Yu Liu ⓘ*,[1] Ph.D. Daniel Attinger,[2] Ph.D.; and Kris De Brabanter,[3] Ph.D.*

Automatic Classification of Bloodstain Patterns Caused by Gunshot and Blunt Impact at Various Distances

ABSTRACT: The forensics discipline of bloodstain pattern analysis plays an important role in crime scene analysis and reconstruction. One reconstruction question is whether the blood has been spattered via gunshot or blunt impact such as beating or stabbing. This paper proposes an automated framework to classify bloodstain spatter patterns generated under controlled conditions into either gunshot or blunt impact classes. Classification is performed using machine learning. The study is performed with 94 blood spatter patterns which are available as public data sets, designs a set of features with possible relevance to classification, and uses the random forests method to rank the most useful features and perform classification. The study shows that classification accuracy decreases with the increasing distance between the target surface collecting the stains and the blood source. Based on the data set used in this study, the model achieves 99% accuracy in classifying spatter patterns at distances of 30 cm, 93% accuracy at distances of 60 cm, and 86% accuracy at distances of 120 cm. Results with 10 additional backspatter patterns also show that the presence of muzzle gases can reduce classification accuracy.

KEYWORDS: forensic science, bloodstain pattern analysis, classification, impact spatters, gunshot spatters, spatter pattern, machine learning, image analysis, random forests, feature engineering

A bloodstain pattern is an ensemble of bloodstains. Stains are marks that can be observed on a solid surface called a target. A blood spatter pattern is produced by drops from a blood source which have traveled through the air before reaching a target surface, such as a wall or the cardstock used in this study. Bloodstain pattern analysis (BPA) has been used in US criminal courts since the second half of last century (1–3). BPA interprets the bloodstain patterns in a crime scene in order to provide evidence to support the crime scene reconstruction (3,4). Typically, BPA uses principles of physics, statistics, biology, and mathematics. Example reconstruction questions are as follows: "What is the mechanism causing the bloodstain patterns? Where is the origin of the blood source?"

Regarding the latter question, established methods based on the assumption of straight trajectories are widely used to predict the region of origin of a blood spatter pattern (5,6). Reference (7) proposed a method to improve the determination of the blood source location, and Varney et al. (8) introduced a plot-based method to locate the blood source. Further, Camana et al. (9) used a probabilistic approach to identify the horizontal projection of the location of the blood source.

While there have been multiple academic efforts to estimate the location of the blood source, less attention has been given to determine the mechanism causing the spatter pattern. A question of interest is whether the spatter patterns were caused by a bullet or blunt impact. Classifying the bloodstain patterns helps support other evidences such as weapons or bullet marks found on the crime scene. Stain size is used in Ref. (3) to classify spatter patterns between high-velocity spatter patterns, usually generated from gunshot, and medium-velocity blood spatter patterns, that is, generated from blunt impacts of club, axe, hammer, fist, and brick. They observed in their experiment that "many" stains in a medium-velocity spatter pattern exhibit 1/8 inch diameter or smaller, while "essentially all" stains in a high-velocity spatter pattern have diameters smaller than 1/8 inch. James et al. (10) presents the traditional method of distinguishing medium- and high-velocity impact spatter patterns based on a hypothetical correlation between the velocity of the generation mechanism and the size of resulting bloodstains. They stated that "medium-velocity impact spatters (MVIS) are bloodstains created when the source of blood is subjected to a force with a velocity in the range of 5 to 25 ft/sec. The diameters of the resulting stains are in the size range of 1 to 3 mm, although smaller and larger stains may be present. Stains in this category were usually associated with beatings and stabbings." On the other hand, "High velocity impact spatters (HVIS) are bloodstains created when the source of blood is subjected to a force with velocity of greater than 100 ft/sec. The diameters of the spatters are predominately less than 1 mm, although smaller and larger stains are often observed...."

There is currently no widely accepted method to discriminate between gunshot and impact patterns, and the forensic community has abandoned the above distinction between medium-

[1]Department of Computer Science, Iowa State University, Atanasoff Hall, 2434 Osborn Dr, Ames, 50011, IA.

[2]Department of Mechanical Engineering, Iowa State University, 2025 Black Engineering, Ames, 50011, IA.

[3]Department of Statistics, Iowa State University, 2438 Osborn Dr, Ames, 50011, IA.

Corresponding author: Yu Liu, Ph.D. E-mail: liuyu0jlu@gmail.com

velocity and high-velocity impact spatter patterns (10,11). A reason is that both gunshot and blunt spatter patterns exhibit a large number of small stains under 1/8 inch (3). In addition, the above classification criterion is not quantitative enough, with expressions such as "many" and "essentially all under 1/8 inch diameter." Also, classification by BPA analysts is viewed as subjective and vulnerable to contextual bias (12). For instance, the phrase "predominately less" used in Ref. (10) depends on the appreciation of the BPA analyst. Taylor et al. (13) assessed the reliability of classification decisions in bloodstain pattern analysis. In that study, well-trained and experienced analysts achieved a classification accuracy of 69.5%. Another challenge is that multiple factors affect the generation of blood spatter patterns and are currently poorly considered. One factor is the horizontal distance from blood to target, called hereafter the BT distance, that is, the horizontal distance between the blood source and the stained target surface. For gunshot spatter patterns, if the BT distance is larger than 3 or 4 feet, the average size of bloodstains will increase because small drops cannot travel as far as larger ones. The reason for this observation stems from basic physics and experience: Small snowballs do not travel as far as larger ones, because drag affects the inertia of smaller objects much more than that of larger ones. Other factors causing variations in the spatter generation process include the speed and shape of the bullet, and these effects are not yet completely understood (14–16). Within these current challenges, there are needs and opportunities for less ambiguous and more objective methods in forensic analysis, and artificial intelligence is certainly a way to assist the analyst in classification (17).

With the development of digital image analysis, a large number of pattern recognition and machine learning methods have found their way into the field of forensic science. De Chazal et al. (18) proposed a system that automatically sorts a database of shoeprint pattern categories in response to a reference shoeprint image. Automatic classification has the potential to minimize bias in forensic investigation and testimony. For instance, classification of bloodstains patterns caused by different atomization processes becomes more objective by applying digital image processing methods. Arthur et al. (19) described in detail how to automatically obtain measurements and features from a digital image, using commercial software (20). Siu et al. (11) proposed a framework to distinguish between forward gunshot and blunt force impact spatter patterns in a quantitative way. In their work, descriptive statistics were used to select the important features; however, they did not show how to classify an unknown spatter pattern and did not mention the performance of their framework. Arthur et al. (21) proposed the first documented attempt to automatically classify impact versus cast-off spatter patterns. This work is related to the classification challenge addressed in this manuscript. The main feature, identified in Ref. (21), to classify between impact and cast-off was the linear alignment of the stains in a given pattern. This feature is likely irrelevant to the case addressed in the present study where the spatter mechanism is radial in both gunshot and blunt impact cases, rather than linear.

It is the belief of the authors and one of the reviewers that BPA needs research to build quantitative, less subjective methods to underpin the pattern classification task. The route to developing computer-assisted bloodstain pattern classification is a long one.

In this paper, we develop a stable, quantitative, and objective framework to classify the *blunt impact* (referred as *impact*) spatter patterns and *backward gunshot* spatter patterns (referred as *gunshot*). The framework combines digital image analysis and machine learning methods. All the spatter patterns used in this study are available to the research community (22,23). The spatters are generated under controlled conditions described in Refs (22,23) and briefly repeated in the next section for the sake of completion. Based on high-quality images of the spatter patterns, we describe how to construct global and local features. A subset of spatter patterns is used for feature engineering, which is the process to construct features (24,25) based on intuition or on an understanding of the physics at hand. Later, we randomly split the remaining spatter patterns into a training and a test set. A training set is used to train the machine learning model, here the random forests model. Then, the model performance is obtained by comparing the predicted mechanism of each test case with its known mechanism. We repeat this procedure 1000 times on randomly selected testing and training cases to obtain the average performance of the proposed model on the given spatter patterns (22,23). Meanwhile, we explore the influence of the BT distance on features and calculate the performance of the proposed framework at different BT distances. Our framework achieves 98.81% accuracy in classifying impact spatter patterns versus gunshot backspatter patterns at BT distances no larger than 30 cm, 93.20% accuracy at BT distances no larger than 60 cm, and 85.96% accuracy at BT distances no larger than 120 cm.

The major contribution of this study consists of three parts. First, we construct a set of novel local or global features from the image of a blood spatter pattern. Second, we propose a new machine learning framework to predict the mechanism causing a blood spatter pattern, either blunt impact or gunshot. Third, we conduct a comprehensive analysis of the performance of the proposed framework at different BT distance ranges. A major difference between the work presented here and the automatic classification study (21) is that the data set used in Ref. (21) is not publicly available, which currently makes it impossible to reproduce the work in Ref. (21). The key contribution of this paper is the methodology itself.

The manuscript starts with a presentation of the used methodology: statistics, production of the bloodstain patterns, image processing, feature developed for the classification, and presentation of the classification method used, called random forests. Results are presented in terms of classification performance and ranking of the most efficient features. Conclusions summarize the novelty of the work, highlight the current limitations, and suggest future work.

Methods

Statistics Terminology and Useful Concepts

We briefly define the statistic terms used in this paper:
Consider a sample of n observed data points $(x_1, \ldots x_n)$, which are independent and identically distributed. Let us define:

1. (sample) mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

2. (sample) SD: $s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$ which measures the spread of the data distribution

3. (sample) median: ordering all data points and selecting the one in the middle (or if there are two middle numbers, taking the mean of those two numbers).

4. (sample) Pearson correlation coefficient: a measurement of the linear correlation between two variables X and Y.

5. Skewness: measure of the asymmetry of the probability distribution about its mean. A Gaussian distribution is symmetrical about its mean; thus, its skewness is zero.

Bloodstain Spatter Patterns and Digital Image Preprocessing

The data set used throughout this paper consists of 94 spatter patterns generated via controlled experiments of gunshot backspatter or blunt impact spatter (22,23). Ten additional patterns from Ref. (23) were then used to specifically test the influence of muzzle gases on the classification accuracy. Table 1a,b describes the names and characteristics of the patterns used in this study. The experimental methods used to generate the spatter patterns are fully described in Refs (22,23) and are only briefly summarized here.

For the 94 impact and backspatter patterns considered in this study, fresh swine blood with anticoagulants was used, at room temperature. Spatter patterns were produced in an indoor room, without external airflow. Spatter patterns were produced on a flat target surface, placed vertically and made of white cardstock for all spatters, except two generated on butcher paper. Spatter patterns were then allowed to dry and scanned at the high resolution of 600 dots per inch. The scanner eliminates the parallax error that is often present in crime scene pictures. Since the area of the scanner is smaller than the maximum target surface of 1.36 m × 1.1 m used to collect the spatter pattern, scans are performed in a piecewise manner and assembled using graphics editor software Adobe Photoshop®. The spatter patterns are listed in Table 1b, and full description of how each pattern was generated is in Refs (22,23).

To generate impact spatter patterns, the following procedure was typically used (Fig. 1*b*). A blood pool with volume 1 mL on a flat immobile hard surface was impacted by either a rod (at velocities between 5 and 9 m/sec, denoted as "hockey puck") or another parallel flat surface (at velocities between 2 and 4 m/sec, denoted as "cylinder"). Horizontal distances between target and blood source were varied between 30 and 200 cm. Table 1b lists all the pattern names, with the few cases where the typical procedure was modified.

