detection of offline handwriting signatures. Accuracy is selected as the primary metric as it reflects as the most frequently reported, and interpretable metric. Accuracy measures the proportion of correctly classified signatures (genuine or forged) relative to the total number of samples.

Table 2. Accuracy comparative analysis

Study	Deep learning method	Dataset	Accuracy result (%)
Joe Harris & Anitha (2023)	Siamese network, Euclidean distance, contrastive loss function	CEDAR	100.0
Chokshi et al. (2023)	Siamese network, scattering wavelets learning approach, Euclidean distance	CEDAR	99.91
Balaji et al. (2024)	CNN with hyperparameter tuning	Kaggle signature verification dataset by RobinReni	99.8
Emberi et al. (2023)	Siamese network	Kaggle signature verification dataset by RobinReni	99.756
Majumder et al. (2023)	Siamese Transformer network, triplet loss function	CEDAR	99.17
Anitha et al. (2024)	Siamese network, Harris corner feature detection	BHSig260 (Hindi only)	98.9
Swamy et al. (2024)	Hybrid of autoencoders and CNN	CEDAR	98.4848
Shirisha et al. (2024)	VGG16 with transfer learning	Kaggle signature verification dataset by RobinReni	98.26
Chokshi et al. (2023)	Siamese network, scattering wavelets learning approach, Euclidean distance	ICDAR 2011 SigComp	97.76
Minh Tram & Chau (2024)	Pre-trained VGG16 with hyperparameter tuning	CEDAR	96.14
Reddy et al. (2024)	Siamese network, Euclidean distance, contrastive loss function	ICDAR 2011 SigComp	96.0
Krishna & Bhuvaneswari (2023)	20 hidden layers of multi-layer perceptron	Kaggle handwritten signatures by Divyansh Rai	92.4
Tehsin et al. (2024)	Siamese network, Euclidean distance, triplet loss function	BHSig260	91.5
Joe Harris & Anitha (2023)	Siamese network, Euclidean distance, contrastive loss function	BHSig260 (Hindi only)	87.13
Tehsin et al. (2024)	Triplet loss Siamese network	CEDAR	86.1
Tarek & Atia (2022)	Pre-trained ResNet50	Kaggle handwritten signatures by Divyansh Rai	82.0
Chaturvedi & Jain (2022)	Ensemble feature extractor and classifier, concatenation of geometrical features and features extracted from pre-trained MobileNet	BHSig260 (Hindi skilled forgery)	69.1
Jain et al. (2021)	Shallow Siamese network, Euclidean distance, contrastive loss function	Kaggle signature verification dataset by RobinReni	73.0 (Precision)

https://doi.org/10.24191/jcrinn.v10i2.532

Table 2 reveals that offline signature forgery detection with deep learning model can be divided into three (3) categories: high, moderate, and low performances. Deep learning models that achieve more than 90% accuracy belongs under high performance category. Meanwhile, deep learning models with accuracy between 80% to 89% falls under the moderate performance category, and finally the models with accuracy below 80% is considered low performance deep learning models. The research conducted by Joe Harris & Anitha (2023) attained flawless accuracy of 100% utilizing a Siamese network with contrastive loss on the CEDAR dataset, ranking among the most effective deep learning model.

Siamese networks dominate the offline signature forgery detection field because of their robustness and adaptability across datasets. However, the choice of dataset has a significant impact on model performance, with CEDAR (Srihari & Leedham, 2001) and Kaggle signature verification dataset by RobinReni (Reni, 2019) provides higher accuracies, but datasets such as BHSig260 (Kathuria, 2022) particularly with skilled forgeries pose larger hurdles. This can be seen through the work by Chaturvedi & Jain (2022) where they reported only 69.1% accuracy despite employing a Siamese network, underscoring how dataset-specific challenges can affect results. The study by Jain et al. (2021) reported the precision results instead of accuracy. On the other hand, integrating advanced components such as scattering wavelets, Transformer blocks, or Autoencoders significantly boost forgery detection performance, suggesting captivating paths for future research.

