Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00264-y

Review Article

Intelligent Bloodstain Pattern Analysis Using Convolutional Neural Networks

Sonia Rajkumari [1], Mridu Sharma [1], Sameer Saharan [1], Tilak Chandrakar [2]

1. Forensic Science, Mody University of Science and Technology, Sikar, IND
2. Forensic Science, National Forensic Science University, Delhi, IND

Received: May 28, 2026 | Review began: June 07, 2026 | Review ended: August 29, 2026 | Published: September 05, 2026

Abstract

Bloodstain pattern analysis is a fundamental part of forensic crime scene reconstruction. However, the samples collected from real crime scenes are often obtained under inconsistent conditions, on different surfaces, and at various impact angles. This makes it difficult to classify them accurately. While research in this area has progressed from rule-based and manual interpretation methods to deep learning-based and convolutional neural network (CNN)-based models, there remains a significant lack of consistency in how their performance is evaluated. This paper presents a systematic review of 76 research papers, organising the field of bloodstain pattern analysis using CNNs into four methodological paradigms, and proposes a structured evaluation framework to address this evaluative gap. It is evident from this review that model-level accuracy metrics dominate the literature, while real-world forensic validation remains rare, and controlled laboratory datasets are overrepresented, limiting their applicability in authentic casework settings. Moreover, this review highlights the need for standardised datasets reflecting real-world forensic variability and emphasises the integration of explainable artificial intelligence to improve interpretability and support legal reliability, thereby strengthening the practical applicability of CNN-based bloodstain pattern analysis. This paper argues that, unlike previous research, the priority for future work in this field must shift from achieving high classification accuracy under ideal conditions to establishing forensically grounded, statistically transparent, and forensically meaningful evaluation standards.

Categories: Ethical AI and Responsible Technology, AI applications, Forensic Analysis

Keywords: convolutional neural networks, image classification, forensic image processing, deep learning, prisma, bloodstain pattern analysis

Introduction And Background

Bloodstain pattern analysis (BPA) is a forensic science discipline that examines and interprets bloodstains deposited at crime scenes involving violent physical contact, including assaults, stabbings, and homicides [1-3]. Through systematic BPA, forensic scientists attempt to reconstruct the sequence of events during a criminal incident by analysing how blood behaves under specific conditions of applied force, impact angle, and surface interaction [4-6]. The discipline has evolved considerably since its modern foundations were established by MacDonell [7] in the 1970s, and today underpins crime scene reconstruction across numerous jurisdictions worldwide.

Despite its established forensic role, conventional BPA relies heavily on the subjective judgment of trained examiners, making outcomes susceptible to interpretative inconsistency. The National Research Council [8] highlighted these limitations explicitly, observing that the uncertainties associated with BPA are enormous and that analyst opinions are frequently more subjective than scientific. This has been empirically confirmed by Hicklin et al. [5], who demonstrated that 11.2% of analyst responses were erroneous in the largest black-box study of BPA to date. Similarly, Taylor et al. [9] and Yuen et al. [10] reported systematic inter-analyst disagreement in pattern classification and directionality determination tasks, respectively.

Many analytical models fail to fully account for the variability introduced by surface type, porosity, and texture [8,11,12]. Studies have been inconsistent in their methodological focus, evaluation criteria, and experimental designs, with most failing to reflect the real-world complexity of crime scene conditions, such as partial impressions, overlapping patterns, and smearing [12-15]. Although there has been growing interest in computational and convolutional neural network (CNN)-based approaches to address these limitations, the existing body of research lacks consistency and integration, particularly regarding the impact of surface variation on model performance [9,10,16].

The emergence of deep learning has created substantial new opportunities for automated forensic image analysis [16,17]. CNNs, in particular, have demonstrated strong classification performance across a range of pattern recognition tasks relevant to forensic science [18,19]. However, as LeCun et al. [16] and Rawat and Wang [17] observe, the translation of laboratory-level deep learning performance into forensically reliable, operationally validated tools requires considerably more than architectural sophistication; it demands rigorous, domain-specific evaluation standards.

This review addresses these gaps by systematically examining 76 research publications spanning the application of CNNs to BPA, with particular emphasis on transfer patterns across variable surface types and conditions. The principal argument advanced is that future research must prioritise the development of evaluation standards that are forensically grounded, statistically transparent, and repeatable in real-world casework settings, rather than focusing exclusively on classification accuracy under idealised experimental conditions.

Review

Review methodology

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [20], which provide a standardised framework for transparent reporting of systematic review methodology. The review protocol, including the search strategy and eligibility criteria, was developed prior to the commencement of the search, though it was not formally registered in a public registry.

Search Strategy

A comprehensive literature search was performed across six electronic databases: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, PubMed, and Google Scholar. The search was restricted to articles published between January 2000 and February 2026. The extended temporal range was selected to capture both foundational analytical BPA literature and the most recent deep learning contributions.

Search strings were constructed using Boolean operators (AND/OR) across three conceptual domains: (i) the forensic domain: bloodstain pattern analysis, blood spatter analysis, BPA, forensic bloodstain, transfer bloodstain; (ii) the methodological domain: convolutional neural network, CNN, deep learning, machine learning, image classification, ResNet, VGG, Inception, MobileNet, transfer learning; and (iii) the contextual domain: forensic image analysis, surface variation, impact angle, crime scene reconstruction, forensic AI. Searches were conducted in English only; no language restrictions were applied to abstract screening.

Inclusion and Exclusion Criteria

Studies were included if they met all of the following criteria:

• Focused on the classification, detection, segmentation, or analysis of bloodstain patterns using computational, machine learning, or deep learning approaches;

• Published in a peer-reviewed journal, conference proceeding, or dataset paper with sufficient methodological detail to enable meaningful synthesis;

• Reported at least one quantitative or qualitative performance measure relevant to BPA classification, feature extraction, or forensic validation;

• Covered bloodstain pattern types consistent with established forensic taxonomies (impact, cast-off, transfer, passive, gunshot).

Studies were excluded if they:

• Applied general image classification methods to non-forensic datasets without any BPA-specific application or validation;

• Were published in non-peer-reviewed sources such as preprints without subsequent publication (exception: preprints cited for dataset transparency);

• Focused exclusively on blood age estimation, DNA analysis, or chemical detection without pattern classification components;

• Lacked sufficient methodological detail to support independent replication or synthesis.