Gunshot backspatters are spatters where blood is atomized by a bullet, and bloodstains are splashed in a direction opposite to that of the bullet (5,26). Figure 1*a* describes the experimental setup to generate the gunshot backspatter patterns. For the spatter patterns investigated in this work, the travel direction of the bullet is perpendicular to both blood source and the target, and a bullet hole is created on the target surface. To generate gunshot backspatter patterns, the following procedure was typically used. A cavity with volume 2.5 or 10 mL was made in a foam board, completely filled with blood, then closed, and held vertically. This blood source was impacted normally by a bullet traveling horizontally. Bullet with pointy, round, or flat nose was shot with a handgun or a rifle, with velocities between 285 and 987 m/sec. Bullet velocities in handgun experiments were about three times lower than the velocities in rifle experiments. Horizontal distances between the vertical target and blood source were varied from 10 cm to about 120 cm. For most experiments, a diffuser plate was used to prevent the muzzle gases of the gunshot to interact with the spattered drops, while this interaction was purposefully present in 10 other experiments. The bullet hole was hidden from the digitally scanned spatter patterns. Table 1b lists all the pattern names, with the few cases where the typical procedure was modified.

The reason why the effect of muzzle gases was investigated in experiments involving 10 additional spatters is that an understanding of the role of muzzle gases is important. For instance, Taylor et al. (27) showed that the blood droplets were observed to change direction under the influence of muzzle gases. Recent fluid dynamic work has quantitatively evaluated the influence of

TABLE 1—*(a) Description of the set of 94 experiments (excluding the ones with muzzle gases) used throughout this study, with mention of distance between blood source and vertical target surface where stains were collected (BT distance). (b) Name of all the patterns (including muzzle gases) used in this study, which are fully described in Refs (22,23). Blood spatter patterns names are preceded with symbols R (rifle), H (handgun), p (pointy bullet), f (flat-tip bullet), and r (round-tip bullet). For most impact experiments, a volume of blood of 1mL deposited on a hard flat surface was impacted. For most gunshot experiments, a volume of blood of 2.5 mL or 10 mL enclosed in a foam cavity was hit by a bullet. Symbols after pattern number describe the following special situations: m (muzzle gases interacting with spatter), # or S (blood source in soaked foam), B (butcher paper used as target), and R (rougher side of cardstock used as target). Blood was at room temperature, which varied between 14.5 and 24.5°C, ±1°C.*

(a)

Generation Mechanism	BT Distance (cm)							
	10	20	30	50	60	90	100	120
Impact: blood on flat surface, hit by another parallel flat surface	0	0	4	0	4	0	0	4
Impact: blood on flat surface, hit by a cylindrical rod	0	0	9	0	16	0	0	12
Handgun (no muzzle gases)	3	2	3	4	3	1	2	3
Rifle (no muzzle gases)	0	0	9	0	7	3	0	5

(b)

| Generation Mechanism	Submechanism	Pattern Name	Number of Patterns
Impact	Blood on flat surface, hit by another parallel flat surface	C1, C10, C11, C12, C2, C3, C4, C5, C6, C7, C8, C9	12
	Blood on flat surface, hit by a cylindrical rod	HP_0, HP_1, HP_10, HP_11, HP_12, HP_2, HP_23, HP_24, HP_25, HP_26, HP_27, HP_28, HP_29, HP_3, HP_30, HP_31, HP_32, HP_33, HP_34, HP_4, HP_5, HP_50$_S$, HP_51$_S$, HP_52, HP_53, HP_54, HP_55, HP_56, HP_57, HP_58$_B$, HP_59$_B$, HP_6, HP_60$_R$, HP_61$_R$, HP_7, HP_8, HP_9	37
Gunshot backspatter	Rifle	Rf13# (T13), Rf14# (T14), Rp10#, Rp101#, Rp11#, Rp12#, Rp15#, Rp16#, Rp41, Rp42, Rp43, Rp44, Rp45, Rp46, Rp47, Rp48, Rp49, Rp50, Rp51m, Rp53m, Rp55m, Rp56m, Rp7#, Rp8#, Rp9#	25
	Handgun	Hf72, Hf73, Hf76, Hf77, Hf81, Hf84, Hf85, Hr105, Hr21#, Hr22, Hr23, Hr24#, Hr25, Hr27, Hr28m, Hr29m, Hr30m, Hr31m, Hr32m, Hr33m, Hr34, Hr35, Hr36, Hr70, Hr71, Hr74, Hr75, Hr78, Hr82, Hr83	30
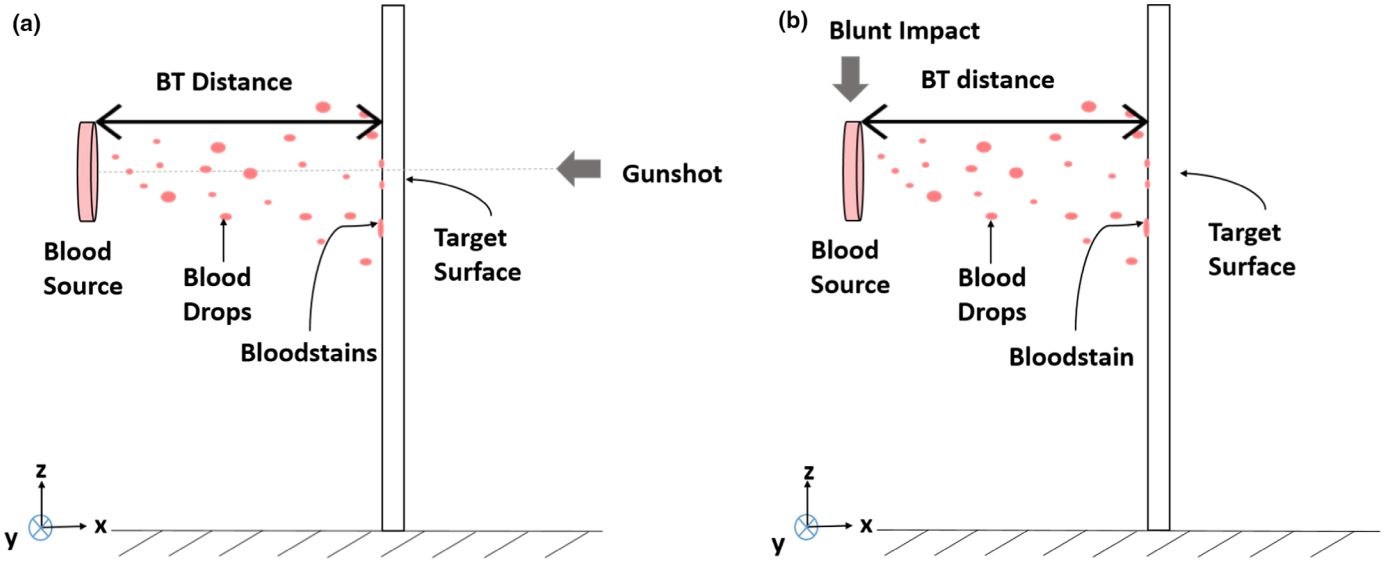
FIG. 1—(a) Experimental setup for gunshot backspatters. (b) Experimental setup for blunt impact spatters. A frame holding the blood source is placed at BT distance away from the target surface. The blood source is either shot or hit to generate the blood spatters. Blood drops from a blood source travel through the air until reaching the target surface, cardstock in this study, resulting in marks called bloodstains on the target cardstock.

muzzle gases on the resulting bloodstain patterns (28), and these effects on classification are discussed in the section Further Discussions.

As shown in Table 1a, the BT distance is one of the main parameters varied across the data set. BT distance affects spatter patterns as the number of stains of a spatter pattern becomes smaller with increasing BT distance, as more drops are deflected by the effect of gravity and drag with increasing BT distance.

In this paper, we define the atomization mechanism as the specific cause of a bloodstain pattern, using the labels *handgun*, *rifle*, *cylinder*, or *hockey puck*. We also define the atomization process as the general cause of a bloodstain pattern, using the labels *gunshot* or *blunt impact*. In order to explore the interaction between BT distance and the mechanism, we generate bloodstains spatter patterns at different BT distances. The data used in the main paper are 94 spatter patterns (excluding muzzle gases) with BT distance no larger than 120 cm. The number of replicate experiments under each BT distance and the mechanism is shown in Table 1a.

There is a wealth of possible ways to create impact or backspatter patterns. A gunshot to a hairy head will give a different spatter pattern than a gunshot on a naked leg. The design of realistic blunt impact and gunshot experiments is still an open research issue. As mentioned in Ref. (23), there is still no consensus on which experimental setup is best to simulate the complexity of gunshot spatter patterns in realistic conditions, where blood is located within a complex structure involving body tissues and blood vessels, covered by skin, clothes, or hair. Head of calves (29), a human cadaver filled with blood (30), foams or sponges soaked in blood (11,26), or cavities filled with blood (14) have been used. Experiments reported in this study used both soaked foams and cavities filled with blood as the blood source, and the information on which blood source was used is specified. Certainly, there is a need for research to produce more realistic models of blood containers for gunshot experiments. We believe, however, that research on classification of blood spatter patterns should not wait until such an elusive model is developed. Furthermore, the models used in this study for impact and for gunshot produce

repeatable results and have allowed a wealth of scientific findings of interest to BPA (14,15,28).

Image Processing

The image analysis is processed using available routines of the commercial software MATLAB® (20). After image segmentation, we apply the image analysis tool *regionprops* to analyze the identified spots (called connected components) in spatter patterns. Not all the components are spatter bloodstains, with a characteristic approximate elliptical shape; some undesired components are pencil marks or overlapping bloodstains that are currently too complex to use for reconstruction. We apply the following criteria to eliminate the undesired components: (i) discard components with area smaller than 0.021 mm^2 (3.4 × 3.4 pixels in scanner resolution); (ii) discard components with a solidity smaller than 0.75, in which solidity is a measure of the density of the pixels within a given component as defined in *regionprops*; (iii) discard components with impact angle approximated from a fitted ellipse (arcsine of the ratio of the minor axis divided by the major axis of fitted ellipse) smaller than $\pi/18$ because the component has a high probability to be a linear mark such as pencil mark or cardstock fold; and (iv) discard components with a ratio of Area over FilledArea smaller than 0.95, in which Area and FilledArea are defined in *regionprops*. The last criterion was found effective at removing the obvious overlapping bloodstains. It is important to note that every image processing method comes with their own choices, which are based on a trial-and-error process focused (in our case) at separating the stains of interest from other marks and from overlapping stains. However, the above choices have been applied consistently throughout the present study.

Global Features

Here, we describe the features that were engineered to describe the spatter patterns. We define the features representing the overall bloodstains in an entire spatter pattern as the global

features, and the ones representing bloodstains in a specific region of the spatter pattern as local features.

Due to the complexity of performing blood spatter experiments, the number of spatter patterns available for classification (22,23) is not as large as other data sets used for classification, such as face pictures of human beings. In order to minimize the risk of overfitting associated with small data sets (24,25), we assign 36 spatter patterns for feature engineering and later obtain the overall performance of the proposed model by repeated random splitting the remaining data set into a training and test set (24,25). Features are created based on intuition and the limited knowledge of fluid dynamics of blood spatter available today.

From Table 1, at least three spatter patterns were generated by every mechanism at BT distances 30, 60, and 120 cm. We sample three images at each experiment setting in {Cylinder, Hockey puck, Handgun, Rifle} × {30 cm, 60 cm, 120 cm} and analyze those 36 spatter patterns to construct features. The features are presented hereafter. Features are either based on an understanding of the fluid dynamics or defined in an arbitrary way. Since the physics of blood spatter generation and propagation is still not well understood (14,15,28), it makes sense that some criteria are set arbitrarily.

