Received: 3 November 2025 | Revised: 9 January 2026 | Accepted: 6 February 2026

DOI: 10.1111/1556-4029.70308

JOURNAL OF FORENSIC SCIENCES AAFS

ORIGINAL PAPER

Anthropology

Assessing the reliability of saw mark features: A comparative study using stereomicroscopy, micro-XCT, and 3D printing

Pieter D. de Wet MSc[1] | Alison F. Ridel PhD[1] | Alieske C. Hagg MSc[1] | Jakobus W. Hoffman MEng[2] | Ericka N. L'Abbé PhD[1]

[1]Department of Anatomy, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa

[2]South African Nuclear Energy Corporation SOC Ltd, Brits, South Africa

Correspondence
Pieter D. de Wet, Department of Anatomy, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa.
Email: pieter.dewet@tuks.co.za

Funding information
Department of Sport, Arts and Culture, Republic of South Africa

Abstract

Saw mark analysis is an important component of forensic anthropology, particularly in cases of postmortem dismemberment. While stereomicroscopy remains the benchmark for evaluating kerf features, recent advances in imaging technologies offer new opportunities for documenting and interpreting bone trauma. Although these modalities have been studied individually, few investigations have directly compared their relative strengths and limitations. Four fresh human femora were sectioned using six teeth per inch ripsaw to produce 32 false starts and 64 complete cuts. Sixteen kerf features were assessed across stereomicroscopy, micro-focus X-ray computed tomography (micro-XCT), and three-dimensional printed bone models, with scoring for presence, absence, or type, and quantitative measurement where applicable. In addition, a subset of 10 false starts and 10 complete cuts were selected for observer error testing. Minimum kerf width and kerf wall shape showed the highest agreement across observers and modalities, while features such as kerf flare, pull-out striae, and residual kerf wall striae demonstrated lower agreement. Stereomicroscopy provided the highest resolution for fine surface features, whereas micro-XCT enhanced non-destructive visualization of internal kerf morphology and fracture-related traits. Three-dimensional printed models reliably reproduced broader qualitative features but lacked sufficient resolution for fine-scale traits and precise measurements. Significant differences between modalities were observed for several features, emphasizing the influence of visualization method on feature detection and measurement. These findings highlight the need for standardized feature definitions and modality-specific scoring protocols. A multimodal approach of the three modalities is recommended to strengthen the reliability, interpretive value, and forensic applicability of saw mark analysis.

KEYWORDS
3D printing, forensic anthropology, micro-focus X-ray computed tomography, observer error, saw mark analysis, stereomicroscope analysis

Presented at the 12th Conference of the African Society of Forensic Medicine (ASFM), Council for Scientific and Industrial Research (CSIR), March 10–13, 2025, in Pretoria, South Africa.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2026 The Author(s). *Journal of Forensic Sciences* published by Wiley Periodicals LLC on behalf of American Academy of Forensic Sciences.

Highlights

- Observer error evaluation and feature comparison using a single saw across three modalities.
- Minimum kerf width, kerf floor, and kerf wall shapes are most reproducible; micro-XCT shows the best agreement.
- Stereomicroscopy provided the highest resolution for fine kerf features and striae.
- Non-destructive imaging improved detection of spalling and internal kerf traits.
- Bone models 3D printed lacked fine detail for analysis but can be used for teaching purposes.

1 | INTRODUCTION

Saw mark analysis is a critical component of forensic anthropology in the assessment of bone trauma, particularly in cases of postmortem dismemberment [1, 2]. Identifying and interpreting tool marks on bone provides critical information about the class of tool used, contributing to both investigative and legal processes [3]. Despite its importance, the field remains underdeveloped, partly due to a lack of standardized methods and error-rate validation, which reduces confidence in its application in forensic casework [1, 4]. The need for reliable approaches is especially urgent in regions such as South Africa, where high rates of violent crime and dismemberment cases [5–7] place additional pressure on forensic practitioners to extract accurate information from skeletal remains.

Saws cause narrow, slow-loaded forces that produce characteristic kerfs from the beveled edges of saw teeth [8]. Although saw marks are typically easier to distinguish from other forms of bone modification, careful analysis is still required, and recent advances, including statistical modeling and digital imaging, have enhanced their documentation and interpretation [9–12]. imaging modalities such as micro-focus X-ray computed tomography (micro-XCT) and three-dimensional (3D) printing provide new ways of documenting and evaluating kerf features, potentially enhancing both accuracy and courtroom presentation [13–15].

Recent research has demonstrated the potential of advanced imaging methods for improving the reproducibility of saw mark analysis. Pelletti et al. [13] used micro-XCT to analyze 170 false starts in human and pig femora created by four handsaws, showing that striae form not only on kerf walls but also on kerf floors, and concluded that high-resolution imaging supports more reliable documentation and interpretation of features. Norman et al. [14] extended this work by assessing 270 saw marks created with eight different tools, reporting that minimum kerf width (MKW) was the most straightforward feature to evaluate, with a 94% accuracy rate in predicting blade width and high inter-observer agreement. Other features, including kerf wall shape, tooth hop, and residual kerf wall striae, were successfully visualized but demonstrated greater variability in observer scoring. These findings highlight that while some features are highly repeatable, others remain less reliable, emphasizing the importance of testing observer error systematically.

Building on these advances, 3D printing offers a novel application by converting digital scans into tangible physical models. These 3D prints are particularly valuable for teaching and courtroom demonstration, where physical manipulation of evidence can aid comprehension among students, legal practitioners, and jurors [13–17]. Human bone structure is well-suited for additive manufacturing due to its rigidity [18], but challenges remain. For example, Baier et al. [15] demonstrated that trauma features larger than 3 mm are generally well preserved in 3D prints, whereas finer features are frequently lost. Variability in printer resolution and post-processing also introduces inconsistencies, raising questions about their forensic admissibility [14, 16–17]. Nonetheless, 3D prints are increasingly recognized as accessible tools for education and communication in forensic anthropology, even if their evidential reliability remains under debate [15, 16, 19].

The validation of saw mark analysis methods remains incomplete, with limited research examining observer error and the influence of analytical modality on the assessment of saw mark features [11, 14, 20]. This study therefore evaluated saw mark features produced by a single six teeth per inch (TPI) ripsaw on human femora using three analytical modalities: stereomicroscopy, micro-XCT, and 3D printed bone models. Sixteen qualitative and quantitative features were assessed to determine intra- and inter-observer agreement, as well as the effects of modality on feature detection, scoring, and measurement. A subset of false starts and complete cuts was analyzed to quantify observer agreement and error rates, while modality-specific performance was evaluated to identify strengths and limitations across approaches. By integrating observer error testing with cross-modality comparison, this research addresses key gaps in saw mark validation, clarifies the reliability of commonly used features, and evaluates the forensic, educational, and courtroom utility of emerging non-destructive and 3D printed models.

2 | MATERIALS AND METHODS

2.1 | Materials

2.1.1 | Sample and sample preparation

Four fresh human femora were obtained from whole-body donors via the National Tissue Bank of South Africa in compliance with the National Health Act no. 61 of 2003 after obtaining ethical clearance from the Research Ethics Committee of the Faculty of Health Sciences, University of Pretoria (ethical clearance reference no. 241/2023). The femora came from two male donors (41 and 59 years

FIGURE 1 Double bench vice setup to clamp the femur at the proximal (left) and distal (right) ends.

of age) with no evidence of trauma or pathology. The femora were disarticulated, stored frozen, and manually defleshed with a scalpel after thawing.

Each femur was secured in a double bench vice with padded clamps to prevent movement [3, 21] (Figure 1). A new six TPI ripsaw was used to create cut marks along the diaphyses of the bones, with the blade progress from lateral to medial and blade stroke from anterior to posterior. The cuts were performed freehand under steady pressure to prevent the blade from binding, with each kerf completed in one sitting. A single saw type was employed to allow for a large sample size and focused assessment of repeatability, providing statistical power while acknowledging that future studies should expand to multiple saw types. After each cut, the blade was withdrawn in a single upward stroke to simulate pull-out striae [21]. False starts were produced first, often extending into the marrow cavity, and were followed by complete cuts. In total, each femur was sectioned into nine pieces, yielding a total of 32 false starts and 64 complete cuts (Figure 2).

Following sectioning, the samples were boiled at 100°C for 7 h with inspections every 30 min to monitor progress. They were then sterilized in a 1:3 bleach–water solution for 30 min, rinsed, and left to air dry.

2.1.2 | Micro-XCT data acquisition

The 36 bone sections were scanned using a Nikon XT H 225L industrial Computed Tomography system (Nikon Metrology, Belgium) micro-XCT scanner housed at the Micro-Focus X-ray Radiography and Tomography facility (MIXRAD) of the South African Nuclear Energy Corporation SOC, Ltd. (Necsa, Pelindaba, South Africa) [22, 23]. The acquisition and reconstruction parameters were as follows: 120-kV beam voltage, 130-μA beam current, and angular increment of 2500 projections per 360° at one frame per second, with final resolutions ranging between 24.23 and 32.44 μm. Each scan took approximately 41 min to complete and was reconstructed using CT Pro 3D V. XT 4.4.3, with beam hardening and noise reduction set at 2, mild image enhancement, and converted into a single 16-bit tag image file format (TIFF) image stack.

2.2 | Methods

2.2.1 | Cut mark analysis

A standard operating procedure was developed to measure and score features across all three modalities based on criteria available in the literature. Observer error data was collected by two observers. Observer 1 was a master's student in forensic anthropology with prior casework experience and comprehensive familiarity with the relevant literature, having reviewed and analyzed published criteria before data collection. The intra-observer error study was conducted over a 6-week period by Observer 1, 2 weeks after conducting data collection for the main dataset. This ensured consistency in measurement and allowed for a direct assessment of intra-observer reliability. Observer 2, who did the inter-observer error, was a forensic case analyst with a master's degree and several years of professional practice in the examination of skeletal material, providing complementary expertise and strengthening the assessment of observer error.

All 16 saw mark features were first assessed for presence or absence. Present features were either categorized and scored or measured quantitatively (Table 1). False start and complete cut exit chipping and entrance shaving were recorded at entrance, exit, or both. Exit chipping and entrance shaving were also counted on both the entrance and exit sides of each kerf to assess whether

FIGURE 2 The total sample size of 36 bone sections consisting of 32 false start and 64 complete cut kerfs.

feature frequency could help infer cutting direction and whether modality type influenced visibility. MKW was measured at three narrowest points per false start, selecting the smallest; stereomicroscope measurements were at the cortical surface, whereas micro-XCT measurements were taken within the kerf perpendicular to the kerf floor. Bone islands, harmonics, tooth hop, and floor dip were measured peak-to-peak and valley-to-valley, while pull-out striae, tooth trough width, and residual kerf wall striae were measured as minima and maxima to capture variability. Residual kerf wall striae types A and B were measured following the descriptions of Saville et al. [24]. This uniform scoring and measurement scheme was applied to stereomicroscope images, micro-XCT 3D reconstructions, and 3D printed models to allow direct modality comparisons.

