

Digital Ballistics and Automated Firearm Identification Systems

Komal Yadav¹, Priyanka Negi¹, Riya Singh¹, Muskan Chauhan¹, Breeti Sharma¹, Riya¹, Aarya Rajora¹, Janhvi¹

Department of Forensic Science, School of Basic and Applied Sciences, K.R. Mangalam University, Gurugram, India

Corresponding author: - Priyanka Negi, priyankasoni.negi@gmail.com

Abstract

Automated identification and comparison of firearms using bullets and cartridge cases has been studied for nearly three decades, but the discipline had trouble in progressing from theoretical research to practical application. This review examines multiple publications on the topic over the last 27 years, beginning with early systems based on 2D imaging and line-scan/profile comparison. It discusses the evolution of the discipline, with later developments including 3D surface topography and statistical evaluation using congruent matching cells, convolutional neural networks, and vision transformer-based methods. Under controlled experimental conditions, the review highlighted that the accuracy rates are usually over 90%, reported on small carefully selected datasets including dozens of similar firearms. The accuracy of such models was mostly tested on the same firearm brands and calibres, which does not always reflect the variety of firearms used by shooters. The latest research increasingly employs deep learning, large data sets and still focuses on particular geographic regions or brands of firearms. Overall, the literature indicates that the field is moving toward more quantitative and less subjective approaches; however, it still depends on Cross-brand and cross-calibre validation, dataset diversity, standardized protocols and independent external validation. This aim to evaluate use of automated imaging and machine learning methods for identifying firearms and assess their effectiveness across various types of firearms and ammunition.

Keywords: Firearm identification; Tool mark comparison; Cartridge-case analysis; Three-dimensional surface topography; Machine learning; Forensic validation, Automated identification

How to cite this article: Komal Yadav¹ Priyanka Negi¹ Riya Singh. Digital Ballistics and Automated Firearm Identification Systems. Int J Drug Deliv Technol. 2026;16(85s):705-714. DOI: 10.25258/ijddt.16.85s.69.

1. Introduction

Firearm and toolmark examination has traditionally relied on comparison microscopy, in which a trained examiner visually compares the striated and impressed marks left on bullets and cartridge cases by the internal surfaces of a firearm. (Senin et al., 2006) This approach has been used in casework for over a century, but it has also been the subject of sustained criticism because the final determination rests on the subjective judgment of the examiner rather than on a quantifiable, repeatable measurement. The 2009 U.S. National Academy of Sciences report on forensic science, and later scientific commentary, called for firearm and toolmark identification to move toward objective, statistically defensible methods with known and reported error rates.

In response, researchers across forensic science, metrology, electrical engineering, and computer science have spent the better part of three decades trying to automate some or all of the firearm-identification workflow of 2006: acquiring a digital image or three-dimensional surface map of a bullet or cartridge case, extracting features from that surface, and computing a similarity score or classification a study of 2012 suggests it that can support (or eventually replace) the examiner's visual comparison. This effort has produced a large and fragmented body of literature that spans several distinct technical eras two-dimensional line-scan imaging, three-dimensional surface topography, classical statistical pattern recognition, and, most recently, deep learning.

This review brings together 60 entries drawn from that literature corresponding to 54 unique studies once companion or duplicate database entries are consolidated published between 1999 and 2026, and organizes them by technological approach in order to (i) trace how the field's methods have changed over time, (ii) summarize the accuracy and validation evidence reported for each approach, and (iii) identify the limitations that recur across studies regardless of the specific technique used. The aim, consistent with the calls for objective firearm identification noted above, is to evaluate the use of automated imaging and machine-learning methods for identifying firearms and to assess how well their demonstrated effectiveness would be expected to generalize across the full diversity of firearms and ammunition encountered in real casework.

2. Review Approach

The 60 database entries reviewed here were compiled from peer-reviewed journal articles (predominantly *Forensic Science International*, the *Journal of Forensic Sciences*, and NIST-affiliated publications), conference proceedings, the AFTE Journal, and a small number of government reports and trade articles, covering the period 1999–2026. For each entry the authorship, year, title, key findings, and stated limitations were extracted, together with the reference or DOI where one was available. Four pairs of entries in the source table are companion or duplicate citations of the same underlying study (Senin et al. 2006), researchers in 2010 following this by a report of 2023, and research done in 2026 shows that each appear more than once

and are treated as single studies in the synthesis below, giving 54 unique works. Papers are grouped into seven broad technological categories that reflect the field's chronological and methodological development; several papers span more than one category and are discussed under whichever theme is most central to their contribution. To provide context for the technological features discussed throughout the review.

Figure 1 provides an overview of the main toolmark features examined in automated firearm identification. It shows the different marks that can be observed on cartridge cases, such as firing-pin, breech-face, extractor, and ejector marks, as well as the land-engraved striations present on fired bullets. These features are important in automated firearm comparison because they provide the microscopic patterns used by imaging and comparison methods to assess similarities between samples.

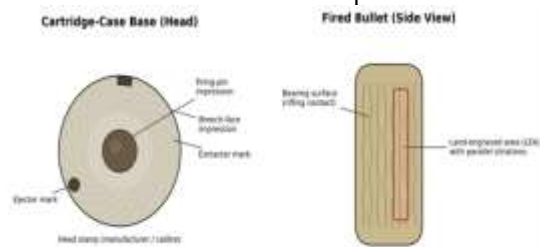


Figure 1. Schematic of the principal toolmark features examined in automated firearm identification: cartridge-case base marks(left) and bullet land-engraved striations(right).