4. DISCUSSIONS AND FUTURE WORK

This section discusses several deep learning models that achieve high performance (accuracy more than 90%) in forgery detection of offline handwriting signatures. Notably, Siamese networks, Siamese networks with scattering wavelets learning approach, Siamese Transformer network, VGG16 with transfer learning, and CNNs integrated with autoencoders have consistently achieved accuracy ranging from 96% to 100% across benchmark datasets. These findings highlight the potential of embedding-based similarity measures (Siamese networks), spatial encoding (Transformers), and convolutional feature extraction (VGG16 and CNNs) in learning offline signature features effectively.

The Siamese networks (Joe Harris & Anitha, 2023; Emberi et al., 2023; Reddy et al., 2024) dominate the landscape, often coupled with enhancements such as scattering wavelets learning approach (Chokshi et al., 2023) or Transformer network (Majumder et al., 2023). Siamese networks outperform the standard deep learning model i.e., standalone CNN, which often inadequately captures particular signature styles, resulting in suboptimal performance. Their success lies in their architecture's ability to learn relative similarity between signature pairs, allowing them to generalize well even with limited samples. Typically, a Siamese network consists of two identical sub-networks sharing weights and parameters, each receiving a different input image. The outputs are compared using a distance metric, usually Euclidean distance to determine its similarity. This architecture is particularly effective when trained with a contrastive loss function, which minimizes the distance between matching pairs while maximizing it for non-matching ones (Hadsell et al., 2006). Another alternative is the triplet loss function. This function compares an anchor image to both a positive and a negative sample and optimizes the network so that the anchor is closer to the positive than the negative by a specified margin (Schroff et al., 2015). While contrastive loss is computationally efficient with fewer samples, triplet loss frequently delivers greater discriminatory power in complex classification circumstances, especially when there are subtle differences in signatures. Table 3 depicts the differences between contrastive and triplet loss functions according to several criterion.

https://doi.org/10.24191/jcrinn.v10i2.532

Table 3. Contrastive loss function vs triplet loss function

Criterion	Contrastive loss function	Triplet loss function	Recommended use case
Input structure	Pairwise input (genuine-genuine or genuine-forgery)	Triplet input (anchor, positive, negative)	Triplet loss is preferred when subtle intra-class differences are present
Loss objective	Minimize distance for similar pairs, maximize for dissimilar ones	Ensures anchor is closer to positive than to negative by a specified margin	Contrastive loss is effective for simpler verification tasks
Training efficiency	Faster training with fewer samples	Requires more structured triplet selection, leading to slower training	Use contrastive when data is limited or less diverse
Discriminative power	Moderate	Higher in complex decision boundaries	Triplet loss for fine-grained classification such as skilled forgery detection
Implementation complexity	Relatively straightforward	More complex due to triplet mining strategies	Contrastive for prototype verification; Triplet for high security systems

The comparisons in Table 3 suggest that in low-resource or low-variation settings, contrastive loss offers a simpler and effective approach. Meanwhile, for high-stakes applications involving skilled forgeries or cross-writer verification, triplet loss provides superior performance and granularity.

Furthermore, recent advancements in offline handwriting signatures forgery detection have investigated hybrid architectures that integrate the strengths of Siamese networks with Transformer-based encoders. This architecture employs a Siamese network, where each branch incorporates a Transformer encoder that analyzes signature embeddings via self-attention methods. In contrast to CNNs that concentrate mostly on local spatial attributes, Transformer-based modules allow the model to comprehend global structural linkages among strokes, curves, and spacing abnormalities, which are frequently essential for identifying competent forgeries. Table 4 shows the advantages and disadvantages of integrating the Transformer encoder into the Siamese network.

Table 4. Advantages and disadvantages of transformer encoder

Advantage(s)	Disadvantage(s)
Captures both local texture and global signature structure	Transformer layers add significant parameters thus, require careful model optimization
Reduces overfitting on writer specific traits	Self-attention mechanism typically performs better with larger training datasets
More adaptable to variable-length inputs	May suffer from convergence issues if not properly regularized or pre-trained

Meanwhile, the VGG16 deep learning model, a well-established architecture in image classification, also proves effective when adapted to offline handwriting signature forgery detection, particularly with transfer learning (Shirisha et al., 2024) and hyperparameter tuning (Minh Tram & Chau, 2024). VGG16's architecture is composed of 13 convolutional layers, followed by three fully connected layers, all utilizing small (3x3) convolutional filters with ReLU activations. This deep and uniform structure enables it to extract rich hierarchical features from input images, rendering it particularly well-suited for tasks such as signature verification, where subtle stroke differences are significant (Simonyan & Zisserman, 2014). Nevertheless, their efficacy is contingent upon the dataset and the fine-tuning strategy that is implemented. VGG16 is susceptible to overfitting and slow training when the dataset size is limited due to its extensive parameter set.