2.2.2 | Stereomicroscope analysis

Each cut surface was examined using a Zeiss Axio Zoom.V16 stereomicroscope (Carl Zeiss Microscopy GmbH, Jena, Germany) equipped with a PlanApo Z 0.5×/0.125 FWD 114 mm objective, providing a magnification range of 3.5× to 56× (Figures 3A and 4A). Built-in illumination was supplemented with angled AmScope® dual illumination to enhance visibility. Surfaces were cleaned with isopropyl alcohol before imaging and positioned perpendicular to the objective. Captured images were post-processed for shading correction where needed, and Z-stacking was used to capture curved surfaces for features such as exit chipping and entrance shaving. The 3D printed bone models were analyzed under the same stereomicroscope conditions.

2.2.3 | Image segmentation

Reconstructed micro-XCT scans were converted from 16-bit to 8-bit TIFF stacks in ImageJ [29], with brightness and contrast adjusted for optimal segmentation. The scans were then imported into VG Studio MAX 3.1 (Volume Graphics GmbH, Heidelberg, Germany) and segmented using the Automatic Surface Determination approach, which applied a global threshold based on gray value distribution to separate bone from air [23]. No smoothing was applied to preserve surface detail. Segmented models were exported as standard triangle language (STL) files.

2.2.4 | Micro-XCT analysis

The STL bone models were imported into Avizo 2019.1 (Thermo Fisher Scientific, Waltham, MA, USA) to score and measure false starts and complete cuts (Figures 3B and 4B). Features were assessed using the orthographic view and 3D length tool, mimicking the viewing, placement, and measuring angles that were used for stereomicroscope analysis. For false starts, kerf walls were examined by digitally removing half of the kerf, allowing unobstructed

TABLE 1 Definitions and methods used to analyze each feature.

Feature	Definition	Magnification	Location	Category	Measurement
Minimum kerf width	The smallest measurement of the kerf from kerf wall to kerf wall [1]	3.5× to 4.0×	False starts	Present/absent	Three measurements of which the smallest was used
Blade drift	The directional change in the markings of the kerf floor due to the push and pull strokes of the saw as each tooth enters the material [2]	3.5× to 4.0×	False starts	Present/absent	N/A
Bone islands	The material in the midline of the kerf encompassed by incised bone. As a result of alternating tooth sets [1]	7.5×	False starts	Present/absent	Measurement at the widest point
Pull-out striae	Striae visible on the kerf wall almost perpendicular to the kerf floor or breakaway spur/notch. Created due to the removal of the blade during cutting [1]	3.5× to 10.0×	False starts and complete cuts	Present/absent	Minimum and maximum measurements
Floor dip	A wave-like pattern seen on the kerf floor due to the saw blade jumping when a new tooth enters the material [1]	10.0×	False starts	Present/absent	Peak-to-peak and trough-to-trough measurements
Kerf floor striae	Striae visible on the kerf floor due to tooth incisions [1]	10.0×	False starts	Present/absent Scored as undulating, straight/undulating, or thin and straight [25]	N/A
Kerf flare	The increase in movement when a flexible blade enters a kerf which will widen the kerf floor on the entrance side [1]	7.5×	False starts	Present/absent Scored as present on the entrance or exit sides	N/A
Kerf wall shape	What the saw mark wall alignment looks like when viewed directly above the surface [1]	5.0×	False starts	Present/absent Scored as alternating between wide and narrow aspects, straight, or necking in the middle [26, 27]	N/A
Kerf floor shape	The shape of the kerf floor in cross-section [1]	5.0×	False starts	Present/absent Scored as rounded, flat, stepped, or W-shaped [26, 28]	N/A
Tooth trough width	The distance between teeth when measuring the drift pattern from the widest-to-widest point or narrowest-to-narrowest point [1]	7.5×	False starts and complete cuts	Present/absent	Minimum and maximum measurements
Harmonics	The 3D distance seen between striation peaks in the kerf wall [2]	3.5× to 10.0×	Complete cuts	Present/absent	Peak-to-peak and trough-to-trough measurements
Tooth hop	Wave appearance seen in the kerf wall caused by the saw moving up and down as each tooth enters the material during the saw motion [2]	3.5× to 10.0×	Complete cuts	Present/absent	Peak-to-peak and trough-to-trough measurements
Breakaway spur	The spur found at the terminal of a saw cut [2]	3.5×	Complete cuts	Present/absent	N/A
Breakaway notch	The notch found at the terminal of a saw cut [2]	3.5×	Complete cuts	Present/absent	N/A

(Continues)

TABLE 1 (Continued)

Feature	Definition	Magnification	Location	Category	Measurement
Type A residual kerf wall striae	Striae visible on the kerf wall caused by the teeth on that side [1]. Type A striae are deep furrows created by the passive stroke, representing the amount of material cut per stroke [24]	10.0×	Complete cuts	Present/absent	Minimum and maximum measurements
Type B residual kerf wall striae	Striae visible on the kerf wall caused by the teeth on that side [1]. Type B striae are smaller striae between the deep furrow borders of type A striae representing the amount of material cut by each saw tooth [24]	20.0×	Complete cuts	Present/absent	Minimum and maximum measurements
Exit chipping	Small chips in the margins of a kerf wall caused by saw teeth exiting the bone [2]	7.5× false starts 5.0× complete cuts	False starts and complete cuts	Present/absent Scored as present at the exit, entrance, or both sides	Counted at the exit and entrance sides
Entrance shaving	An eburnation or shaving defect associated with saw teeth entering a kerf which causes the kerf wall margins to blend with the outer bone surface [1]	7.5× false start 5.0× complete cut	False starts and complete cuts	Present/absent Scored as present at the exit, entrance, or both sides	Counted at the entrance and exit sides

visualization of wall features (Figure 5). All features listed in Table 1 were evaluated under this modality.

2.2.5 | 3D printing

The STL models were prepared in Meshmixer™ to repair mesh artifacts and ensure watertightness. Printing was performed at scale with a Formlabs Form 3L stereolithography (SLA) printer using White V4 resin at 0.05 mm layer thickness. The 3D models were oriented to avoid seam lines and support contact with cut surfaces as much as possible. After printing, they were cleaned in isopropyl alcohol, cured at 60°C, and left under indirect sunlight for additional hardening. Support structures were removed with flush cutters. To improve feature visibility under the stereomicroscope, the models were dyed with a mixture of diluted alcohol inks to create a contrasting light-brown/purple hue that provided sufficient contrast and reduced overexposure when imaged under the stereomicroscope (Figures 3C and 4C).

2.3 | Statistical analysis

2.3.1 | Observer error

All statistical analyses were conducted using R software and the RStudio environment (version 2024.09.1) [30]. A sample of 10 complete cuts and 10 false starts representing all features was selected to evaluate the intra-observer (P1 versus P2) and inter-observer (P1 versus A1) error rates across all three modalities. Cohen's Kappa (*irr* package) was used to assess agreement for presence/absence and categorical scores, with Kappa values and *p*-values interpreted to indicate agreement strength. Proportion of agreement was also calculated to describe inter- and intra-rater concordance, particularly for features with limited variation [31, 32]. Measurement agreement was assessed using technical error of measurement (TEM), relative TEM (%TEM) (*base R* package), and intraclass correlation coefficients (ICC) (*irr* package) with 95% confidence intervals [33]. Measurements with an ICC of >0.8 were considered repeatable [34]. Given sub-millimeter differences, %TEM was the primary reliability metric, with ≤5% deemed acceptable [11, 35]. Significant results across tests indicated strong agreement between observers.

2.3.2 | Influence of modality

Stereomicroscope, micro-XCT, and 3D printed model scores and measurements of false starts and complete cuts were compared with evaluate whether modality influenced feature assessment. Descriptive and inferential statistics were applied to summarize the data and test for differences. For categorical features, Chi-squared tests (*stats* package) were used when expected counts

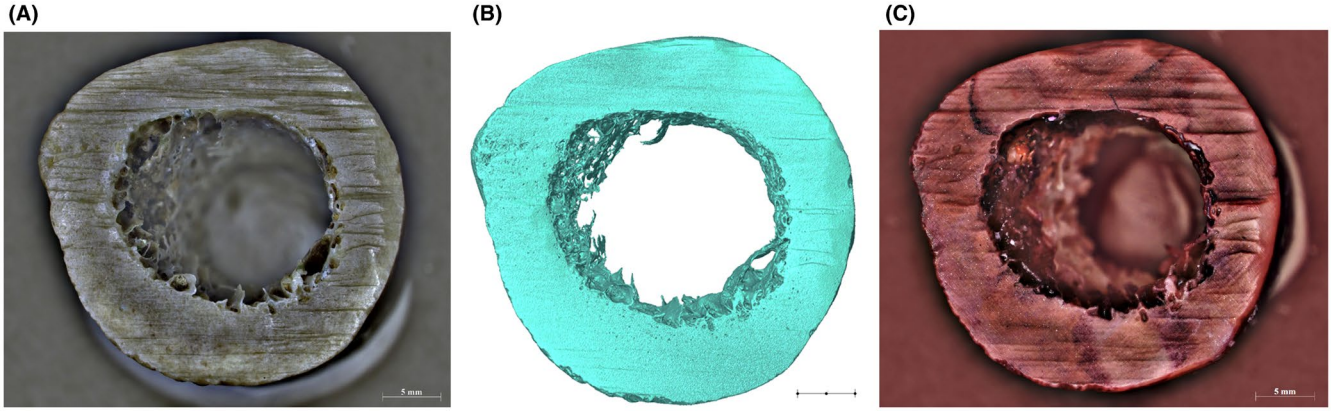

FIGURE 3 The same complete cut surface under stereomicroscope view as the (A) original bone (left), (C) 3D printed model (right), and as a (B) micro-XCT 3D model (middle). Scalebar: 5 mm.

exceeded five, while Fisher's exact tests (*stats* package) were applied to smaller cell counts. For quantitative measurements, normality was assessed visually with Q-Q plots [36]. Normally distributed data were tested using paired Student's *t*-tests (*stats* package), with results reported as mean differences, standard deviations, confidence intervals, and effect sizes (Cohen's *d*) (*effsize* package) [37]. Data not normally distributed were analyzed using Wilcoxon signed-rank tests, with results reported as medians of differences and effect size *r*-values (*rstatix* package). Bland–Altman plots (*blandr* and *ggplot2* packages) were generated for all measurements to visualize agreement and bias between modalities [38]. Exit chipping and entrance shaving counts were specifically tested with Wilcoxon signed-rank tests, given their paired discrete nature [39, 40].

3 | RESULTS

3.1 | Observer error

3.1.1 | Intra-observer error

Across all three modalities, intra-observer agreement for false start scores was generally high (Table 2), with MKW consistently achieving perfect agreement. Blade drift, kerf floor shape, kerf wall shape, and exit chipping also showed near-perfect agreement, while features such as pull-out striae and floor dip were reliably scored as absent. Bone islands, kerf flare, and kerf floor striae type demonstrated more variable agreement, particularly in 3D printed models.

