Steiger, GS and Borrini, M

A proposal for cut marks classification using machine learning: Serrated vs. non-serrated, single vs. double-beveled knives

Article

Received: 19 April 2024 | Revised: 4 June 2024 | Accepted: 3 July 2024

DOI: 10.1111/1556-4029.15588

JOURNAL OF FORENSIC SCIENCES AAFS

ORIGINAL PAPER

Anthropology

A proposal for cut marks classification using machine learning: Serrated vs. non-serrated, single vs. double-beveled knives

Giada Sciâdi Steiger MSc | Matteo Borrini PhD

School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

Correspondence
Matteo Borrini, School of Biological and Environmental Sciences—Liverpool John Moores University, James Parsons Building Byrom Street, Liverpool, L3 3AF, UK.
Email: m.borrini@ljmu.ac.uk

Abstract

In tool mark identification, there is still a lack of characteristics and methodologies standardization used to analyze and describe sharp force trauma marks on skeletal remains. This study presents a classification method for cut marks on human bones, providing an applicable methodology for their examination and the relevant terminology for describing cases of sharp force trauma. A total of 350 cut marks were produced by stabbing pig ribs (*Sus scrofa*) with seven knives. The samples were analyzed under a stereomicroscope with a tangential light source. Through the analysis of cut marks, eleven traits were identified as significantly associated with the type of knife used. These traits included the general morphology of the kerf shape, the entrance and exit cross-profile shapes, the location of the rising on the entrance and exit cross-profile, the presence or absence of feathering, the presence or absence of shards and the location and the general morphology of the mounding. Binary logistic regression models were later trained and tested using nine out of the eleven traits. The first model categorized the cut mark as either produced by a serrated or non-serrated blade, while the second, as either produced by a single- or double-beveled blade. Classification scores of those models ranged between 63%–85% for the serration class and 63%–89% for the blade bevel class. This study proposes a new set of traits and the use of machine learning models to standardize and facilitate the analysis of stab wounds.

KEYWORDS
artificial intelligence, cut mark analysis, forensic anthropology, forensic pathology, machine learning, sharp force trauma, trauma analysis

Highlights

- This study provides a set of traits and their reliability for the classification of cut marks.
- Single rising and Y cross-profile showed a high association with single-beveled knives.
- Shards and feathering were exclusively observed in cut marks produced by serrated knives.
- A logistic regression method has been provided to correctly classify serrated, non-serrated, single- and double-beveled blades.
- Machine learning minimizes the necessity of having an experienced observer if multiple variables are involved.

Presented at the 75th Anniversary Scientific Conference of the American Academy of Forensic Sciences, February 13–18, 2023, in Orlando, FL.

1 | INTRODUCTION

In England and Wales during the last four years, the homicide rate was between 10 and 15 per million, showing an increase between 2001 and 2013. Overall, the homicide rates ranged between 5 and 15 per million in the last 60 years. In terms of homicide methods, Home Office police-recorded statistics reported in 2020 that 35% of homicides in the last 40 years were committed using sharp implements, mostly knives [1]. The recorded numbers suggest an increasing need for knowledge of the sharp force trauma dynamics. Sharp force injuries can be divided into stab, incised, and chop wounds. They are produced by different types of tools, and their identification and classification depend on the different patterns left on the bodies. In the present study, the authors considered and analyzed uniquely stab injuries produced by knives. Stab injuries are produced by the implement penetrating the body and leaving a characteristic deep and narrow cut on the bones. Previous studies in knife trauma analysis have mainly focused on determining the class to which the cuts belong, such as serrated and non-serrated knives [2]. In contrast, only a few authors started to focus on individual characteristics that are unique and distinct for specific implements [3–5]. Metric variables, mostly the width, depth, and length of the kerfs, have been investigated by several studies. However, the results showed that these characteristics are highly influenced by different variables, such as the force or direction of the stabbing or are the result of the composition of the stabbed material [6, 7]. Further research identified numerous morphological variables, focusing on reporting the observable traits, their frequencies within the tool class, and the technique used for the analysis [8–16]. Nevertheless, applying these results when required is highly problematic due to differences in labeling, sometimes lack of unambiguous descriptions, differences in the surfaced analyzed (bones or cartilage) and the overall limited presence of validation studies for knife stab injuries [6, 16–22].

In 2012, Love et al. [17] used the classification tree method to categorize costal cartilage cut marks. Parameters linked to striations, such as presence, type, and size, were used to classify three knives. The results showed low classification rates, and the reliability of the observations was not addressed. In 2013, Crowder et al. [18] analyzed error rates when comparing the knives, the substances they cut into, and how they observed the traits -directly or indirectly- using three different experienced observers. The authors achieved higher classification rates than Love et al. [17]. Furthermore, they highlighted the importance of the observer experience in the analysis of stab injuries produced by knives [18].

In recent years, Ghui et al. [16] developed a flowchart to classify serrated from non-serrated knives. The study also yielded high classification rates and introduced the trait "shards" to distinguish between macro and micro-serrated knives. However, similarly to Love et al.'s method [17], the reliability of the traits' observations was not reported.

One key insight from Crowder et al.'s [18] study is the significant role of the observer's experience in classifying the knife, particularly when multiple characteristics are recorded. To address this limitation, the authors of the present contribution propose using a machine learning algorithm as a novel approach that can potentially overcome the subjectivity of human observation and provide a more objective classification of knife stab wounds.

According to Arthur Samuel [23], machine learning (ML) is defined as the field of study that gives computers the ability to learn without being explicitly programmed. It is mainly used to handle large amounts of data more efficiently and to resolve tasks that the human mind is unable to solve directly [24, 25]. These tasks are the procedures that a machine learning system can use to process a dataset: a collection of features. Different algorithms could be used to process the dataset, and they are categorized by what kind of dataset they can use as training datasets [24]. *Unsupervised algorithms* process a dataset that is not labeled, and their main goal is to learn about the intrinsic properties of the dataset. Clustering algorithms or principal component analysis (PCA) are examples of *unsupervised algorithms*. On the opposite side, *supervised algorithms* process variables associated with a label or a target, usually provided by the human hand, and they learn to associate the inputs x with the outputs y. Two algorithms help classify or predict a value for *supervised learning*: classification models with discrete values as outputs and regression models with continuous values as outputs. Logistic regression is an example of a *supervised learning* algorithm and can be used to predict a categorical dependent variable using a dataset of independent variables. It can classify an observation into two (binary) or more categories (multinomial). For a binary logistic regression, the goal is to train a classifier to give probabilistic values that lie between 0 and 1 or any other dichotomous category, e.g., "Pass" or "Fail", "True" or "False" [26]. The classification of the probabilistic value in one of the two categories is performed using a decision function (threshold) [27, 28]. The binary logistic regression algorithm is well-suited for solving knife classification problems, as the outcome gives the knife's category to which the cut mark belongs (e.g., serrated) using <0.5 or >0.5 as the decision threshold.

Therefore, the present paper aims to address the problems that have emerged during the last few years by proposing a methodology for classifying knife cut marks. It will be accomplished by achieving the following objectives:

1. determining and describing the traits practical to identify knife classes.
2. establishing the use of an accessible technique using binary logistic regression that reduces the necessity of more experienced researchers when more traits are considered.
3. establish the classification accuracies of such a technique.
4. measure the reliability of the observations (inter- and intra- observer error analysis).

2 | MATERIALS AND METHODS

2.1 | Impact medium and implements

Pig ribs (*Sus scrofa*) purchased from a local store were chosen for this study as it has been shown that the chest and the back are the most common sites for stabbing knife trauma [29]. The ribs were defleshed, and any grease and dirt residue were removed with dishwasher soap and water.