https://doi.org/10.24191/jcrinn.v10i2.532

On the other hand, hybrid deep learning models that combine CNNs with autoencoders (Swamy et al., 2024) also demonstrate potential by compressing signature representations and reconstructing them to emphasize distinguishing features. Autoencoders are unsupervised learning models that acquire the ability to encode input data into a lower-dimensional latent space and subsequently decode it back to its original form (Khan et al., 2020). Within the context of CNNs with autoencoders, the CNN acts as a powerful feature extractor, while the autoencoder refines these features by reconstructing the signature image. This reconstruction process forces the hybrid model to preserve the most salient and distinguishing features of a signature, effectively reducing noise and irrelevant variations. As a result, this model becomes more robust in distinguishing between genuine and forged signatures, particularly in datasets with limited labeled samples.

However, despite the encouraging performance in benchmark scenarios, three major challenges persist that impact real-world deployment:

a) Dataset challenges. Most studies relied on limited or specific datasets, which reduces the diversity of signature variations in terms of culture, language, and writing instruments. Subsequently, models trained on these datasets tend to show high performance in controlled environments but struggle to generalize. The implication is a significant gap in model robustness when deployed in practical and diverse contexts.

b) Discriminative model challenges. While discriminative models like Siamese and VGG-based architectures achieve high accuracy, they often operate as black boxes. Their dependency on dataset-specific features makes them prone to overfitting, and the lack of transparency reduces trust in high-stakes environments like banking or legal authentication. Moreover, the absence of interpretability tools limits understanding of model decisions.

c) Deployment challenges. Most of the developed models lack resilience to noise and real-world variability. The accuracy scores reported in the literature assume clean and well-aligned signature samples, while actual verification systems must contend with occlusions, blurred scans, and variable lighting. As a result, model robustness under real-world deployment remains an open challenge.

4.1 Towards end-to-end generative-aware conceptual framework

An end-to-end generative-aware conceptual framework is proposed. It unifies several essential innovations: data augmentation, hybrid architecture, noise resilience, and explainable verification. This proposed framework addresses current limitations in dataset diversity, model generalization, real-world robustness, and interpretability. It is further organized into four interconnected stages as shown on Fig. 2.