In complete cut scores, harmonics and pull-out striae were consistently absent and thus scored with perfect agreement, while breakaway spurs and notches showed substantial agreement across modalities. Type A residual kerf wall striae exhibited near-perfect agreement, though some variability was evident depending on the modality. Type B residual kerf wall striae also had near-perfect agreement, although because the feature was absent under micro-XCT and 3D print analysis. Entrance shaving and tooth trough width had near-perfect agreement; however, entrance shaving side was less consistent, especially in stereomicroscope and micro-XCT analysis.

Overall, in terms of scores, the features with the least observer error across the three modalities were MKW, kerf wall and floor shape, exit chipping, and pull-out striae. In contrast, features that required finer discrimination, such as kerf flare and entrance shaving, showed lower levels of agreement, with the 3D printed models generally yielding the least consistent results.

Intra-observer measurements showed that MKW was highly repeatable across all three modalities, with low error rates and excellent agreement (Table 3). Bone island measurements had lower error rates with the stereomicroscopy and micro-XCT modalities, although 3D prints showed higher error and poor agreement. Tooth trough width proved more problematic, with stereomicroscopy and micro-XCT yielding moderate agreements, but 3D prints demonstrated poor agreement.

For complete cuts, tooth hop and pull-out striae measurements highlighted differences between modalities. Tooth hop was measurable with stereomicroscopy and micro-XCT, but the feature was not readily observable on the 3D prints. Pull-out striae, when present, were measured with excellent agreement and low error rates under both stereomicroscopy and micro-XCT, whereas 3D prints again showed poor agreement. Residual kerf wall striae presented significant challenges: both stereomicroscopy and micro-XCT demonstrated high variability, and 3D prints had high error rates. Type B residual kerf wall striae had high error rates under stereomicroscopy, and the feature was not observed in the other two modalities. Measurements of tooth trough width on breakaway spurs also varied; stereomicroscopy showed high error rates, while micro-XCT produced more acceptable values, and 3D prints showed poor agreement.

Overall, MKW and pull-out striae under stereomicroscopy and micro-XCT had low error rates, while subtle features such as residual kerf wall striae and tooth trough width had higher error rates, especially in 3D printed models.

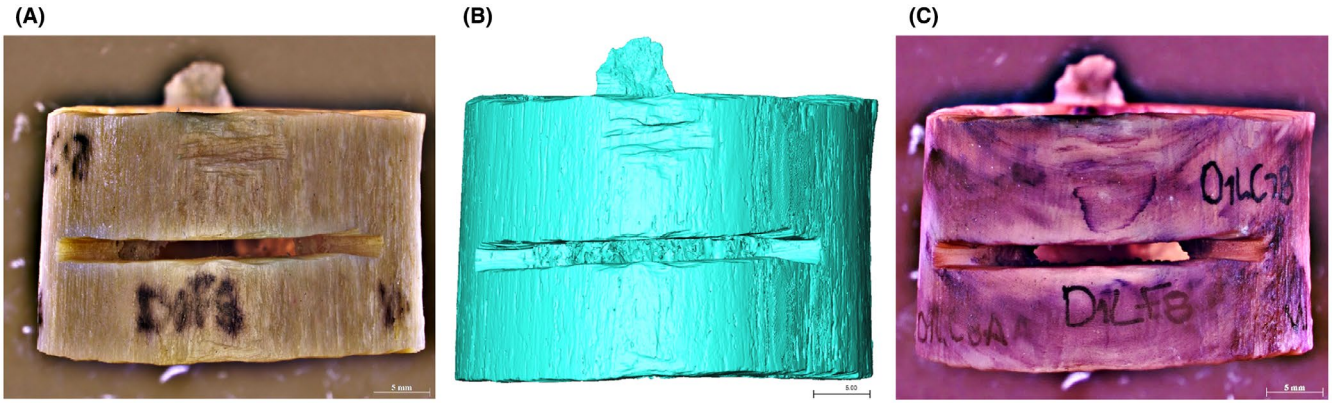

FIGURE 4 Comparative view of the (A) original bone section (left), (B) micro-XCT 3D model (middle), and (C) 3D printed model (right). Scalebar: 5 mm.

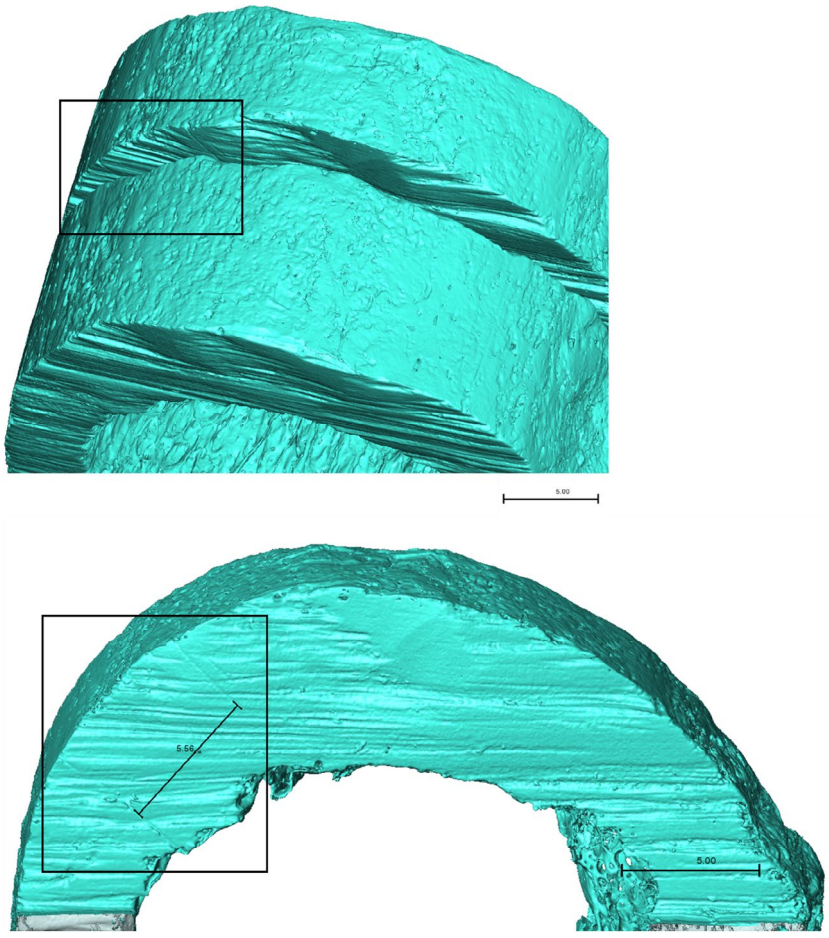

FIGURE 5 An example of pull-out striae measurements (black box) on a micro-XCT scanned 3D model false start kerf wall (top) measurable after temporarily removing the opposite half of the false start kerf (bottom). Scalebar: 5 mm.

3.1.2 | Inter-observer error

Inter-observer agreement for false start scores showed mixed outcomes across the three modalities (Table 4). MKW consistently achieved perfect agreement, along with kerf wall shape, kerf floor shape, and exit chipping. Bone islands and kerf wall shape type demonstrated moderate to substantial agreement, particularly under micro-XCT, whereas features such as kerf flare, floor dip, and entrance shaving side showed poor agreement with all three modalities.

For complete cut scores, observer agreement improved for several features. Breakaway spurs and notches, residual kerf wall striae,

TABLE 2 Reproducibility tests for intra-observer scores of all features on false starts and complete cuts across the three modalities.

	False starts								
	Stereomicroscope			Micro-XCT			3D prints		
Feature	Proportion of agreement	Cohen's kappa	*p*-value	Proportion of agreement	Cohen's kappa	*p*-value	Proportion of agreement	Cohen's kappa	*p*-value
MKW	1	–	–	1	–	–	1	–	–
Blade drift	0.9	–	–	0.9	–	–	1	–	–
Bone islands	0.8	0.58	0.07	0.8	0.60	**0.04**	0.9	**0.78**	**0.01**
Pull-out striae	1	–	–	1	–	–	1	–	–
Floor dip	1	–	–	1	–	–	1	–	–
Kerf floor striae	1	–	–	1	–	–	1	–	–
Kerf floor striae type	0.8	0.11	0.73	-	–	–	–	–	–
Kerf flare entrance	0.5	0.19	0.30	0.8	0.58	0.07	0.7	0.29	0.20
Kerf flare exit	0.5	0.19	0.30	0.9	**0.80**	**0.01**	0.9	**0.78**	**0.01**
Kerf wall shape	1	–	–	1	–	–	1	–	–
Kerf wall shape type	0.9	–	–	0.9	–	–	1	–	–
Kerf floor shape	1	–	–	1	–	–	1	–	–
Kerf floor shape type	0.8	**0.78**	**<0.001**	0.8	**0.63**	**0.03**	0.8	**0.86**	**<0.001**
Tooth trough width	0.9	–	–	1	–	–	0.6	0.09	0.75
Exit chipping	1	–	–	1	–	–	0.9	**0.80**	**0.01**
Exit chipping side	1	–	–	1	–	–	0.7	0.57	**0.03**
Entrance shaving	1	–	–	0.7	0.02	0.60	0.5	0.07	0.78
Entrance shaving side	0.9	–	–	0.4	0.12	0.68	0.3	0.03	0.35
	Complete cuts								
Harmonics	1	–	–	1	–	–	1	–	–
Tooth hop	0.6	0.17	0.60	0.8	0.52	0.10	0.9	–	–
Pull-out striae	1	**1**	**<0.001**	1	**1**	**<0.001**	1	**1**	**<0.001**
Breakaway spur	0.9	**0.78**	**<0.001**	0.9	**0.78**	**0.01**	0.9	**0.78**	**0.01**
Breakaway notch	0.8	0.40	0.20	0.7	**0.62**	**0.04**	1	**1**	**<0.001**
Type A residual kerf wall striae	0.9	–	–	1	–	–	0.8	0.38	0.24
Type B residual kerf wall striae	0.9	**0.62**	**0.04**	1	–	–	1	–	–
Tooth trough width	0.9	**0.78**	**0.01**	1	**1**	**<0.001**	1	–	–
Exit chipping	1	–	–	1	–	–	1	–	–
Exit chipping side	0.7	0.35	0.26	1	–	–	0.7	0.55	**0.04**
Entrance shaving	0.8	0.55	0.05	0.8	0.41	0.11	0.9	**0.78**	**0.01**
Entrance shaving side	0.6	**0.63**	**0.03**	0.6	0.29	0.30	1	**1**	**<0.001**

Note: Bold kappa values: substantial or higher agreement. Bold *p*-values: statistical significance. Dashed lines: no available statistics.

and tooth trough width on breakaway spurs demonstrated high to perfect agreement across modalities, with micro-XCT generally providing the best agreement of the three modalities. Features such as tooth hop, harmonics, and entrance shaving side showed poor agreement. As with intra-observer analysis, the 3D printed models tended to perform poorly for subtle traits, though they showed perfect agreement for macroscopic features such as breakaway spurs and exit chipping.

Overall, the most repeatable features between observers were MKW, kerf wall and floor shape, and breakaway spurs. In contrast, kerf flare, entrance shaving, and tooth hop exhibited poor agreement, emphasizing the difficulty of consistently identifying more subtle or ambiguous traits across the three modalities.