Number of Bloodstains

The total number of bloodstains in a spatter pattern is affected by the BT distance and the mechanism. Box plots in Fig. 2 compare the number of bloodstains of spatter patterns grouped by the BT distance and the mechanism. From the plots, we conclude: First, for the same mechanism, as the BT distance increases, on average the number of bloodstains in a spatter pattern decreases. This is supported by fluid dynamics since drops only travel a given horizontal distance, which increases with their initial diameter and their initial velocity. Second, on average, gunshot spatter patterns exhibit more bloodstains than blunt
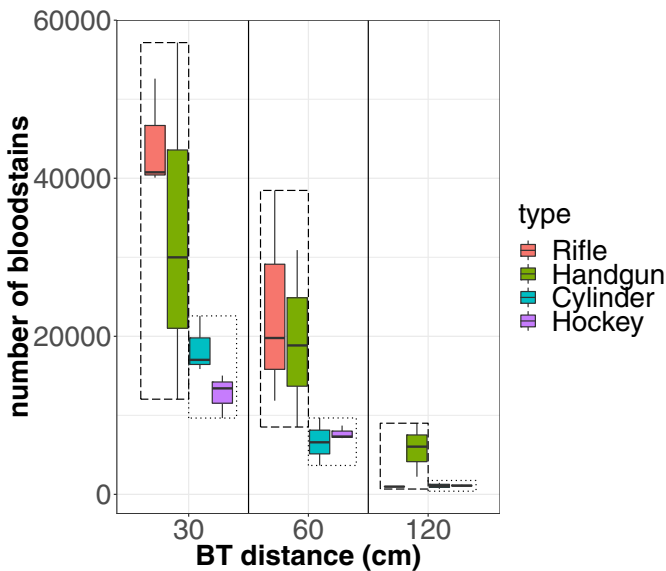
impact spatter patterns, at least for short and medium BT distances ($d \leq 60$ cm).

This latter fact can be associated with the higher velocity and kinetic energy imparted by the bullet to the blood (15). Although gunshot spatters tend to generate more bloodstains than impact spatters, the number of bloodstains is not by itself discriminative enough for classification (see Fig. 2).

Distribution of Stain Diameters

References (3,10,11) showed that higher velocity events produce patterns with smaller stains. The statistical distribution of stain diameters is studied with the following approach based on density function of the stain diameter. Fluid dynamics also dictate that only the larger drops can travel large distances, because they are less affected by drag and gravity than the smaller drops. It is thus expected that both the speed of impact associated with the generation mechanism and the BT distance affect the distribution of stain diameters.

The diameter of a bloodstain is estimated by diameter $= 2\sqrt{\text{Area}/\pi}$, where Area is the area of the bloodstain in mm^2. We use kernel density estimation (KDE) (31,32) to estimate the density of the diameter of the bloodstain in a spatter. Given $(X_1, \ldots, X_n)$, then the KDE at a given point x is as follows:

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where K is, for example, the Gaussian kernel with bandwidth h. Figure 3 compares the KDE of the diameter of the bloodstain of a gunshot backspatter pattern with a blunt impact spatter pattern at BT distance $d = 60$ cm, while omitting bloodstains with diameter larger than 3 mm, which are outliers.

Our results show that both the BT distance and the generation mechanism impact the density of the diameter of the bloodstain in a spatter pattern. At the same BT distance, a gunshot generates smaller bloodstains than a blunt impact and the dispersion
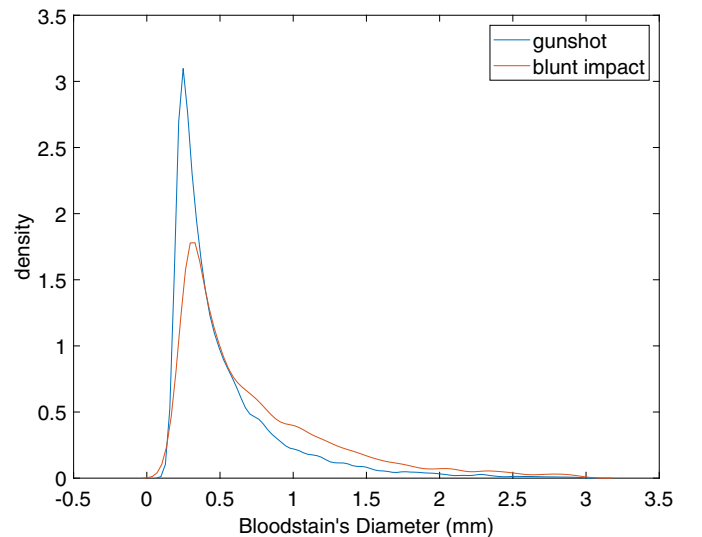

FIG. 2—*Box plots of number of bloodstains in a spatter against BT distance* d *and mechanism based on 36 spatters for feature engineering, where* $d \in \{30, 60, 120\}$ cm. *Colors represent different mechanisms. Box plots belong to either gunshot or blunt impact at the same BT distance, which are grouped by the long dashed and dotted rectangle, respectively.*

FIG. 3—*Comparison of the kernel density estimator (KDE) of the diameter of the bloodstain (≤3 mm) in a gunshot spatter "Hr27" and a blunt impact spatter "C5" at the BT distance* d = 60 cm.

of the diameter of the bloodstain is smaller, suggesting that bullets may produce smaller drops than impacts and that stain sizes produced by bullets have a smaller variance than those from blunt impact. Thus, both the mean and SD of the diameter of the bloodstains can be used to classify gunshot and blunt impact spatter patterns.

Figure 4a,b show box plots of sample mean and sample SD of the diameter of the bloodstain in a spatter pattern. First, sample mean diameter of the bloodstain can be used to classify gunshot patterns from blunt impact patterns at short and medium BT distances ($d \leq 60$ cm). Gunshots generate smaller bloodstains on average than blunt impact. In addition, sample SD of the diameter of the bloodstain distinguishes between impact and gunshot spatter patterns at short BT distance
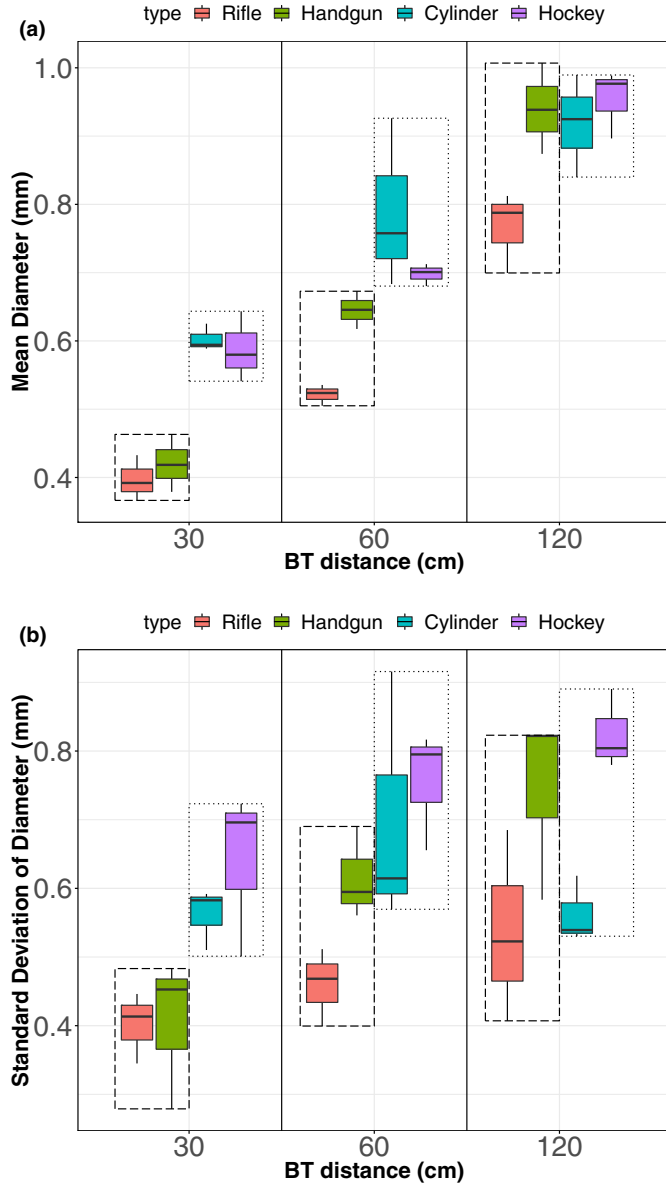
(30 cm). The dispersion of the diameter of the bloodstain in gunshot backspatter patterns is smaller than that of blunt impact spatter patterns at a BT distance of 30 cm. Second, those two features are a function of BT distance. The average diameter of the bloodstains in a spatter pattern becomes larger with increasing BT distance, because small blood drops are more affected by air resistance than heavier drops (33) and thus experience stronger deflections due to gravity, with the result that they do not reach the target. At large BT distance (120 cm), both features cannot differentiate between gunshot and blunt impact spatter patterns.

We also tested features based on higher order moments of the diameter distribution, for example, skewness and kurtosis, but according to the classifier these features are not significant to differentiate between blunt impact and gunshot spatter patterns.

Fraction of Large Stains

Fraction of large bloodstains has been shown to be influenced by the generation process of a spatter pattern (11). For each spatter pattern, we calculate the fraction of bloodstains with area larger than a threshold area s, fraction $= \sum_i \mathbb{I}[\text{Area}_i > s]/n$, where n is the total number of bloodstains in a spatter pattern and $\mathbb{I}[A]$ is the indicator function of an event A, that is, if A is true, $\mathbb{I}[A] = 1$ and zero otherwise.

For $s \in \left\{ \pi\left(\frac{0.3}{2}\right)^2, \pi\left(\frac{0.5}{2}\right)^2, \pi\left(\frac{0.8}{2}\right)^2, \pi\left(\frac{1}{2}\right)^2, \pi\left(\frac{3}{2}\right)^2 \right\}$mm^2, thresholds $\pi\left(\frac{0.8}{2}\right)^2$mm^2 and $\pi\left(\frac{1}{2}\right)^2$mm^2 are found best at discriminating between gunshot and blunt impact patterns based on 36 spatter patterns used for feature engineering. The sample Pearson correlation coefficient (34) between the mean area of the bloodstain and fraction of bloodstains larger than $\pi\left(\frac{0.8}{2}\right)^2$mm^2 is 0.98, indicating the fraction of large bloodstains is highly correlated with average size of the bloodstain. This feature helps distinguish between impact and gunshot spatter patterns at BT distance of 30 cm. By calculating this feature over all the stains of the cases with BT distance of 30 cm, we found that <15% of the bloodstains in a gunshot backspatter pattern are larger than $\pi\left(\frac{0.8}{2}\right)^2$mm^2, while impacts generate more than 15% large bloodstains.

FIG. 4—*Box plots of (a) sample mean diameter of the bloodstains (mm) (b) sample SD of the diameter of the bloodstain (mm) in a spatter against BT distance and mechanism based on 36 spatters for feature engineering. Long dashed and dotted rectangles represent gunshot and blunt impact box plots, respectively.*

Vertical Difference Between Large and Small Bloodstains

In Fig. 1, we define a three-dimensional coordinate system where the z direction is opposite to the gravity direction and the target is in the y-z plane at $x = d$ (BT distance). The origin of the two-dimensional coordinate system is at the lower left of the target cardstock. We construct a feature to measure the average vertical distance between large and small bloodstains. According to fluid dynamics, air resistance reduces the velocity of small blood drops more than that of larger drops. If the initial velocity is the same for small and large blood drops, small blood drops will eventually hit on the cardstock in a position lower than the large blood drops because gravity has had more time to modify their trajectories. Similarly, assuming that the blood drops generated by a bullet have higher initial velocity than the ones generated by impact, they are affected by gravity during a shorter period, possibly leading to smaller vertical distances between large and small bloodstains than in impact spatters.