Study Selection Process (PRISMA)

Table *1* presents the PRISMA 2020-compliant [20] screening and selection process. Following the initial database search, duplicate records were identified and removed using reference management software. Title and abstract screening were performed independently by two reviewers; disagreements were resolved through consensus. Inter-reviewer agreement was not formally quantified using a statistical measure.

Full-text screening was then applied to all records passing the initial screen using the predefined inclusion and exclusion criteria. The final corpus consisted of 76 studies.

Phase	Step	Records (n)
Identification	Records identified through database searches (Scopus, Web of Science, IEEE, ACM, PubMed, Google Scholar)	218
Identification	Duplicate records removed	38
Screening	Records after duplicate removal	180
Screening	Records excluded at the title/abstract level	62
Eligibility	Full-text articles assessed for eligibility	118
Eligibility	Full-text articles excluded (reasons: non-BPA, no peer review, insufficient detail)	42
Included	Studies included in the final synthesis	76

TABLE 1: PRISMA-Based Screening and Selection of Studies

BPA, Bloodstain Pattern Analysis

Data Extraction and Quality Assessment

Data extraction was performed using a standardised form capturing study design, dataset characteristics (size, surface types, number of classes), CNN architecture(s) employed, evaluation metrics reported, key performance findings, and stated limitations. Study quality was assessed using an adapted version of the QUADAS-2 framework [21] tailored for computational forensic studies, evaluating risk of bias across dataset representativeness, performance reporting completeness, and forensic validation adequacy.

Core taxonomy of BPA classification paradigms

The 76 reviewed studies can be organised into four methodological paradigms (Table *2*). This taxonomy provides a structured framework for understanding the evolutionary trajectory of BPA research and the relative strengths and limitations of each approach in forensic contexts.

Paradigm	Description	Distribution	Example References
Traditional Analytical	Rule-based, geometric constraints, manual interpretation	~35%	[1-3,16,24,25]
Optimisation-Based	Feature engineering, constrained classification, statistical methods	~20%	[26-28]
Deep Learning/CNN	Automated feature learning, ResNet/VGG/Inception architectures	~37%	[13,15,19,21,29-31]
Hybrid Approaches	CNN + domain constraints, hyperspectral + DL, XAI-integrated	~8%	[14,44,67,76]

TABLE 2: Distribution of Reviewed Studies Across Methodological Paradigms

CNN, Convolutional Neural Network; DL, Deep Learning; ResNet, Residual Network; VGG, Visual Geometry Group; XAI, Explainable Artificial Intelligence

Traditional Analytical Methods

Traditional analytical methods rely on explicit physical and geometric constraints, stain shape, ellipticity, directionality, and impact angle to classify bloodstain patterns through rule-based interpretation [1-3]. These approaches, grounded in the foundational fluid dynamics work of Attinger et al. [22] and the classical BPA frameworks established by James et al. [1] and Bevel and Gardner [2], are interpretable and consistent under controlled conditions. However, they are fundamentally limited in their ability to handle real-world variability, including surface texture effects, overlapping patterns, and partial transfer impressions [15,23,24].

A critical limitation of traditional approaches is their dependence on human visual perception, which Dror and Cole [4] have shown to be vulnerable to contextual bias, a finding further substantiated by Dror et al. [25] and Hicklin et al. [5]. The reliability studies of Taylor et al. [9] and Yuen et al. [10] demonstrate that even trained analysts produce inconsistent classifications of the same patterns, particularly for transfer and wipe patterns where morphological cues are ambiguous.

Arthur et al. [26] proposed an element-based approach to objective pattern classification that partially addresses this subjectivity by decomposing patterns into quantifiable geometric elements, representing an important transitional step between fully manual and fully automated classification. Peschel et al. [15] provide a detailed taxonomy of traditional BPA classification frameworks that exposes their inherent limitations when confronted with degraded or partial evidence.

Optimisation-Based Methods

Optimisation-based methods employ constrained search and feature engineering techniques, guiding classification through objective functions that impose constraints on stain circularity, spatial distribution, directionality, and morphological consistency [27,28]. Vitiello et al. [28] framed BPA explicitly as an optimisation problem, demonstrating that formal constraint satisfaction approaches could outperform purely analytical methods under moderate degradation conditions. Shoumy et al. [29] extended this to neural network-based feature extraction pipelines, while Arthur et al. [27] developed a systematic image-processing methodology for quantitative feature extraction that provided the groundwork for subsequent automated classification.

The effectiveness of optimisation-based approaches remains heavily reliant on the quality of manually constructed features, the representativeness of training data, and the associated computational complexity, all of which restrict generalisability across diverse surface types and real-world crime scene conditions [30,31]. The likelihood ratio approach proposed by Zou and Stern [31] represents an important conceptual advance, framing pattern classification within a probabilistic forensic evidence evaluation framework that is more compatible with courtroom standards than binary classification outputs.

Deep Learning and CNN-Based Methods

Deep learning methods, and convolutional neural networks in particular, involve the automated learning of classification features directly from raw image data, using architectures including standard CNNs, ResNet [32], VGG [33], Inception [34], and MobileNet [35]. These approaches have demonstrated strong classification performance under complex degradation conditions and are increasingly adopted due to their capacity to learn surface-invariant features without manual feature engineering [16,17].

Bergman et al. [19] achieved 99.73% classification accuracy using Inception V3 on a laboratory-controlled dataset of 965 drip stains and 1,595 blood spatters, the highest accuracy reported in the reviewed literature. However, the study explicitly acknowledged a critical forensic limitation: the model performed substantially worse on wet bloodstains compared to dried samples, highlighting a systematic gap between laboratory performance and operational applicability. Rough et al. [18] demonstrated that ResNet-50 and VGG-16 architectures substantially outperform the 13-23% manual analyst error rates reported by Taylor et al. [9] and Hicklin et al. [5], establishing a clear performance benchmark for future deep learning work. Acampora and Vitiello [36] provided the first multi-class comparison across architectures specifically for BPA, while Jomon et al. [37] demonstrated the superiority of MobileNet transfer learning over custom CNN architectures for small forensic datasets.