Inter-observer measurement error varied notably across modalities, with few select features meeting the criteria for acceptable error margins (Table 5). MKW was one of the most consistent

TABLE 3 Reproducibility tests for intra-observer measurements of features on false starts and complete cuts across the three modalities.

	False starts								
	Stereomicroscope			Micro-XCT			3D prints		
Feature	%TEM	ICC	*p*-value	%TEM	ICC	*p*-value	%TEM	ICC	*p*-value
MKW	**2.78**	**0.94**	**<0.001**	**2.50**	**0.96**	**<0.001**	**4.26**	**0.91**	**<0.001**
Bone islands	**4.24**	**0.95**	**0.03**	**3.75**	0.35	0.33	6.85	**0.80**	0.10
Pull-out striae minimum	–	–	–	–	–	–	–	–	–
Pull-out striae maximum	–	–	–	–	–	–	–	–	–
Floor dip trough-to-trough	–	–	–	–	–	–	–	–	–
Floor dip peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	**4.48**	0.51	0.07	5.79	0.02	0.74	10.09	0.08	0.56
Tooth trough width maximum	**3.70**	0.64	**0.02**	6.80	0.61	**0.02**	9.96	0.64	0.08
	Complete cuts								
Harmonics trough-to-trough	–	–	–	–	–	–	–	–	–
Harmonics peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth hop	8.92	**0.93**	0.12	**1.73**	–	–	–	–	–
Pull-out striae minimum	**2.99**	**1**	**<0.001**	**3.97**	**0.99**	**0.01**	5.33	–	–
Pull-out striae maximum	**4.67**	**0.99**	**0.03**	**1.12**	**0.99**	**0.02**	8.12	–	–
Type A residual kerf wall striae minimum	30.23	0.42	0.11	34.03	0.03	0.81	39.16	0.04	0.47
Type A Residual kerf wall striae maximum	26.63	0.63	**0.03**	7.14	**0.91**	**<0.001**	32.01	0.44	0.14
Type B Residual kerf wall striae minimum	25.67	**0.78**	**0.01**	–	–	–	–	–	–
Type B Residual kerf wall striae maximum	17.95	**0.85**	**<0.001**	–	–	–	–	–	–
Tooth trough width minimum	11.03	0.08	0.91	**3.87**	0.51	0.24	5.85	–	–
Tooth trough width maximum	5.66	0.48	0.26	**3.99**	0.54	0.23	**3.47**	–	–

Note: Bold values: measurement error of ≤5%, good to excellent ICC value, or significant *p*-value. Dashed lines: no available statistics.
Abbreviations: %TEM, relative technical error measurement; ICC, intraclass correlation coefficient.

features, showing moderate agreement across all three modalities, although error rates were higher than those observed in intra-observer analysis. Bone island measurements had high error rates with all three modalities. Tooth trough width, both minimum and maximum, performed poorly across modalities, with high error rates and little statistical agreement, indicating that this feature is not reliably measurable between observers.

For complete cuts, pull-out striae measurements demonstrated excellent agreement with stereomicroscopy and micro-XCT, while 3D prints showed poor agreement. In contrast, type A and B residual kerf wall striae displayed consistently high error rates and poor inter-observer agreement across modalities. Measurements of tooth trough width on breakaway spurs followed a similar pattern: stereomicroscopy showed poor agreement, while micro-XCT provided acceptable error rates and excellent agreement.

In summary, inter-observer measurement error was acceptable for MKW and pull-out striae under stereomicroscopy and micro-XCT, whereas tooth trough width and residual kerf wall striae were highly variable with poor agreement. The 3D printed models had the highest error rates, particularly for features requiring fine resolution.

3.2 | Frequency of features

Across the 32 false starts, MKW was consistently present in all modalities, seen in Table 6. Blade drift was almost always observed, while bone islands and pull-out striae occurred less frequently, with bone islands most often identified under stereomicroscopy. Floor dip was not observed in any sample. Kerf floor striae were only visible under stereomicroscope analysis. Kerf flare and kerf wall shape were regularly observed, though the side of occurrence of kerf flare varied between modalities. Kerf floor shape was most often scored as W-shaped or stepped in stereomicroscope and micro-XCT analyses, while 3D prints tended to show flat profiles. Exit chipping and entrance shaving were generally more distinct in stereomicroscopy and micro-XCT, though both features were variably identified in 3D prints.

For the 64 complete cuts, harmonics were absent across all modalities, while tooth hop and pull-out striae were more commonly identified under stereomicroscopy and less so in micro-XCT and 3D prints as seen in Table 7. Breakaway spurs and notches were frequently observed in stereomicroscopy and micro-XCT but were less distinct in 3D prints. Type A residual kerf wall striae were consistently identifiable across all modalities, but type B striae were only

TABLE 4 Reproducibility tests for inter-observer scores of all features on false starts and complete cuts across the three modalities.

	False starts								
	Stereomicroscope			Micro-XCT			3D prints		
Feature	Proportion of agreement	Cohen's kappa	*p*-value	Proportion of agreement	Cohen's kappa	*p*-value	Proportion of agreement	Cohen's Kappa	*p*-value
MKW	1	–	–	1	–	–	1	–	–
Blade drift	0.5	0.19	0.39	0.5	–	–	0.3	–	–
Bone islands	0.7	0.40	0.20	0.8	0.60	**0.04**	0.9	**0.78**	**0.01**
Pull-out striae	1	–	–	0.7	–	–	1	–	–
Floor dip	0	–	–	0.5	–	–	1	–	–
Kerf floor striae	1	–	–	0.9	–	–	1	–	–
Kerf floor striae type	0.8	0.09	0.73	-	–	–	–	–	–
Kerf flare entrance	0.6	0.20	0.49	0.6	0.23	0.43	0.6	0.09	0.75
Kerf flare exit	0.4	0.12	0.43	0.3	0.04	0.20	0.8	0.58	0.07
Kerf wall shape	1	–	–	1	–	–	1	–	–
Kerf wall shape type	0.9	**0.62**	**0.04**	0.6	–	–	0.6	–	–
Kerf floor shape	1	–	–	1	–	–	1	–	–
Kerf floor shape type	0.8	**0.78**	**<0.001**	0.9	**0.85**	**<0.001**	0.5	0.43	0.11
Tooth trough width	0.8	0.11	0.73	0.9	–	–	0.5	–	–
Exit chipping	1	–	–	1	–	–	0.5	0.09	0.78
Exit chipping side	1	–	–	1	–	–	0.3	0.15	0.63
Entrance shaving	0.9	–	–	0.8	–	–	0.6	0.29	0.20
Entrance shaving side	0.5	–	–	0.4	0.10	0.73	0.5	0.05	0.73
	Complete cuts								
Harmonics	0.3	–	–	0.6	0	1	0.6	0	1
Tooth hop	0.2	0.05	0.07	0.2	0.03	0.10	1	–	–
Pull-out striae	0.5	0.19	0.30	0.5	0.19	0.30	0.8	0.55	0.05
Breakaway spur	1	**1**	**<0.001**	0.9	**0.78**	**0.01**	1	**1**	**<0.001**
Breakaway notch	0.7	0.40	0.11	0.7	0.29	0.20	0.6	0.29	0.20
Type A residual kerf wall striae	0.9	–	–	0.9	–	–	0.7	0.21	0.49
Type B residual kerf wall striae	0.9	–	–	0.9	–	–	1	–	–
Tooth trough width	1	**1**	**<0.001**	0.9	**0.78**	**0.01**	0.9	–	–
Exit chipping	1	–	–	1	–	–	1	–	–
Exit chipping side	0.6	0	1	1	–	–	0.7	0.29	0.20
Entrance shaving	0.6	0	1	0.7	–	–	0.7	0.35	0.26
Entrance shaving side	0.3	0.04	**0.01**	0.7	0.12	0.62	0.6	0.24	0.42

Note: Bold kappa values: substantial or higher agreement. Bold *p*-values: statistical significance. Dashed lines: no available statistics.

TABLE 5 Reproducibility tests for intra-observer measurements of features on false starts and complete cuts across the three modalities.

	False starts								
	Stereomicroscope			Micro-XCT			3D prints		
Feature	%TEM	ICC	*p*-value	%TEM	ICC	*p*-value	%TEM	ICC	*p*-value
MKW	12.44	0.65	**0.01**	6.30	0.72	**0.01**	6.32	0.73	**0.01**
Bone islands	5.91	0.65	0.17	10.59	**0.75**	0.13	17.61	0.03	0.68
Pull-out striae minimum	–	–	–	–	–	–	–	–	–
Pull-out striae maximum	–	–	–	–	–	–	–	–	–
Floor dip trough-to-trough	–	–	–	–	–	–	–	–	–
Floor dip peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	8.25	0	0.50	7.46	0.16	0.33	13.21	0.17	0.64
Tooth trough width maximum	5.46	0.25	0.24	8.59	0.29	0.21	11.38	0.22	0.32
	Complete cuts								
Harmonics trough-to-trough	–	–	–	–	–	–	–	–	–
Harmonics peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth hop	**0.66**	–	–	**4.00**	–	–	–	–	–
Pull-out striae minimum	**0.49**	**1**	**<0.001**	**2.29**	**0.99**	**0.03**	8.94	–	–
Pull-out striae maximum	5.74	**0.99**	**0.05**	**1.02**	**1**	**<0.001**	**3.83**	–	–
Type A residual kerf wall striae minimum	38.02	0.09	0.40	19.20	0.46	0.08	46.38	0.03	0.77
Type A Residual kerf wall striae maximum	32.59	0.07	0.43	35.33	0.08	0.40	48.39	0.03	0.77
Type B Residual kerf wall striae minimum	51.08	0.01	0.51	–	–	–	–	–	–
Type B Residual kerf wall striae maximum	27.51	0.03	0.47	–	–	–	–	–	–
Tooth trough width minimum	6.45	0.04	0.72	**4.94**	0.40	0.30	–	–	–
Tooth trough width maximum	7.01	0.09	0.54	**2.27**	**0.96**	**0.02**	–	–	–

Note: Bold values: measurement error of ≤5%, good to excellent ICC value, or significant *p*-value. Dashed lines: no available statistics.
Abbreviations: %TEM, relative technical error measurement; ICC, intraclass correlation coefficient.

visible under stereomicroscopy. Exit chipping and entrance shaving were most reliably detected under stereomicroscopy and micro-XCT analysis, whereas 3D prints often failed to capture these features.

In summary, larger features such as MKW, kerf wall and floor shape, and exit chipping were reliably present across all modalities, while more subtle features, including kerf floor striae, tooth hop, bone islands, and type B residual kerf wall striae, were mainly observed in stereomicroscope analysis.

3.3 | Influence of modality

3.3.1 | Stereomicroscope versus micro-XCT scores

For false starts, MKW, pull-out striae, kerf flare, kerf wall shape, kerf floor shape, tooth trough width, exit chipping, and exit chipping side did not differ significantly between the stereomicroscope and micro-XCT modalities (Table 8). In contrast, blade drift, bone islands, kerf wall shape type, kerf floor shape type, and entrance shaving showed statistically significant differences between the two modalities. Kerf floor striae were only scored under the stereomicroscope, preventing comparison of this feature and its subtype.