Thus, we rank bloodstains in a spatter pattern by their areas and define the bloodstains whose area falls in the range [12.5th percentile, 37.5th percentile] as small bloodstains and the ones

with area within [62.5th percentile, 87.5th percentile] as large bloodstains. Bloodstains with area smaller than 12.5th percentile or larger than 87.5th percentile are considered as outliers and removed. The thresholds used to define large and small bloodstains are selected based on trial and error on 36 spatter patterns used for feature engineering. Finally, we construct the following feature delta.z to measure the shift in height between the large and small stains:

$$\text{delta}.z \equiv \frac{1}{\lceil 0.875n \rceil - \lceil 0.625n \rceil + 1} \sum_{i=\lceil 0.625n \rceil}^{\lceil 0.875n \rceil}$$

$$Z_i - \frac{1}{\lceil 0.375n \rceil - \lceil 0.125n \rceil + 1} \sum_{i=\lceil 0.125n \rceil}^{\lceil 0.375n \rceil} Z_i,$$

where Z_i indicates the vertical distance between the ith largest bloodstain to the bottom of the spatter, $\lceil x \rceil$ denotes the ceiling function, and n is the number of bloodstains in the spatter. The first term of the right-hand side is the mean vertical position of the large bloodstains, while the second term is that of small bloodstains. From Fig. 5, the proposed feature differentiates between gunshot and impact spatter patterns at BT distance 60 cm. Similar to our earlier discussion, the gunshot spatter patterns have less vertical difference between large and small bloodstains than blunt impact spatter patterns. However, BT distance 30 cm seems too close for gravity to take effect and BT distance 120 cm seems too far, in the sense that most bloodstains are large for both gunshot and blunt impact spatter patterns, leading to difficulties in differentiating spatter patterns at BT distances 30 and 120 cm.

Shape of the Bloodstain

Every bloodstain is automatically fitted with an ellipse as shown in Fig. 6. The measurements include:

1. Estimation of the impact angle α (in degrees) based on the ellipticity of the stain
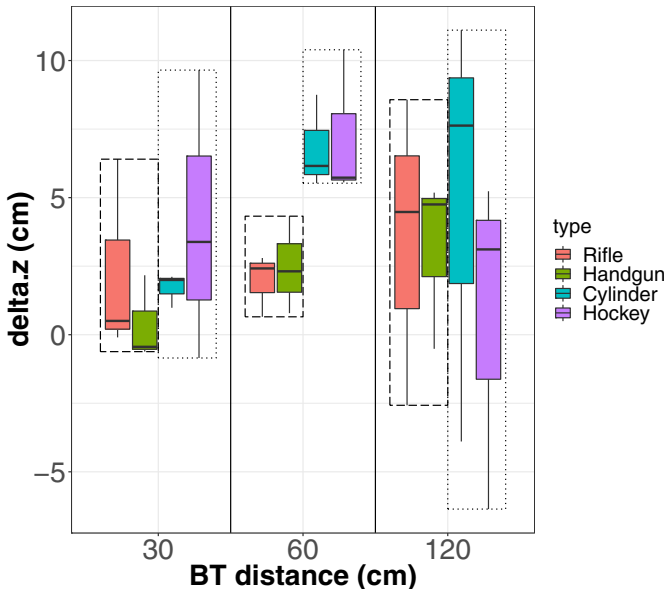
$$\alpha \equiv \frac{180}{\pi} \arcsin \frac{\text{mAxis}}{\text{MAxis}}$$

where *mAxis* and *MAxis* represent the length of the minor axis and the length of the major axis of the fitted ellipse, see Fig. 6. Note that the above formula is an approximation first formulated in Ref. (35) in analogy to bullet impact holes, which leads to systematic errors for specific impact conditions (36).

2. Adjusted impact angle α_ε: Some bloodstains have long tails on their backward part, thus the following ad hoc formula is used to better fit the ellipse to the forward part of the stain as follows

$$\text{Adj}_{\text{MAxis}} = MAxis \left(\frac{\text{FilledArea}}{\text{EllipsArea}}\right)^{\text{tail}_r\text{eduction}}, \quad \alpha_\varepsilon = \frac{\text{mAxis}}{\text{Adj}_{\text{MAxis}}},$$

where *EllipsArea* is the number of pixels in the fitted ellipse. *FilledArea* is the number of pixels in a filled image returned from *regionprops* in Ref. (20), and *tail_reduction* = 3 is an ad hoc parameter.

3. γ: Angle between the vertical reference line and the major axis of the fitted ellipse in Fig. 6. γ = 90°−Orientation, ranging from 0° to 180°, where Orientation is obtained from *regionprops* in Ref. (20). Note that no attempt is made to determine the forward or backward part of the ellipse, so that values of γ and γ + 180° have same meaning in the framework proposed here.

4. Eccentricity: the ratio of the distance between the foci of the fitted ellipse and its major axis length, measuring the circularity of the ellipse.

5. Solidity: proportion of the pixels in the convex hull that are also in the identified region.

We construct the sample mean and sample SD of the above five shape measurements of bloodstains in a spatter pattern as 10 new global features. However, after adding these global shape measurement features into the proposed random forests model
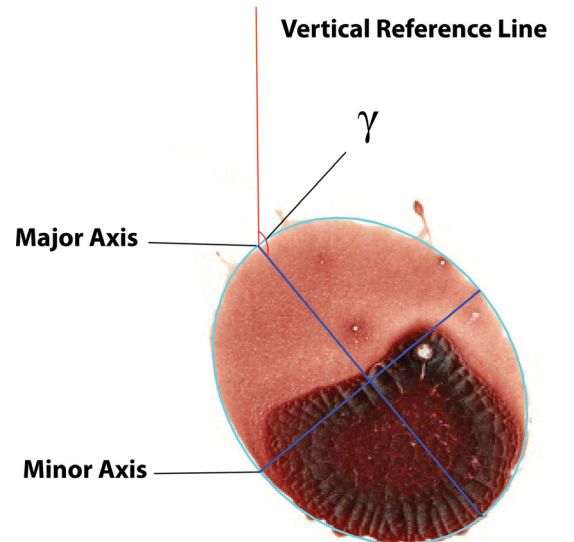

FIG. 5—*Box plots of delta.z in a spatter against BT distance and mechanism based on 36 spatters for feature engineering. Long dashed and dotted rectangles represent gunshot and blunt impact box plots, respectively.*

FIG. 6—*Illustration of the fitting procedure with an ellipse (cyan ellipse). Shape measurement γ of a bloodstain based on the fitted ellipse is defined by the angle between the vertical reference line and the major axis of the fitted ellipse.*

(section Modeling) and ranking the features by the importance, they turned out to be of minor importance for the classification. In contrast, the use of these shape measurements in local features proved relevant to classification, see the section Binning Methodologies and Local Features, below.

Distance of Bloodstains to the Centroid

We construct features which contain the location information of bloodstains in a spatter. Following the coordinate system in Fig. 1, the target is in the two-dimensional vertical plane supported by two perpendicular coordinate axes y and z. Given n bloodstains in a spatter, the observed positions for all bloodstains are (Y_i, Z_i) for $i = 1,\ldots n$ from the distribution of (Y,Z). We first define the centroid of a spatter pattern as the sample median position of all bloodstains in the spatter pattern as (median(Y), median(Z)). Given that there are outliers in the spatter, we choose the median instead of the mean for the sake of robustness.

We then define the variable *distance* as the Euclidean distance between each bloodstain and the centroid in a spatter. Given n bloodstains in a spatter, the observed Euclidean distances between bloodstains and the centroid are denoted as *distance*$_i$, $i = 1,\ldots n$. The sample median of these distances is a new feature median(*distance*) which depicts the average spread of bloodstains in a spatter. The box plots of the proposed feature are shown in Fig. 7a. At BT distance 30 cm, the median(*distance*) of gunshot patterns is typically larger than that of impact patterns. A possible reason is that bullets produce drops with larger radial dispersion than impacts, at least for short BT distances. Figure 7a also shows that the spread of bloodstains in a spatter pattern increases as BT distance increases (except gunshot spatter patterns at 30 cm), a phenomenon that can be explained geometrically as per Fig. 9. We use this finding to construct local features in next section. Note that the spread angle of gunshot spatters can be estimated theoretically as in Ref. (15).

To minimize the effect of BT distance on the spread of a spatter, we construct a new random variable $\frac{distance}{median(distance)}$, which can be interpreted as normalized *distance*. We construct the following two features to reflect the distribution of $\frac{distance}{median(distance)}$ for each spatter: (i) sample mean of $\frac{distance}{median(distance)}$ and (ii) sample SD of $\frac{distance}{median(distance)}$. In Fig. 7b, $\frac{mean(distance)}{median(distance)} > 1$ for impact spatter patterns at BT distance 30 cm and it is larger than gunshot spatter patterns, indicating that the distribution of the distance for the blunt impact spatter is skewed more to right at BT distance 30 cm, that is, the impact spatter pattern has obvious outliers, which are further away from the centroid.

Local Features and Binning Methodologies

All features described above are global features which contain the information for the entire spatter. From here on, we split the whole spatter pattern into small areas and construct features locally in the small area. Each area is called a bin, and so-called bin features represent the spatial information of a spatter pattern. This section first describes binning methodologies and then their use to construct local features.

Binning Methodologies

Fixed Bins: Annulus with Fixed Width—In this process, the bins are concentric annuli, with a common center, taken as the

centroid of the spatter pattern. An annulus is a ring-shaped object, a region bounded by two concentric circles. The width of the annulus is $r_0 = R - r$ with R the radius of the outside circle, and r the radius of the inner circle. Following the convention of Ref. (11), a bloodstain spatter pattern is segregated into 40 annuli with 2.5 cm width that radiate outward from the pattern's centroid. Figure 8 illustrates this process conceptually, and each annulus is labeled with an index $i = 1,\ldots.40$. The innermost annulus is labeled as bin 1 with the centroid as its center and radius r_0. For annulus with index i, the radius of inner circle is $(i-1)r_0$ and radius of the outside circle is $i \cdot r_0$. We set $r_0 = 2.5$ cm for all spatter patterns.