3. Evolution of Automated Firearm Identification Technologies

3.1 Early two-dimensional imaging and line-scan systems (1999–2009)

The earliest work in this review used two-dimensional line-scan imaging and profilometry to digitize the surface of a bullet or cartridge case and compare the resulting signal by correlation. One of the first automated systems of this kind, converting laser-profilometry data into feature vectors that could be compared automatically; the system performed well when firearms produced clearly distinct striation patterns, but it was tested on only six bullets from two pistols (De Kinder & Bonfanti, 1999). A similar early approach combined line-scan imaging with surface profiling to capture land-and-groove striations for digital matching, in an early proof-of-concept that reported no formal error rate (Smith et al., 2000). A different approach using line-scan imaging and Hough-transform circle detection to read the head stamp on a cartridge case rather than the barrel-induced striations in order to classify country, manufacturer, year, and calibre; this was tested on 2,000 head stamps but required controlled lighting and assumed a fixed set of layout templates (Mazumdar, Mitra and Kar (2002)).

A pivotal large-scale evaluation came from a tested reference ballistic-imaging database built from 600

identical pistols and 4,200 test-fired cartridge cases. When queries were restricted to the same ammunition brand, the correct match appeared in the top ten results 72% of the time; however, when queries crossed ammunition brands, that figure fell to only 21%, and accuracy declined further as the database grew (De Kinder et al., 2004). This paper is one of the clearest early demonstrations that same-brand performance figures do not generalize to cross-brand conditions, a finding that reappears throughout the literature reviewed here. The research improved on single-image comparison by fusing several differently lit images of a bullet into one high-quality signature before cross-correlation, reporting better performance than a commercial comparator, though the evaluation used only one commercial baseline and one database (Puentes León, 2006), and proposed an FFT-based line-scan method tested on images from 16 weapons (Li, 2006), argued for full three-dimensional surface topography specifically to overcome the information loss inherent in ranked-list, two-dimensional commercial systems such as IBIS (Baiker-Sørensen et al., 2023). The report of 2006 draws attention to an important conceptual step by introducing a prototype three-dimensional surface-topography system for cartridge cases that combined a "3D virtual comparison microscope" for visual inspection with a cross-correlation score for objective evaluation; notably, the authors (Tong et al., 2015) framed the system as intended to augment rather than replace a trained examiner, a positioning echoed by several later three-dimensional systems, though this early paper reported no large-scale accuracy figures (Senin et al., 2006).

3.2 Three-dimensional surface topography and correlation-cell methods

Following the early three-dimensional prototypes, a substantial body of work much of it centred on the U.S. National Institute of Standards and Technology (NIST) developed three-dimensional topography measurement into a more formal comparison framework. Previous investigations of 2012 documented the shift from optical image comparison toward quantitative two- and three-dimensional surface-topography measurement, reporting improved correlation performance on NIST standard reference bullets and cases (Song et al., 2012) then proposed dividing three-dimensional topography into discrete "correlation cells," a design intended to support high-speed comparison and large-scale database searching, which was further developed by it into open algorithms and statistical models for reporting error rates. This line of work culminated in the Congruent Matching Cells (CMC) method provided its first validation, using 40 cartridge cases fired from ten consecutively manufactured pistol slides a deliberately difficult scenario and reported that matches and non-matches separated cleanly with zero observed errors. Later it was laid out the

statistical theory underpinning CMC, estimating an error rate of roughly one in a million on that same 40-case dataset, by analogy with the way DNA evidence error rates are reported. The report subsequently improved the CMC algorithm by checking correlation scores across all rotation angles and combining forward and backward comparisons, which widened the separation between matching and non-matching score distributions. The literature also extended CMC evaluation to conventional optical/grayscale imagery rather than full three-dimensional topography, again reporting no false identifications or exclusions across 63 matching and 717 non-matching pairs drawn from 40 cartridge cases (Tong et al., 2014). The development of these approaches occurred alongside several overlapping methodological trends in digital ballistics research, as illustrated in Figure 2.

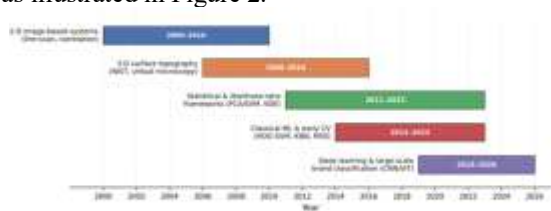


Figure 2. Overlapping methodology eras in digital ballistics research, 2000-2026, showing the gradual, non-sequential accumulation of imaging and analytic approaches.

It is notable that this entire CMC validation programme spanning at least five separate papers was built on the same small, 40-case reference dataset. The study of 2018 attempted to address this by extending the error-rate analysis to an independent 95-case dataset, reporting that a model trained on one dataset generalized successfully to the other with very low false-positive rates; however, a co-author of the CMC programme (Song et al., 2018), a researcher reported to have publicly noted that the study "did not include enough test-fires to calculate realistic error rates for actual casework" and called for larger, more diverse datasets—a caution that is directly relevant to how the very low, headline error-rate figures from this line of research should be interpreted (Tong et al., 2015). Parallel three-dimensional topography work addressed the metrology underpinning these systems rather than matching algorithms as such. The evidence collected in 2012 highlighted the National Ballistics Imaging Comparison (NBIC) project, in which 19 examiners across 13 laboratories performed repeated scans of NIST standard reference materials, uncovering real operational problems such as misaligned bullets, contaminated samples, and a software bug—useful reminder that error can enter at the measurement stage before any comparison algorithm is applied. The above report supported the 2018 research that subsequently proposed quality-assurance guidance for three-dimensional topography instruments, covering

performance specifications, installation validation, and new physical reference standards, and in follow up research of (2015/16) reviewed the broader landscape of optical three-dimensional measurement technologies used in ballistics and toolmark examination (Vorburger et al., 2015).