For complete cuts, the type of modality significantly influenced tooth hop, pull-out striae, breakaway spur, breakaway notch, type A residual kerf wall striae, tooth trough width on breakaway spurs, exit chipping, entrance shaving, and entrance shaving side (Table 8). Type B residual kerf wall striae and exit chipping side did not differ significantly between modalities, even though type B residual kerf wall striae were not present under micro-XCT analysis.

3.3.2 | Stereomicroscope versus micro-XCT measurements

False start measurements showed mixed results (Table 9). MKW did not differ significantly between modalities, with stereomicroscope values slightly larger than micro-XCT. Tooth trough width (minimum and maximum) differed significantly, with stereomicroscope values consistently smaller than micro-XCT. Exit chipping counts on the exit side showed no significant difference, while

TABLE 6 Frequency of false start features on all three modalities.

		Stereomicroscope		Micro-XCT		3D prints	
Feature	**Feature criteria**	**Count**	**Percentage**	**Count**	**Percentage**	**Count**	**Percentage**
MKW	Present	32/32	100%	32/32	100%	32/32	100%
Blade drift	Present	31/32	96.88%	32/32	100%	31/32	69.88%
Bone islands	Present	6/32	18.75%	5/32	15.63%	5/32	15.63%
Pull-out striae	Present	2/32	6.25%	3/32	9.38%	0/32	0%
Floor dip	Present	0/32	0%	0/32	0%	0/32	0%
Kerf floor striae	Undulating	1/32	3.13%	0/32	0%	0/32	0%
	Straight/undulating	31/32	96.88%	0/32	0%	0/32	0%
	Thin and straight	0/32	0%	0/32	0%	0/32	0%
Kerf flare	Entrance	2/32	6.25%	8/32	25%	10/32	31.25%
	Exit	6/32	18.75%	16/32	50%	4/32	12.50%
	Both	23/32	71.88%	5/32	15.63%	3/32	9.38%
	None	1/32	3.13%	3/32	9.38%	15/32	46.88%
Kerf wall shape	Alternating	31/32	96.88%	32/32	100%	31/32	96.88%
	Straight	1/32	3.13%	0/32	0%	1/32	3.13%
	Necking	0/32	0%	0/32	0%	0/32	0%
Kerf floor shape	Round	0/32	0%	0/32	0%	0/32	0%
	Flat	4/32	12.50%	9/32	28.13%	16/32	50%
	Stepped	12/32	37.50%	11/32	34.38%	9/32	28.13%
	W-shaped	16/32	50%	12/32	37.50%	6/32	18.75%
Tooth trough width	Present	30/32	93.75%	31/32	96.88%	22/32	68.75%
Exit chipping	Exit	1/32	3.13%	3/32	9.38%	12/32	37.50%
	Entrance	0/32	0%	0/32	0%	2/32	6.25%
	Both	31/32	96.88%	29/32	90.63%	11/32	34.38%
	None	0/32	0%	0/32	0%	7/32	21.88%
Entrance shaving	Entrance	2/32	6.25%	15/32	46.88%	5/32	15.63%
	Exit	0/32	0%	2/32	6.25%	3/32	9.38%
	Both	30/32	93.75%	13/32	40.63%	9/32	28.13%
	None	0/32	0%	2/32	6.25%	15/32	46.88%

entrance counts and entrance shaving counts (both entrance and exit) were significantly higher under stereomicroscopy, though effect sizes indicated small practical differences. Sample sizes for bone islands and pull-out striae were insufficient for robust statistical testing, and floor dip measurements were absent in both modalities.

For complete cuts, type A residual kerf wall striae minimum values showed no significant difference, while maximum values were significantly smaller under stereomicroscopy (Table 9). Type B residual kerf wall striae measurements were only obtainable under stereomicroscopy. Exit chipping counts (both entrance and exit) were significantly higher in number under micro-XCT, with large effect sizes indicating practical differences. Entrance shaving counts on the exit side also differed significantly between the two modalities, whereas entrance side counts did not. Harmonics were absent, and sample sizes for tooth hop, pull-out striae, and tooth trough width on breakaway spurs were too small for meaningful analyses.

3.3.3 | Stereomicroscope versus 3D print scores

For false starts, most features showed no significant differences between stereomicroscopy and 3D prints (Table 10). Bone islands and pull-out striae differed significantly, while MKW, blade drift, kerf flare, kerf wall and floor shape, and tooth trough width showed no significant differences. Kerf floor striae and striae type could only be scored under stereomicroscopy, preventing a direct comparison to 3D printed models. Exit chipping scores differed significantly, though exit chipping side and entrance shaving (score and side) did not.

For complete cuts, several features were significantly influenced by modality (Table 10). Tooth hop, pull-out striae, breakaway spur, breakaway notch, and tooth trough width on breakaway spurs differed significantly between the two modalities. In contrast, type A and B residual kerf wall striae showed no significant differences, even though type B striae were only present under

TABLE 7 Frequency of complete cut features on all three modalities.

		Stereomicroscope		Micro-XCT		3D prints	
Feature	Feature criteria	Count	Percentage	Count	Percentage	Count	Percentage
Harmonics	Present	0/64	0%	0/64	0%	0/64	0%
Tooth hop	Present	13/64	20.31%	9/64	14.06%	0/64	0%
Pull-out striae	Present	8/64	12.50%	3/64	4.69%	2/64	3.13%
Breakaway spur/breakaway notch	Breakaway spur	27/64	42.19%	30/64	46.88%	31/64	48.44%
	Breakaway notch	22/64	34.38%	26/64	40.63%	14/64	21.88%
	Both	6/64	9.38%	3/64	4.69%	3/64	4.69%
	None	9/64	14.06%	5/64	7.81%	16/64	25%
Type A residual kerf wall striae	Present	57/64	89.06%	64/64	100%	57/64	89.06%
Type B residual kerf wall striae	Present	38/64	59.38%	0/64	0%	0/64	0%
Tooth trough width	Present	8/64	12.50%	5/64	7.81%	2/64	3.13%
Exit chipping	Exit	19/64	29.69%	2/64	3.13%	25/64	39.06%
	Entrance	0/64	0%	0/64	0%	0/64	0%
	Both	43/64	67.19%	62/64	96.88%	38/64	59.38%
	None	2/64	3.13%	0/64	0%	1/64	1.56%
Entrance shaving	Entrance	32/64	50%	18/64	28.13%	12/64	18.75%
	Exit	0/64	0%	4/64	6.25%	0/64	0%
	Both	13/64	20.31%	30/64	46.88%	3/64	4.69%
	None	19/64	29.69%	12/64	18.75%	49/64	76.56%

stereomicroscope analysis. Exit chipping scores did not differ significantly, although exit chipping side did. Entrance shaving and entrance shaving side also showed no significant differences between the two modalities.

3.3.4 | Stereomicroscope versus 3D print measurements

False start measurements showed mixed outcomes (Table 11). MKW differed significantly between modalities, with stereomicroscope measurements consistently larger and showing a substantial effect size. In contrast, minimum and maximum tooth trough width measurements showed no significant difference. Exit chipping (both entrance and exit counts) and entrance shaving (entrance counts) all differed significantly, with stereomicroscope counts generally higher, though the effect sizes ranged from moderate to large. Entrance shaving counts on the side did not differ significantly. Sample sizes for bone islands, pull-out striae, and floor dip were too small for meaningful analysis.

For complete cuts, most features could not be robustly compared due to limited sample sizes (pull-out striae, tooth hop, and tooth trough width on breakaway spurs). Type A residual kerf wall striae minimum measurements differed significantly, with stereomicroscope values smaller than 3D print measurements, whereas maximum measurements showed no significant difference. Exit chipping counts (both exit and entrance), as well as entrance shaving counts on the entrance side, differed significantly, although the effect sizes suggested only moderate to limited practical differences.

Overall, stereomicroscope and 3D print analysis showed poor comparability for several measurements, particularly MKW, exit and entrance exit chipping counts, and entrance shaving entrance counts. However, tooth trough width and several kerf striae measures remained consistent across modalities, suggesting that while finer details are captured more reliably under stereomicroscope analysis, 3D prints can provide broadly comparable results for less complex features.

3.3.5 | Micro-XCT versus 3D print scores

Most false start scores showed no significant difference between micro-XCT and 3D print modalities (Table 12). MKW, floor dip, kerf floor striae, and tooth trough width were consistently comparable. Significant differences were observed for blade drift, bone islands, and pull-out striae. While kerf wall and floor shapes were consistently present across both modalities, both shape types differed significantly between the two modalities. Exit chipping was significantly influenced by modality, though exit chipping side, kerf flare, entrance shaving, and entrance shaving side showed no significant difference.

For complete cuts, several features were significantly influenced by modality. Tooth hop, pull-out striae, breakaway spur, breakaway notch, type A residual kerf wall striae, tooth trough width on breakaway spurs, and exit chipping all differed significantly between micro-XCT and 3D prints. Harmonics and type B residual kerf wall

TABLE 8 Chi-squared and Fisher's exact tests comparing false start and complete cut scores for features between stereomicroscope and micro-XCT.

	False starts			Complete cuts			
Feature	χ^2	χ^2 *p*-value	Fisher's *p*-value	**Feature**	χ^2	χ^2 *p*-value	Fisher's *p*-value
MKW	0	1	–	**Harmonics**	0	1	–
Blade drift	28.13	**<0.001**	–	**Tooth hop**	–	–	**0.01**
Bone islands	–	–	**<0.001**	**Pull-out striae**	–	–	**<0.001**
Pull-out striae	–	–	0.18	**Breakaway spur**	33.04	**<0.001**	–
Floor dip	0	1	–	**Breakaway notch**	29.96	**<0.001**	–
Kerf floor striae	∞	–	–	**Type A residual kerf wall striae**	39.06	**<0.001**	–
Kerf floor striae type	–	–	–	**Type B residual kerf wall striae**	2.25	0.13	–
Kerf flare entrance	–	–	0.20	**Tooth trough width**	–	–	**<0.001**
Kerf flare exit	–	–	1	**Exit chipping**	56.25	**<0.001**	–
Kerf wall shape	0	1	–	**Exit chipping side**	–	–	0.09
Kerf wall shape type	28.13	**<0.001**	–	**Entrance shaving**	–	–	**<0.001**
Kerf floor shape	0	1	–	**Entrance shaving side**	–	–	**<0.001**
Kerf floor shape type	–	–	**<0.001**				
Tooth trough width	–	–	1				
Exit chipping	0	1	–				
Exit chipping side	–	–	0.09				
Entrance shaving	24.50	**<0.001**	–				
Entrance shaving side	–	–	1				

Note: Bold *p*-values: statistical significance. Dashed lines: no available statistics. χ^2: Chi-squared statistic.

striae were absent in both modalities, and entrance shaving (and side) showed no significant difference. Exit chipping side was also not significantly influenced.

Several features were comparable across micro-XCT and 3D prints; however, many complete cut scores and select false start features showed significant differences, highlighting substantial variability between the modalities.