However, many reviewed studies rely on synthetic or laboratory-controlled datasets that may not capture the morphological variability of transfer patterns encountered in authentic forensic casework [12,14,38]. Książek et al. [39] applied multiple deep learning architectures, including 1D, 2D, and 3D CNNs, Recurrent Neural Networks, and Multilayer Perceptrons, to hyperspectral bloodstain classification, demonstrating that deep learning's advantages over Support Vector Machine are most pronounced in the inductive, cross-image scenario most relevant to forensic casework. Foundational work by Bergmann et al. [40] towards substrate-independent bloodstain classification using a common feature set established that cross-substrate generalisation requires deliberate feature engineering beyond standard CNN pipelines, a challenge that subsequent deep learning studies have not consistently resolved. The Segment Anything Model-based segmentation approach developed by Dong and Zhang [41] represents an emerging frontier, demonstrating that foundation models fine-tuned on bloodstain datasets can improve segmentation accuracy by 2.2% over pre-trained models while achieving a 4.7% processing speed improvement.

Hybrid Approaches

Hybrid approaches combine CNN-based feature extraction with domain-specific forensic constraints or conventional image processing pipelines. Acampora et al. [42] demonstrated the utility of one-class classification methods for hyperspectral BPA, while Romaszewski et al. [43] contributed a dataset specifically designed for evaluating blood detection in hyperspectral images. Giulietti et al. [44] demonstrated that substrate-induced spectral distortion in hyperspectral imaging can be corrected using neural network methods, improving classification reliability across diverse background materials. Sheth and Shah [45] applied a discrete wavelet transform combined with a dense CNN for hyperspectral bloodstain classification, while Giulietti et al. [46] extended hyperspectral neural networks to substrate-independent bloodstain age estimation. Beyond pattern classification, Alkhuder [47] demonstrated attenuated total reflection-Fourier transform infrared spectroscopy for bloodstain identification, and Doty et al. [48] and Bremmer et al. [49] established machine learning approaches for bloodstain age estimation extending up to two years. Attinger et al. [50] showed that determining the region of blood spatter origin requires incorporating fluid dynamic modelling alongside statistical uncertainty quantification, a standard not met by classification-only CNN approaches. Home et al. [51] further demonstrated that 3D crime scene scanning significantly enhances juror comprehension of BPA evidence, underscoring that forensic utility extends beyond analyst classification to legal communication of results.

Despite their potential, hybrid techniques remain underrepresented in the literature, constituting fewer than 8% of reviewed studies, and lack standardised evaluation protocols. The integration of explainable AI (XAI) methods, such as SHAP (SHapley Additive exPlanations), into hybrid architectures, as demonstrated by Rough et al. [52] for feature importance ranking, represents a particularly promising direction for improving forensic interpretability [53].

CNN-based BPA evaluation framework

A central finding of this review is the absence of a unified, forensically grounded evaluation framework across the reviewed literature. To address this gap, we propose a five-level hierarchical evaluation framework for CNN-based BPA systems, adapted from forensic validation principles [54] and structured to address both computational performance and operational forensic utility (Table *3*).

Level	Dimension	Focus	Metrics/Parameters	Priority
L1	Condition	Dataset design and experimental condition modelling	Surface types, impact angles, drying stages	HIGH – controls generalisability
L2	Image	Visual and structural quality of classified images	Accuracy, Precision, Recall, F1, AUC	MODERATE – baseline performance only
L3	Feature	Forensically relevant structural properties	Ellipticity, directionality, morphology preservation	HIGH – forensic pattern integrity
L4	Utility	Real-world crime scene reconstruction outcomes	Impact angle, AO estimation, pattern type accuracy	CRITICAL – operational applicability
L5	Reproducibility	Statistical transparency and replication standards	Train-test splits, augmentation, hyperparameters	HIGH – scientific validity

TABLE 3: Proposed Five-Level Hierarchical Evaluation Framework for CNN-Based BPA Systems

AO, Area of Origin; AUC, Area Under the Curve; BPA, Bloodstain Pattern Analysis; CNN, Convolutional Neural Network

Level 1: Condition Modelling and Dataset Design

A comprehensive evaluation of CNN-based BPA systems requires careful consideration of dataset construction and experimental condition modelling. Most reviewed studies rely on laboratory-generated datasets under controlled, idealised conditions, which offer reproducibility but fail to reflect the variability of real crime scenes [23,43,52]. A meaningful evaluation framework must: (i) distinguish between synthetic laboratory data, semi-controlled experimental data, and operationally representative data; (ii) document surface material, porosity, texture, impact angle, drying stage, and projection distance for each sample; and (iii) assess model performance across multiple condition scenarios rather than a single standardised setting.

The gunshot backspatter dataset of Attinger et al. [55], producing 68 patterns under systematically varied conditions of distance, ammunition type, and bullet velocity, represents a methodological exemplar for condition-documented BPA dataset construction. Kannan et al. [56] demonstrated the value of multi-surface datasets spanning paper, grass, ceramic tile, and snow, while Laan et al. [24] provided a fluid dynamics-grounded framework for understanding how physical parameters translate into pattern morphology. The theoretical and experimental work of Comiskey et al. [57] on droplet-stain formation across porous and non-porous surfaces further established that substrate porosity fundamentally alters stain morphology in ways that single-surface laboratory datasets cannot capture. Building on this, Attinger et al. [50] demonstrated that determining the region of origin of blood spatter must incorporate fluid dynamic modelling and statistical uncertainty quantification to produce forensically defensible estimates, a standard that purely CNN-based approaches rarely meet.

Level 2: Image-Level Evaluation

Image-level evaluation assesses the visual and structural quality of classified or reconstructed bloodstain images using standard computational metrics. Classification accuracy, precision, recall, F1-score, area under the ROC curve (AUC), and confusion matrices remain valid tools for assessing the baseline performance of CNN models across pattern categories [5,9,58]. Osborne et al. [59] demonstrated that differentiating impacted from spattered blood using classification frameworks requires carefully defined decision boundaries that standard accuracy metrics alone cannot validate. Siu et al. [60] provided quantitative differentiation of gunshot versus blunt force trauma patterns, establishing performance benchmarks on physically distinct pattern types against which purely image-level metrics must be interpreted. Esaias et al. [61] showed that 3D scanning improves area-of-origin estimation accuracy beyond what 2D image classification alone can achieve, while McCleary et al. [62] demonstrated automated cast-off reconstruction using Euclidean geometry and statistical likelihood, both highlighting the limitations of treating classification accuracy as the sole performance criterion. The prediction of gunshot backspatter by Comiskey et al. [63] using fluid dynamic modelling further illustrates that forensically meaningful outputs require physical grounding that image-level CNN metrics do not inherently provide.