Seven knives were selected to create cut marks in the ribs, each with slightly different characteristics to have heterogeneity in the sample (see Table 1). They were newly purchased to avoid a possible production of artifacts on the cut marks from broken tips of previously used knives. Some knives were purchased from the local supermarket, while most single-beveled knives were purchased via amazon.co.uk and anythinglefthanded.co.uk websites. Three different types of grinds were selected, chisel, saber, and full flat. All the grinds besides the chisel grind have a bevel edge (see Figure 1), hence the sharpness on both sides. All the grinds with this sharpness were called double-beveled knives instead of single-beveled ones. The right side of the blades was defined as the side towards the right of the person handling the knife while having the point aimed away from them.

280 cut marks (40 per each of the knives) were produced with the knives' blades placed perpendicular to the superior surface of the bones. The cut marks were made using one forward and one backward motion to mimic a stabbing motion into the chest. An attempt was made to keep the force and movement range constant throughout the marks' production. The action was, however, operator-dependent: for consistency, the same operator performed all the cuts. A right-handed person produced the cut marks, and the direction of the movement was marked with an arrow, making the entrance and the exit extremities identifiable. Additionally, a separate validation dataset consisting of 70 cut marks (10 cuts per knife) was later produced by the same person with the same modalities as the previous dataset.

2.2 | Microscopic analysis

The first author (Observer 1) performed microscopic analysis on the training and validation datasets. From the training dataset, Observer 1 collected the categories of traits that were used as input variables to train the logistic regression algorithms. From the second, the categories collected were used to evaluate the models' performances. The analysis of the cut surface was conducted using a Zeiss Stemi DV4 Stereo microscope with an 8× -32× zoom magnification range, and the analyst was allowed to use the full range for each cut mark. A tangential white light source was used in addition to the stereomicroscope's light to exhibit the cut mark characteristics more clearly. Figure 2 depicts the labels based on Capuani et al. [5] and Tennick [8] that have adopted by the authors in the results section of the present study. The variables and their categories were determined with a combination of different approaches: some of them were previously chosen from the literature and then observed in the data collection, some of them were recorded and later identified in the literature based on the characteristics presented during the data collection process while some variables were only observed during the data collection.

The kerf and cross-profile shapes were two variables identified during the literature review. The kerf shape, which is the cut produced by the knife, was first recorded by Lewis in 2008 [9]. Capuani and colleagues [8] determined that serrated knives leave

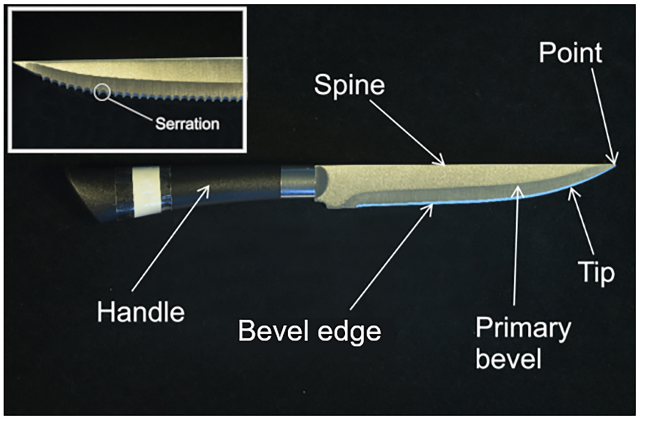

FIGURE 1 Anatomy of a knife and a blade: Close up on a serrated blade (Top left).

TABLE 1 Knives used in the study.

Knife n°	Type of knife	Type of serration	Type of bevel
1	Sushi knife Chisel-beveled	Non-serrated	Single-beveled Bevel edge on the right side
2	Paring knife Chisel-beveled	Non-serrated	Single-beveled Bevel edge on the right side
3	Paring knife Saber grind	Non-serrated	Double-beveled
4	Utility knife Full flat grind	Non-serrated	Double-beveled
5	Meat knife Chisel-beveled	Serrated	Single-beveled Bevel edge on the left side
6	Meat/vegetable knife Chisel-beveled	Serrated	Single-beveled Bevel edge on the right side
7	Meat knife Chisel-beveled	Serrated	Single-beveled Bevel edge on the right side

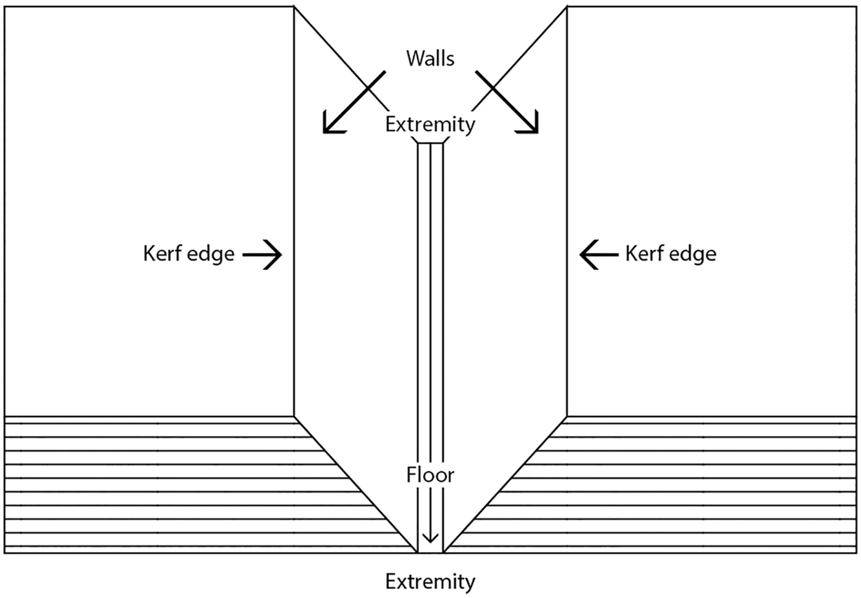

FIGURE 2 Anatomy of a kerf.

a wider kerf than a narrower mark from non-serrated knives while Vachirawongsakorn et al. [10] and Ghui et al. [16] described specific shapes: linear, ellipse, rectangular, trapezoid, B-shaped, crescent-shaped and irregular. Regarding the shape cross-profile, different authors have frequently described the characteristic V shape produced by knives [10–12, 16, 19]. However, more studies identified a difference between the cross-profile left by serrated and non-serrated knives. In 2012, Tegtmeyer [12] identified that non-serrated knives leave a V cross-profile, while serrated knives leave a Y cross-profile, characterized by a vertical wall corresponding to the side with the serration. Although many studies reported these traits as distinctive for serrated and non-serrated knives, this study hypothesized that a non-serrated single-beveled knife could also produce a Y shape.

The bone formation on the kerf's edges, produced by the passage of the blade on the bone, was considered a trait in this study. Several studies have documented the trait, though labeled with different names: edge ridging, mounding, kerf margin, and edge mound [4, 5, 10, 11]. A specific distinction in the terminology used to describe this bone formation was made by Capuani et al. [8]. When the ribs were tilted by 90°, and the extremities were observed, the variable rising was recorded. The differentiation between the trait mounding and the rising was considered in this study, and it was hypothesized that the side of the bevel edge corresponded to the type of rising, unilateral for single-beveled and bilateral for double-beveled.

The last characteristic considered was the presence of tiny bony shards in the cut mark. Three authors identified it as a significant diagnostic trait to differentiate between non-serrated and serrated knives [9, 16, 19, 21].