3.3.6 | Micro-XCT versus 3D print measurements

False start measurements showed mixed outcomes (Table 13). MKW differed significantly between modalities, with micro-XCT values larger on average and a moderate effect size, indicating systematic measurement discrepancies. Tooth trough width (minimum and maximum) also differed significantly, with micro-XCT yielding consistently larger measurements and large effect sizes, highlighting practical discrepancies. Exit chipping counts (entrance and exit) were significantly higher for micro-XCT, though effect sizes ranged from small to moderate. Entrance shaving counts on the entrance side were also significantly higher under micro-XCT, albeit with a small effect size, with no significant difference for the exit side. Bone islands, pull-out striae, and floor dip could not be robustly compared due to insufficient data.

For complete cuts, most measurements also differed significantly between modalities. Type A residual kerf wall striae minimum values were significantly smaller under micro-XCT, with a small-to-moderate effect size, whereas maximum values were not significantly different, suggesting general consistency. Exit chipping counts (entrance and exit) were markedly higher for micro-XCT, with large effect sizes supporting strong practical differences. Entrance shaving counts on the entrance side were also significantly higher in micro-XCT, though with moderate effect size. Harmonics, type B residual kerf wall striae, and entrance shaving counts on the exit side could not be tested due to limited data, while tooth hop, pull-out striae, and tooth trough width on breakaway spurs were measurable only in micro-XCT, precluding comparison.

False start measurements such as MKW and tooth trough width, and complete cut measurements including type A residual kerf wall striae minimum, exit chipping, and entrance shaving, differed significantly between modalities, with micro-XCT generally yielding larger or higher counts. While some features showed limited practical differences, others demonstrated substantial divergence, emphasizing important variability between the two modalities. Features with inadequate data remain inconclusive.

4 | DISCUSSION

This study evaluated observer agreement and modality-related effects for 16 saw mark features produced by a six TPI ripsaw on human

TABLE 9 Paired Student's *t*-tests and Wilcoxon signed-rank tests comparing false start and complete cut measurements for features between stereomicroscope and micro-XCT.

	False starts								
	Paired Student's *t*-test						Wilcoxon signed-rank test		
Feature	*p*-value	CI lower	CI upper	SD	M	Cohen's d	*p*-value	MOD	*r*
MKW	0.10	−0.01	0.10	0.16	0.05	0.30	–	–	–
Bone islands	–	–	–	–	–	–	–	–	–
Pull-out striae minimum	–	–	–	–	–	–	–	–	–
Pull-out striae maximum	–	–	–	–	–	–	–	–	–
Floor dip trough-to-trough	–	–	–	–	–	–	–	–	–
Floor dip peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	**<0.001**	−0.10	−0.05	0.07	−0.07	−1.06	–	–	–
Tooth trough width maximum	**<0.001**	−0.14	−0.09	0.07	−0.11	−1.68	–	–	–
Exit chipping count exit side	–	–	–	–	–	–	0.46	0	0.15
Exit chipping count entrance side	–	–	–	–	–	–	**0.02**	2	0.21
Entrance shaving count entrance side	–	–	–	–	–	–	**<0.001**	1	0.22
Entrance shaving count exit side	–	–	–	–	–	–	**<0.001**	1	0.20
	Complete cuts								
Harmonics trough-to-trough	–	–	–	–	–	–	–	–	–
Harmonics peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth hop	–	–	–	–	–	–	0.13	−0.24	0.91
Pull-out striae minimum	–	–	–	–	–	–	0.18	−0.45	0.93
Pull-out striae maximum	–	–	–	–	–	–	1	−0.06	0
Type A residual kerf wall striae minimum	0.97	−0.25	0.24	0.92	0.00	−0.01	–	–	–
Type A residual kerf wall striae maximum	**0.01**	−0.67	−0.11	1.07	−0.39	−0.37	–	–	–
Type B residual kerf wall striae minimum	–	–	–	–	–	–	–	–	–
Type B residual kerf wall striae maximum	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	–	–	–	–	–	–	–	–	–
Tooth trough width maximum	–	–	–	–	–	–	–	–	–
Exit chipping count exit side	–	–	–	–	–	–	**<0.001**	−3	0.66
Exit chipping count entrance side	–	–	–	–	–	–	**<0.001**	−3	0.76
Entrance shaving count entrance side	–	–	–	–	–	–	0.14	1	0.17
Entrance shaving count exit side	–	–	–	–	–	–	**<0.001**	0	0.78

Note: Bold *p*-values: statistical significance. Dashed lines: no available statistics.

Abbreviations: CI, confidence interval; M, mean; MOD, median of differences; *r*, effect size; SD, standard deviation.

femora, assessed using stereomicroscopy, micro-XCT, and 3D printed bone models. Although limited to a single saw type, the findings provide an important foundation for understanding which features can be reliably observed, scored, and measured, and which remain problematic due to observer subjectivity or technical constraints. Taken together, the results demonstrate that observer agreement and feature visibility are strongly influenced by feature definition, measurement strategy, and modality-specific limitations, reinforcing the need for standardized definitions, measurement protocols, and modality-specific considerations in forensic saw mark analysis.

MKW emerged as one of the most repeatable features, showing consistently high intra- and inter-observer agreement across all three modalities. However, subtle but meaningful differences were observed between observers and modalities, reflecting the influence of measurement strategy. Under stereomicroscopy, MKW was identified visually at the cortical surface, whereas micro-XCT

TABLE 10 Chi-squared and Fisher's exact tests comparing false start and complete cut scores for features between stereomicroscope and 3D prints.

	False starts			Complete cuts			
Feature	χ^2	χ^2 *p*-value	Fisher's *p*-value	**Feature**	χ^2	χ^2 *p*-value	Fisher's *p*-value
MKW	0	1	1	**Harmonics**	0	1	1
Blade drift	–	–	1	**Tooth hop**	22.56	**<0.001**	–
Bone islands	–	–	**0.03**	**Pull-out striae**	–	–	**0.01**
Pull-out striae	24.5	**<0.001**	–	**Breakaway spur**	35.99	**<0.001**	–
Floor dip	0	1	1	**Breakaway notch**	21.16	**<0.001**	–
Kerf floor striae	–	–	–	**Type A residual kerf wall striae**	–	–	0.57
Kerf floor striae type	–	–	–	**Type B residual kerf wall striae**	2.25	0.13	–
Kerf flare entrance	–	–	0.67	**Tooth trough width**	–	–	**0.01**
Kerf flare exit	–	–	1	**Exit chipping**	–	–	1
Kerf wall shape	0	1	1	**Exit chipping side**	–	–	**<0.001**
Kerf wall shape type	–	–	1	**Entrance shaving**	–	–	0.20
Kerf floor shape	0	1	1	**Entrance shaving side**	–	–	0.35
Kerf floor shape type	–	–	0.41				
Tooth trough width	–	–	1				
Exit chipping	10.13	**0.001**	–				
Exit chipping side	–	–	1				
Entrance shaving	0.13	0.72	–				
Entrance shaving side	–	–	0.18				

Note: Bold *p*-values: statistical significance. Dashed lines: no available statistics. χ^2: Chi-squared statistic.

measurements captured the narrowest point within the kerf, effectively functioning as an internal caliper measurement. These methodological differences likely account for small discrepancies and underscore the importance of standardized measurement approaches. As recommended by VanBaarle and Garvin [11], taking multiple measurements along the kerf and selecting the minimum value may improve both accuracy and reproducibility, particularly in forensic contexts where methodological transparency and repeatability are essential for admissibility.

Kerf wall shape and kerf floor shape also demonstrated high observer agreement, particularly for intra-observer assessments. These features benefit from relatively clear morphological boundaries and categorical definitions, reducing subjective interpretation. However, variability increased in 3D printed models, where subtle transitions were often lost during segmentation, mesh repair, and printing. Inter-observer agreement also decreased in some instances, especially when assessed on 3D printed models, where subtle morphological transitions were more difficult to visualize. Despite this limitation, the consistency observed across stereomicroscopy and micro-XCT suggests that these features remain robust indicators when appropriate modalities are used. The reproducibility of these traits across modalities is encouraging, as these features are frequently relied upon in tool mark analysis.

In contrast, residual kerf wall striae (types A and B) consistently exhibited poor observer agreement and high error rates. This finding aligns with previous research demonstrating the inherent variability of intra-kerf striae, both within and between cuts [41]. These challenges are compounded by modality limitations, particularly the inability of micro-XCT, at the resolution used in this study, to reliably capture fine striae and the loss of surface texture during 3D model conversion and printing. Variability in striae expression, coupled with ambiguity in selecting representative measurement locations, makes standardization challenging. Differences in observer interpretation of striae prominence and continuity further contribute to low inter-observer agreement. Without substantial refinement of definitions or digitally assisted measurement protocols, residual kerf wall striae should be interpreted with caution.

Other features, including bone islands, tooth trough width, exit chipping, breakaway spurs, and breakaway notches, were generally recognizable across modalities, though inter-observer measurement error frequently exceeded intra-observer error. This pattern indicates that while these features are identifiable, their boundaries are often indistinct, leading to measurement variability. Tooth trough width measurements on breakaway spurs, for example, showed improved agreement under micro-XCT, likely due to enhanced visualization of kerf floor morphology and fracture boundaries. By contrast, stereomicroscope and 3D print assessments were more susceptible to variability, illustrating how modality resolution and visualization techniques directly influence observer agreement.

TABLE 11 Paired Student's *t*-tests and Wilcoxon signed-rank tests comparing false start and complete cut measurements for features between stereomicroscope and 3D prints.

	False starts								
	Paired Student's *t*-test						Wilcoxon signed-rank test		
Feature	*p*-value	CI lower	CI upper	SD	M	Cohen's d	*p*-value	MOD	r
MKW	**<0.001**	0.06	0.15	0.13	0.11	0.80	–	–	–
Bone islands	–	–	–	–	–	–	0.25	−0.17	0.93
Pull-out striae minimum	–	–	–	–	–	–	–	–	–
Pull-out striae maximum	–	–	–	–	–	–	–	–	–
Floor dip trough-to-trough	–	–	–	–	–	–	–	–	–
Floor dip peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	0.51	−0.03	0.07	0.11	0.02	0.15	–	–	–
Tooth trough width maximum	0.51	−0.02	0.05	0.08	0.01	0.15	–	–	–
Exit chipping count exit side	–	–	–	–	–	–	**<0.001**	3	0.58
Exit chipping count entrance side	–	–	–	–	–	–	**<0.001**	4	0.62
Entrance shaving count entrance side	–	–	–	–	–	–	**<0.001**	3	0.53
Entrance shaving count exit side	–	–	–	–	–	–	0.24	1	0.11
	Complete cuts								
Harmonics trough-to-trough	–	–	–	–	–	–	–	–	–
Harmonics peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth hop	–	–	–	–	–	–	–	–	–
Pull-out striae minimum	–	–	–	–	–	–	–	–	–
Pull-out striae maximum	–	–	–	–	–	–	–	–	–
Type A residual kerf wall striae minimum	**0.02**	−0.63	−0.05	1.03	−0.34	−0.33	–	–	–
Type A residual kerf wall striae maximum	0.13	−0.56	0.07	1.13	−0.25	−0.22	–	–	–
Type B residual kerf wall striae minimum	–	–	–	–	–	–	–	–	–
Type B residual kerf wall striae maximum	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	–	–	–	–	–	–	–	–	–
Tooth trough width maximum	–	–	–	–	–	–	–	–	–
Exit chipping count exit side	–	–	–	–	–	–	**<0.001**	2	0.34
Exit chipping count entrance side	–	–	–	–	–	–	**0.04**	0	0.30
Entrance shaving count entrance side	–	–	–	–	–	–	**0.001**	4	0.09
Entrance shaving count exit side	–	–	–	–	–	–	1	0	0.78

Note: Bold *p*-values: statistical significance. Dashed lines: no available statistics.

