Available online at www.medicinescience.org

REVIEW ARTICLE

Medicine Science International Medical Journal

Medicine Science 2026;15(3):1451-64

Digitization in forensic document examination: A systematic review of artificial intelligence-based handwriting and signature analysis

Ekin Bahar Aytuglu, Dilek Salkim Islek

İstanbul University-Cerrahpaşa, Institute of Forensic Sciences and Legal Medicine, İstanbul, Türkiye

Received 17 April 2026; Accepted 12 June 2026
Available online 31 August 2026 with doi: 10.5455/medscience.2026.04.077

Abstract

Forensic document examination depends heavily on expert opinion, often leading to subjective reports. This review systematically evaluated image processing and artificial intelligence (AI)-based for offline signature and author verification. The evaluation encompasses model architectures, explainability components, and compliance with forensic reporting standards. Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines, a systematic review was conducted across PubMed, Web of Science, IEEE Xplore, and Google Scholar for records published until February 1, 2026. Inclusion criteria covered offline signature verification, document preprocessing, and explainability approaches. Online/dynamic signature studies or those with insufficient methodology were excluded. From 200 initial records, strict eligibility assessments yielded 22 studies. Due to methodological diversity, findings were evaluated using a narrative synthesis. Post-2020 studies predominantly rely on AI and transfer learning. Common limitations include non-writer-independent designs, lack of subject-disjoint training/test sets, data leakage risks, and limited external validation. Performance metrics varied significantly, and reporting based on interpretability and likelihood ratio (LR) frameworks remains notably scarce. The synthesis also showed that similarity scores require calibration before they can be interpreted as evidential strength, and that explainability outputs are valuable only when integrated with expert-based forensic reasoning. While AI-based methods show promise for handwriting analysis, significant gaps remain in external validation and reporting standardization. Establishing quantifiable, standardized approaches is urgently required to support expert opinions and ensure the robust forensic validity of evidence in legal proceedings. Future forensic applications should therefore prioritize subject-disjoint validation, calibrated score interpretation, uncertainty reporting, and transparent integration of AI outputs into expert opinion.

Keywords: Forensic document examination, offline signature verification, handwriting analysis, image processing, artificial intelligence

Introduction

Handwriting and signature verification serve as critical defense mechanisms against fraud in diverse domains, encompassing financial transactions, legal agreements, testamentary documents, and digitized physical records [1,2]. In the traditional nine-point method [3], used in practice today, morphological indicators such as stroke shape, speed and rhythm, letter and connection structures, writing habits and pressure marks are qualitatively compared by an expert. Since this comparison primarily evaluates documents obtained via photocopies, scans, or camera images [4–6], physical factors such as paper type, resolution, lighting and device variations can create discrepancies even among samples produced by the same writer. This limitation has made it necessary to support findings with numerical criteria rather than basing evaluations solely on qualitative observation [7,8].

Not all references cited in this section are primary studies; some are included to support the conceptual framework and the context of application.

CITATION
Aytuglu EB, Salkim Islek D. Digitization in forensic document examination: A systematic review of artificial intelligence-based handwriting and signature analysis. Med Science. 2026;15(3):1451-64.

Corresponding Author: Ekin Bahar Aytuglu, İstanbul University-Cerrahpaşa, Institute of Forensic Sciences and Legal Medicine, İstanbul, Türkiye
Email: ekinbahargezer@gmail.com

A review of the literature reveals that post-2020 artificial intelligence (AI)-based approaches have gained significant prominence in signature verification and handwriting comparison research [1,2,9,10]. In particular, transfer learning strategies using deep neural networks [11], network architectures focused on directly learning the similarity between two samples [12], and modules that reduce alignment errors across different writing scales [11] are among the other approaches that enhance verification performance. It has been emphasized in the literature that such approaches should address key forensic requirements, including the recognition that accuracy alone is insufficient for forensic applications, the clear specification of evaluation criteria [13], the standardized exchange of evidence [14,15], and the secure presentation of outputs [16], a contribution highly emphasized in recent literature [1,7,8,11,17–20]. Therefore, the ultimate objective in forensic document examination is not merely to improve accuracy but also to ensure that findings can be reported and traced in court with the support of additional approaches [20–22].

Materials and Methods: Literature Review Strategy and Scope

Review Design

This review was executed systematically in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 reporting guidelines [4]. Due to methodological diversity across the studies, a quantitative meta-analysis was not planned, resulting in the findings being reported as a narrative synthesis. While publications from January 1, 2020 to February 1, 2026, were the primary focus, some pre-2020 studies used solely for methodological background were not included in the qualitative synthesis but were used as contextual references in the introduction.

Research Question and Scope

The aim of this review was to systematically examine image processing and AI-based approaches used for offline signature and author verification in forensic document examinations in the post-2020 literature. The evaluation was based on data sets, model architectures, performance metrics, explainability components, and suitability for forensic reporting. Additionally, the study aimed to identify the strengths and weaknesses of the methods in the literature, as well as the limitations that affect the reliability and reportability of decisions in forensic science.

Suitability Criteria

Studies not intended for forensic use were evaluated at a secondary level, with priority given to publications related to forensic reporting, explainability, or evidential value. Within this framework, the inclusion criteria were established as (i) studies on offline signature verification, author identification, or forensic handwriting comparison; (ii) studies on document preprocessing; and (iii) publications that methodologically address explainability in forensic reporting. Exclusion criteria included (i) studies focusing on online signatures or dynamic signature data; (ii) studies involving only document classification that do not include handwriting or signatures; and (iii) studies where the full text is inaccessible or the methodology and results reporting is insufficient.

Only articles published in English or Turkish were included in the review.

Information Sources

The literature review was conducted in the PubMed, Web of Science, IEEE Xplore, and Google Scholar databases. All searches were completed on February 1, 2026. All retrieved records were transferred to Mendeley for reference management.

Search Strategy

The core search terms were formulated using the following phrases: “offline signature verification,” “offline handwritten signature,” “writer identification,” “forensic handwriting,” “document examination,” “artificial intelligence,” “machine learning,” “deep learning,” “transfer learning,” “explainable artificial intelligence,” and “likelihood ratio.” Equivalent terms were used in the Web of Science, IEEE Xplore, and Google Scholar databases. Google Scholar was used as a supplementary source to search for grey literature. Due to the variability in algorithmic ranking of results, only the first 100 records sorted by date were searched and this choice was considered a methodological limitation. All detailed search strings are presented in Supplementary Table S1.

Research Process

Title/abstract and full-text evaluations were conducted independently by two researchers, and any disagreements were resolved through consensus. Sources were evaluated in three groups: (i) studies directly relevant to the qualitative synthesis, (ii) studies excluded during the full-text review due to scope or methodological suitability, and (iii) supplementary sources not included in the synthesis but supporting the methodological and conceptual framework. Ultimately, 22 studies were included in the qualitative synthesis, 21 studies were excluded during the full-text review, and 11 additional sources were retained in the main text for contextual support. Thus, the final reference list consists of a total of 54 references. All these data are summarized in Table 1. The studies excluded during the full-text assessment and the corresponding reasons for exclusion are presented in Table 2.