3.3 Statistical pattern recognition before deep learning

A separate methodological thread applied classical statistical and machine-learning techniques geometric moments, principal component analysis (PCA), support vector machines (SVM), and related classifiers to features extracted from bullet or cartridge-case images. As research progressed in 2010 researchers extracted 16 geometric-moment features from firing-pin impressions on 747 cartridge cases fired from five pistols of the same brand and model and found that the six most informative features gave 74.0% cross-validated classification accuracy a relatively modest figure that the authors themselves noted leaves considerable room for error in a forensic context. Further report of 2010 applied a similar geometric-moment approach to the same type of firing-pin data (747 cases from five identical 9 mm pistols) but reduced the feature set to 11 dimensions and reported markedly higher accuracy of 96.7%; because all five firearms were the same model, this result mainly demonstrates that the method can separate those specific pistols rather than firearms in general, and the authors also flagged that higher-order moments are noise-sensitive (Md Ghani et al., 2010).

Extended this approach of 2011, three-dimensional surface metrology with PCA, SVM, and conformal prediction to objectively discriminate firearm toolmarks, reporting a 3.5% empirical error rate at 95% confidence but on a small, homogeneous dataset of 58 primer shear marks from only four Glock 19 pistols, and requiring high-resolution confocal microscopy that would limit routine deployment (Gambino, C., 2011). Later another report of 2012 published two related contributions in the same year: a broader statistical framework for representing toolmarks and firearm impressions mathematically, and a more specific method in which partial least-squares discriminant analysis (PLS-DA) achieved 97% or better correct classification using eight dimensions, while a PCA-SVM approach matched that performance with only four dimensions, with conformal prediction used to assign confidence to individual identifications; both studies used idealized, reproducible striation patterns from a relatively small set of screwdrivers rather than firearms evidence collected under casework conditions (Petraco, Shenkin, Speir et al., 2012).

Rotation- and orientation-invariant feature design was a recurring concern in this era, since cartridge cases and bullets can be presented to a scanner at

arbitrary angles. Building on this work, subsequent research proposed a rotation-invariant feature derived from Fourier coefficients of the primer image in polar coordinates, explicitly to reduce the effect of case rotation, although the approach was limited to the primer region and pre-dated deep-learning methods (Thumwarin et al. 2008). Following these findings, later research of 2011 focused on using white-light interferometry to extract firing-pin-impression and aperture-outline features for firearm class identification (that is, narrowing down the type or model of firearm rather than identifying an individual firearm), which is a materially different and less demanding task than individual-source identification (Kamalakkannan et al., 2011)

3.4 Toolmark statistics and likelihood-ratio approaches

A related strand of research, often working with screwdrivers and other tools rather than firearms directly, developed statistical frameworks for expressing the strength of a toolmark match as a likelihood ratio rather than a binary decision. More recent reports of (2014) showed that an automated method could strongly discriminate known matches from non-matches even with angle differences of up to 30° between compared marks and found that three-dimensional data performed slightly better than two-dimensional data, though the automated method sometimes gave more incorrect support for common-origin hypotheses than a human examiner would, underscoring the need for careful statistical validation. As the field progressed, researchers increasingly focused on developing a similarity-score and likelihood-ratio framework specifically for breech-face, firing-pin, and aperture-shear toolmarks on cartridge-case primers, and showed that likelihood ratios derived from different toolmark types on the same case could be statistically combined an important step toward integrating multiple sources of evidence from a single item but the framework was built around a single pistol model and a single brand of ammunition. Cuellar (2024) proposed a similarity-score, clustering, and likelihood-ratio method for three-dimensional toolmark comparison evaluated on 560 marks from consecutively manufactured screwdrivers, reporting 98% sensitivity and 96% specificity, while noting that marks under 1.5 mm could not be reliably compared and that the method's applicability to tools other than slotted screwdrivers remains untested (Baiker-Sørensen et al., 2023).

3.5 Commercial systems and database-scale operational validation

Several studies evaluated commercial automated ballistic-comparison systems principally Evofinder® and the Integrated Ballistic Identification System (IBIS) under conditions closer to operational deployment. Feng Dong and colleagues (2019) scored 2,996 bullets fired from

1,000 Chinese Norinco QSZ-92 pistols using Evofinder, reporting areas under the ROC curve (AUC) between 0.93 and 0.9996, with the best performance obtained for land-engraved areas, and used this analysis to propose cut-off scores for common-origin determinations; the study was, however, limited to one firearm model and one ammunition type, and some critics have questioned whether the similarity-log-ratio measure used truly reflects statistical "typicality" rather than simple resemblance. Thereafter, a fired five QSZ-92 pistols 3,070 rounds each and found that both human examiners and Evofinder correctly matched early- and late-fired cases despite a slight drift in similarity scores as barrels wore, although the automated test design compared only each gun's first case against the database rather than every case against every other case (Zhang, K., 2017). In the follow up research of 2019, tested a much larger operational resource, China's Registered Ballistic Database of roughly 1,000 QSZ-92 pistols and 3,000 bullets and cases, by searching 1,000 randomly selected items against it; every successful identification ranked first, with high match rates across all mark types, though this measures ranking accuracy only for items that are already known to be present in the database, and again involves a single firearm model and a single national database structure (Li, et al., 20). Recent work of 2021 took a cross-border perspective, testing a prototype French-Swiss Evofinder network for distributed versus centralized data exchange; most mock trials were correctly matched regardless of query direction, supporting distributed exchange as a cost-effective option, but a handful of problems an unmatched bullet, settings requiring adjustment, and an inter-laboratory marking discrepancy indicated uneven performance across sites (Sautier, C., 2021).