Abbreviations: CI, confidence interval; M, mean; MOD, median of differences; *r*, effect size; SD, standard deviation.

Several features demonstrated only moderate agreement, including kerf flare and pull-out striae. Pull-out striae were infrequently observed, consistent with previous studies [20, 21], and their sporadic occurrence suggests that lateral blade movement during withdrawal, rather than blade wear, may be a primary contributing factor. Kerf flare inconsistency likely reflects the absence of precise scoring criteria, as the feature's appearance can be subtle and dependent on viewing angle and illumination. Reduced visibility from stereomicroscopy to micro-XCT and 3D printing further highlights modality-dependent detection limits.

Entrance shaving and exit chipping were consistently identified but difficult to assign reliably to entrance, exit, or both sides of the kerf. Analysis under the micro-XCT modality captured these features more frequently than stereomicroscopy, particularly exit chipping, likely due to its superior ability to visualize fracture morphology and spalling. However, as reported previously [21], neither feature provided a reliable indicator of cutting direction, emphasizing the need for caution when interpreting these traits in forensic contexts.

Observer subjectivity was especially evident in features such as blade drift, floor dip, and harmonics. While intra-observer agreement

TABLE 12 Chi-squared and Fisher's exact tests comparing false start and complete cut scores for features between micro-XCT and 3D prints.

	False starts			Complete cuts			
Feature	χ^2	χ^2 *p*-value	Fisher's *p*-value	**Feature**	χ^2	χ^2 *p*-value	Fisher's *p*-value
MKW	0	1	1	**Harmonics**	0	1	1
Blade drift	28.13	**<0.001**	–	**Tooth hop**	33.06	**<0.001**	–
Bone islands	–	–	**0.001**	**Pull-out striae**	–	–	**0.001**
Pull-out striae	21.13	**<0.001**	–	**Breakaway spur**	35.99	**<0.001**	–
Floor dip	0	1	1	**Breakaway notch**	14.93	**<0.001**	–
Kerf floor striae	0	1	1	**Type A residual kerf wall striae**	39.06	**<0.001**	–
Kerf floor striae type	–	–	–	**Type B residual kerf wall striae**	0	1	1
Kerf flare entrance	2.64	0.10	–	**Tooth trough width**	–	–	**0.005**
Kerf flare exit	–	–	0.66	**Exit chipping**	60.06	**<0.001**	–
Kerf wall shape	0	1	1	**Exit chipping side**	–	–	0.18
Kerf wall shape type	28.13	**<0.001**	–	**Entrance shaving**	–	–	1
Kerf floor shape	0	1	1	**Entrance shaving side**	–	–	0.21
Kerf floor shape type	–	–	**0.02**				
Tooth trough width	–	–	0.31				
Exit chipping	10.13	**0.001**	–				
Exit chipping side	–	–	0.81				
Entrance shaving	–	–	1				
Entrance shaving side	–	–	0.81				

Note: Bold *p*-values: statistical significance. Dashed lines: no available statistics. χ^2: Chi-squared statistic.

was often good to excellent, inter-observer agreement ranged from moderate to poor, suggesting that individual observers applied internally consistent but divergent interpretation thresholds. In forensic practice, such discrepancies highlight the risk of over- or under-interpreting features, particularly when they are rare, faint, or easily confused. Tooth hop is a good example: although the few measurable cases suggested high observer agreement, the rarity of its occurrence meant the overall agreement was uncertain and results must be interpreted with caution.

Each analytical modality contributed distinct strengths and limitations. Stereomicroscopy remains the benchmark for saw mark analysis, offering high-resolution visualization of surface features, including kerf floor striae and fine residual kerf wall striae [13, 14, 24]. The micro-XCT modality provided a non-destructive, 3D visualization of kerf morphology, excelling in the assessment of macroscopic fracture-related features such as breakaway spurs, notches, and exit chipping. However, the resolution used in this study (24-32 µm) was insufficient to reliably capture finer striae, illustrating the trade-off between resolution, scan time, specimen size, and model processing [42]. Targeted high-resolution scans of specific kerf regions may offer a practical compromise for future research.

The 3D printed models offered tangible, manipulable replicas of bone cuts, which proved valuable for teaching, training, and courtroom demonstration [15, 16]. Larger, well-defined features such as MKW and kerf wall shape were consistently identifiable, supporting the use of 3D prints as complementary tools. Nevertheless, finer and measurable features, including kerf floor striae, pull-out striae, and tooth hop, were poorly reproduced, consistent with prior findings that features smaller than approximately 3 mm are inadequately captured through current printing workflows [15]. Quantitative discrepancies, particularly for MKW, further limit the suitability of 3D prints for precise measurement, reinforcing their role as explanatory rather than primary analytical tools.

Several methodological limitations should be acknowledged. The use of a single saw limits generalization across saw classes and tooth configurations, although it provided a controlled framework for assessing observer agreement and modality effects. Maceration protocols may influence surface preservation, and future studies should consider enzymatic detergents in combination with simmering to improve cleaning while minimizing damage [43]. Advances in micro-XCT resolution, segmentation workflows, and 3D printing technology are likely to improve feature visibility and measurement fidelity. Larger-scale prints may further enhance the value of 3D models for educational and courtroom applications, even if one-to-one replicas remain unsuitable for quantitative forensic analysis.

Future research should expand analyses to include multiple saw types, tooth configurations, and kerf depths to better understand how feature expressions vary across cutting conditions. Integrating quantitative classification methods, such as decision trees and random forest models, may further strengthen the objectivity of saw class estimation and reduce reliance on subjective feature interpretation.

TABLE 13 Paired Student's *t*-tests and Wilcoxon signed-rank tests comparing false start and complete cut measurements for features between micro-XCT and 3D prints.

	False starts								
	Paired Student's *t*-test						Wilcoxon signed-rank test		
Feature	*p*-value	CI lower	CI upper	SD	M	Cohen's d	*p*-value	MOD	r
MKW	**0.03**	0.01	0.11	0.14	0.06	0.42	–	–	–
Bone islands	–	–	–	–	–	–	0.13	−0.14	0.91
Pull-out striae minimum	–	–	–	–	–	–	–	–	–
Pull-out striae maximum	–	–	–	–	–	–	–	–	–
Floor dip trough-to-trough	–	–	–	–	–	–	–	–	–
Floor dip peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	**<0.001**	0.04	0.13	0.09	0.08	0.92	–	–	–
Tooth trough width maximum	**<0.001**	0.09	0.15	0.07	0.12	1.73	–	–	–
Exit chipping count exit side	–	–	–	–	–	–	**<0.001**	3	0.23
Exit chipping count entrance side	–	–	–	–	–	–	**<0.001**	2	0.45
Entrance shaving count entrance side	–	–	–	–	–	–	**0.01**	1	0.18
Entrance shaving count exit side	–	–	–	–	–	–	0.40	0	0.53
	Complete cuts								
Harmonics trough-to-trough	–	–	–	–	–	–	–	–	–
Harmonics peak-to-peak	–	–	–	–	–	–	–	–	–
Tooth hop	–	–	–	–	–	–	–	–	–
Pull-out striae minimum	–	–	–	–	–	–	–	–	–
Pull-out striae maximum	–	–	–	–	–	–	–	–	–
Type A residual kerf wall striae minimum	**0.004**	−0.61	−0.12	0.92	−0.37	−0.40	–	–	–
Type A residual kerf wall striae A maximum	1	−0.10	0.37	0.90	0.14	0.15	–	–	–
Type B residual kerf wall striae minimum	–	–	–	–	–	–	–	–	–
Type B residual kerf wall striae maximum	–	–	–	–	–	–	–	–	–
Tooth trough width minimum	–	–	–	–	–	–	–	–	–
Tooth trough width maximum	–	–	–	–	–	–	–	–	–
Exit chipping count exit side	–	–	–	–	–	–	**<0.001**	6	0.76
Exit chipping count entrance side	–	–	–	–	–	–	**<0.001**	3	0.43
Entrance shaving count entrance side	–	–	–	–	–	–	**0.001**	5	0.33
Entrance shaving count exit side	–	–	–	–	–	–	–	–	–

Note: Bold *p*-values: statistical significance. Dashed lines: no available statistics.

Abbreviations: CI, confidence interval; M, mean; MOD, median of differences; *r*, effect size; SD, standard deviation.

Based on observed intra- and inter-observer agreement, several recommendations can be made. Residual kerf wall striae (types A and B) and kerf flare showed consistently low agreement and are not recommended as standalone diagnostic features and should rather just be scored as kerf wall striae to indicate direction of blade stroke. Pull-out striae, entrance shaving, floor dip, and harmonics showed high intra- but low inter-observer agreement and should be restricted to research or specialist contexts unless clearer definitions and standardized training are implemented. Blade drift should be interpreted only in conjunction with kerf wall shape. The remaining features demonstrated good to excellent agreement for the saw class tested and are appropriate for forensic use, keeping limitations in mind; though validation across additional saw types, tooth configurations, and observer experience levels is required.

It is also important to contextualize the relative forensic value of these characteristics. Features such as MKW, kerf wall and floor shape, kerf floor striae, floor dip, tooth trough width, and tooth hop

are critical for inferring saw type and blade characteristics, making them central to tool identification in dismemberment cases. Complimentary features, like breakaway spurs and notches, exit chipping, entrance shaving, and kerf wall striae provide information on direction and progress of the saw, indicating how the tool was used. By contrast, features such as bone islands, blade drift, kerf flare, or harmonics provide limited or inconsistent information about tool use and therefore have less evidentiary value. Documenting observer agreement alongside forensic relevance helps distinguish which traits should be prioritized in casework and which should be treated as supplementary or excluded altogether.

5 | CONCLUSION

This study evaluated observer agreement for 16 saw mark features produced by a six TPI ripsaw on human femora using stereomicroscopy, micro-XCT, and 3D printed models. The results demonstrate that observer agreement is strongly feature-dependent and influenced by both modality-specific limitations and subjective interpretation. Features with clear morphological boundaries and objective measurement criteria, such as MKW, kerf wall shape, kerf floor shape, and exit chipping, showed the highest and most consistent agreement across observers and modalities. In contrast, features with ambiguous definitions or subtle expression were associated with lower reproducibility and higher error rates.

Residual kerf wall striae and kerf flare consistently exhibited poor intra- and inter-observer agreement and should not be used as standalone diagnostic features in forensic casework. Pull-out striae, entrance shaving, floor dip, and harmonics demonstrated good intra-observer but weak inter-observer agreement, indicating that these traits should be restricted to research or specialist applications unless supported by explicit scoring criteria, rater training, or corroborating imaging. Blade drift should be interpreted only in conjunction with kerf wall shape, while the remaining features are appropriate for routine scoring for the tested saw class and tooth size, pending validation across additional saw types and larger rater samples.