The shape and mounding observations were performed by placing the cut perpendicular to the observer. In contrast, the cross-profile observations were conducted by tilting the cut by 90° (Figure 3). To define the mounding variable based on the observations, the cut mark was qualitatively divided into three different zones: entrance, center, and exit, following the similar division proposed by Tennick [5] (Figure 3). The variables and the categories were first described for clarity, and each trait was labeled according to their location.

A subset of the validation sample consisting of 20 cut marks (10 for knife numbers 3 and 5) was used to test the intra- and inter-observer reliability. Two additional researchers with different training experience in cut marks analysis examined the cuts and classified the category for each trait. Observer 2 was a master's student in forensic anthropology, while Observer 3 was a professional forensic anthropologist. The tests were conducted 12 months after the analysis of Observer 1.

2.3 | Statistical analysis and machine learning

The traits identified in the literature and the further microscopic analysis were used as independent variables to train binary logistic regression algorithms.

Two separate models were developed to determine which knife class produced the analyzed cut mark. The former predicted the type of serration—serrated vs. non-serrated—, while the latter predicted the type of bevel—single- vs. double-beveled—.

Chi-squared tests of independence and a multicollinearity test were performed to decide which independent variables to include. Chi-squared tests of independence evaluated which traits depended on the type of serration or blade bevel ($\alpha = 0.05$ significance level), while multicollinearity was used to test whether the input variables showed a high correlation between each other. A VIF of more than 5 indicates high collinearity, which can influence the model's performance [30].

Machine learning models can be prone to overfitting; they perform exceptionally well on the training dataset while performing worse on the validation dataset, failing to generalize to unseen data. To detect models' overfitting, 10-fold cross-validation was used on

JOURNAL OF FORENSIC SCIENCES AAFS

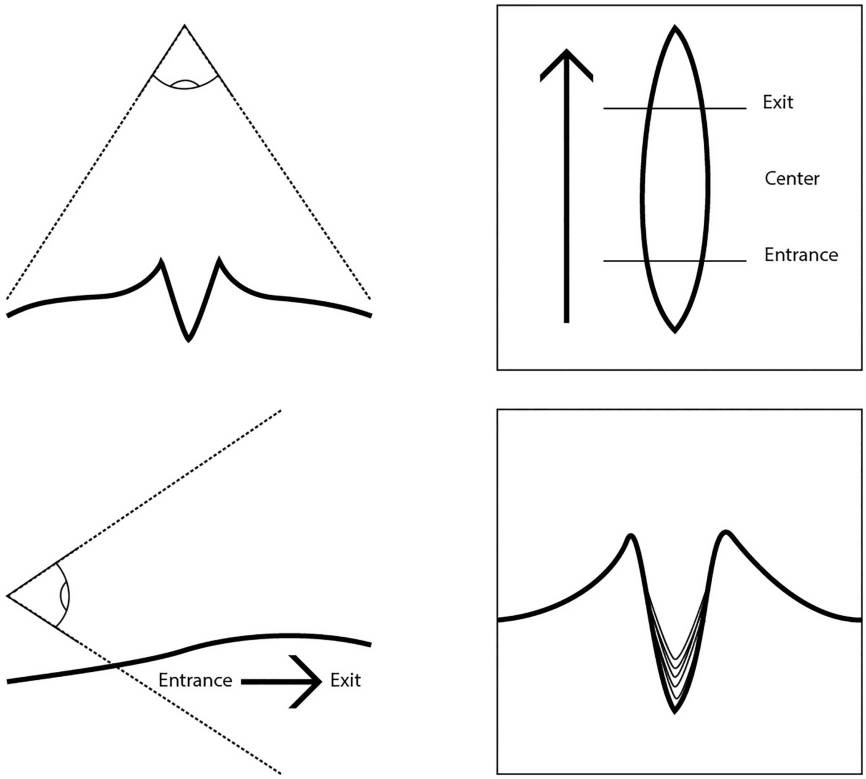

FIGURE 3 Representation of the two observation's point of views. Perpendicular to the cut, showing the cut division (Top) and tilted by 90° (Bottom).

the training dataset. The dataset was partitioned into ten equally sized subsets (fold). The model was then trained and evaluated ten times, using each time a different fold as the test set and the remaining nine folds as the training set. The performance metrics were then averaged, providing reliable estimates and helping operate hyperparameter tuning. The validation dataset was then used to evaluate how well the model generalizes on unseen data.

During the evaluation process, the authors compared various performance metrics such as accuracy, F1 score, recall, and precision. These metrics were measured across the train sets, the cross-validation test sets, and the validation sets, providing a comprehensive view of the model's performance.

Classification accuracies for each knife and for each class of knife were reported.

In this case, accuracy denotes the capability of Observer 1 or the algorithms to correctly classify cut marks into their respective knife classes.

Fleiss' kappa was used as a metric for inter- and intra-rater reliability as it assesses the degree of agreement among two or more observers when evaluating a categorical response variable. This metric expresses the strength of agreement between raters beyond what would be expected by chance alone by placing the resulting value between −1 and 1. The resulting *k*s were then classified based on Altman's adaptation of Landis and Koch guidelines [31, 32].

All the statistical analysis was conducted using SPSS 29.0.1.0 [33] and Python 3.10.5 [34]. The classification algorithms were trained and evaluated using Python 3.10.5 and Jupyter Lab 3.4.3 [34].

3 | RESULTS

3.1 | Shape

Three categories of the shape variable (general morphology) were identified during the microscopy study. They were observed by looking at the surface of the kerf. Three diagnostic shapes were identified: Ellipse, D, and Indeterminate. The Ellipse shape was the same as identified in the literature [10, 16].

- An Ellipse is characterized by the cut mark having an elliptic shape: both the edges of the cut mark are rounded. The entrance and the exit extremities of the ellipse have the same angle (Figure 4A).
- A D shape is characterized by one border of the cut mark being straight. The rounded border, similar to the belly of the letter D, can be on either side, depending on which side of the blade there is beveling on (Figure 4B).
- If neither an Ellipse nor a D is identifiable, then the shape is classified as Indeterminate.

Table 2 reports the frequency counts of each shape recorded for both classes. The D shape was primarily observed in cut marks left by single-beveled knives, while the E shape was mostly in non-serrated knives.

3.2 | Cross-profile shape

The two categories of the cross-profile variable (general morphology), previously suggested from the literature [12], were observed during the microscopic data analysis by tilting the rib by 90°. The two categories are V shape and Y shape. This trait was recorded on both extremities—entrance, and exit—of the cut mark. The following characteristics can identify the two diagnostic shapes:

- A V shape cross-profile is characterized by having both the walls of the kerf obliquely oriented, with an acute angle (Figure 5A).
- A Y shape cross-profile is characterized by having one wall of the kerf vertical at 90°, while the second wall is inclined with an acute angle (Figure 5B).

The frequency counts for the cross-profile are shown in Table 3. The cross-profile Y shape was observed exclusively for single-beveled knives, while the V shape was found in both types of knives. For the recording of the cross-profile shape, it should be noted that for knife 5, due to the high amount of shards present and the irregularity of the extremities of the walls, if the Y shape was not distinguishable, the cross-profile was classified as V shape. From the analysis conducted on the 280 cut marks, it was observed that knife 1 and 5 showed the Y cross-profile on the entrance.

3.3 | Rising

The categories of the rising variable (location) were chosen from the literature [8]. They were observed by looking at the cross-profile. The three categories are: single, bilateral, and absent. This trait was recorded for both extremities of the cut mark. Three diagnostic locations can therefore be identified as follows:

- A single rising is characterized by having either the left or the right edge extremity stretching out above the surface of the cut (Figure 6A).