The details of the study selection process are illustrated in the PRISMA 2020 flowchart presented in Figure 1.

DOI: 10.5455/medscience.2026.04.077

Table 1. Primary and direct methodological studies included in the qualitative synthesis (n = 22)

No.	Reference no	Reference abbreviation	Final category	DOI
1	7	Meena R et al., 2025	Offline signature verification / a comparative pilot study	10.1002/bsl.70000
2	8	Crawford AM et al., 2023	Forensic handwriting analysis	10.1111/1556-4029.15337
3	11	Xiao W and Wu H, 2025	Offline signature verification	10.1038/s41598-025-92704-3
4	12	Xiao W and Ding Y, 2022	Offline signature verification	10.3390/sym14061216
5	13	Diaz M et al., 2024	Explainable offline signature verification	10.1007/s00521-023-09192-7
6	20	Veneri F and Ommen DM, 2023	LR / score-based forensic reporting method	10.1002/sam.11637
7	21	Hicklin RA et al., 2022	Forensic handwriting comparison	10.1073/pnas.2119944119
8	22	Johnson MQ and Ommen DM, 2022	Forensic handwriting + lr	10.1002/sam.11566
9	23	Banerjee D et al., 2021	Offline signature verification	10.1016/j.eswa.2021.115756
10	24	Özyurt F et al., 2023	Offline signature verification	10.18280/ts.400623
11	25	Suteddy W et al., 2023	Author identification	10.30630/joiv.7.1.1293
12	26	Semma A et al., 2021	Author identification	10.1016/j.eswa.2021.115473
13	27	Javidi M and Jampour M, 2020	Author identification	10.1016/j.engappai.2020.103912
14	28	Muhtar Y et al., 2023	Offline signature verification	10.3390/app13148022
15	29	Kim J et al., 2024	Handwritten document comparison	10.1002/sam.11660
16	30	Manna S et al., 2022	Author-independent offline signature verification	No DOI
17	31	Vergeer P et al., 2021	LR (forensic reporting)	10.1016/j.forsciint.2021.110722
18	32	Hannig J and Iyer H, 2022	LR test	10.1111/rssa.12747
19	33	Zhao H and Li H, 2023	Handwriting recognition/verification	10.1038/s41598-023-48789-9
20	34	Sharma N et al., 2022	Offline signature verification	10.1117/1.JEI.31.4.041210
21	35	Goher AS and Vaccarone P, 2024	Forensic signature feature	10.26735/WXIF8532
22	36	Kao HH and Wen CY, 2020	Explainable offline signature verification	10.3390/app10113716

This table includes only primary studies included in the qualitative synthesis as a result of the systematic review, as well as methodological studies that directly contribute to forensic reporting within the scope of the inclusion criteria; Contextual or general-purpose supporting sources have not been included in this table

Table 2. Studies excluded during the full-text review phase (n = 21)

No	Reference no	Reference abbreviation	DOI	Reason for exclusion
1	1	Priya GS et al., 2025	10.4103/jfsm.jfsm_18_25	Not a primary study / review [Insufficient reporting]
2	2	Jones K and Bird CL, 2025	10.1002/wfs2.1540	Online / digital / print (off-focus) [Online/dynamic signature]
3	6	Geistová Čakovská B et al., 2021	10.1111/1556-4029.14627	Methodological guide for digital capture; lacks offline signature evaluation [Online/dynamic signature]
4	17	Lambert B et al., 2024	10.1016/j.artmed.2024.102830	Non-clinical / medical imaging-focused review [Out of scope]
5	18	Kurz A et al., 2022	10.2196/36427	Non-clinical / medical imaging-focused review [Out of scope]
6	19	Hasanzadeh F et al., 2025	10.1038/s41746-025-01503-7	Non-medical / healthcare AI [Out of scope]
7	37	Heckeroth J et al., 2021	10.1016/j.forsciint.2020.110587	Digital signature-focused [Online/dynamic signature]
8	38	Kalantzis N and Platt, 2022	10.1111/1556-4029.14927	Digital signature-focused [Online/dynamic signature]
9	39	Jasuja OP and Sharma P, 2024	10.1504/IJESDF.2024.10055680	Not a primary study / review [Insufficient reporting]
10	40	Sharma P et al., 2021	10.19195/2084-5065.59.11	Focused on electronic signatures [Online/dynamic signature]
11	41	Zimmer J et al., 2021	10.1016/j.forsciint.2021.110945	Digital signature-focused [Online/dynamic signature]
12	42	Bataineh B et al., 2025	10.3390/jimaging11050133	General document binarization compilation [Out of scope]
13	43	Lins RD et al., 2021	10.1007/978-3-030-86337-1_47	General document preprocessing, no forensic signature [Out of scope]
14	44	Garrido-Munoz C et al., 2021	10.3233/FAIA210289	General document preprocessing, no forensic signature [Out of scope]
15	45	Diaz M et al., 2019	10.1145/3274658	Pre-2020 and review/perspective [Insufficient reporting]
16	46	He K et al., 2016	10.1109/CVPR.2016.90	General-purpose baseline model, no direct forensic work [Out of scope]
17	47	Hafemann LG et al., 2018	10.1007/s10032-018-0301-6	Prior to 2020, only methodological background [Insufficient reporting]
18	48	Hafemann LG et al., 2020	10.1109/TIFS.2019.2949425	Outside scope / background source [Out of scope]
19	49	Rodriguez AM et al., 2022	10.1016/j.forsciint.2022.111239	Biometric but not handwriting/signature [Online/dynamic signature]
20	50	Merlino ML et al., 2020	10.69525/jasqde.266	Focuses on human factors; lacks core algorithmic processing [Insufficient reporting]
21	51	Swofford H et al., 2024	10.1016/j.fsisyn.2024.100472	Focuses on human factors; lacks core algorithmic processing [Insufficient reporting]

Exclusion criteria have been standardized into three main groups: (i) online/digital signature or out-of-scope data type, (ii) lack of direct relevance to forensic handwriting-signature comparison, (iii) lack of primary study status or provision of only methodological/contextual background