For IBIS, a study of (2017) conducted an independent, non-NIST evaluation of the IBIS Heritage™ system using NIST standard reference cartridge cases, finding minimal variability between and within examiners and excellent overall discrimination; the reliance on clean, standardized reference material rather than worn or damaged casework evidence, and the testing of only one system generation, limit how far this finding can be extrapolated (Morris et al., 2017). Within the developing field some researchers evaluated a different kind of tool a "virtual microscopy" platform allowing examiners to inspect three-dimensional scans on screen rather than under a physical comparison microscope in a proficiency-test format involving 56 examiners at 15 laboratories, who correctly identified 100% of known matches with no false positives; this remains a controlled proficiency exercise for one company's software rather than a blind test on casework (Duez, P., 2018).

3.6 Feature-based firing-pin and cartridge-case comparison methods

A further group of studies focused specifically on firing-pin impressions and related cartridge-case features using two- and three-dimensional spatial descriptors that sit methodologically between the classical statistical methods above and full deep learning. Collectively, these developments demonstrated the feasibility of automatically matching firing-pin impressions using digitally extracted two- and three-dimensional spatial characteristics (Fischer, R., 2014). The same authors subsequently developed the Multiple-Slice-Shape (MSS) technique, which automated firing-pin comparison without requiring manual point-labelling and achieved 100% accuracy distinguishing two firearms, though accuracy fell to 73% (using a maximum-a-posteriori scoring rule) or 92% (using a different scoring rule) across a slightly larger set of six pistols, with the authors describing the six-gun study as preliminary and noting that one gun's pin design was poorly discriminable (Fischer & Vielhauer, 2015a,b). A companion paper proposed rotation-invariant features based on confocal microscopy, achieving 100% accuracy for two similar 9 mm firearms and 86% across all six tested, again with a small sample and reference circles that had to be annotated manually. Against this background a report developed a phase-correlation image-registration method that identifies "effective correlation areas" via cross-covariance and edge density, deliberately excluding irrelevant image regions to improve discrimination over whole-image correlation, though these risks discarding potentially useful information and was validated on only one dataset (Yammen & Muneesawang, 2013). Based on this work research applied a Histogram-of-Oriented-Gradients (HOG) feature extractor with an SVM classifier to firing-pin impressions cast from nine Ruger 10/22 pistols, reaching 93% accuracy, but again on a single firearm type. Recent work applied K-nearest-neighbour classification to segmented bullet striation patterns from just six bullets belonging to three firearms, achieving 84.67% accuracy, while noting that the segmentation step generated an excess of spurious striation detections. With the emergence of further evidence developed two novel scoring methods for bullet identification based on three-dimensional pattern matching and line counting, tested on 406 bullets and reporting 98% sensitivity, though the authors acknowledged a lack of data on extreme-tail non-matches and noted the method cannot be applied to bullet fragments (Changmai, P., 2019).

3.7 The deep-learning era: segmentation, region detection, and classification (2019–2026)

The most recent phase of the literature applies convolutional neural networks (CNNs) and related deep-learning architectures, generally to two distinct sub-tasks: locating or segmenting the relevant

regions of a cartridge case or bullet image and classifying or matching firearms or firearm brands from those regions.

On the segmentation and region-detection side, as subsequent evidence emerged a trained U-Net model was developed to locate breech-face and firing-pin regions on 1,195 nine-millimetre cartridge-case images, matching expert-drawn boundaries with 95–96% intersection-over-union (IoU) and over 99% Dice coefficient a strong result for what is explicitly a preprocessing step rather than full identification, and one obtained on cases from a single scanner and a single calibre (Le Bouthillier et al., 2023). With continued research, a standard U-Net against two residual variants for segmenting bullet striations across 528 images (132 base images before augmentation) was compared, finding that a Residual U-Net achieved the highest accuracy at 89.70%, again addressing segmentation rather than whole-firearm matching and without comparison to prior systems. More recent findings benchmarked five lightweight CNN architectures for striation-mark segmentation, all of which exceeded 95% accuracy, including a quantized "QD-AMSN" model as small as 6 MB an encouraging result for eventual field or mobile deployment (Baruah, S. et al. 2025) though the paper notes that Dice and IoU scores were comparatively modest despite the high headline accuracy figure. In the follow up investigation three compared CNN-based segmentation architectures U-Net, Dense U-Net, and a modified DeepLabv3+ on NIST cartridge-case images, with the modified DeepLabv3+ achieving Dice scores above 95%; because the models were trained and tested on the NIST dataset specifically, the authors call for external validation on images from different cameras, illumination setups, and cartridge types. The subsequent study reported a foundational feature-extraction and similarity-measurement study across firing-pin, breech-face, and annulus impressions, explicitly describing it as an early step that will require larger datasets and a wider variety of weapons before any forensic application (Baruah et al 2025).

On the classification side, recent work has built a smartphone-based system for capturing bullet-marking images and trained three CNN architectures (DenseNet121, ResNet50, and Xception) on 898 bullets from eight firearm brands, with DenseNet121 performing best at 91–98% accuracy; the underlying dataset was imbalanced across brands, which can bias a classifier's apparent performance, and the system was validated only on intact 9 mm bullets from eight brands (Pisantanroj, P., 2020). The largest single dataset identified in this review used the BALISTIKA system and more than 350,000 cartridge cases from 89 firearm brands (21 leading brands accounting for roughly 97% of the Turkish dataset used), augmenting the data to over one million samples and finding that ResNet and

Vision Transformer architectures reached up to 92% accuracy, outperforming classical SVM and Random Forest baselines; because the dataset is drawn overwhelmingly from firearms encountered in Türkiye, its generalizability to other regions and firearm populations is untested, and, as the authors themselves caution, a 92% brand-classification accuracy is a different quantity from a 92% probability that a specific firearm fired a specific crime-scene cartridge case brand classification and individual-source identification are not equivalent tasks (Meral & Akyüz, 2026). The approach was further developed through application of CNNs and Random Forest classifiers to firing-pin markings and secondary rifling marks, and reported that assembling the top three CNNs raised firing-pin detection accuracy from 82.6% to 92.8%, with bullet-marking classification reaching 84–86%; performance depended heavily on dataset quality and the composition of the "unknown-item" class, and the training data required augmentation because of its limited initial size (Herke, C., 2026).