Adaptive Bins: Annulus with Width Proportional to the Median Distance—Similar to the previous feature, a bloodstain pattern is segregated into 40 equidistant annuli that radiate outward from the pattern center as shown in Fig. 8. However $r_0 = \frac{2}{40}$ median(distance), which varies among spatter patterns. The definition of *median(distance)* is discussed in the above

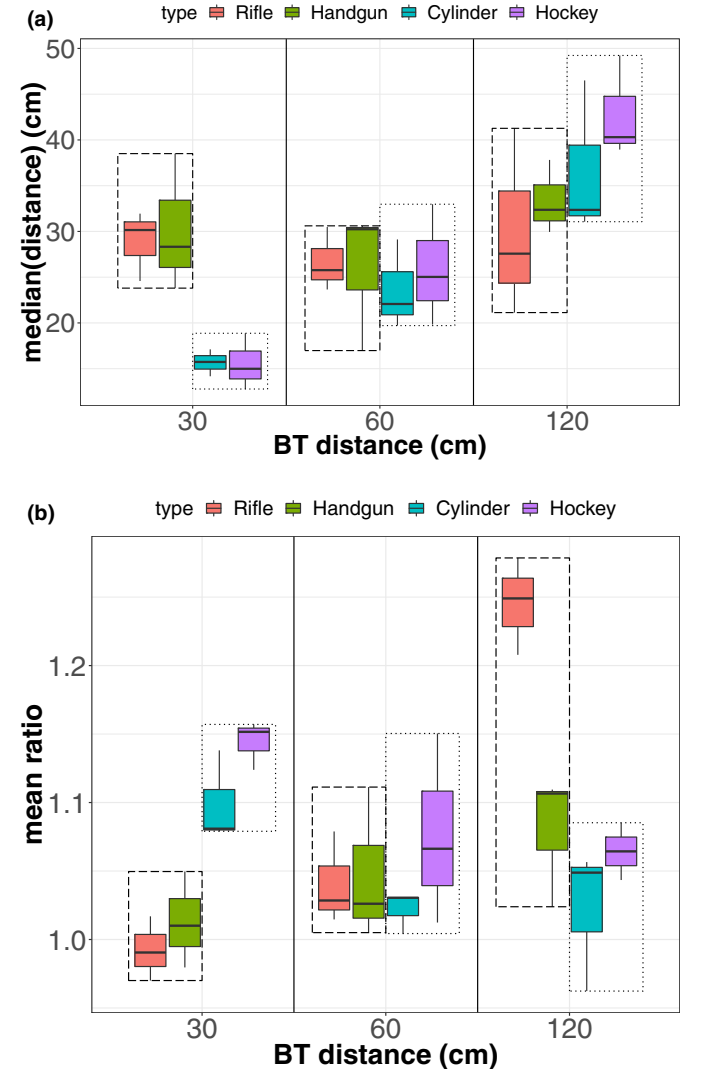FIG. 7—*Box plots of (a) median(distance) and (b) $\frac{mean(distance)}{median(distance)}$ in a spatter against BT distance and mechanism based on 36 spatters for feature engineering. Long dashed and dotted rectangles represent gunshot and blunt impact box plots, respectively.*
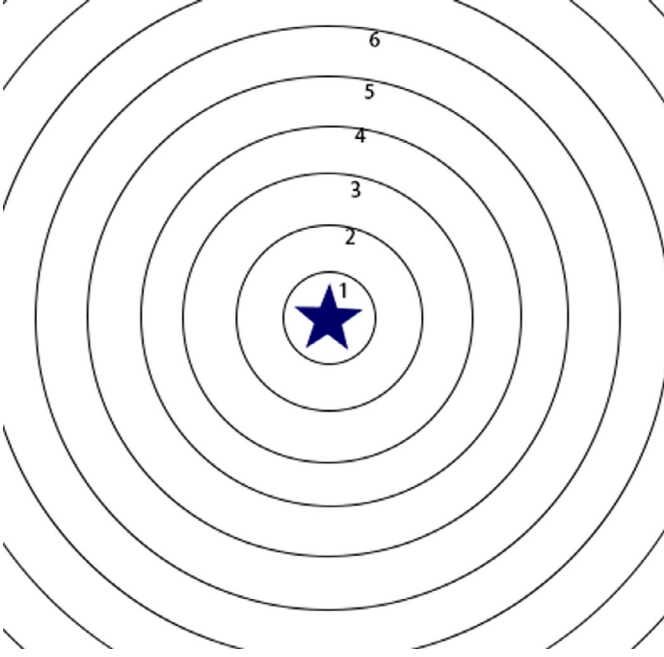
FIG. 8—*Concentric ring bins. The star represents the position of the centroid in the spatter.*
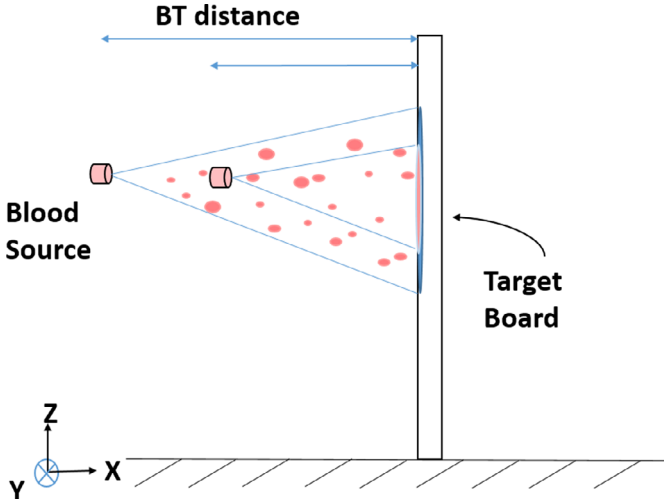

FIG. 9—*Cone shape bloodstain trajectory if gravity and air resistance are ignored. Spatters close to the target cardstock generate spatter patterns with smaller spread.*

section Distance of Bloodstains to the Centroid. The total number of annuli equates to 40, and the constant 2 is chosen such that more than 98% of bloodstains fall within the distance of $2 * median(distance)$ away from the centroid.

Figure 9 illustrates why we vary r_0 among spatter patterns. If the gravity and air resistance are ignored, the trajectory of drops is contained in a cone (15,16), whose base is roughly circular on the target surface. Moving the blood source toward the target surface, the area of the base progressively becomes smaller. In order to have a fair amount of bloodstains in bin i, $i = 1,\ldots40$ for spatter patterns at different BT distances, r_0 is expressed as a function of the area of the base. The median(distance) is used to reflect the area of the base. In this way, the bin with index i is comparable among spatter patterns at different BT distances.

Figure 10 shows an example comparing the two bin definitions. Handgun backspatter patterns at BT distances 30 and 120 cm are split into 40 bins according to the definition of fixed bins and adaptive bins. The fraction of bloodstains in bin i in spatter pattern j at BT distance d is the ratio of number of bloodstains in the ith bin to the total number of bloodstains in spatter pattern j and denoted as fraction_{ijd}, where $i = 1,\ldots.40$, $j = 1, 2, 3$, and $d = 30$ cm, 120 cm. The point (circle) of error bar plots at $x = i$ in Fig. 10 is defined as $\overline{\text{fraction}_{i.d}} = \frac{1}{3}\sum_{j=1}^{3} \text{fraction}_{ijd}$, which is over 3 replicated spatter patterns ($j = 1, 2, 3$) at distance d and
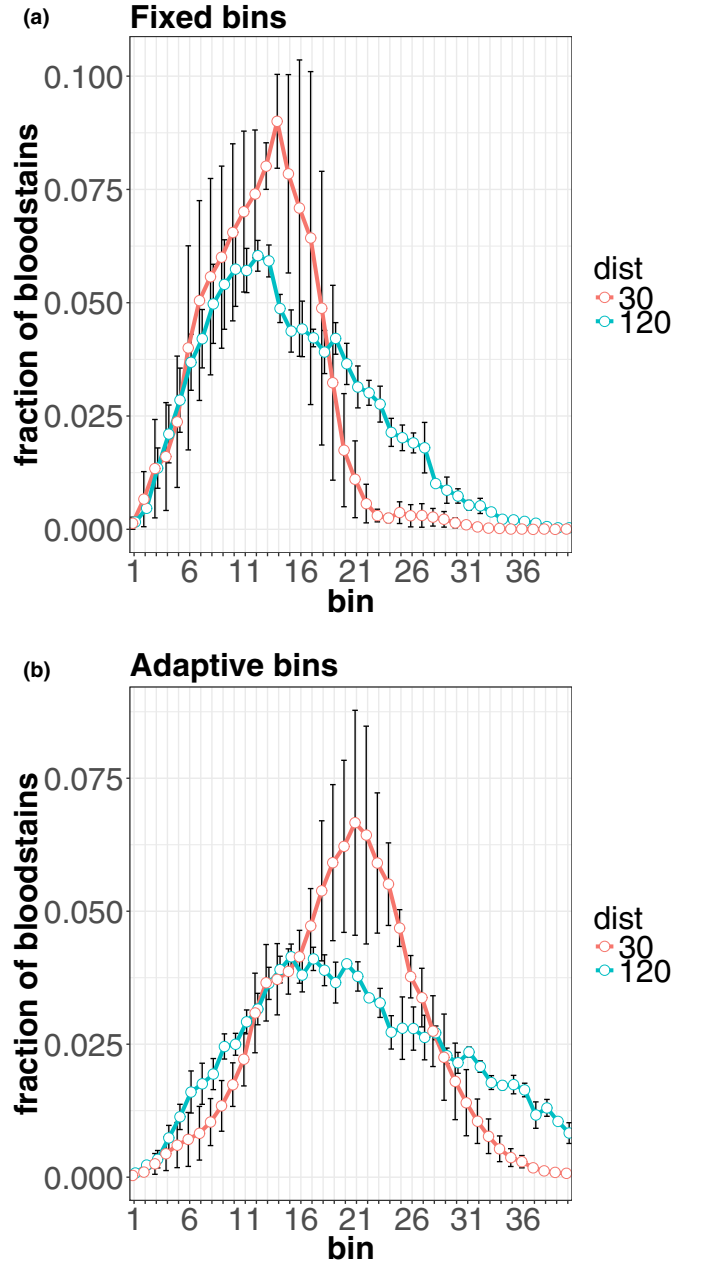

FIG. 10—*(a) Fixed bins. (b) Adaptive bin. Error bar plots of the fraction of bloodstains in each bin against the index of the bin grouped by BT distances 30 and 120 cm for handgun backspatters in the feature engineering data set. Points (circle) and half-length of the bars in error bar plots represent the mean and the standard error of the mean over three trial replicates.*

bin i. The half-length of each error bar is the standard error of the mean over three trial replicates. We connect points with curves to easily observe the curvature information.

Figure 10a illustrates that when fixed bins are used at the short BT distance 30 cm, almost no bloodstains fall further than bin 22. For BT distance 120 cm, there is a fair amount of bloodstains appearing in bin 22–30. BT distance has thus a significant impact on features with fixed bin size. Therefore, adaptive binning is used in Fig. 10b. The curves at BT distances 30 and 120 cm are reshaped such that most bins contain at least a few bloodstains. In addition, the curves are smoothed (especially for small BT distance), to facilitate the comparison of derivatives over bins between gunshot and blunt impact spatter patterns, as done in the section Shape of Curve Over Adaptive Bins.

Rectangular Bins—A different asymmetric definition of bins is to segregate the spatter pattern area into equal-size rectangular areas along the vertical direction. We define the random variable Z as the vertical distance between the bloodstain to the bottom of the spatter. Given n bloodstains, the observed vertical distances are Z_i, $i = 1, \ldots n$. The width of each rectangle is the same as the width of the spatter pattern and the length of rectangle is $\frac{6}{40}\text{median}(|Z - \text{median}(Z)|)$, where $\text{median}(|Z - \text{median}(Z)|)$ is the median absolute deviation of vertical distance (in z direction) between the bloodstain to the centroid. The total number of bins equals 40, and the factor 6 is chosen so that more than 81% of the bloodstains are falling within a distance of $3*\text{median}(|Z - \text{median}(Z)|)$ from the centroid in the vertical direction and are covered by rectangular bins (see Fig. 11). Both the lower side of rectangle 21 and the top side of rectangle 20 pass through the centroid.

Construction of Local Features

Local features are typically constructed by defining features in each bin. These are called local features; comparison between

FIG. 11—*Rectangle bins. The star represents the position of centroid in the spatter.*

bins helps describe how the features are spatially distributed within a given pattern. We illustrate the construction of local features with the following example: the fraction of bloodstains with diameter larger than 1 mm in each bin.

Fraction of Large Stains in Bins—The construction procedure is

---

ALGORITHM—*Local Feature Construction Example*

For BT distance $d = 30$, 60, and 120 cm:

1. Choose a binning method, for example, fixed bins with a width of 2.5 cm.
2. For bin i, $i = 1, \ldots 40$, calculate the ratio of bloodstains with diameter larger than 1 mm in bin i to the number of bloodstains in bin i, which is ratio_i. We select the threshold 1 mm because it is the best at discriminating between gunshot and blunt impact spatter patterns based on feature engineering data set.
3. Plot error bars of the fraction of bloodstains with diameter larger than 1 mm against the index of the bin over 3 replicated spatter patterns (see Fig. 12a).
4. Select a set of bins S in which ratio_i shows a difference between gunshot and blunt impact spatter patterns.
5. construct a new local feature LF by simple average: $\text{LF} = \frac{1}{|S|}\sum_{i \in S}\text{ratio}_i$, where $|S|$ is the cardinality of the set S.