DOI: 10.5455/medscience.2026.04.077

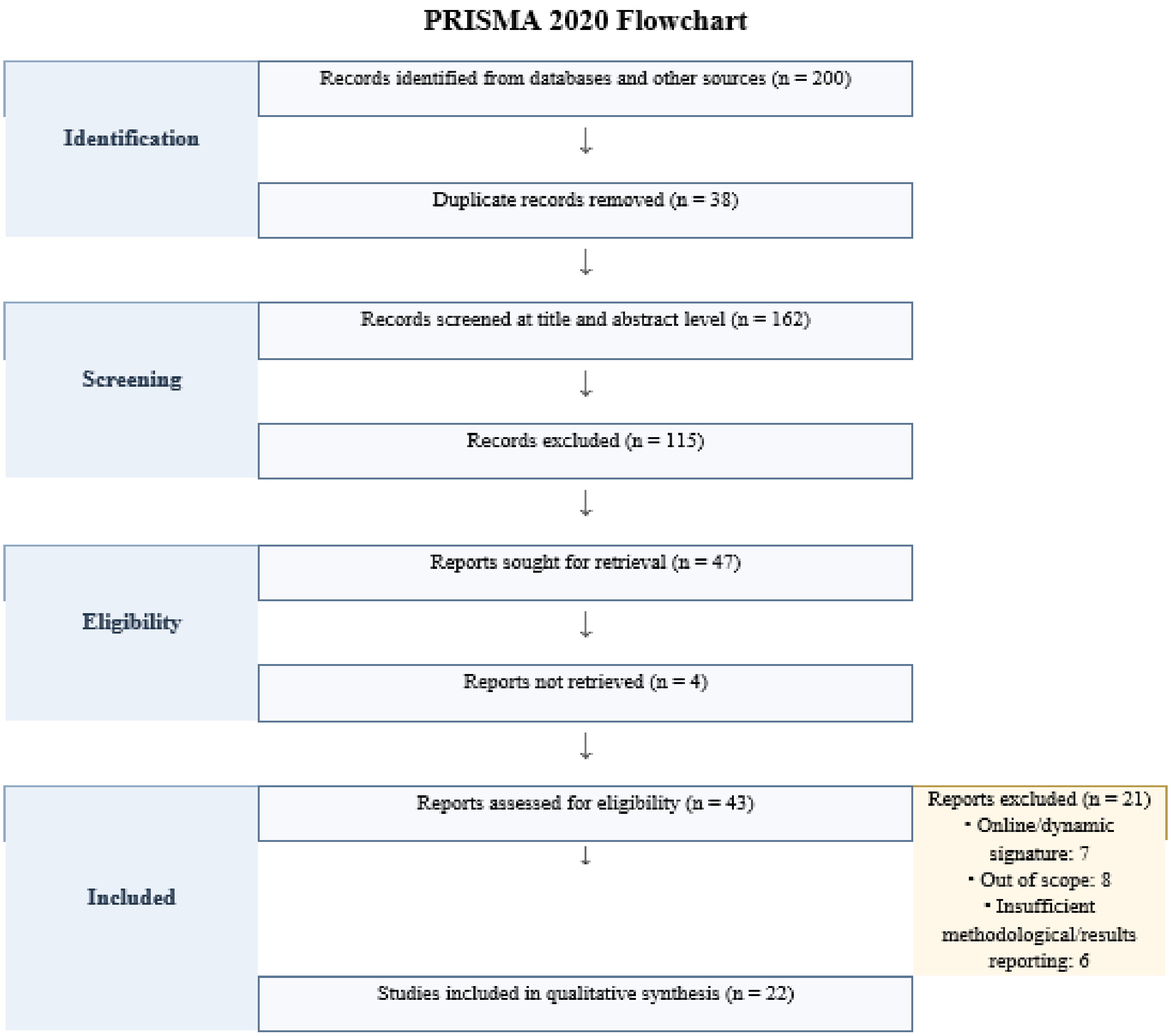

Figure 1. Systematic review of relevant studies

Data Extraction

A data extraction form was used for the data from the included studies. This form includes for each study: author and publication year, study type (offline signature verification/author identification or forensic handwriting comparison), dataset used, sample size, type of forgery (if applicable), image preprocessing methods, model or algorithm type, data splitting approach (writer-dependent/writer-independent), performance metrics, use of external validation, components of explainability, and data related to outputs suitable for forensic reporting.

The key characteristics of the 22 studies included in the qualitative synthesis and the main findings are presented in Table 1. The reported performance measures, evaluation strategies, and forensic suitability of the included studies are compared in Table 3.

Methodological Quality and Bias

The methodological strength of the included studies and potential sources of bias were evaluated based on the clear separation of training and test data, the risk of writer-level data leakage, the adequate definition of the dataset used, the specification of forgery types, the use of external validation or independent test data, the adequate, and balanced reporting of performance metrics, the level of interpretability of the results and their suitability for generating forensic interpretations or evidentiary value. This evaluation was conducted using a framework structured by the authors based on criteria for data separation, potential data leakage, external validation, and reporting adequacy, tailored to the review's objectives, without the use of a standard external quality scale. A study-level summary of methodological quality and potential sources of bias is provided in Table 4.

DOI: 10.5455/medscience.2026.04.077

Table 3. Comparative summary of key performance metrics and forensic suitability across included studies

Reference no	Study	Main task	Main method / model type	Key performance metrics reported	Main performance finding	Evaluation strategy / validation issue	Forensic suitability
7	Meena et al., 2025	Real/forged signature classification	SVM and Random Forest Classifier	Accuracy	RFC: 92%; SVM: 89.64%	Dataset-specific classification; external validation not clearly established	Useful as a comparative pilot study, but forensic admissibility requires stronger validation and reporting standards
8	Crawford et al., 2023	Forensic handwriting analysis	Statistical / score-based approach	LR-related performance and calibration indicators	Provides a quantitative framework to support examiner interpretation	More directly aligned with forensic propositions than simple classification studies	Methodologically relevant for forensic reporting due to evidential interpretation and examiner-support orientation
11	Xiao and Wu, 2025	Offline signature verification	Siamese / spatial transformer-based deep model	Verification accuracy	Reported improved verification performance across offline signature datasets	Potential risk if writer-independent or subject-disjoint splitting is not fully clear	Technically promising, but forensic use depends on transparent split design and external validation
12	Xiao and Ding, 2022	Offline signature verification	Two-stage Siamese network with Focal Loss	Accuracy / verification performance	Reported better performance than comparative models on multiple datasets	Possible limitation if same-writer data separation is not clearly subject-disjoint	Strong technical contribution, but forensic suitability remains limited without leakage-resistant validation
13	Diaz et al., 2024	Explainable offline signature verification	Explainable automatic signature verifier	Verification performance and explainability outputs	Combines verification with visual/model explanation	External validation and forensic reporting remain limited	Forensically relevant because it explicitly supports examiner reasoning
20	Veneri and Ommen, 2023	LR / score-based forensic reporting	Ensemble learning for score likelihood ratios	LR calibration-related measures	Focuses on score-to-LR interpretation rather than simple accuracy	Designed around evidential propositions and calibration	Methodologically relevant for forensic reporting for translating scores into evidential strength
21	Hicklin et al., 2022	Forensic handwriting comparison	Human examiner accuracy / reliability study	Accuracy, reliability, error-related measures	Provides empirical evidence on examiner performance	Directly relevant to forensic reliability and reporting	Methodologically relevant for forensic reporting as a benchmark for expert-based comparison
22	Johnson and Ommen, 2022	Handwriting identification with LR	Random forests and score-based LR	LR-related performance / calibration	Supports handwriting identification through score-based likelihood ratios	More suitable for forensic interpretation than binary classification alone	Methodologically relevant for forensic reporting due to LR-based evidential reporting
23	Banerjee et al., 2021	Offline signature verification	Wrapper feature selection / machine learning	Accuracy or classification metrics	Reported performance based on selected feature sets	External validation not clearly specified	Technically useful, but forensic reliability is limited if evaluated only within dataset-specific conditions
24	Özyurt et al., 2023	Offline signature verification	MobileNetV2 feature extraction and feature selection	Accuracy	Reported 91.3% without feature selection and 97.7% with NCA-based feature selection	Strong performance, but forensic reliability depends on dataset split and external validation	Promising technical performance; forensic suitability requires subject-disjoint validation and calibrated reporting
25	Suteddy et al., 2023	Writer identification	Comparative deep learning architectures	Accuracy / comparative classification metrics	Compares end-to-end deep architectures for writer identification	More relevant when tested under writer-independent conditions	Moderately to highly relevant if person-level separation is maintained