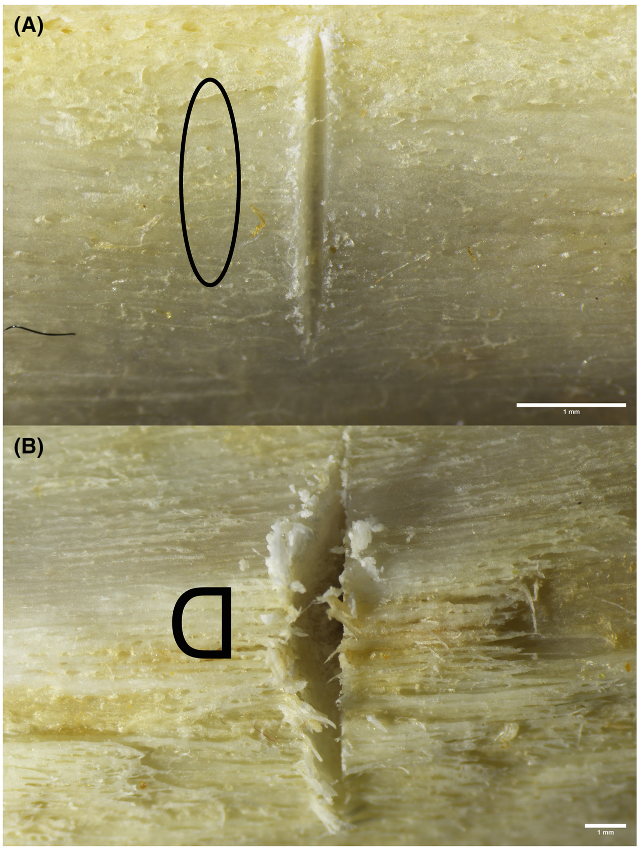

FIGURE 4 Example of the two kerf shapes Ellipse (A) and D shape (B).

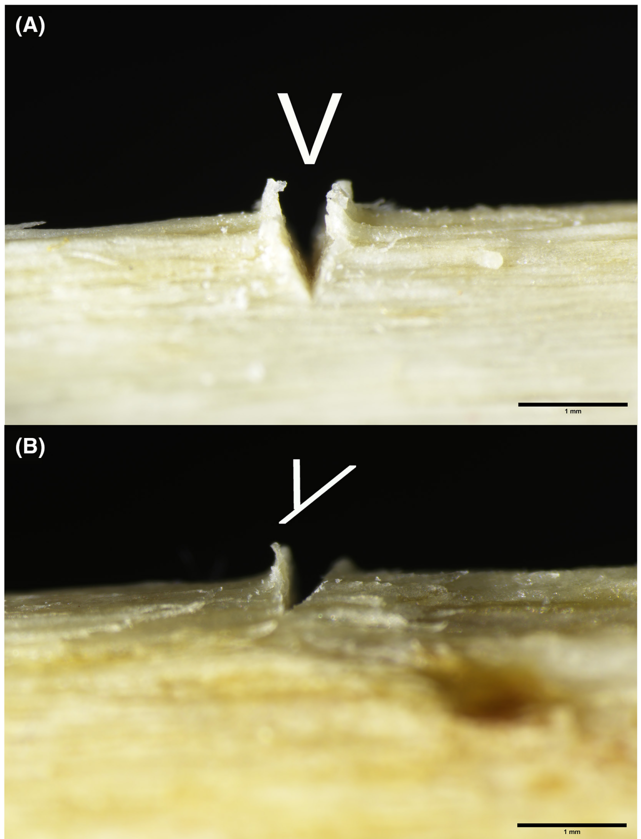

FIGURE 5 Example of the two cross-profile shapes V shape (A) and Y shape (B).

TABLE 2 Frequency counts of the shape's categories in the serration class and in the blade bevel class.

	Non-serrated	Serrated	Single-beveled	Double-beveled
E	109 (68.1%)	29 (24.2%)	67 (33.5%)	71(88.8%)
D	31 (19.4%)	88 (73.3%)	119 (59.5%)	0 (0%)
Indeterminate	20 (12.5%)	3 (2.5%)	14 (7%)	9 (11.3%)

JOURNAL OF FORENSIC SCIENCES AAFS

TABLE 3 Frequency counts of the cross-profile's categories of the entrance and exit areas in the serration class and in the blade bevel class.

	Non-serrated	Serrated	Single-beveled	Double-beveled
Entrance				
V	127 (79.4%)	49 (40.8%)	97 (48.5%)	79 (98.8%)
Y	33 (20.6%)	71 (59.2%)	103 (51.5%)	1 (1.3%)
Exit				
V	159 (99.4%)	65 (54.2%)	144 (72%)	80 (100%)
Y	1 (0.6%)	55 (45.8%)	56 (28%)	0 (0%)

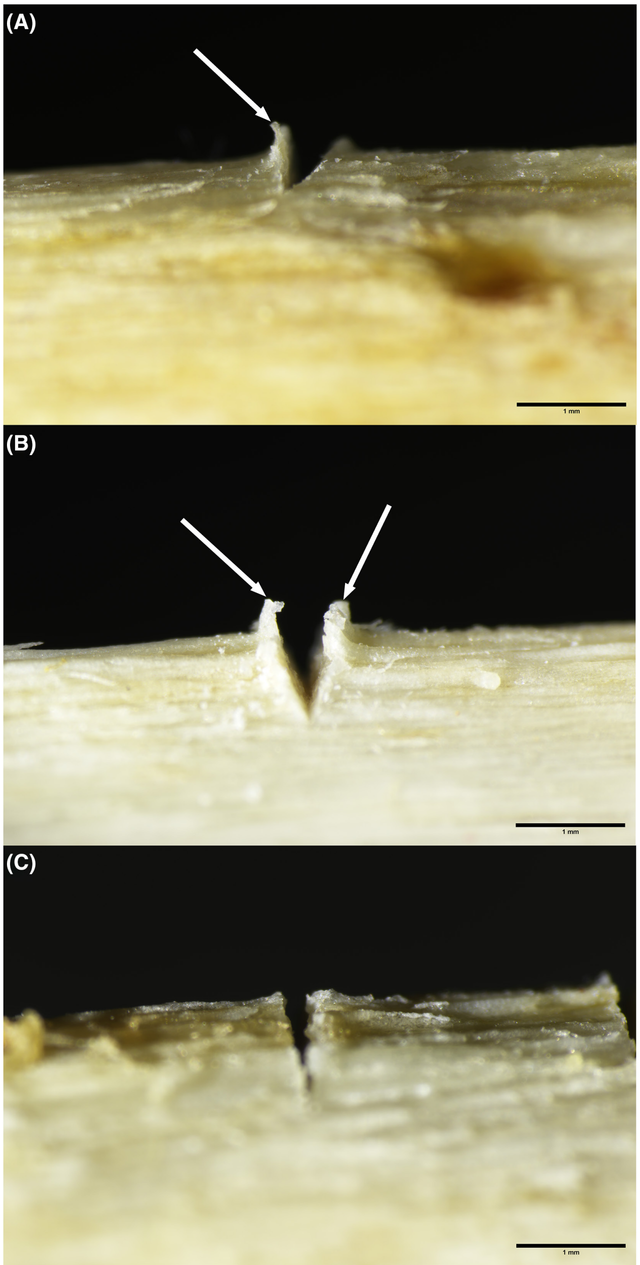

FIGURE 6 Example of the three rising types recorded in the experiment and indicated by the arrows. Single rising (A), bilateral rising (B) and absent rising (C).