The performance hierarchy established by Li et al. [38], where XGBoost achieved 92.89% on the same dataset and random forest achieved lower accuracy [64], illustrates how metric selection and dataset characteristics can substantially influence reported performance. Liu et al. [64] further demonstrated that accuracy decreases systematically with distance (99% at 30 cm, 93% at 60 cm, 86% at 120 cm), highlighting the critical importance of reporting performance across experimental conditions rather than at a single optimised setting.

Level 3: Feature-Level Evaluation

Feature-level evaluation examines whether CNN models preserve and correctly classify the forensically relevant structural properties of bloodstain patterns: stain shape, directionality, satellite formation, ellipticity, and transfer morphology across varying surface types [10,12,27,65]. This level is particularly critical for transfer patterns, where subtle morphological differences arising from surface porosity and contact dynamics are the primary discriminating features and are easily lost under generalised classification approaches.

Arthur et al. [27,65,66] established that structured image-processing methodologies can reliably extract quantitative pattern features that align with forensic analyst decision-making, as confirmed through their eye-tracking study [65]. Rough et al. [52] applied SHAP analysis to identify circularity (60% contribution), mean intensity (28%), and stain area (28%) as the three most discriminative features for drip pattern classification, a finding with direct implications for forensic feature priority in CNN architectures. Extending this line of work, Pachong et al. [67] applied unsupervised machine learning to identify key observable features in drip patterns without requiring labelled training data, demonstrating that automated feature discovery aligns with established forensic classification taxonomies while also revealing previously uncharacterised morphological properties.

Level 4: Forensic Utility Evaluation

Forensic utility evaluation is the most operationally relevant assessment level, examining whether CNN-based classification contributes meaningfully to crime scene reconstruction outcomes [6,8,25,31]. This includes testing whether model outputs allow appropriate determinations of impact angle, directionality, area of origin, and pattern type in real-world scenarios. Without this evaluation layer, claims of practical forensic value remain provisional and cannot be translated into operational use.

The persistent gap between digital tool development and forensic integration identified by Balbudhe et al. [68], despite HemoVision achieving area-of-origin estimation errors as low as 1.3 mm, exemplifies the challenge of translating computational performance into forensically accepted validation. The National Research Council [8] identified this gap as systemic across forensic science disciplines, and Hook et al. [58] confirm that it remains unresolved a decade later, despite substantial methodological advances.

Level 5: Statistical Reporting and Reproducibility

Evaluation frameworks lack completeness without rigorous statistical reporting. Reviewed studies must document data-splitting procedures, train-test isolation, preprocessing steps, data augmentation strategies, and hyperparameter selection in sufficient detail to enable independent replication [5,9,36,42,52,58]. These standards are essential for distinguishing genuine performance improvements from results reflecting experimental convenience or dataset-specific overfitting. Collectively, these five levels define an evaluation logic in which CNN-based BPA systems are held to standards that are forensically grounded, statistically transparent, and reproducible across diverse real-world conditions.

Results, quantitative synthesis, and critical gaps

Overview of Reviewed Studies

The 76 studies included in this systematic review span foundational analytical frameworks, experimental CNN-based classification studies, dataset contributions, and deep learning architecture evaluations from 2000 to February 2026. Tables *4-5* summarise key experimental studies with performance data. Given the methodological heterogeneity of the reviewed literature, a purely quantitative meta-analysis is not appropriate; this review therefore adopts a qualitative thematic synthesis supported by selective quantitative evidence where available.

Study	Method	Dataset	Performance	Surfaces	Key Limitation
Bergman et al. [19]	Inception V3 (CNN)	2,560 drip/spatter images	99.73%	Single (lab)	Wet stains failed
Jung et al. [14]	CNN, DT, RF	14 pattern types	80% (RF)	Single	Small dataset
Li et al. [38]	XGBoost, RF	169 spatter patterns	92.89%	Single (poster board)	Public dataset only
Liu et al. [64]	RF	94 patterns, 3 distances	99%@30 cm	Single	Distance-dependent

TABLE 4: Summary of Key Experimental Studies (Part 1)

CNN, Convolutional Neural Network; DT, Decision Tree; RF, Random Forest

Study	Method	Dataset	Performance	Surfaces	Key Limitation
Rough et al. [18]	ResNet-50, VGG-16	Impact/cast-off/expired	Accuracy: 78.2% to 100% (26 model/dataset combinations; best: 100% ResNet-50, largest crop window)	Single	Lab conditions
Jomon et al. [37]	MobileNet (transfer)	Multi-height, 5 substrates	Superior to a custom CNN	5 types	Passive patterns only
Acampora and Vitiello [36]	Multi-architecture DL comparison	Multi-class BPA dataset	Accuracy: varies by architecture (per-model results reported in source study)	Controlled	No operational validation
Rough et al. [52]	SHAP + ML pipeline	635 drip patterns, 4 surfaces	60% circularity contribution	4 types	Drip patterns only

TABLE 5: Summary of Key Experimental Studies (Part 2)

BPA, Bloodstain Pattern Analysis; CNN, Convolutional Neural Network; DL, Deep Learning; ML, Machine Learning; SHAP, SHapley Additive exPlanations

Methodological Distribution

Across the 76 reviewed studies, approximately 35% employed traditional analytical or manual BPA methods, 20% applied optimisation or classical machine learning techniques, 37% used CNN or deep learning architectures, and 8% adopted hybrid approaches combining deep learning with forensic domain constraints. This distribution confirms that while computational adoption is growing, the field has not yet converged on a dominant evaluation paradigm. The evolutionary trajectory from rule-based analytical methods through optimisation techniques to CNN-based deep learning is consistent with broader trends in forensic image analysis documented by Dargan and Kumar [13] and LeCun et al. [16]. Critically, despite the growing prevalence of CNN-based approaches, the proportion of studies employing multi-surface evaluation remains low (17.5%), and operational forensic validation is even rarer (12.5%). This disconnect between architectural sophistication and forensic applicability represents the central unresolved tension in the field.