AI: artificial intelligence, CBAM: convolutional block attention module, EER: equal error rate, LR: likelihood ratio, NCA: neighborhood component analysis, RFC: random forest classifier, and SVM: support vector machine; Only the performance metrics explicitly reported or clearly identifiable from the included studies were summarized; Because the reviewed studies differ in datasets, sample structures, preprocessing steps, model architectures, and validation protocols, the reported values should not be interpreted as directly comparable measures of forensic reliability; When external validation, writer-independent testing, subject-disjoint splitting, or forensic reporting procedures were not clearly described, the forensic suitability of the study was interpreted cautiously; This table is intended to support methodological comparison rather than to rank the studies solely by numerical performance

DOI: 10.5455/medscience.2026.04.077

Table 3. Comparative summary of key performance metrics and forensic suitability across included studies

Reference no	Study	Main task	Main method / model type	Key performance metrics reported	Main performance finding	Evaluation strategy / validation issue	Forensic suitability
26	Semma et al., 2021	Writer identification	Deep learning with FAST keypoints and Harris corner detector	Identification accuracy	Reports deep feature-based writer identification performance	External validation partially available	Useful for authorship-related analysis, but forensic reporting framework remains limited
27	Javidi and Jampour, 2020	Text-independent writer identification	Deep learning framework	Identification accuracy	Demonstrates performance for text-independent writer identification	Requires careful separation between training and test writers	Relevant to forensic handwriting comparison if evaluated on unseen writers
28	Muhtar et al., 2023	Multilingual signature verification	Improved ResNet with CBAM	Accuracy	Reported high accuracy across multilingual and public signature datasets	External validation not clearly established as forensic casework validation	Technically strong, but forensic use requires uncertainty and calibration reporting
29	Kim et al., 2024	Handwritten document comparison	Deep latent feature vectors	Score-based comparison / LR-oriented interpretation	Uses latent representations for document comparison	More closely aligned with same-source/different-source forensic propositions	Methodologically relevant for forensic reporting due to representation-based comparison and evidential framing
30	Manna et al., 2022	Writer-independent offline signature verification	Self-supervised representation learning	EER / verification performance	Specifically addresses writer-independent verification	Stronger forensic relevance if subject-disjoint testing is maintained	Suitable methodological direction for forensic applications
31	Vergeer et al., 2021	LR calibration	Calibration metric comparison	LR calibration metrics	Evaluates calibration quality of LR systems	Not a signature-specific classifier but methodologically important	High relevance for forensic reporting and LR reliability
32	Hannig and Iyer, 2022	LR calibration testing	Statistical calibration discrepancy testing	LR calibration discrepancy measures	Provides tools for assessing whether reported LRs are well calibrated	Methodological rather than dataset-specific validation	Methodologically relevant for forensic reporting for checking reliability of LR-based evidence
33	Zhao and Li, 2023	Handwriting identification / verification	AI-assisted textural features	Accuracy, precision or related classification metrics	Reports improved handwriting identification/ verification using textural features	External verification generally limited	Useful for feature-based handwriting analysis, but forensic reporting requires stronger standardization
34	Sharma et al., 2022	Offline signature verification	Deep neural network / Inception-based models	Training and test accuracy	INCV I: 97% train and 92% test accuracy; INCV II: 98% train and 96% test accuracy	Dataset-specific testing; external validation limited	Technically promising, but high performance should be interpreted cautiously without subject-disjoint forensic validation
35	Goher and Vaccarone, 2024	Forensic signature feature analysis	Line-length feature measurement	Feature-level measurement / comparative analysis	Proposes a measurable signature feature for forensic examination	Feature-focused rather than full automated verification	Forensically relevant as an interpretable measurable feature, but not a complete decision system
36	Kao and Wen, 2020	Offline signature verification and forgery detection	Explainable deep learning approach	Accuracy / explainability-related outputs	Combines forgery detection with explainable model outputs	External validation not clearly specified	Relevant for expert support, but explainability should not be treated as independent proof

AI: artificial intelligence, CBAM: convolutional block attention module, EER: equal error rate, LR: likelihood ratio, NCA: neighborhood component analysis, RFC: random forest classifier, and SVM: support vector machine; Only the performance metrics explicitly reported or clearly identifiable from the included studies were summarized; Because the reviewed studies differ in datasets, sample structures, preprocessing steps, model architectures, and validation protocols, the reported values should not be interpreted as directly comparable measures of forensic reliability; When external validation, writer-independent testing, subject-disjoint splitting, or forensic reporting procedures were not clearly described, the forensic suitability of the study was interpreted cautiously; This table is intended to support methodological comparison rather than to rank the studies solely by numerical performance

DOI: 10.5455/medscience.2026.04.077

Table 4. Summary of methodological quality and potential bias for included studies (n = 22)