https://doi.org/10.24191/jcrinn.v10i2.532

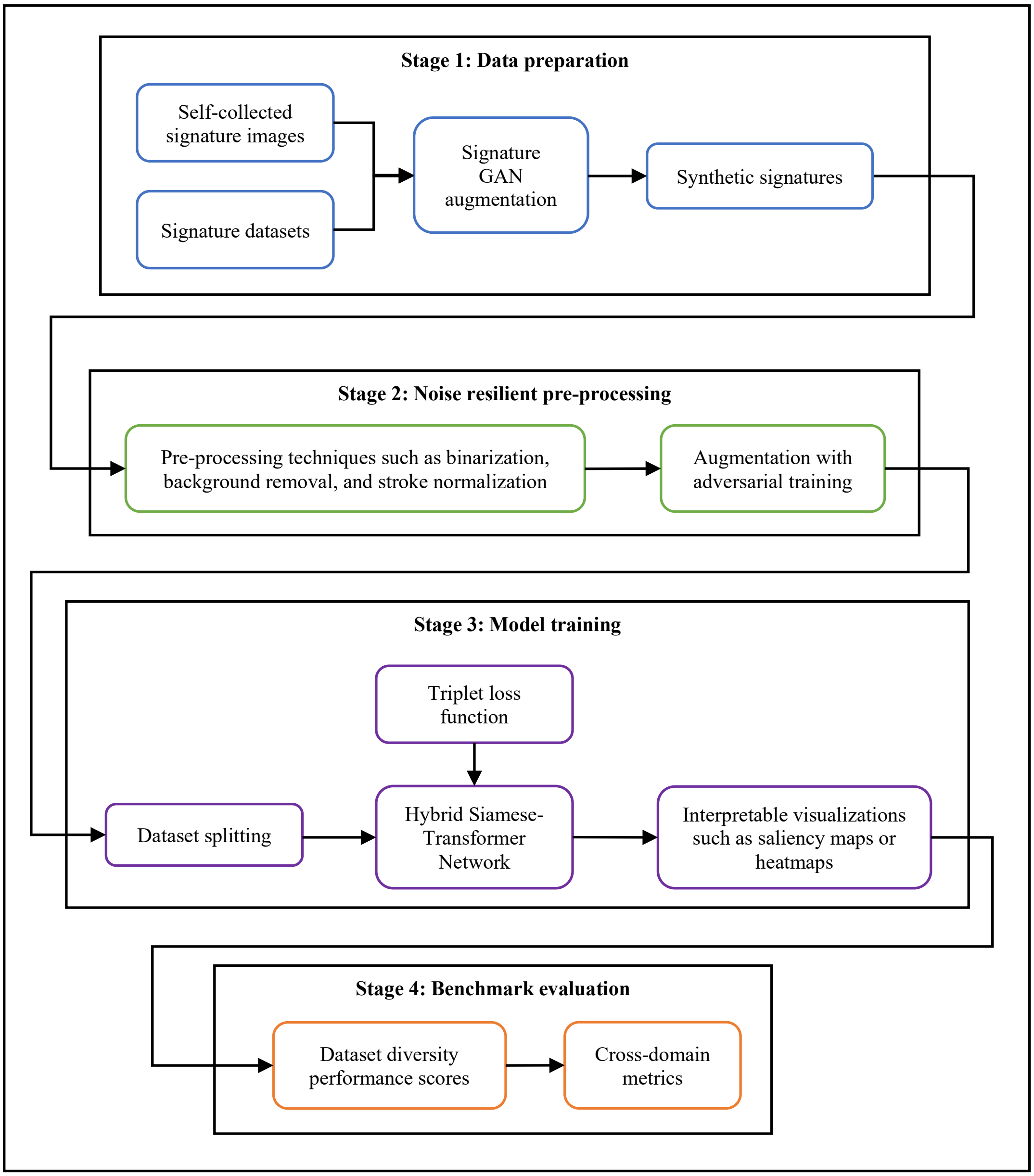
Stage 1: Data preparation
Self-collected signature images
Signature datasets
Signature GAN augmentation
Synthetic signatures
Stage 2: Noise resilient pre-processing
Pre-processing techniques such as binarization, background removal, and stroke normalization
Augmentation with adversarial training
Stage 3: Model training
Triplet loss function
Dataset splitting
Hybrid Siamese-Transformer Network
Interpretable visualizations such as saliency maps or heatmaps
Stage 4: Benchmark evaluation
Dataset diversity performance scores
Cross-domain metrics

Fig. 2. Proposed end-to-end generative-aware conceptual framework

Stage 1: Data Preparation. Self-collected signature images as well as signature datasets such as CEDAR (Srihari & Leedham, 2001) and BHSig260 (Kathuria, 2022) are composed as foundational data. A specialized Signature Generative Adversarial Network (GAN) is trained to generate synthetic forgeries that simulate variations in signatures' style and structure. It utilizes Generative Adversarial Networks (GANs) to synthetically produce realistic signature samples by learning the intricate variations found in handwriting, such as pressure, stroke curvature, and pen lifts. Introduced by Goodfellow et al. (2017), GANs consist of two core components: a generator that aims to produce synthetic data samples resembling real data, and a discriminator that evaluates whether a given sample is real or generated. These two networks are trained

https://doi.org/10.24191/jcrinn.v10i2.532

simultaneously in a minimax game, where the generator strives to deceive the discriminator, and the discriminator attempts to accurately identify genuine versus forged samples. This adversarial training process allows the Signature GAN to model complex data distributions more effectively than conventional augmentation methods, such as rotation, flipping, and scaling. The visual process of GANs is shown in Fig. 3.