Algorithm 1.

Although it is difficult to classify spatter patterns at BT distance 120 cm, for BT distances 30 and 60 cm, in bins with index $i \in \{15, 16, \ldots 25\}$, the fraction of large bloodstains of gunshot spatter patterns is significantly smaller than that of blunt impact spatter patterns. The simple average of the fraction of bloodstains with diameter larger than 1 mm from bin 15 to 25 in a spatter pattern is a new local feature, shown in Fig. 12b. The proposed feature is useful to classify impact from gunshot spatter patterns at BT distances 30 and 60 cm.

Following this method, one can construct additional features by selecting different combinations of the set of bins, binning method, and averaging method (e.g., weighted average).

Shape of Stains in Bins—Our study found that shape measurements in local features are proved relevant to classification, for example, the average mean adjusted impact angle of the bloodstain over rectangle bin 27,…, 30 in Fig 15a.

Shape of Density Curve Over Adaptive Bins—Figure 13a shows the error bar plots of the fraction of bloodstains in each adaptive bin for spatter patterns at BT distance 30 cm grouped by process (gunshot or blunt impact). We plot a curve connecting the points which are the mean fractions over 3 replicated spatter patterns, with red and blue curves representing the blunt impact and gunshot spatter patterns, respectively.

To smooth the data, we use local polynomial regression (37). Consider the data $(X_1, Y_1), \ldots, (X_n, Y_n)$ which form an independent and identically distributed (i.i.d.) sample from a population (X, Y), where X_i is the index of the ith adaptive bin, Y_i is the fraction of bloodstains in ith adaptive bin, and n is the total number of bins. The local polynomial regression estimator in an arbitrary point x is obtained by minimizing the following

(a)

BT distance 60cm

type ∘ Rifle ∘ Handgun ∘ Cylinder ∘ Hockey

(b)

Local feature

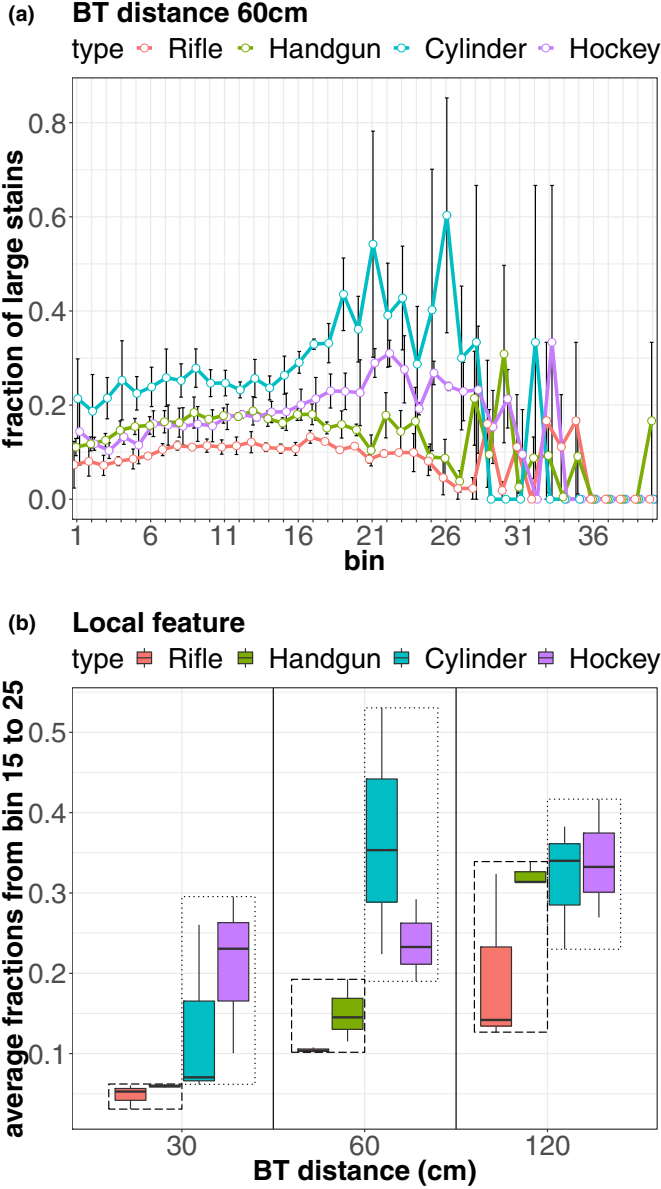type ⊟ Rifle ⊟ Handgun ⊟ Cylinder ⊟ Hockey

FIG. 12—*(a) Error bar plots of the fraction of bloodstains with diameter larger than 1mm against the index of the bin for spatters at BT distance 60 cm grouped by mechanism. Points and half-length of the error bar represent the mean and the standard error of the mean over three trial replicates. (b) Box plots of the average of the fraction of large stains with the diameter larger than 1mm from bin 15 to bin 25 against BT distance grouped by mechanism. Long dashed and dotted rectangles represent gunshot and impact box plots, respectively. Both (a) and (b) are based on the 36 spatters used for feature engineering.*

weighted least squares problem (37):

$$\min_{\beta_j \in \mathbb{R}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right\}^2 K_h(X_i - x).$$

where β_j are the solutions to the weighted least squares problem. We select the Gaussian kernel $K_h(u) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{u^2}{2h^2}}$. The bandwidth h is selected from $\{1, 1.006, \ldots, 4\}$. According to the properties of local polynomial regression, $\hat{m}^{(q)}(x) = q! \hat{\beta}_q$ is an estimator

for the qth order derivative where $q = 0, \ldots, p$. In Fig. 13b, the red and blue points represent the fraction of bloodstains in each adaptive bin for the respective impact and gunshot spatter pattern at BT distance 60 cm. The fitted regression curves ($q = 0$) are estimated using the above equation with $p = 1$ for the respective impact and gunshot patterns. From both plots in Fig. 13, it is obvious that the regression function for gunshot patterns is convex over bins 1,…, 10 and 30,…., 40, while concave for blunt impact spatter patterns. The physical meaning of this difference is not clear at the present time. In addition, the absolute values of the first-order derivatives for the gunshot curve are larger than the blunt impact curve over bins 10–15 and 25–30. We construct the estimated first- and second-order derivatives at different adaptive bins using local quadratic ($p = 2$) and cubic ($p = 3$) regression as new features to measure the shape of curves.
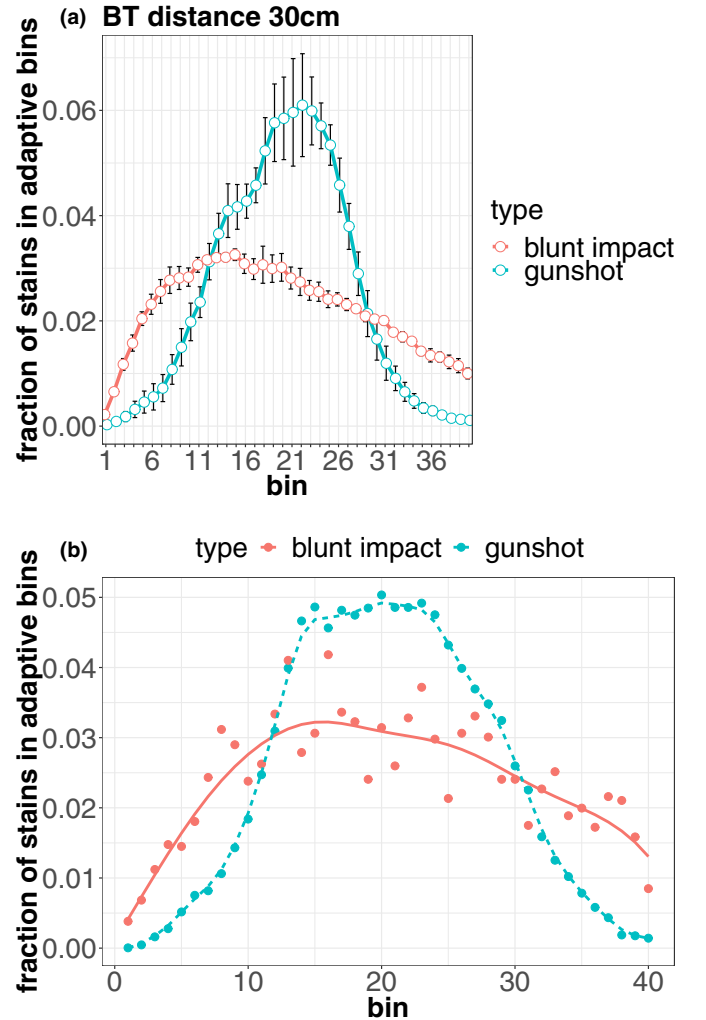
(a)

BT distance 30cm

type ∘ blunt impact ∘ gunshot

(b)

type • blunt impact • gunshot

FIG. 13—*(a) Error bar plots show the fraction of bloodstains in each adaptive bin against the index of the bin for spatters at BT distance 30 cm grouped by process based on feature engineering data. Point and half-length of the error bar represent the mean and the standard error of the mean over three replicates. (b) Blue points are fractions of bloodstains in adaptive bins for the gunshot spatter "Rp45" and Red points are for the blunt impact spatter "C6." Both spatters are generated at BT distance 60 cm. Blue dashed line and red solid line are the local linear regression curves for the gunshot and blunt impact spatter, respectively.*

Modeling

Random Forests

Random forests is an automatic learning method for classification, which can determine which features are most useful for classification. Random forests apply bagging, which averages the prediction over a collection of bootstrap samples, to decision tree (38–40), while only selecting a part of the predictors (features), say $m \approx \sqrt{\text{number of features}}$, at random in each split of the tree. Algorithm 2 illustrates the training and prediction using the random forests in the classification problem.

ALGORITHM—*Random Forests for Classification*

For $b = 1$ to B:

1. Bootstrap sample from the training set. (The bootstrap sample is obtained by sampling with replacement from the training set with sample size equal to the training set.)
2. Fit a classification tree to bootstrapped sample by splitting every terminal node whose size is larger than the minimum size of terminal nodes:
   a Randomly sample m features.
   b Pick the best split among m and split node into two children nodes.

The predictions for an unseen sample are made by taking the majority votes of the classification trees.

Features

A brief description of each feature is shown in Table 2. Dedicating 36 spatter patterns to feature engineering, 58 spatter patterns are left for training and testing. Thus, the feature matrix for the remaining spatter patterns is 58 × 58, which is composed of 58 features from Table 2 for 58 spatter patterns.

We use the following procedure to determine the average performance and variability of the model:

1. Randomly split spatter patterns into a training (75%) and test (25%) set. Values of the label $y \in \{\text{blunt impact, gunshot}\}$.
2. Use random forests (38–40) on the training set and obtain predictions on the test set.
3. Compare the predicted labels and true label for test set.
4. Repeat (1) – (3) 1000 times to obtain the average performance and variability of the model.

In step (2), the number of features randomly sampled as candidates at each split of each tree is set to $\sqrt{58}$, and the minimum size of the terminal nodes in each tree is 1. The number of trees is set to 5000.