No	Reference no	Reference abbreviation	Study type	Data coding / risk of bias	External verification	Reporting	Overall risk
1	7	Meena R et al., 2025	Signature verification/comparison	Partially open	Generally not available	Basic metrics available	Moderate
2	8	Crawford AM et al., 2023	Forensic handwriting comparison	Open	Partially available	Strong	Low
3	11	Xiao W and Wu H, 2025	Wet signature verification	Partially open / may pose a risk of leakage	None / Not specified	Basic metrics available	High
4	12	Xiao W and Ding Y, 2022	Wet signature verification	Partially open / may be at risk of leakage	None / Not specified	Basic metrics available	High
5	13	Diaz M et al., 2024	Explainable signature verification	Partially open	Generally not available	Explainability present, forensic reporting limited	Moderate
6	20	Veneri F and Ommen DM, 2023	LR method	Open	Partially available	Strong	Low
7	21	Hicklin RA et al., 2022	Forensic handwriting comparison	Open	Partially available	Strong	Low
8	22	Johnson MQ and Ommen DM, 2022	Forensic handwriting + LR	Open	Partially available	Strong	Low
9	23	Banerjee D et al., 2021	Handwritten signature verification	Partially open	None / Not specified	Basic metrics available	High
10	24	Özyurt F et al., 2023	Handwritten signature verification	Partially open	None / Not specified	Basic metrics available	High
11	25	Suteddy W et al., 2023	Author identification	Open	Partially available	Sufficient	Low
12	26	Semma A et al., 2021	Author identification	Open	Partially available	Sufficient	Low
13	27	Javidi M and Jampour M, 2020	Author identification	Open	Partially available	Sufficient	Low
14	28	Muhtar Y et al., 2023	Offline signature verification	Partially open	None / Not specified	Basic metrics available	High
15	29	Kim J et al., 2024	Handwritten document comparison	Partially available	Partially available	Strong	Low
16	30	Manna S et al., 2022	Offline signature verification	Partially open	None / Not specified	Basic metrics available	High
17	31	Vergeer P et al., 2021	LR calibration	Open	Partially available	Strong	Low
18	32	Hannig J and Iyer H, 2022	LR calibration test	Open	Partially available	Strong	Low
19	33	Zhao H and Li H, 2023	Handwriting recognition/ verification	Partially open	Generally not available	Sufficient but limited	Moderate
20	34	Sharma N et al., 2022	Offline signature verification	Partially open	None / Not specified	Basic metrics available	High
21	35	Goher AS and Vaccarone P, 2024	Forensic signature feature	Partially open	Generally not available	Feature-focused, limited	Moderate
22	36	Kao HH and Wen CY, 2020	Explainable signature verification	Partially open	None / Not specified	Explainability present	Moderate

The assessment of methodological quality and potential bias was conducted only for the included studies listed in Table 1; The classification has been standardized into low, moderate, and high risk levels, taking into account the transparency of data disaggregation, the likelihood of person-based data leakage, the presence of independent external validation, and the adequacy of performance and reporting.

Protocol

No pre-registered protocol was established for this review. However, eligibility criteria, information sources, search strategy, data extraction fields, and the qualitative synthesis approach were defined prior to the search. The PRISMA 2020 flow diagram is presented in the main text, and the detailed search strategies are provided in the appendix.

Findings Related to Data Acquisition and Standardization

In this section, the findings of the qualitative synthesis are evaluated alongside sources that support the contextual framework of the application. Therefore, the references used do not correspond exactly to the studies listed in Table 1. While sample collection via mobile applications offers a quick and easily accessible option in the field, factors related to camera specifications (such as issues with automatic exposure and focus, uneven lighting, and perspective distortion) can cause shifts in image alignment [2,4-6,10,37-41]. In this context, the current best practice guidelines of the European Network of Forensic Science Institutes (ENFSI) [52], state that the sample collection process must be standardized [40]. For this standardization, they emphasize the need to use the same writing instrument and paper, ensure an adequate number of samples and maintain a traceable/reportable chain of custody under controlled lighting conditions [6,40]. To this end, mobile applications are recommended to provide a reference scale and color scale, automatic cropping and perspective correction, a blur measurement with a retake alert and metadata recording (e.g., device model and shooting distance) in the form of user alerts. On the other hand, dynamic (tablet/pen) scenarios may enable these features but this review is based on offline handwriting and signature image data in alignment with the relevant research stream [2,4,5].

Image Preprocessing: Binarization, Noise Reduction and Segmentation

In this section, the findings of the qualitative synthesis are discussed alongside sources that support the methodological background. Preprocessing of text and signature images is the first step that affects the success of the entire system [23,24]. Studies published after 2020 have highlighted the importance of AI-driven binarization approaches over classical image processing [42,43]. The purpose of binarization is to segment the background and preserve only the relevant text or signature area while also reducing shading and color inconsistencies caused by data acquisition [42,44]. In the literature, these requirements are used to achieve the highest accuracy across different document types and varying capture conditions [44]. While the necessary level of segmentation varies depending on the application, signature verification can work with entire signature samples, whereas text analysis can operate at the line/word/character level [23]. AI-based hybrid approaches aim to automatically detect text and signature regions irrespective of the actual text content by combining the learning of relevant representations with AI neural networks [25-27,53].

Detection and Representation Learning Approaches Reported in the Literature: Segmentation Using You Only Look Once (YOLO) and Analysis Using Residual Network 18 (ResNet18)

The technical framework presented in this section includes not only studies directly incorporated into the qualitative synthesis but also contextual sources referenced to demonstrate methodological continuity.

When the digitization of morphological features based on the nine-point rule [3], image preprocessing, automatic region detection, and representation-based analysis are combined, a multi-stage hybrid decision structure can be established. This approach can provide an effective framework for forensic evaluation [2,11,28].

In recent years, offline signature verification studies have relied on first automatically detecting and cropping the signature region, followed by analyzing the representations obtained from this region. For this reason, single-stage detectors have been used in some studies [12,45,53]. The preprocessing steps applied after detection in turn provide cleaner data for the representation learning phase [23,42].

When applied to clean, verified signature images during the representation learning phase, convolutional neural network architectures such as ResNet18 [25–27], offer fast and highly accurate processing even on small datasets, facilitated by their transfer learning capabilities [28]. Indeed, the advantages of ResNet18 in signature verification such as reduced parameter and inference costs, have also been reported [28]. During the representation learning phase, ResNet18 measures similarity between signatures by comparing them using cosine similarity or Euclidean distance [29]. Furthermore, ResNet-based architectures have been preferred in various studies involving small to medium-sized datasets [30,46]. However, these findings should not be interpreted as indicating the existence of a single superior architecture.

Scoring and Reporting

In forensic applications, presenting the evidence's scores expressed quantitatively is more appropriate in terms of reliability than simply labeling a document as genuine or forgery [8,20,22]. For this purpose, the Likelihood Ratio (LR) framework provides direct input for decision evaluation by comparing the likelihood of the data under two hypotheses (same-source versus different-source) [8,20,31,32]. LR calculation primarily relies on interpreting cosine similarity or Euclidean distance scores derived from representation vectors obtained via deep neural networks [20,42–44]. Confidence interval and distribution-based analyses are recommended for reporting the validity and reliability of the interpreted scores [31,32].

The similarity score, in this setting, represents the technical output of the computational system rather than a final forensic conclusion. A high or low value cannot, on its own, establish

 DOI: 10.5455/medscience.2026.04.077

whether two samples were produced by the same writer. For forensic interpretation, the score must be evaluated in relation to competing propositions, such as whether the questioned and reference samples originate from the same source or from different sources. The LR framework provides a way of expressing how strongly the observed data support one proposition over the alternative. The score therefore functions as an intermediate quantitative measure. It becomes forensically meaningful only when it is calibrated, accompanied by uncertainty assessment, and interpreted by the examiner in relation to the case circumstances [8,20,31,32].

From a forensic admissibility perspective, numerical outputs should be transparent, reproducible, and traceable. A reportable AI-supported result should include not only the final similarity score but also the preprocessing steps, model or algorithm used, comparison strategy, calibration approach, uncertainty range, and the limitations of the available material. In legal contexts, a numerical score without methodological explanation may be difficult to evaluate. Therefore, reporting should avoid categorical language based solely on the system output and should instead present the computational result as one component of the expert's overall assessment.