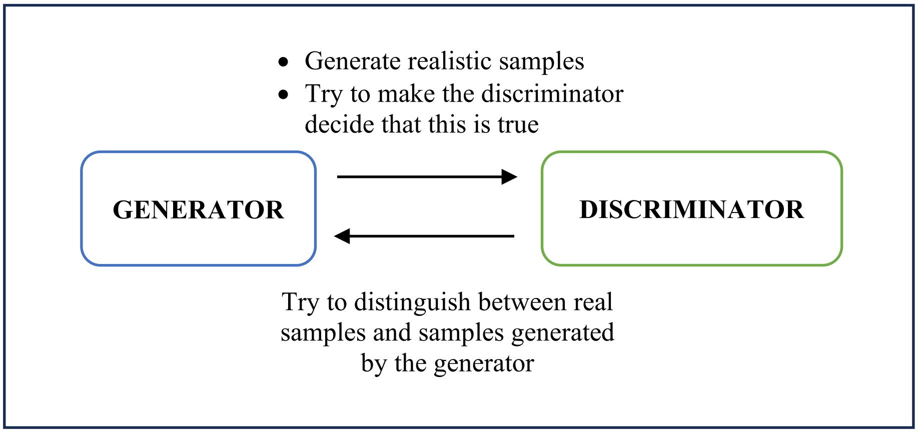
Generate realistic samples
Try to make the discriminator decide that this is true
GENERATOR
DISCRIMINATOR
Try to distinguish between real samples and samples generated by the generator

Fig. 3. GANs process between a generator and a discriminator

From a computational standpoint, while GANs require extended training durations and meticulous tuning to ensure stability, they significantly reduce the necessity for repetitive, hand-crafted augmentation techniques. This optimizes the data enrichment procedure and reduces design intricacy. Researchers like Dash et al. (2023) have emphasized the significance of employing GANs in tasks where data authenticity and variety are critical, such as signature forgery detection. The use of Signature GANs, in particular enable the generation of scalable and balanced datasets that encompass diverse forgery scenarios, eliminating the necessity for manual intervention or additional data collection. This technique efficiently mitigates data scarcity while incorporating controlled variability during training, essential for creating robust and generalizable forgery detection model.

Stage 2: Noise resilient pre-processing. As the real-world variability involve noises such as occlusions, blurred scans, and variable lighting, this stage ensures the dataset quality through preprocessing steps such as binarization (converting grayscale or color signature images into binary black-and-white format to enhance contrast), background removal (eliminating unnecessary background textures or lines that may interfere with signature contours), and stroke normalization (adjusting stroke width, continuity, and alignment to standardize writing patterns across samples). It is further strengthened with adversarial training to increase robustness and resilience to unexpected input degradation.

Stage 3: Model training. The core architecture is a Hybrid Siamese-Transformer Network, which merges the strengths of Siamese structures and Transformer-based encoders. The Siamese network allows for effective use of triplet loss to measure similarity between anchor-positive-negative pairs, while the Transformer encoder captures long-range dependencies in the stroke sequence. In addition, an auxiliary binary classification submodule is introduced at the end of this hybrid architecture, so that it can strongly improve the detection between genuine and forged signatures, particularly in skilled forgery detection. Interpretable visual outputs such as saliency maps and heatmaps are added to contribute to the system's transparency and trustworthiness.

Stage 4: Benchmark evaluation. This final stage aims to assess the model's generalizability and robustness across diverse real-world conditions. This stage consists of two key components: dataset diversity performance scores and cross-domain metrics. The first component evaluates how well the model

https://doi.org/10.24191/jcrinn.v10i2.532

performs across multiple signature datasets with varying languages, writing styles, cultural characteristics, and acquisition conditions. Metrics such as accuracy, precision, and recall are used to determine the model's adaptability and resistance to overfitting. The second component, cross-domain metrics, focuses on how well the model transfers to unseen domains such as new user populations or degraded documents. Together, these metrics ensure the model is not only benchmark-strong but also practically deployable, robust, and fair in real-world applications.

5. CONCLUSIONS

This review paper provided an extensive examination of AI-driven offline handwriting signature forgery detection, focusing specifically on deep learning and novel generative adversarial networks methodologies. A rigorous analysis of existing models, datasets, and performance indicators revealed that Siamese-based architecture prevails in the field due to their capacity to learn relative similarity between pairs of signatures. Although great accuracy has been attained on controlled datasets like CEDAR (Srihari & Leedham, 2001), deployment issues remain in real-world settings because of differences in dataset complexity, noise, and extraneous fluctuations. Meanwhile, hybrid Siamese-Transformer models and Generative-Aware frameworks exhibit considerable potential for addressing existing constraints, particularly by improving contextual learning and resilience while mitigating overfitting concerns.