3.8 Visualization tools and synthesis reviews

Two entries in the dataset fall outside the imaging-and-classification pipeline as such. Based on this earlier work developed a virtual-reality application (using an Oculus Rift S headset) was developed that lets an examiner manipulate three-dimensional models of cartridge cases with hand gestures; fourteen general users and three domain experts rated the system highly for usability, though the study measured usability and comfort rather than identification accuracy, and the sample was small and drawn largely from a gaming-experienced population. These findings provide a narrative review of the field's progress from early two-dimensional microscopy and Consecutive Matching Striae through three-dimensional techniques, Congruent Matching Cells, and likelihood-ratio methods, and consistent with the synthesis in this review, highlights multimodal data fusion and the identification of non-standard firearms (for example, home-made or three-dimensional-printed firearms) as open challenges for the field. The distribution of the reviewed studies across methodological approaches and publication periods further illustrates this progression, as shown in Figure 3.

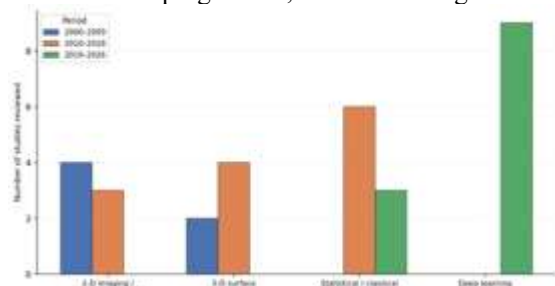


Figure 3. Distribution of methodological approach by publication period among the studies discussed, illustrating the progressive shift from 2-D imaging

toward 3-D topography, statistical machine learning, and most recently, deep learning.

4. Synthesis of Reported Accuracy and Performance

Read across the 54 studies, reported accuracy figures for automated firearm and toolmark identification are frequently high often above 90%, and in several Congruent-Matching-Cells and proficiency-test studies, effectively at or near 100% with zero observed errors. However, three patterns recur across almost every high-accuracy result reported here:

4.1 Small sample sizes. This study aims to understand that many of the strongest accuracy figures come from datasets of well under a hundred firearms (Md Ghani et al., 2010), and several from fewer than ten.

4.2 Same-brand, same-model, or same-calibre testing. Where a study specifies the firearms used, they are frequently identical or near-identical in make and model (for example, five identical 9 mm pistols in both Leng et al., 2010, and Ghani et al., 2010; four Glock 19 pistols in Gambino et al., 2011); ten consecutively manufactured slides in (Chu et al., 2013), which tests the method's ability to discriminate among very similar items rather than its performance across the full diversity of firearms encountered in casework.

4.3 Reuse of the same reference datasets. The analysis from 2013-2018 in the Congruent Matching Cells research programme (Song et al., 2018), in particular, produced at least five publications built substantially on the same 40-case NIST dataset, and the widely cited "one-in-a-million" CMC error rate derives specifically from that dataset rather than from an independently replicated, casework-representative sample.

Where studies have deliberately tested cross-brand or cross-calibre conditions, or worked with larger and more heterogeneous datasets, reported performance is consistently lower and more variable: The 2004 study show that accuracy falling from 72% (same brand) to 21% (cross-brand) in a 600-pistol database (De Kinder et al., 2004); 2010 study reported only 74.0% accuracy even within a same-model dataset; and 2020 and 2026 both year reported accuracy in the low-to-mid 80s to low 90s once brand imbalance and larger (Meral & Akyüz, 2026), more heterogeneous data were involved. This pattern is consistent with the review's central conclusion: automated firearm identification performs very well

under narrow, controlled conditions but its performance has not yet been convincingly demonstrated across the breadth of firearms, ammunition, and casework conditions that a forensic laboratory would actually encounter.

5. Recurring Limitations Across the Literature

Six limitations recur across the studies reviewed here, regardless of the specific imaging modality or algorithm used:

5.1 Limited dataset diversity. The great majority of studies use firearms of a single brand, model, or even a single set of consecutively manufactured parts (De Kinder et al., 2004), and a single ammunition type; only a small minority test across multiple brands, and even these are geographically or demographically concentrated (Meral & Akyüz, 2026).

5.2 Absence of cross-brand and cross-calibre validation. Where cross-brand testing has been performed at all, the study reveals substantially lower performance than same-brand testing suggesting that same-brand accuracy figures reported elsewhere in the literature should not be assumed to generalize (De Kinder et al., 2004).

5.3 Reliance on a small number of shared reference datasets. NIST's standard reference materials and the 40-case CMC dataset underpin a large cluster of publications (Song et al., 2018), which limits the independence of the field's accumulated evidence base.

5.4 Segmentation or region-detection treated as a proxy for identification. A substantial and growing share of the deep-learning literature from 2023-2025 addresses the relevant regions of cartridge-case image rather than performing or validating full firearm-source identification; high segmentation accuracy (Dice, IoU) does not by itself establish comparable identification accuracy (Le Bouthillier et al., 2023; Mookiah et al., 2025).