This distinction also requires a clear reporting pathway between the computational score and the final expert opinion.

Score-to-Evidence-to-Expert Opinion Pipeline: For forensic reporting, this process can be described in three linked stages. First, the system produces a similarity score based on the comparison of feature representations extracted from the questioned and reference samples. Second, this score is interpreted as evidential information after calibration against relevant same-source and different-source distributions. At this stage, the score should be accompanied by uncertainty estimates, threshold information, and a clear explanation of the reference population or dataset used for calibration. Third, the calibrated evidential information is incorporated into the expert opinion. The expert does not merely repeat the numerical output of the system; rather, the expert evaluates whether the computational finding is consistent with observable handwriting or signature characteristics, image quality, sample limitations, and case-specific conditions. In this structure, the AI-based output supports the reasoning process but does not replace the forensic document examiner [8,20,31,32].

Explainability and Expert Interaction

Within forensic science, studies [11] on explainable handwriting and signature verification have been developed. There is a study that particularly emphasizes the goal of supporting the expert [54]. In expert decision-support systems, explainability [11] is necessary not only to build trust but also for error analysis and reportability. In forensic signature verification, the ability to explain how a decision was reached based on specific physical characteristics helps the expert substantiate their findings [11]. A hybrid architecture based on YOLO and ResNet18 can provide a strong starting point for explainability. However, explainability alone is not sufficient. A recent study proposing a framework applicable at the forensic level also emphasizes the need to define practical components for the secure and auditable development of such systems.

Common explainability approaches in handwriting and signature analysis include saliency maps, heatmap-based visualization, Gradient-weighted Class Activation Mapping (Grad-CAM), attention maps, and feature importance analysis [13,36]. In image-based systems, these outputs may indicate which parts of a signature or handwriting sample contributed more strongly to the similarity or dissimilarity assessment [13,36]. For example, highlighted regions may correspond to stroke endings, curvature, connection points, letter formations, alignment, or local variations in line quality [13,35,36,54].

In forensic reporting, such outputs can support expert reasoning by making the computational process more transparent. They may help the examiner compare whether the model's focus is compatible with document examination criteria such as form, proportion, rhythm, connection, spacing, and overall writing habit. However, explainability outputs should not be treated as independent proof of authorship or forgery. A heatmap or attention region does not by itself establish evidential value. It is useful only when interpreted together with the original image, the quality of the questioned material, the known samples, the model limitations, and the examiner's professional assessment.

Potential Forensic Workflow Derived from the Literature

This workflow should be considered not as a single mandatory model proposal but as an example of a forensic decision support pipeline derived from prominent detection, preprocessing, and representation learning approaches in the literature:

Input Quality: Given that image quality (e.g., resolution, lighting variations and noise) can affect performance depending on data acquisition conditions during offline signature verification, image preprocessing steps (such as background segmentation, noise reduction) are recommended in the literature [6,23,33,34,42,47,48].

Detection and Cropping: Automatic detection of the signature region using YOLO-based detectors provides a more standardized input for the representation learning phase, and working exclusively with the relevant region is considered an advantageous strategy [12,22].

Representation and Comparison: ResNet-based representation learning approaches are frequently preferred for small to medium-sized datasets, and similarity scores can be efficiently calculated using cosine or Euclidean-based metrics [25-29].

Reporting: Calibrating the calculated scores using methods such as LR and translating them into an interpretable framework can contribute to a more standardized reporting of evidential strength in a forensic context [8,20,22,29,31,32,35,41].

Threshold Calibration and Uncertainty Estimation: Thresholds used to classify or interpret similarity scores should not be treated as fixed universal values. They must be calibrated

DOI: 10.5455/medscience.2026.04.077

according to the dataset, comparison task, image quality, and the intended forensic use. In addition, uncertainty should be reported whenever possible, particularly when the score lies close to the decision threshold or when image quality, sample number, or writing variability limits the reliability of the comparison. This stage prevents the numerical output from being presented as a categorical decision without sufficient forensic context.

Expert-Based Interpretation and Reporting Format: The final stage of the workflow should be an expert-based reporting process. The report should include the nature and quality of the questioned and reference samples, preprocessing steps, the comparison method, the similarity score, threshold or calibration information, uncertainty considerations, and the examiner's interpretation. The computational result should be presented as decision-support evidence rather than an autonomous conclusion. In this sense, the proposed workflow is intended to strengthen the consistency and transparency of forensic document examination, not to replace the examiner.

Discussion

The strongest finding of this review is that post-2020 studies on AI-based handwriting and signature analysis have moved beyond classical image processing and increasingly rely on deep learning, transfer learning, and representation-based comparison strategies. However, the forensic value of these methods remains uneven. High experimental performance does not automatically translate into forensic reliability, particularly when evaluation strategies, data splitting procedures, external validation, uncertainty estimation, and reporting standards are insufficiently described.

The reviewed literature indicates that AI-based models may improve the consistency of handwriting and signature comparison by extracting measurable features from image data. Nevertheless, current studies differ substantially in terms of datasets, preprocessing procedures, model architectures, reported metrics, and validation strategies. This methodological heterogeneity limits direct comparison between studies and prevents a single model architecture or performance value from being accepted as generally superior. Therefore, the contribution of AI in forensic document examination should be evaluated not only through accuracy but also through generalizability, transparency, reproducibility, and reportability.

Common limitations of post-2020 studies include [1,2,6,19,31,32,36,49-51] data scarcity, barriers to inter-institutional sharing, difficulties in generalization due to physical conditions, the wide variety of forgery types, the forensic validity of scores and challenges in reporting.

However, to ensure that these methods can be reliably applied in forensic science, it is critical to develop hybrid approaches [7,19,51], enhanced testing protocols, and improved reporting features. The examination of similarity scores and, where possible, reporting the strength of evidence as a LR is a fundamental requirement for forensic validity. At this point, the utilization of an established LR-based approach is crucial for future studies [8,20,29,31,32].

Accordingly, AI-based systems should be positioned as decision-support tools rather than autonomous decision-makers. Their value in forensic document examination depends on whether the numerical and visual outputs can be interpreted together with the examiner's assessment of handwriting characteristics, sample quality, and case-specific limitations [11].

The recent studies and current limitations addressed in this review suggest that the hybrid system may provide a viable framework for use in the field of forensic science. It is anticipated that future studies [5,6], by incorporating solutions that account for physical variability in mobile devices, could enhance the reliability of the system in real-world field scenarios [6].

These considerations are particularly important for forensic admissibility. A system that produces a high accuracy value in experimental conditions may still be insufficient for forensic use if it does not provide traceable outputs, calibrated thresholds, uncertainty information, and a reporting structure that can be examined by another expert. For this reason, forensic suitability should be evaluated not only by performance metrics but also by the extent to which the method can be explained, audited, and integrated into expert-based reporting.