Furthermore, the role of data augmentation is equally significant, since the integration of Signature GANs produces realistic, high-fidelity synthetic samples that mitigate the ongoing issue of data scarcity. In contrast to conventional augmentation techniques that utilize fixed, manually designed transformations, Signature GANs acquire intricate handwriting distributions directly from authentic examples. This facilitates the creation of sophisticated forgeries that more accurately represent real-world variances. Thus, this augmentation technique not only improves training diversity but also strengthens model robustness under degraded or adversarial conditions.

Future research should prioritize the integration of generative adversarial networks for both data augmentation and forgery detection, development of noise-tolerant models, and embedding explainable AI mechanisms to ensure transparency and trustworthiness in forensic applications. By solving these crucial shortcomings, AI-driven offline handwriting signature forgery detection can move closer to reaching forensic-grade reliability suited for implementation in high-risk legal, financial, and security situations.

6. ACKNOWLEDGEMENTS/FUNDING

This research was not funded by any grant.

7. CONFLICT OF INTEREST STATEMENT

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

8. AUTHORS' CONTRIBUTIONS

Safura Adeela Sukiman: Conceptualisation, methodology, formal analysis, investigation, data curation, writing-original draft, and visualization; **Nor Azura Husin:** Conceptualisation, methodology, validation, resources, writing-review and editing, and supervision; **Hazlina Hamdan:** Conceptualisation, validation, resources, and supervision; **Masrah Azrifah Azmi Murad:** Conceptualisation, validation, resources, and supervision.

https://doi.org/10.24191/jcrinn.v10i2.532

9. REFERENCES

Anitha, A., Priya, M., Kamaraj, B., Yadav, N. K., & Srivastav, S. (2024). Handwritten signature forgery identification and prediction using Harris corner detection and Siamese network. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-5). IEEE. https://doi.org/10.1109/ic-etite58242.2024.10493595

Balaji, Y., Vikas Kedar, R., Virupakshi, H., & Anandan, P. (2024). Crafting trust with convolutional neural networks and hyperparameter tuning for precision signature verification in insurance claim. In *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)* (pp. 1-6). IEEE. https://doi.org/10.1109/adics58448.2024.10533503

Chaturvedi, P., & Jain, A. (2022). Feature ensemble based method for verification of offline signature images. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)* (pp. 710-714). IEEE. https://doi.org/10.1109/com-it-con54601.2022.9850628

Chokshi, A., Jain, V., Bhope, R., & Dhage, S. (2023). SigScatNet: A siamese + scattering based deep learning approach for signature forgery detection and similarity assessment. In *2023 International Conference on Modeling, Simulation &amp;Amp; Intelligent Computing (MoSICom)* (pp. 480-485). IEEE. https://doi.org/10.1109/mosicom59118.2023.10458765

Dash, R., Bag, M., Pattnayak, D., Mohanty, A., & Dash, I. (2023). Automated signature inspection and forgery detection utilizing VGG-16: A deep convolutional neural network. In *2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)* (pp. 01-06). IEEE. https://doi.org/10.1109/icaihc59020.2023.10431430

Emberi, N. B., Mohan, A., Naphade, C. A., & Ransing, R. (2023). Harnessing deep neural networks for accurate offline signature forgery detection. In *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 619-626). https://doi.org/10.1109/iciccs56967.2023.10142593

Engin, D., Kantarci, A., Arslan, S., & Ekenel, H. K. (2020). Offline signature verification on real-world documents. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp 806-808). https://doi.org/10.1109/cvprw50498.2020.00412

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2017). GAN (Generative Adversarial Nets). *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, *29*(5), 177. https://doi.org/10.3156/jsoft.29.5_177_2

Hadsell, R., Chopra, S., & LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*, 1735–1742. https://doi.org/10.1109/cvpr.2006.100

Hafemann, L. G., Sabourin, R., & Oliveira, L. S. (2017). Offline handwritten signature verification - Literature review. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-8). https://doi.org/10.1109/ipta.2017.8310112

Jain, S., Khanna, M., & Singh, A. (2021). Comparison among different CNN architectures for signature forgery detection using Siamese neural network. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (pp. 481–486). https://doi.org/10.1109/icccis51004.2021.9397114

Joe Harris, S. K., & Anitha, J. (2023). An improved signature forgery detection using modified CNN in siamese network. In *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)* (pp. 1–5).