Results

Simulation Results

Two metrics are used to evaluate the performance of the classifier (38,39):

1. Misclassification (MSC) rate:

$$\frac{\sum \mathbb{I}[\text{predicted label} \neq \text{actual label}]}{\text{size of test set}}$$

2. Out-of-bag (OOB) error:

For each tree in the random forests, only 2/3 of the bootstrap sample from the original data is used to construct the tree and the remaining samples are referred to as the out-of-bag (OOB) samples. We can predict the label for ith sample using each of trees in which that observation was OOB and take a majority vote, leading to the OOB prediction for ith sample. The OOB error is the classification error for all samples with predication obtained in this way and OOB error is asymptomatic approximation to the leave-one-out cross-validation (LOO-CV) error (25).

Both MSC rate and OOB error are obtained at each iteration. The performance of the proposed classifier is shown in Fig. 14 based on the MSC rate and OOB error over 1000 replications. In order to study the performance of the proposed classifier at different BT distance ranges among 58 spatter patterns, we perform a Monte Carlo simulation on the spatter patterns with BT distance no further than 30, 60, and 120 cm separately. Table 3 shows the total number of spatter patterns used for training (75%) and testing (25%), the mean MSC error, the mean OOB error, and the mean accuracy, where accuracy = 1 − OOB error in each replication. For spatter patterns at short BT distances $d \leq 30$ cm, the proposed classifier achieves 98.81% accuracy, and for spatter patterns with BT distances $d \leq 60$ cm, the proposed classifier achieves 93.20% accuracy on average. For spatter patterns with BT distances $d \leq 120$ cm, the average accuracy

TABLE 2—*All features used for classification.*

| Feature Name	Brief Description
number.stains	Number of bloodstains
Diameter.mean	Mean diameter of the bloodstain
Diameter.std	SD diameter of the bloodstain
fc	Ratio of number of bloodstains with area larger than $\pi\left(\frac{0.8}{2}\right)^2$mm^2 to the total number of bloodstains
fd	Ratio of number of bloodstains with area larger than $\pi\left(\frac{1}{2}\right)^2$mm^2 to the total number of bloodstains
delta.z	Vertical difference between large and small bloodstains
mean.ratio	Sample mean of $\frac{\text{distance}}{\text{median(distance)}}$, where distance is the Euclidean distance between the bloodstain and the centroid
sd.ratio	Sample SD of $\frac{\text{distance}}{\text{median(distance)}}$
fraction1_bin_15_25	Average of the fraction of bloodstains with area larger than $\pi\left(\frac{1}{2}\right)^2$mm^2 in fixed bins 15, 16,..., 25
fraction1_bin_35_40	Average of the fraction of bloodstains with area larger than $\pi\left(\frac{1}{2}\right)^2$mm^2 in fixed bins 35, 36,..., 40
fraction075_bin_1_17	Average of the fraction of bloodstains with area larger than $\pi\left(\frac{0.75}{2}\right)^2$mm^2 in fixed bins 1, 2,..., 17
fraction075_bin_20_21	Average of the fraction of bloodstains with area larger than $\pi\left(\frac{0.75}{2}\right)^2$mm^2 in fixed bins 20, 21
fraction075_adp_25_31	Average of the fraction of bloodstains with area larger than $\pi\left(\frac{0.75}{2}\right)^2$mm^2 in adaptive bins 25, 26,..., 31
fraction1_adp_27_30	Average of the fraction of bloodstains with area larger than $\pi\left(\frac{1}{2}\right)^2$mm^2 in adaptive bins 27, 28,..., 30
alpha_adp_23_30	Average of mean impact angle of the bloodstain over adaptive bins 23, 24,...30
epsilon_adp_23_30	Average of mean adjusted impact angle of the bloodstain over adaptive bins 23, 24,...30
epsilon_rec_27_30	Average of mean adjusted impact angle of the bloodstain over rectangle bins 27, 28,...30
derivatives2bin_i	Second-order derivatives of the regression function of the fraction of bloodstains at the adaptive bin $i \in \{1, 2,...9, 30, 31,...39\}$
derivatives1bin_i	First-order derivatives of the regression function of the fraction of bloodstains at the adaptive bin $i \in \{5, 6,...15, 25, 26,...35\}$

(a)

(b)
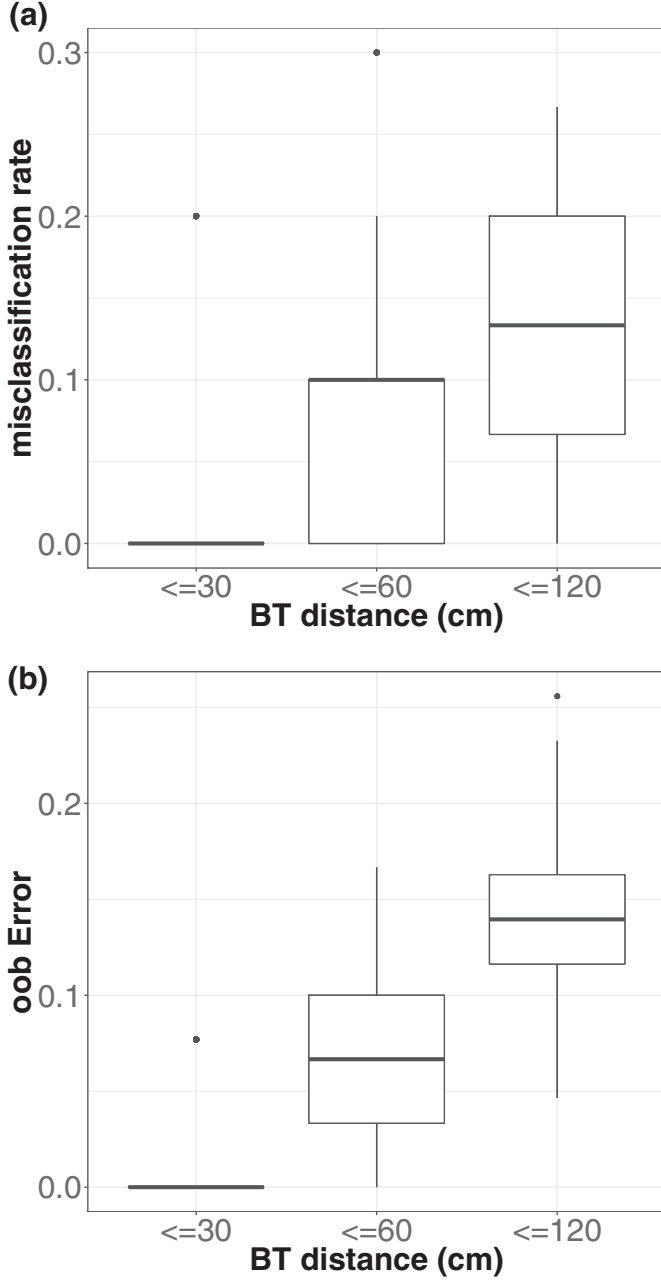
FIG. 14—*Monte Carlo study with 1000 runs for the proposed classifier at different BT distance ranges using (a) misclassification rate and (b) Out-of-bag error.*

TABLE 3—*Performance of the model.*

BT Distance (cm)	Number of Spatter Patterns	OOB Error (Average)	MSC Rate (Average)	Accuracy (Average)
$d \leq 30$ cm	18	1.19%	1.04%	98.81%
$d \leq 60$ cm	40	6.80%	7.08%	93.20%
$d \leq 120$ cm	58	14.04%	14.00%	85.96%

drops to 85.96%. From the discussion in the section Global Features and section Binning Methodologies and Local Features, many features, "Diameter.mean," "Diameter.std," "fc," "mean.ratio" (in Table 2), etc, could distinguish between blunt impact and gunshot spatter patterns at BT distance 30 cm based on feature engineering data with a high accuracy, while it is difficult

to find such a feature or a set of features at BT distance 120 cm. In addition, since the BT distance d is treated as unknown, it is challenging to classify the spatter patterns at different BT distances correctly.

Figure 15*a* shows the six most important features selected by random forests based on a one-time simulation for spatter patterns at BT distances $d \leq 120$ cm. The features are sorted in decreasing order of importance. Figure 15*b* shows OOB error based on that one-time simulation.

Based on 1000 simulations, the distribution of the size of bloodstains is important (i.e., mean and SD of the diameter of the bloodstain), which coincides with the observations in Refs (3,10). The "average size" feature has also been found important for classification between cast-off and impact spatter patterns (21). In our data
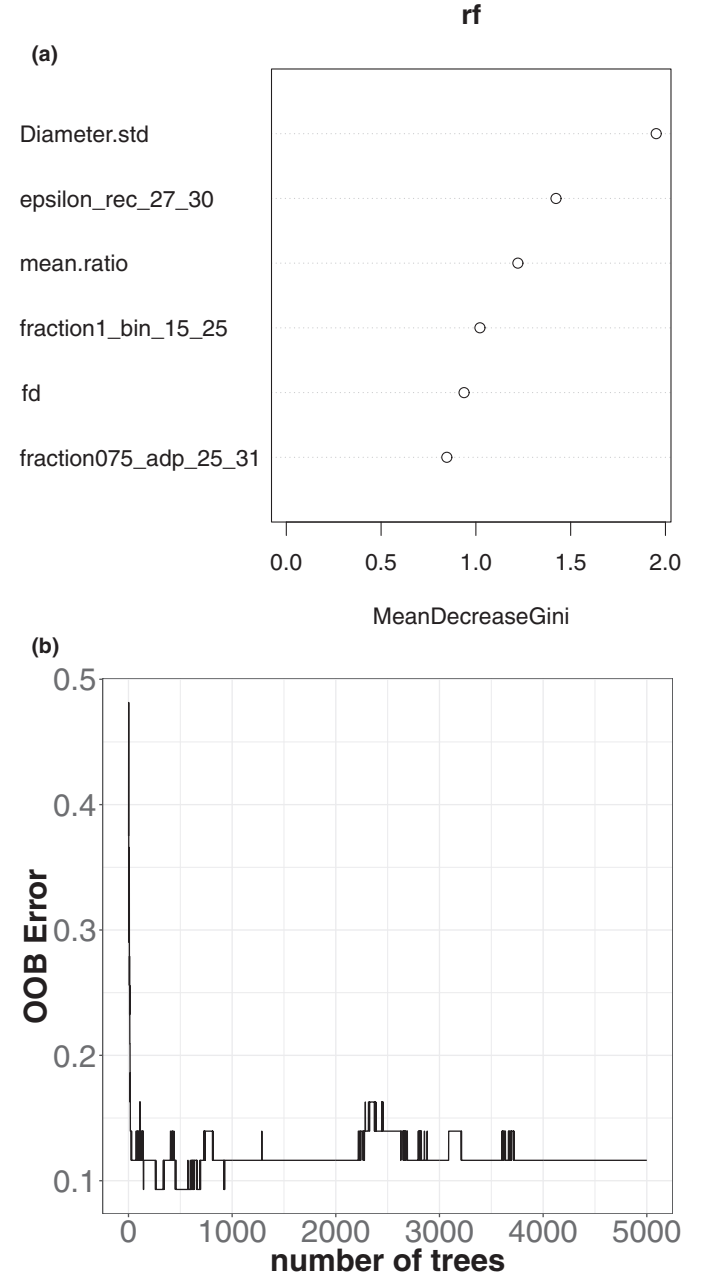
rf

(a)

(b)

FIG. 15—*(a) Important feature. (b) OOB error of random forests against number of trees. (a) Top six important features and (b) OOB error for all trees up to the ith in one simulation.*

set, the SD of the bloodstain diameter is the most important. The feature "fraction of large stains" is also an important feature as per Fig. 15a, as shown in Ref. (11). Figure 15a also shows that "mean.ratio" and other local shape features are also important. When the procedure is repeated 1000 times, the rank of important features depends on the choice of the training and testing data set, possibly indicating that some of these features carry the same information. Due to the data size limitation, a further study on a larger data set would be recommended.