Classification Performance Across Experimental Studies

Classification performance metrics reported in the reviewed literature show significant variation depending on architecture, dataset, and experimental setting. The performance range spans from 80% (Random Forest on a diverse multi-class dataset [14]) to 99.73% (Inception V3 on a homogeneous laboratory dataset [19]). This variance is not primarily attributable to architectural differences; it primarily reflects dataset composition and experimental condition homogeneity. Bergman et al. [19] achieved 99.73% accuracy with Inception V3 on a dataset comprising 965 drip stains and 1,595 blood spatters, all evaluated under controlled laboratory settings. This result must be interpreted cautiously: it was obtained on a homogeneous dried-bloodstain dataset where the model was unlikely to encounter the morphological variability that defines real crime scene evidence. Jung et al. [14] found that Random Forest outperformed CNNs on a diverse 14-class dataset when training data were limited, suggesting that deep learning advantages are not universal and depend critically on dataset size and diversity. Li et al. [38] demonstrated the substantial potential of boosting algorithms, with XGBoost achieving 92.89% accuracy on a public 169-pattern dataset. Liu et al. [64] showed that classification accuracy for gunshot versus impact patterns decreases systematically with distance (99% at 30 cm, 93% at 60 cm, 86% at 120 cm), a finding of direct operational relevance that most studies fail to investigate.

Surface Coverage and Dataset Design

Approximately 70% of reviewed studies used exclusively laboratory-controlled datasets on single surface types, predominantly paper or poster board, under idealised conditions. Only 17.5% of studies tested models across more than one surface type, and only 12.5% incorporated data from real or operational crime scene conditions. This systematic underrepresentation of surface diversity is the most consequential methodological gap in the reviewed literature. Kannan et al. [56] used their automated BPA technique across 132 cast-off, 208 impact, and 132 expired patterns on paper targets, identifying circularity, mean intensity, and stain area as the three most discriminative features. Boos et al. [69] systematically characterised how droplet impact velocity and droplet count influence static drip pattern features, providing quantitative evidence that these physical parameters must be controlled and documented in any dataset intended for machine learning classification. Balbudhe and Babar [70] further demonstrated that even within a single material category, differences in paper composition and texture produce measurable variation in stain diameter, edge regularity, and shape, underscoring the need for surface-stratified dataset design. The influence of environmental conditions on stain formation is similarly documented by Benabdelhalim and Brutin [71], who showed that relative humidity significantly alters blood drip stain morphology on fabric substrates. Rough et al. [52] further contributed a dataset of 635 drip patterns across paper, grass, ceramic tile, and snow under both indoor and outdoor conditions. Jomon et al. [37] demonstrated across five substrate types that MobileNet transfer learning significantly outperformed custom architectures, establishing the value of pretrained features for small forensic datasets.

Evaluation of Digital Versus Traditional Methods

Balbudhe et al. [68], in a scoping review of 36 studies on digital versus non-digital BPA tools, found that digital systems including HemoSpat, FARO Scene, and HemoVision consistently outperformed traditional manual methods, with HemoVision 3D achieving area-of-origin estimation errors as low as 1.3 mm. However, the review also identified a persistent gap between digital tool development and operational forensic integration, echoing the National Research Council [8] findings regarding the broader need for validated and standardised forensic methodologies. This gap is particularly significant because operational adoption depends not only on technical performance but also on legal admissibility standards, analyst training requirements, and institutional acceptance.

Critical Gaps

GAP 1 - Homogeneous Laboratory Datasets: Despite growing computational sophistication, approximately 70% of studies generate and test models on idealised single-surface datasets that fail to capture real-world morphological variability [5,6,69]. The operational failure mode, where a model achieves 99%+ accuracy in the laboratory but cannot generalise to porous surfaces, overlapping impressions, or partially degraded evidence, represents the most pressing practical limitation in the field.

GAP 2 - Absence of Multi-Condition Evaluation: Fewer than 13% of reviewed studies systematically varied more than one experimental condition simultaneously. This prevents assessment of interaction effects between variables and significantly limits generalisability [12,37,64].

GAP 3 - Inconsistent Performance Reporting: Heterogeneous evaluation metrics, inconsistent train-test split protocols, and incomplete hyperparameter reporting prevent direct cross-study comparison. Without standardised reporting, the field cannot accumulate knowledge in a cumulative, scientifically rigorous manner [18,29,58].

GAP 4 - Underrepresentation of Real-World Forensic Complexity: Partial transfer impressions, overlapping patterns, degraded stains, and contaminated surfaces are consistently encountered at actual crime scenes but rarely incorporated into experimental designs [8,25,67,69]. This is compounded by the near-total absence of operational validation studies that test CNN-based BPA tools on actual casework evidence.

Future directions

The path forward for CNN-based BPA depends on aligning model development more closely with forensic reliability requirements. Three principal avenues emerge from the reviewed literature. First, the development of realistic and diverse datasets must be prioritised. Future datasets should incorporate transfer patterns across broader ranges of surface materials, impact angles, drying stages, and projection distances, and include operationally representative conditions such as partial impressions, overlapping stains, and background contamination [6,8,37]. The contributions of Rough et al. [52,56] and Attinger et al. [55] provide methodological templates for rigorous dataset construction that future work should build upon.

Second, standardisation of evaluation protocols is essential. Future studies should adopt multi-level evaluation frameworks such as the five-level hierarchy proposed in Section 4, assessing not only image-level accuracy but also feature-level preservation, forensic interpretability, and real-world applicability. Reporting standards should mandate disclosure of train-test splits, augmentation procedures, hyperparameter selection, and cross-validation methodology [5,18,58].

Third, the development of explainable and uncertainty-aware models is critical for legal admissibility. CNN-based classification systems must provide confidence measures alongside classification outputs so that forensic analysts can assess the reliability of automated decisions rather than treating model outputs as absolute determinations [13,42,67]. The integration of XAI frameworks, such as SHAP [53], Gradient-weighted Class Activation Mapping [72], and Local Interpretable Model-agnostic Explanations [73], into BPA classification pipelines represents a particularly promising direction for improving interpretability while maintaining classification performance. Solanke [74] explicitly argues that interpretable models are essential for mitigating distrust in AI-based forensic analysis, particularly in courtroom contexts where black-box outputs are legally and ethically problematic. The emerging SAM-based segmentation framework [41] and 3D CNN hyperspectral approaches [39] suggest that foundation model fine-tuning may offer a practical pathway to operational deployment within existing forensic workflows. Kirillov et al. [75] established the foundational architecture underlying SAM, demonstrating that promptable segmentation generalises across domains with minimal fine-tuning. Acampora et al. [76] further demonstrated that cognitive robotic systems can be integrated with BPA workflows for semi-automated crime scene analysis, representing an early but important step toward fully operationalised AI-assisted forensic investigation.