- A bilateral rising is characterized by having both extremity edges stretching out above the surface of the cut (Figure 6B).
- An absent rising is characterized by not having either extremity edges stretching out (Figure 6C).

The single rising was primarily found in cut marks produced by single-beveled knives (see Table 4).

3.4 | Feathering

The categories of the feathering variable were chosen from the literature as they were observed by looking at the surface of the cut [9, 16]. The two categories are: present and absent. This category is characterized by the edge having bone spurs with a jagged appearance which is almost "feather-like" (Figure 7).

The feathering variable was most frequently recorded for single-beveled and serrated knives (see Table 5).

3.5 | Shards

The categories of shards variable were chosen from the literature [9, 16, 19, 21]. They were observed by looking at the surface of the cut. The two categories are: present and absent. The category is characterized by the surface and the surrounding area of the cut presenting with flake-like bone spurs (Figure 8).

The presence of shards was primarily recorded in single-beveled, serrated knives, as shown in Table 6.

3.6 | Mounding

The categories of the mounding variable were identified during the microscopy study. They were observed by looking at the cuts' surface. Two variables of mounding were observed during the data analysis.

The first variable (location) was recorded on both extremities and the center area of the cut mark. It included three diagnostic location categories: single, bilateral, and absent, which could be identified by the following characteristics:

- A single mounding is characterized by a rounded, "wave-like" formation of the bone spur on either one of the edges of the cut (Figure 9A).
- A rounded, "wave-like" formation of a bone spur on both edges of the cut characterizes a bilateral mounding.
- An absent mounding is characterized by the absence of any rounded, "wave-like" bone spur formation on either edge of the cut (Figure 9B).

The second variable, which was recorded as general morphology, included three categories that are: marked, not marked, and

TABLE 4 Frequency counts of the rising's categories of the entrance and exit areas in the serration class and in the blade bevel class.

	Non-serrated	Serrated	Single-beveled	Double-beveled
Entrance				
Single	93 (58.1%)	57 (47.5%)	110 (55%)	40 (50%)
Bilateral	60 (37.5%)	10 (8.3%)	37 (18.5%)	33 (41.3%)
Absent	7 (4.4%)	53 (44.2%)	53 (26.5%)	7 (8.8%)
Exit				
Single	81 (58.1%)	67 (47.5%)	108 (55%)	40 (50%)
Bilateral	59 (37.5%)	7 (8.3%)	34 (18.5%)	32 (41.3%)
Absent	20 (4.4%)	46 (44.2%)	58 (26.5%)	8 (8.8%)

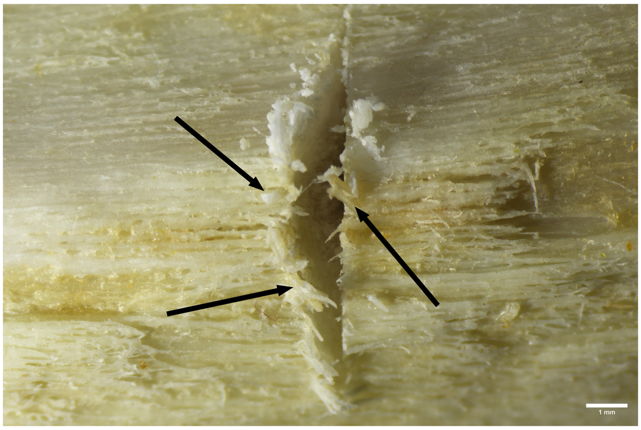

FIGURE 7 Example of feathering (arrows).

absent. The following characteristics can identify the three diagnostic categories:

- A marked mounding is characterized by an evident presence of the rounded, "wave-like" bone spur on the edges of the cut mark (Figure 9A).
- A not marked mounding is characterized by a not-so-evident presence of the rounded, "wave-like" bone spur formation on the edges of the cut mark (Figure 9B).
- An absent mounding is characterized by the complete absence of the rounded, "wave-like" bone spur formation on the edges of the cut mark (Figure 9C).

Table 7 shows no particular pattern for any of the knives.

3.7 | Logistic regression models

Following the results of the Chi-square tests of independence, all eleven variables were found to be significantly associated with the serration class, while only ten of the eleven variables were found to be significantly associated with the blade bevel class. The VIF statistical analysis showed a VIF higher than 5 for the variables cross-profile exit and mounding.

These variables were removed as input variables for the serration and the blade bevel model as they showed a high correlation. Therefore, nine out of eleven variables were used as predictor variables. Liblinear was chosen as the solver for the serration logistic regression model, and the C and penalty parameters were tuned on 100 and L2, respectively. The evaluation of the performance metrics showed that the training accuracy (86.62%), precision (85.70%), recall (82.69%), and F1 (84.18%) were slightly higher than the test sets (82.14%, 83.08%, 78.33%, and 78.65%).

Between the test sets and the validation dataset, a loss of 19.5% in accuracy and 23.08% in precision was observed. Validation recall (85.19%9) was higher than test recall, and the validation F1 score dropped from 78.65% to 70.77%.

For the second model, saga was chosen as the solver, 1 as the C and L2 as the penalty. The evaluation showed that the training metrics were slightly higher than the test metrics. Specifically, training accuracy (84.97%) and precision (89.66%) were slightly higher than the test ones (77.50% and 83.92%). Both recall and F1 metrics showed consistency across sets (89.35% vs. 85.50%) and (89.43% vs. 81.95%).

Test and validation metrics yielded closer results indicating that the model's performance generalizes reasonably well to unseen data; around 77% for both test and validation accuracies, and between 81% and 84% for precision, recall, and F1.

Classification reports for the specific classes are shown in Table 8.

3.8 | Classification accuracies

Out of the seven knives, the first author achieved a 70% accuracy in classifying them. Only two knives were not correctly classified: knives 4 and 6. Knife 4 was incorrectly labeled as serrated and 6 as non-serrated. For the blade bevel class, knife 2 was incorrectly classified as double-beveled, and 4 was incorrectly classified as single-beveled.

The classification accuracies for each of the seven knives were determined using two models: serration and blade bevel. Under the serration model, Knives 1, 2, and 4 each achieved an accuracy of

TABLE 5 Frequency counts of the feathering's categories in the serration class and in the blade bevel class.

	Non-serrated	Serrated	Single-beveled	Double-beveled
Present	6 (3.8%)	40 (33.3%)	40 (20%)	6 (7.5%)
Absent	154 (96.3%)	80 (66.7%)	160 (80%)	74 (92.5%)

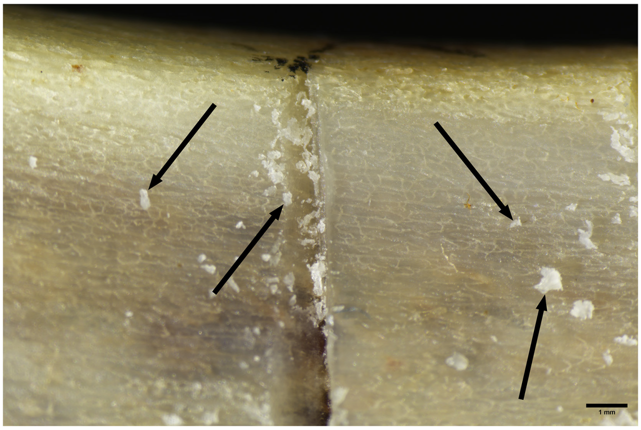

FIGURE 8 Example of shards (arrows).