Modality comparisons highlight that stereomicroscopy remains the benchmark for high-resolution assessment of surface features, while micro-XCT offers a valuable non-destructive complement for visualizing internal kerf morphology and fracture-related traits. 3D printed models, although limited in reproducing fine-scale features and precise measurements, provide practical value for education, training, and courtroom communication. These findings emphasize that modalities should not be treated interchangeably and that feature scoring protocols must be modality-specific.

Overall, observer agreement must be a prerequisite for forensic admissibility. The development of standardized feature definitions, reference images, and measurement protocols is essential to reduce subjectivity and improve reproducibility. Features that cannot be consistently scored across observers and modalities should be excluded from forensic interpretation. By integrating observer agreement testing with modality-aware protocols and broader validation across saw types, saw mark analysis can progress toward a more robust, reliable, and legally defensible forensic discipline.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Marius Loots from the Department of Anatomy, Faculty of Health Sciences, University of Pretoria, for assisting with the 3D printers, and the National Tissue Bank, Pretoria, South Africa, for providing the femora. The 3D printer and associated equipment utilized in this study was acquired through the *Dirisana* + grant (Erasmus+ project no. 618489-EPP-1-2020-1-ZA-EPPKA2-CBHE-JP). This publication reflects the views of only the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

FUNDING INFORMATION

This research was partially funded by the Department of Sport, Arts, and Culture, Republic of South Africa.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Pieter D. de Wet https://orcid.org/0000-0002-0055-2193
Ericka N. L'Abbé https://orcid.org/0000-0002-6722-8814

REFERENCES

1. Symes SA, Chapman EN, Rainwater CW, Cabo LL, Myster SM. Knife and saw toolmark analysis in bone: a manual designed for the examination of criminal mutilation and dismemberment. Washington, DC: National Institute of Justice; 2010.
2. Love JC. Sharp force trauma analysis in bone and cartilage: a literature review. Forensic Sci Int. 2019;299:119–27. https://doi.org/10.1016/j.forsciint.2019.03.035
3. Menschel M, Pokines JT, Reinecke G. Correlation between saw blade width and kerf width. J Forensic Sci. 2021;66(1):25–43. https://doi.org/10.1111/1556-4029.14556
4. Grosso AR. Tissue variability effects on saw mark evidence in bone [dissertation]. Pittsburgh, PA: University of Pittsburgh; 2020.
5. South African Police Service annual crime statistics 2023/2024. Available from: https://www.saps.gov.za/services/crimestats.php. Accessed 7 Dec 2024
6. Labuschagne G. Features and investigative implications of muti murder in South Africa. J Investig Psychol Offender Profil. 2004;1(3):191–206. https://doi.org/10.1002/jip.15
7. Steyn M, Brits D. Dismemberment in South Africa: case studies. In: Ross AH, Cunha E, editors. Dismemberments: perspectives in forensic anthropology and legal medicine. Cambridge, MA: Academic Press; 2019. p. 69–83.
8. Scientific Working Group for Forensic Anthropology (SWGANTH). Trauma analysis. Available from: https://www.nist.gov/system/

files/documents/2018/03/13/swganth_trauma.pdf. Accessed 18 May 2022
9. Love JC, Derrick SM, Wiersema JM, Peters C. Microscopic saw mark analysis: an empirical approach. J Forensic Sci. 2015;60(s1):S21–S26. https://doi.org/10.1111/1556-4029.12650
10. Bailey JA, Wang Y, van de Goot FRW, Gerretsen RRR. Statistical analysis of kerf mark measurements in bone. Forensic Sci Med Pathol. 2011;7(1):53–62. https://doi.org/10.1007/s12024-010-9185-6
11. VanBaarle AL, Garvin HM. Investigating error in saw mark minimum kerf width measurements. J Forensic Sci. 2023;68(1):257–66. https://doi.org/10.1111/1556-4029.15178
12. Dupouy DLM, Bolton MS, Berry TP, Raymond J, Meakin GE. Saw marks in bone: a preliminary empirical study to inform decision making and best practice. Forensic Sci Int. 2023;353:111857. https://doi.org/10.1016/j.forsciint.2023.111857
13. Pelletti G, Viel G, Fais P, Viero A, Visentin S, Miotto D, et al. Micro-computed tomography of false starts produced on bone by different hand-saws. Leg Med. 2017;26:1–5. https://doi.org/10.1016/j.legalmed.2017.01.009
14. Norman DG, Baier W, Watson DG, Burnett B, Painter M, Williams MA. Micro-CT for saw mark analysis on human bone. Forensic Sci Int. 2018b;293:91–100. https://doi.org/10.1016/j.forsciint.2018.10.027
15. Baier W, Norman DG, Donnelly MJ, Williams MA. Forensic 3D printing from micro-CT for court use- process validation. Forensic Sci Int. 2021;318:110560. https://doi.org/10.1016/j.forsciint.2020.110560
16. Blau S, Phillips E, O'Donnell C, Markowsky G. Evaluating the impact of different formats in the presentation of trauma evidence in court: a pilot study. Aust J Forensic Sci. 2019;51(6):695–704. https://doi.org/10.1080/00450618.2018.1457717
17. Salazar D, Thompson M, Rosen A, Zuniga J. Using 3D printing to improve student education of complex anatomy: a systematic review and meta-analysis. Med Sci Educ. 2022;32(5):1209–18. https://doi.org/10.1007/s40670-022-01595-w
18. Lai J, Wang M. Developments of additive manufacturing and 5D printing in tissue engineering. J Mater Res. 2023;38(21):4692–725. https://doi.org/10.1557/s43578-023-01193-5
19. Carew RM, Morgan RM, Rando C. A preliminary investigation into the accuracy of 3D modeling and 3D printing in forensic anthropology evidence reconstruction. J Forensic Sci. 2019;64(2):342–52. https://doi.org/10.1111/1556-4029.13917
20. Love JC, Derrick SM, Wiersema JM. Independent validation test of microscopic saw mark analysis. Washington, DC: National Institution of Justice; 2013.
21. Martlin B, Rando C. An assessment of the reliability of cut surface characteristics to distinguish between hand-powered reciprocating saw blades in cases of experimental dismemberment. J Forensic Sci. 2021;66(2):444–55. https://doi.org/10.1111/1556-4029.14628
22. Hoffman JW, de Beer FC. Characteristics of the micro-focus X-ray tomography system at the MIXRAD facility at NECSA in South Africa. Proceedings of the 18th world conference on nondestructive testing; 2012 Apr 16–20; Durban, South Africa. Red Hook, NY: Curran Associates, Inc.; 2013.
23. Theye CEG. The effects of aging and tooth loss on the microstructure of the mandible in south Africans [dissertation]. Pretoria, South Africa: University of Pretoria; 2022.
24. Saville PA, Hainsworth SV, Rutty GN. Cutting crime: the analysis of the "uniqueness" of saw marks on bone. Int J Leg Med. 2007;121(5):349–57. https://doi.org/10.1007/s00414-006-0120-z
25. Nogueira L, Alunni V, Bernardi C, Quatrehomme G. Saw marks in bones: a study of "secondary features" of false start lesions. Forensic Sci Int. 2018;290:157–61. https://doi.org/10.1016/j.forsciint.2018.06.023
26. Nogueira L, Quatrehomme G, Rallon C, Adalian P, Alunni V. Saw marks in bones: a study of 170 experimental false start lesions. Forensic Sci Int. 2016;268:123–30. https://doi.org/10.1016/j.forsciint.2016.09.018
27. Bernardi C, Nogueira L, Cabusat-Mailliet C, Carle G, Alunni V, Quatrehomme G. Analysis of false starts lesions on human bones produced by two hand saws with high TPI. Int J Leg Med. 2020;134(2):613–8. https://doi.org/10.1007/s00414-020-02251-x
28. Symes SA. Morphology of saw marks in human bone: identification of class characteristics [dissertation]. Knoxville, TN: The University of Tennessee; 1992.
29. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5. https://doi.org/10.1038/nmeth.2089
30. R Core Team. R: a language and environment for statistical computing [computer program]. Version 2025.09.2; 2025. Vienna, Austria: R Foundation for Statistical Computing; 2013.
31. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med. 2012;22(3):276–82.
32. Waugh SM, He J. Inter-rater agreement estimates for data with high prevalence of a single response. J Nurs Meas. 2019;27(2):152–61. https://doi.org/10.1891/1061-3749.27.2.152
33. MedCalc® Statistical Software, Schoonjans F [computer program]. Intraclass correlation coefficient. Ostend, Belgium: MedCalc; 2020.
34. Liljequist D, Elfving B, Skavberg-Roaldsen K. Intraclass correlation – a discussion and demonstration of basic features. PLoS One. 2019;14(7):e0219854. https://doi.org/10.1371/journal.pone.0219854
35. Franklin D, Cardini A, Flavel A, Kuliukas A, Marks MK, Hart R, et al. Concordance of traditional osteometric and volume-rendered MSCT interlandmark cranial measurements. Int J Leg Med. 2013;127(2):505–20. https://doi.org/10.1007/s00414-012-0772-9
36. Scrucca L. Assessing multivariate normality through interactive dynamic graphics. Quad Stat. 2000;2:221–40.
37. Almeid L, Barros N, Ferrari-Piloni C, Torres EM, Estrela C, Valladares-Neto J. Effect size: a statistical basis for clinical practice. Rev Odonto Cienc. 2019;33(1):84–90. https://doi.org/10.15448/1980-6523.2018.1.29437
38. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–60. https://doi.org/10.1177/096228029900800204
39. Ahad NA, Yahaya SSS, MdYusof Z, Abdullah S, Fung LY. Performance of the modified Wilcoxon signed rank test. AIP Conf Proc. 2014;1635(1):690–4. https://doi.org/10.1063/1.4903657
40. Hawkins DM. Tests of difference: two related samples. In: Hawkins DM, editor. Biomeasurement: a student's guide to biological statistics. Oxford, UK: Oxford University Press; 2019. p. 166–88.
41. Greer S. Quantifying variation in kerf wall striations created by hacksaws and reciprocating saws. Forensic Anthropol. 2018;1(3):160–9. https://doi.org/10.5744/fa.2018.0017
42. Yu H, He F, Pan Y. A novel segmentation model for medical images with intensity inhomogeneity based on adaptive perturbation. Multimed Tools Appl. 2019;78(9):11779–98. https://doi.org/10.1007/s11042-018-6735-5
43. Keyes CA, Giltrow KR, Mahon T-J. A comparison of maceration methods for the preparation of infant skeletal remains for forensic anthropological analysis. Int J Leg Med. 2024;138(3):1085–92. https://doi.org/10.1007/s00414-023-03137-4

How to cite this article: de Wet PD, Ridel AF, Hagg AC, Hoffman JW, L'Abbé EN. Assessing the reliability of saw mark features: A comparative study using stereomicroscopy, micro-XCT, and 3D printing. J Forensic Sci. 2026;00:1–22. https://doi.org/10.1111/1556-4029.70308