Building upon the gaps and directions identified in this review, ongoing research is exploring the application of CNN and deep learning architectures to bloodstain pattern classification across multiple real-world surface types, including both porous and non-porous substrates encountered at actual crime scenes. Such work aims to move beyond single-surface, laboratory-controlled evaluations toward a more ecologically representative classification framework and seeks to contribute both a multi-surface forensic dataset and an empirically grounded set of performance benchmarks that could inform future standardisation efforts in CNN-based BPA research.

Conclusions

BPA using CNNs has emerged as a rapidly developing interdisciplinary field at the intersection of forensic science, fluid dynamics, and deep learning. The 76 reviewed studies demonstrate a clear transition from traditional manual interpretation towards computational approaches, with CNN-based architectures achieving strong classification performance under controlled laboratory conditions, ranging from 80% with ensemble methods to 99.73% with Inception V3. However, the assessment paradigm across the field remains inconsistent and insufficiently grounded in forensic utility, with the majority of performance claims resting on idealised experimental conditions that do not reflect the operational realities of crime scene evidence.

The principal conclusion of this review is that the performance of CNN-based BPA systems cannot be evaluated by laboratory accuracy alone. For forensic purposes, models must demonstrate: (i) robustness across varied surface types and real-world conditions; (ii) preservation of forensically relevant pattern features; (iii) outputs that are interpretable and explainable for courtroom presentation; and (iv) reproducibility in authentic casework settings. The most pressing requirement in the field is therefore not more sophisticated architectures, but a standardised, forensically grounded validation framework connecting dataset realism, feature-level integrity, and operational applicability. Without this framework, the gap between laboratory performance and operational utility will continue to limit the adoption of CNN-based tools in real forensic investigations. Future empirical work focusing on CNN-based BPA classification across systematically varied real-world surface conditions represents a promising and necessary step towards closing the divide between laboratory performance and operational forensic applicability. Automated classification approaches that demonstrate successful differentiation of impact spatter and cast-off patterns provide a foundation upon which multi-surface, operationally validated systems can be built, a direction that this review identifies as among the most critical unmet needs in the field.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Sonia Rajkumari, Mridu Sharma

Acquisition, analysis, or interpretation of data: Sonia Rajkumari, Mridu Sharma, Sameer Saharan, Tilak Chandrakar

Drafting of the manuscript: Sonia Rajkumari, Mridu Sharma, Sameer Saharan, Tilak Chandrakar

Critical review of the manuscript for important intellectual content: Sonia Rajkumari, Mridu Sharma

Supervision: Sonia Rajkumari, Mridu Sharma, Sameer Saharan, Tilak Chandrakar

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following:

- **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work.
- **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work.
- **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. James SH, Kish PE, Sutton TP: Principles of Bloodstain Pattern Analysis: Theory and Practice. CRC Press, Boca Raton; 2005. 10.1201/9781420039467
2. Bevel T, Gardner RM: Bloodstain Pattern Analysis with an Introduction to Crime Scene Reconstruction. CRC Press, Boca Raton; 2008. 10.1201/9781420052725
3. De Forest PR, Gaensslen RE, Lee HC: Forensic Science: An Introduction to Criminalistics. McGraw-Hill, New York; 1983.

4. Dror IE, Cole SA: The vision in "blind" justice: Expert perception, judgment, and visual cognition in forensic pattern recognition. Psychonomic Bulletin & Review. 2010, 17:161-167. 10.3758/pbr.17.2.161

5. Hicklin RA, Winer KR, Kish PE, et al.: Accuracy and reproducibility of conclusions by forensic bloodstain pattern analysts. Forensic Science International. 2021, 325:110856. 10.1016/j.forsciint.2021.110856

6. Home PH, Norman DG, Palmer A, Field P, Williams MA: Quantifying forensic investigations involving bloodstain pattern analysis within the UK. Forensic Science International. 2022, 339:111424. 10.1016/j.forsciint.2022.111424

7. MacDonell HL: Flight Characteristics and Stain Patterns of Human Blood. U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice, Washington, DC; 1971.

8. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council: Strengthening Forensic Science in the United States: A Path Forward. National Academies Press, Washington, DC; 2009.

9. Taylor MC, Laber TL, Kish PE, Owens G, Osborne NKP: The reliability of pattern classification in bloodstain pattern analysis—Part 2: Bloodstain patterns on fabric surfaces. Journal of Forensic Sciences. 2016, 61:1461-1466. 10.1111/1556-4029.13191

10. Yuen SKY, Taylor MC, Owens G, Elliot DA: The reliability of swipe/wipe classification and directionality determination methods in bloodstain pattern analysis. Journal of Forensic Sciences. 2017, 62:1037-1042. 10.1111/1556-4029.13298

11. Hulse-Smith L, Mehdizadeh NZ, Chandra S: Deducing drop size and impact velocity from circular bloodstains. Journal of Forensic Sciences. 2005, 50:54-63. 10.1520/jfs2003224

12. Chen L, Zhu Y, Ma C, Lyu Z: Automated identification of impact spatters and fly spots with a residual neural network. Forensic Science International. 2025, 367:112354. 10.1016/j.forsciint.2024.112354

13. Dargan S, Kumar M, Ayyagari MR, Kumar G: A survey of deep learning and its applications: A new paradigm to machine learning. Archives of Computational Methods in Engineering. 2020, 27:1071-1092. 10.1007/s11831-019-09344-w

14. Jung H, Jo YS, Ahn YJ, Jeong J, Lim SK: A first step towards a machine learning-based framework for bloodstain classification in forensic science. Forensic Science International. 2024, 365:112278. 10.1016/j.forsciint.2024.112278

15. Peschel O, Kunz SN, Rothschild MA, Mutzel E: Blood stain pattern analysis. Forensic Science, Medicine and Pathology. 2011, 7:257-270. 10.1007/s12024-010-9198-1

16. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature. 2015, 521:436-444. 10.1038/nature14539

17. Rawat W, Wang Z: Deep convolutional neural networks for image classification: A comprehensive review. Neural Computation. 2017, 29:2352-2449. 10.1162/neco_a_00990

18. Rough R, Bhagat S, Bainbridge-Smith A, Batchelor O, Green R: Classification of bloodstain spatter patterns with convolutional neural networks. 2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ), Christchurch, New Zealand. 2024, 1-6. 10.1109/IVCNZ64857.2024.10794486

19. Bergman T, Kloden M, Dressler J, Labudde D: Automatic classification of bloodstains with deep learning methods. KI - Künstliche Intelligenz. 2022, 36:135-141. 10.1007/s13218-022-00760-y