Further Discussions

Include Feature Engineering Data

Using 94 spatter patterns (i.e., including the 36 feature engineering spatter patterns) in the Monte Carlo study to split for the training (75%) and test (25%) set, error rates obtained after 1000 simulations (Table 4) are similar to the ones obtained in Table 3, without including the 36 patterns used for feature engineering.

Based on the Monte Carlo study with 1000 runs, the proposed model achieves 99.72% accuracy at BT distance $d \leq 30$ cm.

Performance on Gunshot Spatter patterns Involving Muzzle Gases

As presented in the section "Bloodstain Spatter patterns and Digital Image Preprocessing," the data set (23) contains 10 gunshot backspatter patterns in which muzzle gases interact with the backspatter. We did not use those spatter patterns in previous sections for the sake of simplicity. This section determines the performance of the proposed classification model on gunshot backspatter patterns with muzzle gases. Among the 10 gunshot backspatter patterns involving muzzle gases, the number of spatter patterns at each BT distance is given in Table 5:

We use random forests on 94 spatter patterns with 58 features listed in Table 2 to train the model and classify the 10 gunshot spatter patterns involving muzzle gases. The misclassification rate is 0% for spatter patterns at BT distances 30 and 60 cm, and 100% at BT distance 120 cm. This indicates that the proposed model and features also work on gunshot backspatter patterns with muzzle gases at short and medium BT distances (30, 60 cm); however, they do not work for gunshot backspatter patterns with muzzle gases at BT distance $d = 120$ cm. A possible cause for this is that interactions between muzzle gases and spatter drops are more important for $d = 120$ cm than for shorter distances. Certainly, the production of drops during gunshot and the propagation of muzzle gases have different time scales, which is a topic of current research (14,16,27,28).

Conclusion and Future Work

This study proposes a machine learning method to classify between blood spatter patterns generated by blunt impact and

TABLE 5—*Number of gunshot backspatter patterns involving muzzle gases.*

	BT Distance (cm)		
	30	60	120
Number of spatter patterns	2	4	4

gunshot backspatter for cases where muzzle gases do not interfere with the spatter. The patterns are generated at various horizontal distances between blood source and stained target surface. Ninety-four spatter patterns are considered, in which 36 spatter patterns are used for feature engineering and the rest are used for training and testing. We engineer 58 features and several binning methods to segment images to construct local features. Random forests is used to fit to the training set and to predict the label on the test set. Based on 1000 replications of random splitting training and test data, the proposed method achieves 98.81% accuracy in classifying generation mechanisms at BT distance no larger than 30 cm, 93.20% accuracy to classify spatter patterns at BT distance no larger than 60 cm, and 85.96% accuracy on spatter patterns at BT distances no larger than 120 cm.

A main contribution of this paper is to propose a formal quantitative and physics-based method to engineer features relevant to classification, especially the local features. Another contribution is to propose a novel and quantitative method to classify bloodstain spatter patterns by combining digital image processing method and machine learning methods.

The classification model is then used to classify gunshot backspatter patterns where muzzle gases are present and possibly interact with the motion of drops. In that case, the model achieves a classification accuracy of 100% at BT distance of 30 and 60 cm. However, at BT distance 120 cm, all the backspatter patterns are classified as impact patterns, corresponding to a classification accuracy of 0%.

We emphasize that all the bloodstain patterns studied here were generated under controlled laboratory conditions, which may differ from actual crime scenes because of more complicating and compound factors involved. The effect of those complicated factors needs further study. However, the key point of this paper is to demonstrate a methodology that can be adapted to classification of spatter patterns created with different mechanisms.

Despite limitations on the availability of open access spatter data and real crime scene patterns, it would be important to validate this method on other spatter data.

The results of our study demonstrate the importance on classification accuracy of the distance (BT distance) between blood source and the target where stains are collected. Features such as the mean diameter of the bloodstains, the SD of diameter of the bloodstains, and the ratio of large bloodstains are found important and effective to classify gunshot vs. impact spatter patterns at a short and medium BT distance (≤60 cm). However, the same features may not work well for larger BT distances especially when muzzle gases exist. This important result suggests that classification between impact and backspatter pattern is only reliable at BT distances no larger than 60 cm.

Future work might focus on applying the same classification method to other spattering mechanisms, such as the expiration spatter patterns recommended by one of the reviewers. A prerequisite of this effort certainly is the availability of more open access data sets of spatter patterns produced under a variety of controlled conditions representative of actual crime scenes.

TABLE 4—*Performance of the model including the feature engineering data.*

BT distance (cm)	Number of Spatter Patterns	OOB Error (Average)	MSC Rate (Average)	Accuracy (Average)
$d \leq 30$ cm	30	0.28%	0.14%	99.72%
$d \leq 60$ cm	64	8.00%	8.40%	92.00%
$d \leq 120$ cm	94	13.90%	13.60%	86.10%

References

1. Kirk PL. Affidavit of Paul Leland Kirk. State of Ohio vs. Samuel H. Sheppard, 1955. https://static1.squarespace.com/static/543841fce4b0299b22e1956a/t/54be8edee4b0d16cfa14fd21/1421774558208/KirkAffadavit-Shepherd+Case.pdf (accessed December 4, 2019).

2. Kirk PL. Crime investigation: physical evidence and the police laboratory. New York, NY: Interscience Publishers Inc., 1953.

3. MacDonell HL, Bialousz LF. Flight characteristics and stain patterns of human blood. Washington, DC: National Institute of Law Enforcement and Criminal Justice, 1971.

4. Karger B, Rand S, Fracasso T, Pfeiffer H. Bloodstain pattern analysis case work experience. Forensic Sci Int 2008;181:15–20.

5. Bevel T, Gardner RM. Bloodstain pattern analysis with an introduction to crime scene reconstruction, 3rd edn. Boca Raton, FL: CRC Press, 2008;165–97, 212–4.

6. Knock C, Davison M. Predicting the position of the source of blood stains for angled impacts. J Forensic Sci 2007;52(5):1044–19.

7. de Bruin KG, Stoel RD, Limborgh J. Improving the point of origin determination in bloodstain pattern analysis. J Forensic Sci 2011;56(6):1476–82.

8. Varney CR, Gittes F. Locating the source of projectile fluid droplets. Am J Phys 2011;79(8):838–42.

9. Camana F. Determining the area of convergence in bloodstain pattern analysis: a probabilistic approach. Forensic Sci Int 2013;231(1–3):131–6.

10. James SH, Kish PE, Sutton TP. Principles of bloodstain pattern analysis: theory and practice. Boca Raton, FL: CRC Press, 2005;7–8.

11. Siu S, Pender J, Springer F, Tulleners F, Ristenpart W. Quantitative differentiation of bloodstain patterns resulting from gunshot and blunt force impacts. J Forensic Sci 2017;62(5):1166–79.

12. Laber T, Kish P, Taylor M, Owens G, Osborne N, Curran J. Reliability assessment of current methods in bloodstain pattern analysis. Washington, DC: National Institute of Justice, US Department of Justice, 2014.

13. Taylor MC, Laber TL, Kish PE, Owens G, Osborne N. The reliability of pattern classification in bloodstain pattern analysis, part 1: bloodstain patterns on rigid non-absorbent surfaces. J Forensic Sci 2016;61(4):922–7.

14. Comiskey PM, Yarin AL, Attinger D. Theoretical and experimental investigation of forward spatter of blood from a gunshot. Phys Rev Fluid 2018;3(6):063901.

15. Comiskey PM, Yarin AL, Attinger D. Hydrodynamics of back spatter by blunt bullet gunshot with a link to bloodstain pattern analysis. Phys Rev Fluid 2017;2(7):073906.

16. Comiskey PM, Yarin AL, Attinger D. High-speed video analysis of forward and backward spattered blood droplets. Forensic Sci Int 2017;276:134–41.

17. Arthur RM, Cockerton SL, de Bruin KG, Taylor MC. A novel, element-based approach for the objective classification of bloodstain patterns. Forensic Sci Int 2015;257:220–8.

18. De Chazal P, Flynn J, Reilly RB. Automated processing of shoeprint images based on the Fourier transform for use in forensic science. IEEE Trans Pattern Anal Mach Intel 2005;27(3):341–50.

19. Arthur RM, Humburg PJ, Hoogenboom J, Baiker M, Taylor MC, de Bruin KG. An image-processing methodology for extracting bloodstain pattern features. Forensic Sci Int 2017;277:122–32.

20. MATLAB. R2018b. Natick, MA: The MathWorks Inc., 2018.

21. Arthur RM, Hoogenboom J, Baiker M, Taylor MC, de Bruin KG. An automated approach to the classification of impact spatter and cast-off bloodstain patterns. Forensic Sci Int 2018;289:310–9.

22. Attinger D, Liu Y, Bybee T, De Brabanter K. A data set of bloodstain patterns for teaching and research in bloodstain pattern analysis: impact blunt impact spatters. Data Brief 2018;18:648–54.

23. Attinger D, Liu Y, Faflak R, Rao Y, Struttman BA, De Brabanter K, et al. A data set of bloodstain patterns for teaching and research in bloodstain pattern analysis: gunshot backspatters. Data Brief 2019;22:269–78.

24. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction, 2nd edn. New York, NY: Springer, 2009;219–57.

25. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. New York, NY: Springer, 2013;15–37.

26. Stephens BG, Allen TB. Back spatter of blood from gunshot wounds-Observations and experimental simulation. J Forensic Sci 1983;28(2):437–9.

27. Taylor MC, Laber TL, Epstein BP, Zamzow DS, Baldwin DP. The effect of firearm muzzle gases on the backspatter of blood. Int J Legal Med 2011;125(5):617–28.

28. Comiskey PM, Yarin AL. Self-similar turbulent vortex rings: interaction of propellant gases with blood backspatter and the transport of gunshot residue. J Fluid Mech 2019;876:859–80.

29. Grabmuller M, Cachee P, Madea B, Courts C. How far does it get?–The effect of shooting distance and type of firearm on the simultaneous analysis of DNA and RNA from backspatter recovered from inside and outside surfaces of firearms. Forensic Sci Int 2016;258:11–8.

30. Rossi C, Herold LD, Bevel T, McCauley L, Guadarrama S. Cranial backspatter pattern production utilizing human cadavers. J Forensic Sci 2018;63(5):1526–32.

31. Parzen E. On estimation of a probability density function and mode. Ann Math Stat 1962;33(3):1065–76.

32. Rosenblatt M. Remarks on some nonparametric estimates of a density function. Ann Math Stat 1956;27(3):832–7.

33. Attinger D, Moore C, Donaldson A, Jafari A, Stone HA. Fluid dynamics topics in bloodstain pattern analysis: comparative review and research opportunities. Forensic Sci Int 2013;231(1–3):375–96.

34. Pearson K. Note on regression and inheritance in the case of two parents. Proc R Soc Lond 1895;58:240–2.

35. Rizer C. Police mathematics. Springfield, IL: Charles C. Thomas, 1955;72–3.

36. Adam CD. Fundamental studies of bloodstain formation and characteristics. Forensic Sci Int 2012;219(1–3):76–87.

37. Fan J, Gijbels I. Local polynomial modelling and its applications: monographs on statistics and applied probability 66. London, U.K.: Chapman & Hall/CRC Press, 2003;57–90.

38. Liaw A, Wiener M. Classification and regression by random forest. R News 2002;2(3):18–22.

39. Breiman L. Random forests. Mach Learn 2001;45(1):5–32.

40. Cutler A, Cutler DR, Stevens JR. Random forests. In: Zhang C, Ma YG, editors. Ensemble machine learning. New York, NY: Springer, 2012;157–75.