5.5 Limited or absent independent, external validation. Several of the field's headline results have been directly questioned (Song et al., 2018) from 2018-2025 by co-authors or independent researchers or explicitly caveated by their own authors as preliminary or foundational.

5.6 Lack of standardized protocols across studies and laboratories. Feature sets, scoring thresholds, imaging conditions, and evaluation metrics differ from study to study (Kamalakaran et al.,

2011), which, together with the National Ballistics Imaging Comparison project's finding that measurement-stage errors (misalignment, contamination, software bugs) (Song et al., 2018) can occur even under controlled NIST protocols that makes it difficult to compare results across the literature or to certify a single method as ready for widespread casework deployment.

6. Future Scope

Future research in the field of automated firearm identification should be focused on expanding ballistic databases to include various brands, models, calibres, ammunition types and degrees of firearm wear (De Kinder et al., 2004; Meral & Akyüz, 2026). Most of the previous studies were performed under highly controlled laboratory settings with limited data subsets, which might not necessarily represent crime scene data.

Additional studies are needed to validate performance on independent data sets from different manufacturers, laboratories and imaging instruments. Inter-brand, inter-caliber and inter-laboratory experiments should be conducted to evaluate performance on non-training sets and establish generalizability. Acquisition, calibration, processing and reporting protocols should also be standardised (Song et al., 2018).

Furthermore, the combination of 2D information with 3D surface topography data may be beneficial for the analysis of striations, firing-pin impressions, breech-face impressions, and other details found on the cartridge case. Future systems should aim to analyse deformed, fragmented, corroded or otherwise compromised ballistic evidence, as such evidence is often encountered in practice (Senin et al., 2006; Vorbürger et al., 2015).

The analysis should not be limited to an automated classification but should also attempt to identify relevant features used in the comparison for visual inspection by an examiner. Artificial-intelligence software should report not only the classification but also the features used for comparison, confidence levels, likelihood ratios, and error rates. Blind testings and external validation are critical for the field before such systems can be considered reliable enough for forensic applications (Song et al., 2018). The future of forensic firearms examinations in this domain should include the development of standardized inter-laboratory ballistic databases with open access, algorithms for reduced resource requirements in smaller forensic labs, and identification of three-dimensionally printed firearms. Overall, such automated systems should help human examiners by reducing the time required for analysis and adding another layer of consistency, but results should always be subjected to human evaluation.

7. Conclusion

The development of automated and computer-assisted firearm identification has undergone substantial technological transformation over the last twenty-seven years. The literature reviewed in this study indicates an evolution from early two-dimensional imaging and line-scan techniques, to three-dimensional surface topography, statistical pattern-recognition methods, likelihood-ratio frameworks, and, more recently, deep-learning-based approaches. Such developments have progressively shifted firearm and toolmark examination from predominantly subjective visual comparison towards more quantitative and reproducible methods (Senin et al., 2006; Song et al., 2012). Automated imaging has demonstrated considerable potential for the analysis of bullets and cartridge cases through extraction, comparison, and statistical evaluation of firearm-produced marks. A major finding of the reviewed literature is that many automated systems report high levels of accuracy, with several studies achieving results above 90% and some controlled experiments reporting zero observed errors. However, such performance figures must be interpreted in relation to the conditions under which they were obtained. A considerable portion of the studies relied on relatively small datasets consisting of firearms from the same brand, model, calibre, or indeed consecutively manufactured components (Md Ghani et al., 2010). Such controlled datasets are useful for demonstrating the technical capability of a method, but they do not necessarily represent the wide diversity of firearms, ammunition, environmental conditions, and damaged or contaminated evidence encountered in actual forensic casework (De Kinder et al., 2004; Song et al., 2018). The decrease in performance observed in cross-brand and cross-calibre testing further demonstrates the importance of evaluating automated systems under more challenging and representative conditions. The review also highlights the increasing presence of artificial intelligence and deep learning in forensic ballistics. Convolutional neural networks, U-Net-based architectures, residual networks, and Vision Transformers have demonstrated promising results in firearm-related image segmentation, feature extraction, and classification. The availability of larger datasets has also improved the potential of such methods. However, high segmentation or classification accuracy should not automatically be interpreted as successful individual firearm identification (Meral & Akyüz, 2026). In particular, brand classification and individual-source identification represent different forensic tasks, and evidence supporting one cannot be directly treated as evidence for the other. Another crucial limitation is the lack of independent, external validation. A number of the most highly cited publications used the same reference sets across the Congruent Matching Cells research programme. While these

papers represent tremendous value to both the fields of statistical analysis and error estimation, the collective power of these papers is reduced by the fact that all of them rely on the same set of references. Additional measurement-related problems, including misalignment, contamination, variability of imaging and software, illustrates that a variety of errors could come before the comparison algorithm (Vorbürger et al., 2015). At this point in time, automated firearm identification should be considered as a burgeoning field with significant forensic value, but one which requires significantly more empirical research to confidently take on real-world casework. Future studies should collect larger, more diverse, and more geographically representative sets of reference materials that incorporate a wide range of firearms manufacturers, models, calibres, ammunition types, and realistic casework circumstances. Cross-validation between different brands and calibres should be a standard practice in performance evaluation. Furthermore, standardization of imaging practices, feature extraction, scoring, quality-assurance criteria, and reporting must be established (Kamalakaran et al., 2011) to allow meaningful comparisons between systems and across laboratories. Finally, external, independent validation should be required before exceptionally low error rates or high accuracy scores are extrapolated to real-world casework. Thus, the future of the field of automated firearms identification should look to building algorithms which possess demonstrable value under forensic conditions rather than simply pushing the field forwards to more sophisticated patterns and methods of analysis. By combining cutting-edge three-dimensional imaging, statistical evaluation and machine learning, along with systematic standardization and validation, the field of firearms identification can bring order to chaos and apply scientific rigor to the field.