20. Page MJ, McKenzie JE, Bossuyt PM, et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021, 372:n71. 10.1136/bmj.n71

21. Whiting PF, Rutjes AW, Westwood ME, et al.: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine. 2011, 155:529-536. 10.7326/0003-4819-155-8-201110180-00009

22. Attinger D, Moore C, Donaldson A, Jafari A, Stone HA: Fluid dynamics topics in bloodstain pattern analysis: Comparative review and research opportunities. Forensic Science International. 2013, 231:375-396. 10.1016/j.forsciint.2013.04.018

23. Attinger D, Liu Y, Bybee T, De Brabanter K: A data set of bloodstain patterns for teaching and research in bloodstain pattern analysis: Impact beating spatters. Data in Brief. 2018, 18:648-654. 10.1016/j.dib.2018.02.070

24. Laan N, de Bruin KG, Slenter D, Wilhelm J, Jermy M, Bonn D: Bloodstain pattern analysis: implementation of a fluid dynamic model for position determination of victims. Scientific Reports. 2015, 5:11461. 10.1038/srep11461

25. Dror IE, Charlton D, Peron AE: Contextual information renders experts vulnerable to making erroneous identifications. Forensic Science International. 2006, 156:74-78. 10.1016/j.forsciint.2005.10.017

26. Arthur RM, Cockerton SL, de Bruin KG, Taylor MC: A novel, element-based approach for the objective classification of bloodstain patterns. Forensic Science International. 2015, 257:220-228. 10.1016/j.forsciint.2015.08.028

27. Arthur RM, Humburg PJ, Hoogenboom J, Baiker M, Taylor MC, de Bruin KG: An image-processing methodology for extracting bloodstain pattern features. Forensic Science International. 2017, 277:122-132. 10.1016/j.forsciint.2017.05.022

28. Vitiello A, Di Nunzio C, Garofano L, Saliva M, Ricci P, Acampora G: Bloodstain pattern analysis as optimisation problem. Forensic Science International. 2016, 266:e79-e85. 10.1016/j.forsciint.2016.06.022

29. Shoumy NJ, Ehkan P, Yaakob SN, Ali MS, Khatun S: Feature extraction for neural network pattern recognition for bloodstain analysis. International Journal of Applied Engineering Research. 2016, 11:8583-8589.

30. Krizhevsky A, Sutskever I, Hinton GE: ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017, 60:84-90. 10.1145/3065386

31. Zou T, Stern HS: Towards a likelihood ratio approach for bloodstain pattern analysis. Forensic Science International. 2022, 341:111512. 10.1016/j.forsciint.2022.111512

32. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016, 770-778. 10.1109/CVPR.2016.90

33. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition [PREPRINT]. arXiv. 2015, 10.48550/arXiv.1409.1556

34. Szegedy C, Liu W, Jia Y, et al.: Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA. 2015, 1-9. 10.1109/CVPR.2015.7298594

35. Howard AG, Zhu M, Chen B, et al.: MobileNets: Efficient convolutional neural networks for mobile vision applications [PREPRINT]. arXiv. 2017, 10.48550/arXiv.1704.04861

36. Acampora G, Vitiello A: A comparison of deep learning methods for multi-class classification of bloodstain patterns. Technologies and Applications of Artificial Intelligence. Chu WT, Shen CY, Shuai HH (ed): Springer, Singapore; 2025. 2415:275-288. 10.1007/978-981-96-4596-1_21

37. Jomon S, Bastian BT, Joseph MM: Analysis of passive bloodstain morphology across surface textures and drop heights using deep learning. Forensic Science International. 2026, 383:112890. 10.1016/j.forsciint.2026.112890

38. Li Y, Shen W: From images to detection: Machine learning for blood pattern classification. Forensic Science International. 2025, 375:112558. 10.1016/j.forsciint.2025.112558

39. Książek K, Romaszewski M, Głomb P, Grabowski B, Cholewa M: Blood stain classification with hyperspectral imaging and deep neural networks. Sensors. 2020, 20:6666. 10.3390/s20226666

40. Bergmann T, Heinke F, Labudde D: Towards substrate-independent age estimation of blood stains based on dimensionality reduction and k-nearest neighbor classification of absorbance spectroscopic data. Forensic Science International. 2017, 278:1-8. 10.1016/j.forsciint.2017.05.023

41. Dong Z, Zhang ZD: Enhancing bloodstain analysis through AI-based segmentation: Leveraging segment anything model for crime scene investigation [PREPRINT]. arXiv. 2023, 10.48550/arXiv.2308.13979

42. Acampora G, Santoro M, Vitiello A: A comparative study of one-class classification methods for bloodstain detection in hyperspectral forensic imaging. Neural Computing and Applications. 2026, 38:52. 10.1007/s00521-025-11825-y

43. Romaszewski M, Glomb P, Cholewa M, Sochan A: A dataset for evaluating blood detection in hyperspectral images. Forensic Science International. 2021, 320:110701. 10.1016/j.forsciint.2021.110701

44. Giulietti N, Discepolo S, Castellini P, Martarelli M: Correction of substrate spectral distortion in hyper-spectral imaging by neural network for blood stain characterization. Sensors. 2022, 22:7311. 10.3390/s22197311

45. Sheth TB, Shah MS: A discrete wavelet transform and dense CNN for hyperspectral imaging-based bloodstain classification. Indian Journal of Science and Technology. 2023, 16:3795-3802. 10.17485/ijst/v16i42.1682

46. Giulietti N, Discepolo S, Castellini P, Martarelli M: Neural network based hyperspectral imaging for substrate independent bloodstain age estimation. Forensic Science International. 2023, 349:111742. 10.1016/j.forsciint.2023.111742

47. Alkhuder K: Attenuated total reflection-Fourier transform infrared spectroscopy: a universal analytical technique with promising applications in forensic analyses. International Journal of Legal Medicine. 2022, 136:1717-1736. 10.1007/s00414-022-02882-2

48. Doty KC, Muro CK, Lednev IK: Predicting the time of the crime: Bloodstain aging estimation for up to two years. Forensic Chemistry. 2017, 5:1-7. 10.1016/j.forc.2017.05.002

49. Bremmer RH, de Bruin KG, van Gemert MJC, van Leeuwen TG, Aalders MCG: Forensic quest for age determination of bloodstains. Forensic Science International. 2012, 216:1-11. 10.1016/j.forsciint.2011.07.027

50. Attinger D, Comiskey PM, Yarin AL, De Brabanter K: Determining the region of origin of blood spatter patterns considering fluid dynamics and statistical uncertainties. Forensic Science International. 2019, 298:323-331. 10.1016/j.forsciint.2019.02.003

51. Home PH, Norman DG, Wade K, Spearing E, Williams MA: 3D scanning a crime scene to enhance juror understanding of Bloodstain Pattern Analysis evidence. Science & Justice. 2024, 64:333-338. 10.1016/j.scijus.2024.04.007

52. Rough R, Batchelor O, Green R, Bainbridge-Smith A: An automated method for the generation of bloodstain pattern metrics from images of blood spatter patterns. Forensic Science International. 2024, 363:112200. 10.1016/j.forsciint.2024.112200

53. Lundberg SM, Lee SI: A unified approach to interpreting model predictions. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017, 4765-4774.