50%, Knife 3 achieved a perfect 100%, Knives 5 and 6 both had an accuracy of 88.88%, and Knife 7 had an accuracy of 77.77%. In contrast, using the blade bevel model, Knife 4 had the lowest accuracy at 40%, Knife 1 had an accuracy of 70%, Knife 2 at 60%, Knife 3 at 90%, Knives 5 and 6 each reached a perfect 100%, and Knife 7 had an accuracy of 88.89%.

3.9 | Inter- and intra-observer reliability

Inter-reliability results demonstrated a high level of agreement for shape (k=0.803), cross-profile entrance and cross-profile exit (k=0.859) alongside a moderate level of agreement for general morphology (k=0.421). However, some traits exhibited lower levels of agreement among the three observers. A fair amount of consensus was observed for rising entrance (k=0.288), mounding entrance (k=0.231) and mounding center (k=0.211) while rising exit (k=0.143) feathering (k=0.193), shards (k=0.156), and mounding exit (k=0.173) displayed poor levels of agreement. The results for intra-observer reliability revealed good and very good levels of agreement for cross-profile exit (k=0.894) and for shape (k=0.799). Moderate levels of consensus were observed for cross-profile entrance (k=0.568), mounding center (k=0.492), mounding exit (k=0.447), and general morphology (k=0.410). Rising exit (k=0.286) displayed a fair level of agreement, while rising entrance (k=0.190), feathering (k=0.124) and mounding entrance (k=0.038) exhibited a poor level of agreement. The variable shards showed disagreement among the same observer (k=−0.086).

4 | DISCUSSION

The main goal of this study was to propose standardized terminology and methodology for classifying knives based on the imprints left on the cut marks. The preliminary determination and description of the traits was achieved by combining the literature review conducted before starting the data collection and the data collection itself, confirming and unifying the terminology and description of each variable. For the shape variable, two new categories were defined, expanding the few ones found in the literature, while the Ellipse shape was already identified by Vachirawongsakorn et al. [10, 16]. These categories were chosen as the most diagnostic and recognizable categories. A possible explanation for the Indeterminate category is that the authors did not use a clamp to keep the rib steady when producing the cut marks causing the knife to create different shapes and artifacts in the cut mark. The results for the shape of the cross-profile confirmed the hypothesis that the Y shape is not strictly associated with the serrated knife but more with the single-beveled ones. Indeed, the Y shape was more prominent on single-beveled knives than on double-beveled ones, while the V shape was almost equally distributed. It was also noticed that knife 1 had the vertical wall on the left side while knife 5 had it on the right. The same thing was noticed when looking at the position of the belly of the D shape. The belly was on the right side for knife 1, while knife 5 exhibited a belly on the other. These traits therefore relate to the side where the bevel edge is, indicating a right or left single-beveled blade/knife. Similarly to the cross-profile shape, the location of the rising of the walls showed a pattern and most of the single-beveled knives were associated with unilateral rising.

The shards variable was exclusively visible in serrated knives and single-beveled knives, confirming what was observed by Sandras et al. [19] and Capuani et al. [21].

The feathering variable was included even though it was not frequently reported in the literature as it was observed during the analysis [9, 16].

The recording of the variable mounding in such a manner is the first of its kind, and therefore, further investigation should focus on how this trait is produced. No distinctive pattern was observed in the knife classes. However, these variables were necessary and used to predict the classes using the logistic regression model. At first, the mounding variable (location) and the variable rising seemed to be the same trait observed from two different angles, as the two "shoulders" protruding above the surface could result from the mounding on the surface of the kerf. However, the VIF results showed that they must be considered as two separate and distinct variables.

The second objective of this research was achieved. All variables were observed using the stereomicroscope, confirming what Crowder et al. [18] found in their study. The classification could be performed by the observer without the use of the algorithms tested in this study. However, it was noticed that besides the shape,

TABLE 6 Frequency counts of the shards' categories in the serration class and in the blade bevel class.

	Non-serrated	Serrated	Single-beveled	Double-beveled
Present	0 (0%)	26 (21.7%)	26 (13%)	0 (0%)
Absent	160 (100%)	94 (78.3%)	174 (87%)	80 (100%)

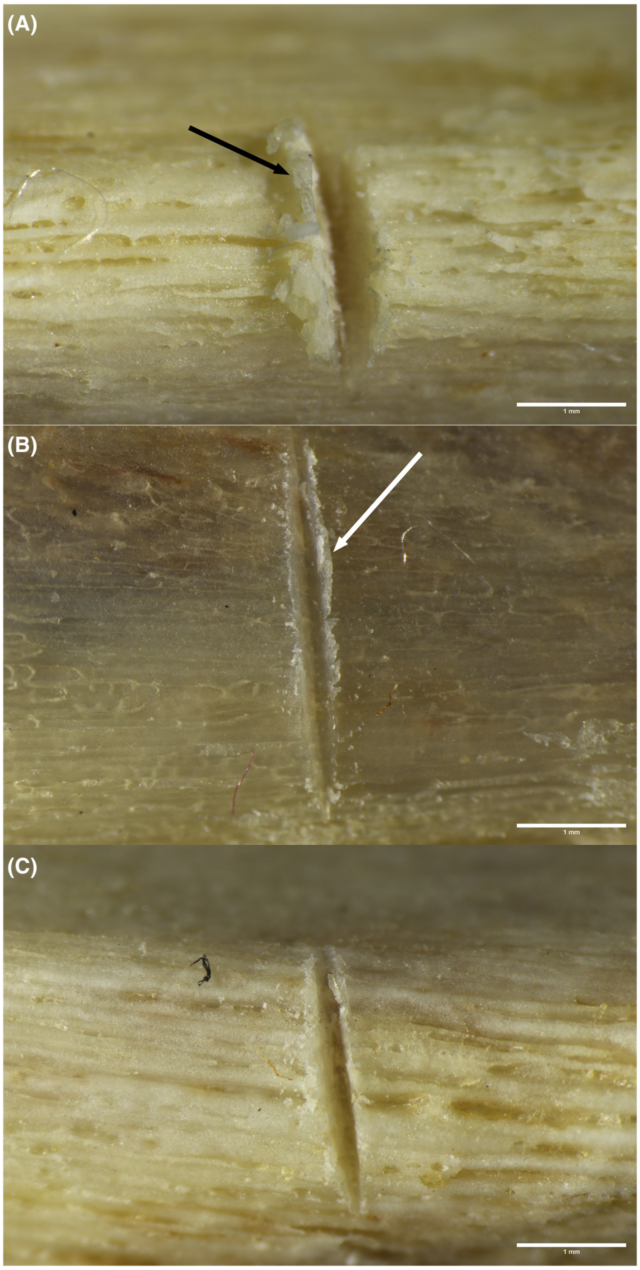

FIGURE 9 Example of the three mounding types (general morphology) indicated by arrows. Marked mounding and single (A), not marked mounding (B) and absent mounding (C).

TABLE 7 Frequency counts of the mounding categories (location) of the entrance, center, and exit areas in the serration class and blade bevel class.

	Non-serrated	Serrated	Single-beveled	Double-beveled
Entrance				
Single	86 (53.8%)	53 (44.2%)	98 (49%)	41 (51.3%)
Bilateral	9 (5.6%)	0 (0%)	1 (0.5%)	8 (10%)
Absent	65 (40.6%)	67 (55.8%)	101 (50.5%)	31 (38.8%)
Center				
Single	67 (41.9%)	46 (38.3%)	76 (38%)	37 (46.3%)
Bilateral	14 (8.8%)	1 (0.8%)	3 (1.5%)	12 (15%)
Absent	79 (49.4%)	73 (60.8%)	121 (60.5%)	31 (38.8%)
Exit				
Single	92 (57.5%)	67 (55.8%)	115 (57.5%)	44 (55%)
Bilateral	16 (10%)	2 (1.7%)	6 (3%)	12 (15%)
Absent	52 (32.5%)	51 (42.5%)	79 (39.5%)	24 (30%)
General morphology				
Marked	86 (53.8%)	55 (45.8%)	95 (47.5%)	46 (57.5%)
Not marked	50 (31.3%)	25 (20.8%)	53 (26.5%)	22 (27.5%)
Absent	24 (15%)	40 (33.3%)	52 (26%)	12 (15%)

Note: Frequency of mounding categories (general morphology) in the serration and in the blade bevel class.