References

- [1] Baiker-Sørensen, M., Alberink, I., Granell, L. B., van der Ham, L., Mattijssen, E. J. A. T., Smith, E. D., Soons, J., Vergeer, P., & Zheng, X. A. (2023). Automated interpretation of comparison scores for firearm toolmarks on cartridge case primers. *Forensic Science International*, 353, 111858. <https://doi.org/10.1016/j.forsciint.2023.111858>
- [2] De Kinder, J., & Bonfanti, M. (1999). Automated comparisons of bullet striations based on 3D topography. *Forensic Science International*, 101(2), 85–93. [https://doi.org/10.1016/S0379-0738\(98\)00212-6](https://doi.org/10.1016/S0379-0738(98)00212-6)
- [3] De Kinder, J., Tulleners, F., & Thiebaut, H. (2004). Reference ballistic imaging database performance. *Forensic Science International*, 140(2), 207–215. <https://doi.org/10.1016/j.forsciint.2003.12.002>
- [4] Kamalakannan, S., Mann, C. J., Bingham, P. R., Karnowski, T. P., & Gleason, S. S. (2011).

- Automatic firearm class identification from cartridge cases. *Image Processing: Machine Vision Applications* IV, 7877, 207. <https://doi.org/10.1117/12.872414>
- [5] Le Bouthillier, M.-E., Hrynkiw, L., Beauchamp, A., Duong, L., & Ratté, S. (2023). Automated detection of regions of interest in cartridge case images using deep learning. *Journal of Forensic Sciences*, 68(6), 1958–1971. <https://doi.org/10.1111/1556-4029.15319>
- [6] Li, D. (2006). Ballistics Projectile Image Analysis for Firearm Identification. *IEEE Transactions on Image Processing*, 15(10), 2857–2865. <https://doi.org/10.1109/TIP.2006.877530>
- [7] Md Ghani, N. A., Liong, C.-Y., & Jemain, A. A. (2010). Analysis of geometric moments as features for firearm identification. *Forensic Science International*, 198(1–3), 143–149. <https://doi.org/10.1016/j.forsciint.2010.02.011>
- [8] Meral, E., & Akyüz, A. O. (2026). Firearm brand classification using deep learning on cartridge case images. *Forensic Science International*, 378, 112671. <https://doi.org/10.1016/j.forsciint.2025.112671>
- [9] Mookiah, M. R. K., Puch-Solis, R., Farhan, S., Ajala, B., & Daeid, N. N. (2025). Automated segmentation of the breech and firing pin faces of fired cartridge case images. *Forensic Science International*, 375, 112554. <https://doi.org/10.1016/j.forsciint.2025.112554>
- [10] Puente León, F. (2006). Automated comparison of firearm bullets. *Forensic Science International*, 156(1), 40–50. <https://doi.org/10.1016/j.forsciint.2004.12.016>
- [11] Senin, N., Groppetti, R., Garofano, L., Fratini, P., & Pierni, M. (2006). Three-Dimensional Surface Topography Acquisition and Analysis for Firearm Identification. *Journal of Forensic Sciences*, 51(2), 282–295. <https://doi.org/10.1111/j.1556-4029.2006.00048.x>
- [12] Smith, C. L., Robinson, M., & Evans, P. (2000). Line-scan imaging for the positive identification of ballistics specimens. *Proceedings IEEE 34th Annual 2000 International Carnahan Conference on Security Technology (Cat. No.00CH37083)*, 269–275. <https://doi.org/10.1109/CCST.2000.891199>
- [13] Song, J., Vorburger, T. V., Ballou, S., Thompson, R. M., Yen, J., Renegar, T. B., Zheng, A., Silver, R. M., & Ols, M. (2012). The National Ballistics Imaging Comparison (NBIC) project. *Forensic Science International*, 216(1), 168–182. <https://doi.org/10.1016/j.forsciint.2011.09.016>
- [14] Song, J., Vorburger, T. V., Chu, W., Yen, J., Soons, J. A., Ott, D. B., & Zhang, N. F. (2018). Estimating error rates for firearm evidence identifications in forensic science. *Forensic Science International*, 284, 15–32. <https://doi.org/10.1016/j.forsciint.2017.12.013>
- [15] Tong, M., Song, J., & Chu, W. (2015). An Improved Algorithm of Congruent Matching Cells (CMC) Method for Firearm Evidence Identifications. *Journal of Research of the National Institute of Standards and Technology*, 120, 102. <https://doi.org/10.6028/jres.120.008>
- [16] Tong, M., Song, J., Chu, W., & Thompson, R. M. (2014). Fired Cartridge Case Identification Using Optical Images and the Congruent Matching Cells (CMC) Method. *Journal of Research of the National Institute of Standards and Technology*, 119, 575. <https://doi.org/10.6028/jres.119.023>
- [17] Vorburger, T. V., Song, J., & Petraco, N. (2015). Topography measurements and applications in ballistics and tool mark identifications*. *Surface Topography: Metrology and Properties*, 4(1), 013002. <https://doi.org/10.1088/2051-672X/4/1/013002>
- [18] Yammen, S., & Muneesawang, P. (2013). Cartridge case image matching using effective correlation area-based method. *Forensic Science International*, 229(1), 27–42. <https://doi.org/10.1016/j.forsciint.2013.03.015>
- [19] De Kinder, J., Tulleners, F., & Thiebaut, H. (2004). Reference ballistic imaging database performance. *Forensic Science International*, 140(2), 207–215. <https://doi.org/10.1016/j.forsciint.2003.12.002>
- [20] Morris, K. B., Law, E. F., Jefferys, R. L., Dearth, E. C., & Fabyanic, E. B. (2017). An evaluation of the discriminating power of an Integrated Ballistics Identification System® HeritageTMSsystem with the NIST standard cartridge case (Standard Reference Material 2461). *Forensic Science International*, 280, 188–193. <https://doi.org/10.1016/j.forsciint.2017.09.004>
- [21] Song, J., Vorburger, T. V., Ballou, S., Thompson, R. M., Yen, J., Renegar, T. B., Zheng, A., Silver, R. M., & Ols, M. (2012). The National Ballistics Imaging Comparison (NBIC) project. *Forensic Science International*, 216(1), 168–182. <https://doi.org/10.1016/j.forsciint.2011.09.016>
- [22] Song, J., Vorburger, T. V., Chu, W., Yen, J., Soons, J. A., Ott, D. B., & Zhang, N. F. (2018). Estimating error rates for firearm evidence identifications in forensic science. *Forensic Science International*, 284, 15–32. <https://doi.org/10.1016/j.forsciint.2017.12.013>
- [23] Mazumdar, D., Mitra, S., & Kar, A. (2002). [Line-scan/image-processing approach for cartridge-case headstamp recognition]. *Proceedings of the 5th Asian Conference on Computer Vision (ACCV 2002)*.
- [24] Gambino, C., McLaughlin, P., Kuo, L., et al. (2011). [3D metrology/PCA-SVM approach to firearm toolmark discrimination]. *Scanning*, 33(5), 272–278. <https://doi.org/10.1002/sca.20251>
- [25] Petraco, N. D. K., Shenkin, P., Speir, J., et al. (2012). [Statistical classification of firearm/toolmark impressions]. *Journal of Forensic Sciences*, 57(4), 900–911. <https://doi.org/10.1111/j.1556-4029.2012.02115.x>