Evaluation Strategy and Forensic Reliability: A central methodological distinction among the reviewed studies concerns the evaluation strategy. In writer-dependent designs, samples from the same writer may be represented during both model development and testing. Such designs may be useful for controlled biometric verification scenarios, but they may overestimate performance in forensic contexts. In contrast, writer-independent or subject-disjoint evaluation requires that the individuals used for testing are not represented in the training phase. This design more closely reflects real forensic casework, where the system may be applied to questioned material from individuals not previously encountered by the model. For this reason, subject-disjoint evaluation should be considered a minimum requirement for studies that claim forensic applicability.

Potential data leakage may occur when samples from the same writer are randomly divided between training and test sets. In such a case, the model may learn writer-specific visual patterns during training and then encounter visually related samples from the same person during testing. This does not necessarily measure the ability of the system to generalize to unseen writers. Instead, it may produce inflated performance values and give an overly optimistic impression of forensic reliability. In the reviewed literature, studies with unclear or non-subject-disjoint splitting procedures should therefore be interpreted with caution, even when their reported accuracy appears high.

For forensic document examination, the critical question is not only whether the model performs well on a dataset, but whether it remains reliable when applied to a new case, a new writer, a different acquisition condition, or a previously unseen forgery pattern.

In this review, studies in which writer-independent or fully subject-disjoint splitting was not clearly specified, or in which external validation was limited, were interpreted cautiously because such designs may increase the risk of inflated performance estimates in forensic applications [11,12,23,24,28,30,34].

Strengths of the Review: A major strength of this review is that it evaluates AI-based handwriting and signature analysis from a forensic perspective rather than limiting the synthesis to technical performance. The review considers not only model architectures and reported metrics but also data splitting strategies, potential data leakage, explainability, LR-based interpretation, and reporting suitability. This approach is important because forensic applicability depends on whether a method can support expert reasoning under real casework conditions, not merely on whether it performs well in a controlled dataset.

Limitations: This review also has several limitations. First, because the included studies used different datasets, model architectures, evaluation strategies, and performance metrics, a quantitative meta-analysis was not appropriate. Second, some studies did not clearly report whether data splitting was writer-dependent, writer-independent, or fully subject-disjoint. This limited the ability to compare forensic reliability across studies. Third, the review focused on offline handwriting and signature image data; therefore, findings related to dynamic or online signature acquisition were not evaluated in detail. Finally, although Google Scholar was used as a supplementary source, only the first 100 records sorted by date were screened, which may have limited the retrieval of some relevant grey literature.

Forensic Implications: The practical implication of this review is that AI-based handwriting and signature analysis should be positioned as a structured decision-support tool in forensic document examination. Such systems may help standardize comparison procedures, quantify similarity, identify relevant image regions, and support more transparent reporting. However, their outputs should not be presented as standalone determinations of authorship or forgery. Forensic use requires subject-disjoint validation, calibrated scoring, uncertainty reporting, explainable outputs, and integration with expert-based morphological assessment. When these conditions are met, AI-based methods may strengthen the consistency, auditability, and evidential communication of forensic document examination.

Conclusion

The results are primarily based on the studies included in the qualitative synthesis. Additional references were used solely to support the conceptual and practical context.

This review has provided a conceptual framework derived from the literature by comprehensively evaluating studies that combine the ease of mobile acquisition with AI-based analysis in forensic handwriting and signature analysis post-2020. In the reviewed literature, it was observed that the limitations in real-world usage scenarios cannot be resolved through a single process [1,2,6,19,51]. It is anticipated that a quantitative sequential pipeline encompassing data collection, preprocessing, feature extraction, scoring, and reporting could provide a more structured response to existing challenges. This approach may not be confined merely to boosting accuracy rates; it critically reinforces the consistency, auditability and reportability strictly required for forensic applications.

A harmonized reporting approach such as ENFSI [52] will enhance the forensic validity, reportability and practical field application of the hybrid system, while also contributing to the comparability of results obtained across different institutions and conditions.

Overall, the reviewed evidence suggests that AI-based handwriting and signature analysis has meaningful potential for forensic document examination, but this potential depends on the way the systems are evaluated and reported. Methods that lack subject-disjoint validation, calibrated scoring, uncertainty estimation, and explainable reporting should be interpreted cautiously, even when they report high performance. Future studies should therefore move from accuracy-centered evaluation toward forensic reliability-centered validation.

Conflict of Interest
The authors confirm the absence of any conflicts associated with this work.

Funding
The authors declare that they have received no financial support for the study.

References

1. Priya GS, Ahemad U, Mundotiya N. A review of multilingual handwriting analysis techniques in forensic science. J Forensic Sci Med. 2025;11:170-7.
2. Jones K, Bird CL. Non-original and digitally captured handwriting: considerations for forensic handwriting examinations. WIREs Forensic Sci. 2025;7:e1540.
3. Aşıcıoğlu F. El yazısı incelemesinde temel tanı unsurları (Dokuzlar Kaidesi). In: Adli Belge İnceleme ve Sahtecilik Alanındaki Bilirkişi Raporlarının İncelikleri. Ankara: Seçkin Publishing. 2019;25-45.
4. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
5. Kim G, Hong T, Yim M, et al. OCR-free document understanding transformer. 17th European Conference on Computer Vision, 23-27 October 2022. Tel Aviv, Israel, 498-517.
6. Geistová Čakovská B, Kalantzis N, Dziedzic T, et al. Recommendations for capturing signatures digitally to optimize their suitability for forensic handwriting examination. J Forensic Sci. 2021;66:743-7.
7. Meena R, Siwan D, Krishan P, et al. Manual and machine learning approaches for classifying real and forged signatures—a comparative study and forensic implications. Behav Sci Law. 2025;43:489-500.
8. Crawford AM, Ommen DM, Carriquiry AL. A statistical approach to aid examiners in the forensic analysis of handwriting. J Forensic Sci. 2023;68:1768-79.
9. Bhavani SD, Bharathi RK. A multi-dimensional review on handwritten signature verification: strengths and gaps. Multimed Tools Appl. 2024;83:2853-94.

10. Wang Z, Muhammat M, Yadikar N, et al. Advances in offline handwritten signature recognition research: a review. IEEE Access. 2023;11:120222-36.

11. Xiao W, Wu H. Learning features for offline handwritten signature verification using spatial transformer network. Sci Rep. 2025;15:9453.

12. Xiao W, Ding Y. A two-stage Siamese network model for offline handwritten signature verification. Symmetry (Basel). 2022;14:1216.

13. Diaz M, Ferrer MA, Vessio G. Explainable offline automatic signature verifier to support forensic handwriting examiners. Neural Comput Appl. 2024;36:2411-27.

14. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: optimal speed and accuracy of object detection. arXiv. 2020;arXiv:2004.10934.