TABLE 8 Classification report for the serration and blade bevel models.

	Precision	Recall	F1	Accuracy
Serration				
Serrated	61%	85%	71%	85%
Non-serrated	86%	62%	72%	63%
Blade bevel				
Single	85%	83%	84%	89%
Double	62%	65%	63%	63%

cross-profile, and rising variables for single or double knives, and the feathering and shards variables for serrated, the other variables were not distributed in a manner that could be directly used for classification. The models included these variables, highlighting the importance of incorporating machine learning algorithms in cut marks analysis. The authors have developed a user-friendly tool that can be easily run locally in the terminal. Comprehensive installation guides for Python 3 on Windows, MacOS, and Linux, along with the launcher scripts, are available on the popular cloud-based service Github at the following link: https://github.com/sciadi98/dissertation-steiger-2022.

The results of the blade bevel model highlight the model's effectiveness in classifying cut marks produced by single-beveled blades more accurately than by double blades. Single blade predictions were accurate, and the model successfully identified the majority of actual single-beveled blades. The serration model identified the cut marks produced by non-serrated knives more precisely than serrated ones, but it was less effective in capturing all actual non-serrated cases. This trade-off between precision and recall for the two classes can be due to a class imbalance in the training data. Consequently, the model showed some overfitting, struggling to generalize well to unseen data. Techniques such as regularization were employed to help reduce the overfitting, but the use of more training data could be considered for future improvements.

Nevertheless, the results of the present classification align with other validation studies. For serrated vs. non-serrated knives, the authors achieved similar accuracies to Crowder's (63%–85% vs. 93%–97%) and Ghui's (63%–85% vs. 61%–90%) [16, 18]. The models also outperformed Crowder's for bevel classification (63%–89% vs. 65%) specifically for single-bevel classification. Love's error rates using the classification trees algorithm ranged between 50% and 65% [17], while the highest error rate obtained in the present study was 37%. It is important to note that both Love's [17] and Crowder's [18] work primarily focused on the diagnostic potential of striations left on the cut mark. However, their diagnostic significance varies in different studies [11, 13–15, 17, 35]. In the present work and Ghui's paper [16], striations were not taken into consideration for the classification, demonstrating that this trait is not paramount for classifying serrated and non-serrated knives.

The misclassification of knives 2, 4, and 6 by the first author and the logistic regression models could be due to different reasons. The input variables fed to the algorithms were derived from observations, so the absence or ambiguity of some traits can cause difficulties in the classification process. Furthermore, knife 4 was not as sharp as the others, hence not producing the same quality cuts, which could be behind the misclassification. It is, therefore, paramount to consider the variability within the blades of the same classes when applying this methodology to real-world scenarios.

A significant contribution of the present study is the inter- and intra-observer reliability of each of the identified traits. The absence of this approach in previous publications did not allow a comparison; nevertheless, the results showed promising outcomes, as several traits yielded scores between moderate and very good. The variables related to the mounding trait showed more consistent agreement within the same rater. This should be expected because although clear photographs and descriptions were provided, subjective interpretation can always affect the traits. The time lapse between the original observations and the validation tests could be a possible reason for poor kappa scores, specifically for the traits shards, feathering, and rising. During this period, which lasted several months, some traits may have been modified due to the constant handling of the samples and the storage conditions.

In this respect, recent studies showed that the rising, shape and cross-profile variables are more resistant and less prone to be impacted by surface and heat exposure [36, 37]. On the other hand, these traits are more impacted by corrosive chemicals [38].

It is also important to remark that Fleiss' kappa considers the possibility of guessing, and they cannot be directly interpreted. Consequently, they may underestimate the true agreement between the observers, leading to the observed lower scores of inter- and intra-rater reliability [39].

This research was carried out under controlled conditions, maintaining consistent factors like the direction and position of the person making the cuts. In real-world situations, numerous variables can come into play, affecting the direction, size, or length of cut marks [7, 22, 40, 41]. Moreover, the presence of soft tissues and clothing can dampen the impact force and alter the characteristics of the marks [42, 43].

Finally, even though pig bones serve as a reasonable substitute for human bones, their distinct composition may still influence how bones respond to trauma, potentially changing the characteristics of the cut marks.

5 | CONCLUSION

This study develops and proposes standardization in the terminology for the variables used in cut marks analysis as the basis for cut mark description and classification standards.

The use of the binary logistic regression for cut mark analysis was found to be reliable for this research, therefore showing the potential of assisting the analysts when facing challenges in the decision-making process. These models aimed to reduce the dependence on experienced observers and streamline the use of multiple variables for classification. Future research should explore using computer vision to extract traits from images, lowering human errors caused by subjectivity or bone surface modifications. Furthermore, the authors suggest extending the typologies of knives (e.g. partially serrated knives, hunting knives, sharp tools), testing different target materials and other types of cut movements (e.g. slicing, stabbing, chopping).

In conclusion, the results indicate the variables' reliability, suggesting the possibility of considering the proposed terminology as a ground for developing future standards. Similarly, the authors recommend incorporating machine learning models for more accurate, direct, and rapid classification of knives in sharp force trauma analysis.

ACKNOWLEDGMENTS

The authors would like to thank Nathan Lepori for his contribution to the programming of the machine learning models and Timothy Hofmann for illustrating the "Anatomy of a kerf" and "Representation of the two observation's point of views" pictures.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