- [26] Thumwarin, P., Prasit, C., Boonbumroong, P., & Matsuura, T. (2008). *[Rotation-invariant feature approach for cartridge-case/firing-pin image analysis]*. International Conference on Computer and Information Technology / IVCNZ. <https://doi.org/10.1109/IVCNZ.2008.4762085>
- [27] Leng, J., Huang, Z., & Li, D. (2010). *[Geometric-moment approach to firearm identification]*. IEA/AIE 2010, Springer Lecture Notes in Artificial Intelligence, 6098, 653–662. https://doi.org/10.1007/978-3-642-02481-8_88
- [28] Herke, C. (2026). *[CNN and Random Forest approaches for firearm-mark identification]*. Forensic Sciences, 6(3), 69.
- [29] Duez, P., Weller, T., Brubaker, M., Hockensmith, R. E., & Lilien, R. (2018). *Development and validation of a virtual examination tool for firearm forensics*. Journal of Forensic Sciences, 63(4), 1069–1084. <https://doi.org/10.1111/1556-4029.13668>
- [30] Zhang, K., Luo, Y., & Zhou, P. (2017). *[Evofinder comparison of cartridge cases from QSZ-92 pistols]*. Forensic Science International, 278, 78–86. <https://doi.org/10.1016/j.forsciint.2017.06.032>
- [31] Sautier, C., Christen, S., & Chidiac, F. (2021). *[French–Swiss Evofinder network/cross-border ballistic data exchange study]*. Forensic Science International.
- [32] Baruah, S., Suresh, R., Govindarajulu, R. B., Kumar, C. J., Chanda, B., Dugar, L., & Saikia, M. J. (2025). *Feature Extraction and Comparative Analysis of Firing Pin, Breech Face, and Annulus Impressions from Ballistic Cartridge Images*. Forensic Sciences, 5(4), 62. <https://doi.org/10.3390/forensicsci5040062>
- [33] Fischer, R., & Vielhauer, C. (2014). *Digital crime scene analysis: Automatic matching of firing pin impressions on cartridge bottoms using 2D and 3D spatial features*. Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '14), 77–82. <https://doi.org/10.1145/2600918.2600930>
- [34] Fischer, R., & Vielhauer, C. (2015). *Automated firearm identification: On using a novel Multiple-Slice-Shape (MSS) approach for comparison and matching of firing pin impression topography*. Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '15), 161–171. <https://doi.org/10.1145/2756601.2756619>
- [35] Fischer, R., & Vielhauer, C. (2015). *Towards automated firearm identification based on high resolution 3D data: Rotation-invariant features for multiple line-profile-measurement of firing pin shapes*. Proceedings of SPIE, 9393, 93930Q. <https://doi.org/10.1117/12.2077567>
- [36] Changmai, P., Bora, K., Suresh, R., Deb, N., & Mahanta, L. B. (2019). *[Image-processing/KNN approach for bullet-striation classification]*. 31st International Symposium on Ballistics
- [37] Baruah, S., Suresh, R., Govindarajulu, R. B., Kumar, C. J., Chanda, B., Dugar, L., & Saikia, M. J. (2025). *Feature Extraction and Comparative Analysis of Firing Pin, Breech Face, and Annulus Impressions from Ballistic Cartridge Images*. Forensic Sciences, 5(4), 62. <https://doi.org/10.3390/forensicsci5040062>
- [38] Pisantanaroj, P., Tanpisuth, P., Sinchavanwat, P., et al. (2020). *[Smartphone/CNN-based firearm identification from bullet markings]*.
- [39] Chu, W., Tong, M., & Song, J. (2013). *Validation tests for the congruent matching cells (CMC) method using cartridge cases fired with consecutively manufactured pistol slides*. AFTE Journal, 45(4), 361–366.