54. Ypma RJ, Ramos D, Meuwly D: AI-based forensic evaluation in court: The desirability of explanation and the necessity of validation. Artificial Intelligence (AI) in Forensic Sciences. Geradts Z, Franke K (ed): Wiley, Hoboken, NJ; 2023. 3-20.

55. Attinger D, Liu Y, Faflak R, et al.: A data set of bloodstain patterns for teaching and research in bloodstain pattern analysis: Gunshot backspatters. Data in Brief. 2019, 22:269-278. 10.1016/j.dib.2018.11.075

56. Kannan S, Pachong SM, Lewis PR, Stotesbury T: A dataset of drip patterns for teaching and research purposes in forensic bloodstain pattern analysis. Data in Brief. 2025, 59:111352. 10.1016/j.dib.2025.111352

57. Comiskey PM, Yarin AL, Kim S, Attinger D: Prediction of blood back spatter from a gunshot in bloodstain pattern analysis. Physical Review Fluids. 2016, 1:043201. 10.1103/physrevfluids.1.043201

58. Hook E, Fieldhouse S, Flatman-Fairs D, Williams G: Bloodstain classification methods: A critical review and a look to the future. Science & Justice. 2024, 64:408-420. 10.1016/j.scijus.2024.06.004

59. Osborne NKP, Taylor MC, Healey M, Zajac R: Bloodstain pattern classification: Accuracy, effect of contextual information and the role of analyst characteristics. Science & Justice. 2016, 56:123-128. 10.1016/j.scijus.2015.12.005

60. Siu S, Pender J, Springer F, Tulleners F, Ristenpart W: Quantitative differentiation of bloodstain patterns resulting from gunshot and blunt force impact. Journal of Forensic Sciences. 2017, 62:1166-1179. 10.1111/1556-4029.13418

61. Esaias O, Noonan GW, Everist S, Roberts M, Thompson C, Krosch MN: Improved area of origin estimation for bloodstain pattern analysis using 3D scanning. Journal of Forensic Sciences. 2020, 65:722-728. 10.1111/1556-4029.14250

62. McCleary S, Liscio E, De Brabanter K, Attinger D: Automated reconstruction of cast-off blood spatter patterns based on Euclidean geometry and statistical likelihood. Forensic Science International. 2021, 319:110628. 10.1016/j.forsciint.2020.110628

63. Comiskey PM, Yarin AL, Attinger D: Hydrodynamics of back spatter by blunt bullet gunshot with a link to bloodstain pattern analysis. Physical Review Fluids. 2017, 2:073906. 10.1103/physrevfluids.2.073906

64. Liu Y, Attinger D, De Brabanter K: Automatic classification of bloodstain patterns caused by gunshot and blunt impact at various distances. Journal of Forensic Sciences. 2020, 65:729-743. 10.1111/1556-4029.14262

65. Arthur RM, Hoogenboom J, Green RD, Taylor MC, de Bruin KG: An eye tracking study of bloodstain pattern analysts during pattern classification. International Journal of Legal Medicine. 2018, 132:875-885. 10.1007/s00414-017-1711-6

66. Arthur RM, Humburg A, Attinger D, De Brabanter K: An automated approach to the classification of impact spatter and cast-off bloodstain patterns. Forensic Science International. 2018, 289:310-319. 10.1016/j.forsciint.2018.05.019

67. Pachong SM, Alavi A, Kannan S, Stotesbury T, Lewis PR: Unsupervised machine learning for the detection and interpretation of key features in drip patterns. Forensic Science International. 2026, 378:112669. 10.1016/j.forsciint.2025.112669

68. Balbudhe MS, Kumar Shetty BS, Kumar NA, Holla R, Render R, Mithal D: Comparison between digital and non-digital bloodstain pattern analysis methods: a scoping review. Egyptian Journal of Forensic Sciences. 2026, 16:13. 10.1186/s41935-026-00517-w

69. Boos K, Orr A, Illes M, Stotesbury T: Characterizing drip patterns in bloodstain pattern analysis: An investigation of the influence of droplet impact velocity and number of droplets on static pattern features. Forensic Science International. 2019, 301:55-66. 10.1016/j.forsciint.2019.05.002

70. Balbudhe MS, Babar SP: A comparative study of bloodstain pattern analysis on different types of paper surfaces. Advances in Data-Driven Computing and Intelligent Systems. Bansal JC, Saha S, Coello CAC, Rathore H (ed): Springer, Singapore; 2025. 1333:35-47. 10.1007/978-981-96-4536-7_4

71. Benabdelhalim H, Brutin D: Influence of the relative humidity on the blood drip stain formation on jeans fabrics. Forensic Science International. 2023, 350:111808. 10.1016/j.forsciint.2023.111808

72. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-CAM: Visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy. 2017, 618-626. 10.1109/ICCV.2017.74

73. Ribeiro MT, Singh S, Guestrin C: "Why should I trust you?": Explaining the predictions of any classifier. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 1135-1144. 10.1145/2939672.2939778

74. Solanke AA: Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models. Forensic Science International: Digital Investigation. 2022, 42:301403. 10.1016/j.fsidi.2022.301403

75. Kirillov A, Mintun E, Ravi N, et al.: Segment anything. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023. 2023, 3992-4003. 10.1109/ICCV51070.2023.00371

76. Acampora G, Vitiello A, Di Nunzio C, Saliva M, Garofano L: Towards automatic bloodstain pattern analysis through cognitive robots. 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China. 2015, 2447-2452. 10.1109/SMC.2015.428