15. Ge Z, Liu S, Wang F, et al. YOLOX: exceeding YOLO series in 2021. arXiv. 2021;arXiv:2107.08430.

16. Cubuk ED, Zoph B, Shlens J, Le QV. RandAugment: practical automated data augmentation with a reduced search space. 34th Conference on Neural Information Processing Systems, 6-12 December 2020. Virtual Conference, 18613-24.

17. Lambert B, Forbes F, Doyle S, et al. Trustworthy clinical AI solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis. Artif Intell Med. 2024;150:102830.

18. Kurz A, Hauser K, Mehrtens HA, et al. Uncertainty estimation in medical image classification: systematic review. JMIR Med Inform. 2022;10:e36427.

19. Hasanzadeh F, Josephson CB, Waters G, et al. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. NPJ Digit Med. 2025;8:154.

20. Veneri F, Ommen DM. Ensemble learning for score likelihood ratios under the common source problem. Stat Anal Data Min. 2023;16:528-46.

21. Hicklin RA, Eisenhart L, Richetelli N, et al. Accuracy and reliability of forensic handwriting comparisons. Proc Natl Acad Sci U S A. 2022;119:e2119944119.

22. Johnson MQ, Ommen DM. Handwriting identification using random forests and score-based likelihood ratios. Stat Anal Data Min. 2022;15:357-75.

23. Banerjee D, Chatterjee B, Bhowal P, et al. A new wrapper feature selection method for language-invariant offline signature verification. Expert Syst Appl. 2021;186:115756.

24. Özyurt F, Majidpour J, Rashid TA, Koç C. Offline handwriting signature verification: a transfer learning and feature selection approach. Traitement du Signal. 2023;40:2613-22.

25. Suteddy W, Agustini DAR, Adiwilaga A, Atmanto DA. End-to-end evaluation of deep learning architectures for off-line handwriting writer identification: a comparative study. JOIV: Int J Inform Vis. 2023;7:178-85.

26. Semma A, Hannad Y, Siddiqi I, et al. Writer identification using deep learning with FAST keypoints and Harris corner detector. Expert Syst Appl. 2021;184:115473.

27. Javidi M, Jampour M. A deep learning framework for text-independent writer identification. Eng Appl Artif Intell. 2020;95:103912.

28. Muhtar Y, Muhammat M, Yadikar N, et al. FC-ResNet: a multilingual handwritten signature verification model using an improved ResNet with CBAM. Appl Sci (Basel). 2023;13:8022.

29. Kim J, Park S, Carriquiry A. A deep learning approach for the comparison of handwritten documents using latent feature vectors. Stat Anal Data Min. 2024;17:e11660.

30. Manna S, Chattopadhyay S, Bhattacharya S, Pal U. SWIS: self-supervised representation learning for writer independent offline signature verification. 2022 IEEE International Conference on Image Processing, 16-19 October 2022. Bordeaux, France, 1411-5.

31. Vergeer P, van Schaik Y, Sjerps M. Measuring calibration of likelihood-ratio systems: a comparison of four metrics, including a new metric devPAV. Forensic Sci Int. 2021;321:110722.

32. Hannig J, Iyer H. Testing for calibration discrepancy of reported likelihood ratios in forensic science. J R Stat Soc Ser A Stat Soc. 2022;185:267-301.

33. Zhao H, Li H. Handwriting identification and verification using artificial intelligence-assisted textural features. Sci Rep. 2023;13:21739.

34. Sharma N, Gupta S, Mehta P, et al. Offline signature verification using deep neural network with application to computer vision. J Electron Imaging. 2022;31:041210.

35. Goher AS, Vaccarone P. Measuring line length: a new signature feature in forensic examination. Arab J Forensic Sci Forensic Med. 2024;6:10-8.

36. Kao HH, Wen CY. An offline signature verification and forgery detection method based on a single known sample and an explainable deep learning approach. Applied Sciences. 2020;10:3716.

37. Heckeroth J, Kupferschmid E, Dziedzic T, et al. Features of digitally captured signatures vs. pen and paper signatures: similar or completely different?. Forensic Sci Int. 2021;318:110587.

38. Kalantzis N, Platt AWG. Digitally captured signatures: A method for the normalization of force through calibration and the use of the zeta function. J Forensic Sci. 2022;67:651-68.

39. Sharma P, Jasuja OP. Forensic examination of digitally captured handwriting—a review of contemporary tools and techniques. Int J Electron Secur Digit Forensics. 2024;16:357-71.

40. Sharma P, Singh M, Jasuja OP. Forensic examination of electronic signatures: a comparative study. Nowa Kodyfikacja Prawa Karnego. 2021;59:149-84.

41. Zimmer J, Kalantzis N, Dziedzic T, et al. The challenge of comparing digitally captured signatures registered with different software and hardware. Forensic Sci Int. 2021;327:110945.

42. Bataineh B, Tounsi M, Zamzami N, et al. A comprehensive review on document image binarization. J Imaging. 2025;11:133.

43. Lins RD, Bernardino RB, Smith EB, Kavallieratou E. ICDAR 2021 competition on time-quality document image binarization. 16th International Conference on Document Analysis and Recognition, 5-10 September 2021. Lausanne, Switzerland, 708-22.

44. Garrido-Munoz C, Sánchez-Hernández A, Castellanos FJ, Calvo-Zaragoza J. Domain adaptation for document image binarization via domain classification. In: Frontiers in Artificial Intelligence and Applications. Vol 341. IOS Press BV; 2021:569-82.

45. Diaz M, Ferrer MA, Impedovo D, et al. A perspective analysis of handwritten signature technology. ACM Comput Surv. 2019;51:117.

46. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016:770-8.

47. Hafemann LG, Oliveira LS, Sabourin R. Fixed-sized representation learning from offline handwritten signatures of different sizes. Int J Doc Anal Recognit. 2018;21:219-32.

48. Hafemann LG, Sabourin R, Oliveira LS. Meta-learning for fast classifier adaptation to new users of signature verification systems. IEEE Trans Inf Forensics Secur. 2020;15:1735-45.

49. Rodriguez AM, Geradts Z, Worring M. Calibration of score based likelihood ratio estimation in automated forensic facial image comparison. Forensic Sci Int. 2022;334:111239.

50. Merlino ML, Dahir VB, Edwards CP, et al. Cognitive human factors and forensic document examiner methods and procedures: writing characteristics, visual context, and handwriting examination decision accuracy. J Am Soc Quest Doc Exam. 2020;23:3-20.

51. Swofford H, Lund S, Iyer H, et al. Inconclusive decisions and error rates in forensic science. Forensic Sci Int Synerg. 2024;8:100472.

52. European Network of Forensic Science Institutes. Best Practice Manual for the Forensic Handwriting Examination. 4th ed. Wiesbaden, Germany: ENFSI; 2022.

53. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-22 June 2023. Vancouver, Canada, 7464-7475.

54. Crown N, Marquis R, Kupferschmid E, et al. Error mitigation in forensic handwriting examination: the examiner's perspective. Forensic Sci Res. 2024;9:owae065.