https://doi.org/10.24191/jcrinn.v10i2.532

https://doi.org/10.1109/iceeict56924.2023.10156951

Kathuria, I. (2022). *Handwritten signature datasets* [Data set]. Kaggle. https://www.kaggle.com/datasets/ishanikathuria/handwritten-signature-datasets.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional Neural Networks. *Artificial Intelligence Review*, *53*(8), 5455–5516. https://doi.org/10.1007/s10462-020-09825-6

Krishna, C. A., & Bhuvaneswari, R. (2023). Offline signature forgery detection using Multi-Layer Perceptron. In *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-4). https://doi.org/10.1109/asiancon58793.2023.10269874

Liwicki, M., Blumenstein, M., van den Heuvel, E., Berger, C. E. H., Stoel, R. D., Found, B., Chen, X., & Malik, M. I. (2011). *Hugging Face* [Data set]. ICDAR 2011 Signature Verification Competition (SigComp2011). https://huggingface.co/datasets/1aurent/ICDAR-2011

Majumder, P., Joaa, A. M., Rahman Rhythm, E., Kabir Mehedi, M. H., & Alim Rasel, A. (2023). Siamese-transformer network for offline handwritten signature verification using few-shot. In *2023 26th International Conference on Computer and Information Technology (ICCIT)* (pp. 1–6). IEEE. https://doi.org/10.1109/iccit60459.2023.10441035

Minh Tram, P. H., & Chau, D. N. (2024). Offline signature forgery detection using image processing. In *2024 13th International Conference on Control, Automation and Information Sciences (ICCAIS)* (pp. 1-6). IEEE. https://doi.org/10.1109/iccais63750.2024.10814324

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372.

Rai, D. (2020). *Handwritten signatures dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/divyanshrai/handwritten-signatures

Reddy, M. S., Lakshmi, A. A., Reddy, G. S., Madhavi, B. K., Panigrahi, B. S., & Mohan, V. (2024). Signature forgery detection using siamese-convolutional neural network. In *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)* (pp. 1–5). IEEE. https://doi.org/10.1109/ic-cgu58078.2024.10530761

Reni, R. (2019). *Signature verification dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/robinreni/signature-verification-dataset.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 815-823). IEEE. https://doi.org/10.1109/cvpr.2015.7298682

Shirisha, N., Pranitha, V. P., Balaram, A., Kalpana Chowdary, M., & Shekar, K. (2024). A learning-based intelligent method for signature forgery detection using pre-trained deep models. In *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)* (pp. 1337-1345). IEEE. https://doi.org/10.1109/icoici62503.2024.10696616

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*

Soelistio, E. A., Hananto Kusumo, R. E., Martan, Z. V., & Irwansyah, E. (2021). A review of signature recognition using machine learning. In *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)* (pp. 219-223). https://doi.org/10.1109/iccsai53272.2021.9609732

Srihari, S. N., & Leedham, G. (2001). *CEDAR signature verification dataset* [Data set]. Center of Excellence for Document Analysis and Recognition (CEDAR), University at Buffalo. http://www.cedar.buffalo.edu/NIJ/

Swamy, B. N., Kumar, D. L., Jyothi, K. L., Harika, M. V., Kancharla, G., & Karthik, B. R. (2024). Offline signature verification with AUTOENCODERCNN hybrid feature extraction for improved fake signature detection. In *2024 International Conference on Social and Sustainable Innovations in Technology and Engineering (SASI-ITE)* (pp. 418–423). IEEE. https://doi.org/10.1109/sasi-ite58663.2024.00085

Tarek, O., & Atia, A. (2022). Forensic handwritten signature identification using Deep Learning. In *2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)* (pp. 185–190). IEEE. https://doi.org/10.1109/setit54465.2022.9875697

Tehsin, S., Hassan, A., Riaz, F., Nasir, I. M., Fitriyani, N. L., & Syafrudin, M. (2024). Enhancing signature verification using triplet Siamese similarity networks in digital documents. *Mathematics*, *12*(17), 2757. https://doi.org/10.3390/math12172757

Wang, S., & Jia, S. (2019). Signature handwriting identification based on generative adversarial networks. *Journal of Physics: Conference Series*, *1187*(4), 042047. https://doi.org/10.1088/1742-6596/1187/4/042047

https://doi.org/10.24191/jcrinn.v10i2.532