ORCID

Giada Sciâdi Steiger https://orcid.org/0009-0002-1269-257X

REFERENCES

1. Morgan N, Shaw O, Mailley J, Channing R, Sweiry A, Kruithof K, et al. Trends and drivers of homicide: Main findings. Longon, UK: Home Office; 2020 Research Report No.: 113.
2. Rao VJ, Hart R. Tool mark determination in cartilage of stabbing victim. J Forensic Sci. 1983;28(3):794–9. https://doi.org/10.1520/JFS11577J
3. Houck MM. Skeletal trauma and the individualization of knife marks in bone. In: Reichs K, editor. Forensic osteology: advances in the identification of human remains. 2nd ed. Springfield, IL: Charles C. Thomas; 1998. p. 410–24.
4. Ofele AS, Eliopoulos C, Borrini M. Sharp force trauma: the effects of blade damage on cut mark characteristics. SciFed J Forensics. 2018;1(2):1–8.
5. Tennick CJ. The identification and classification of sharp force trauma on bone using low power microscopy [dissertation]. Preston, England: University of Central Lancashire 2012.
6. 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
7. Cerutti E, Magli F, Porta D, Gibelli D, Cattaneo C. Metrical assessment of cutmarks on bone: is size important? Leg Med. 2014;16(4):208–13. https://doi.org/10.1016/j.legalmed.2014.03.010
8. Capuani C, Telmon N, Moscovici J, Molinier F, Aymeric A, Delisle MB, et al. Modelling and determination of directionality of the kerf in epifluorescence sharp bone trauma analysis. Int J Leg Med. 2014;128(6):1059–66. https://doi.org/10.1007/s00414-014-1022-0
9. Lewis JE. Identifying sword marks on bone: criteria for distinguishing between cut marks made by different classes of bladed weapons. J Archaeol Sci. 2008;35(7):2001–8. https://doi.org/10.1016/j.jasrep.2023.103843
10. Vachirawongsakorn V, Painter J, Márquez-Grant N. Knife cut marks inflicted by different blade types and the changes induced by heat: a dimensional and morphological study. Int J Leg Med. 2022;136(1):329–42. https://doi.org/10.1007/s00414-021-02726-5
11. Kooi RJ, Fairgrieve SI. SEM and stereomicroscopic analysis of cut marks in fresh and burned bone. J Forensic Sci. 2013;58(2):452–8. https://doi.org/10.1111/1556-4029.12050
12. Tegtmeyer CE. A comparative analysis of serrated and non-serrated sharp force trauma to bone [dissertation]. San Marcos, TX: Texas State University-San Marcos 2012.
13. Pounder DJ, Reeder FD. Striation patterns in serrated blade stabs to cartilage. Forensic Sci Int. 2011;208:91–4. https://doi.org/10.1016/j.forsciint.2010.11.015
14. Pounder DJ, Sim LJ. Virtual casting of stab wounds in cartilage using micro-computed tomography. Am J Forensic Med Pathol. 2011;32:97–9. https://doi.org/10.1097/PAF.0b013e3182186f37
15. Pounder DJ, Cormack L, Broadbent E, Millar J. Class characteristics of serrated knife stabs to cartilage. Am J Forensic Med Pathol. 2011;32:157–60. https://doi.org/10.1097/PAF.0b013e3181db7ee4
16. Ghui M, Eliopoulos C, Borrini M. A proposed method for differentiating knives from cut marks on bone: a forensic anthropological approach. Med Sci Law. 2024;64(3):199–203. https://doi.org/10.1177/00258024231198912
17. Love JC, Derrick SM, Wiersema JM, Peters C. Validation of tool mark analysis of cut costal cartilage. J Forensic Sci. 2012;57(2):306–11. https://doi.org/10.1111/j.1556-4029.2011.01977.x
18. Crowder C, Rainwater CW, Fridie JS. Microscopic analysis of sharp force trauma in bone and cartilage: a validation study. J Forensic Sci. 2013;58:1119–26. https://doi.org/10.1111/1556-4029.12180
19. Sandras A, Guilbeau-Frugier C, Savall F, Telmon N, Capuani C. Sharp bone trauma diagnosis: A validation study using epifluorescence microscopy. Int J Leg Med. 2019;133(2):521–8. https://doi.org/10.1007/s00414-018-1944-z
20. Weber M, Niehoff A, Rothschild MA. Insights to enhance the examination of tool marks in human cartilage. Int J Leg Med. 2021;135(5):2117–34. https://doi.org/10.1007/s00414-021-02609-9
21. Capuani C, Rouquette J, Payré B, Moscovici J, Delisle MB, Telmon N, et al. Deciphering the elusive nature of sharp bone trauma using epifluorescence microscopy: a comparison study multiplexing classical imaging approaches. Int J Leg Med. 2013;127(1):169–76. https://doi.org/10.1007/s00414-012-0678-6
22. Bergman L, Brock F, Errickson D. Use of different imaging techniques in stab wound analysis. Sci Justice. 2024;64(1):50–62. https://doi.org/10.1016/j.scijus.2023.11.002
23. Samuel AL. Some studies in machine learning using the game of checkers. IBM J Res Dev. 1959;3(3):210–29. https://doi.org/10.1147/rd.33.0210
24. Goodfellow I, Bengio Y, Courville A. Deep learning, illustrated edn. Cambridge, MA: The MIT Press; 2016. p. 96–118.
25. Mahesh B. Machine learning algorithms—a review. Int J Sci Res. 2019;9(1):381–6. https://doi.org/10.21275/ART20203995
26. Chandrasekaran M. Logistic regression for machine learning. Capital One. 2021. https://www.capitalone.com/tech/machine-learning/what-is-logistic-regression/. Accessed 14 June 2022
27. JavaTpoint. Logistic regression in machine learning. https://www.javatpoint.com/logistic-regression-in-machine-learning. Accessed 13 June 2022.
28. Tutorialspoint. Logistic regression in Python—Limitations. https://www.tutorialspoint.com/logistic_regression_in_python/logistic_regression_in_python_limitations.htm. Accessed 14 June 2022.
29. Swann IJ, MacMillan R, Watson AA. A study of stab wounds. Emerg Med J. 1985;2(1):31–6. https://doi.org/10.1136/emj.2.1.31
30. Daoud JI. Multicollinearity and regression analysis. J Phys Conf Ser. 2017;949:12009. https://doi.org/10.1088/1742-6596/949/1/012009
31. Altman DG. Practical statistics for medical research. New York, NY: Chapman & Hall/CRC Press; 1999.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74. https://doi.org/10.2307/2529310
33. IBM. SPSS statistics for windows. Version 28.0. Armonk, NY: IBM; 2022.
34. Python. Python. Version 3.10.5. Amsterdam, NL: Python; 2022.
35. Komo L, Grassberger M. Experimental sharp force injuries to ribs: multimodal morphological and geometric morphometric analyses using micro-CT, macro photography and SEM. Forensic Sci Int. 2018;288:189–200. https://doi.org/10.1016/j.forsciint.2018.04.048
36. Vachirawongsakorn V, Márquez-Grant N, Painter J. Survival of sharp force trauma in burnt bones: effects of environmental factors. Int J Leg Med. 2023;137(3):809–23. https://doi.org/10.1007/s00414-022-02916-9
37. Griffiths K, Borrini M. The resilience of cut mark characteristics: An experimental study on the effect of fire exposure. In: Proceedings of the 76th Annual Scientific Conference of the American Academy of Forensic Sciences; Denver, CO. Colorado Springs, CO: American Academy of Forensic Sciences; 2024. p. 87.
38. Jordan M, Borrini M. The resilience of cut mark characteristics: An experimental study on the effect of corrosive chemicals. In: Proceedings of the 76th Annual Scientific Conference of the

American Academy of Forensic Sciences; Denver, CO. Colorado Springs, CO: American Academy of Forensic Sciences; 2024. p. 86.

39. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med. 2012;22(3):276–82. https://doi.org/10.11613/bm.2012.031
40. Bartelink EJ, Wiersema JM, Demaree RS. Quantitative analysis of sharp-force trauma: an application of scanning electron microscopy in forensic anthropology. J Forensic Sci. 2001;46(6):1288–93. https://doi.org/10.1520/JFS15148J
41. Shaw KP, Chung JH, Chung FC, Tseng BY, Pan CH, Yang KT, et al. A method for studying knife tool marks on bone. J Forensic Sci. 2011;56(4):967–71. https://doi.org/10.1111/j.1556-4029.2011.01741.x
42. Ferllini R. Macroscopic and microscopic analysis of knife stab wounds on fleshed and clothed ribs. J Forensic Sci. 2012;57(3):683–90. https://doi.org/10.1111/j.1556-4029.2012.02087.x
43. Lynn KS, Fairgrieve SI. Microscopic indicators of axe and hatchet trauma in fleshed and defleshed mammalian long bones. J Forensic Sci. 2009;54(4):793–7. https://doi.org/10.1111/j.1556-4029.2009.01062.x

How to cite this article: Steiger GS, Borrini M. A proposal for cut marks classification using machine learning: Serrated vs. non-serrated, single vs. double-beveled knives. J Forensic Sci. 2024;00:1–13. https://doi.org/10.1111/1556-4029.15588