

Comparison and Statistical Analysis of Land Impressions from Consecutively Rifled Barrels

In: AFTE J. 45(1):3-20 (2013)

Jeremy Monkres¹, Christopher Luckie², Nicholas D. K. Petraco³, and Allison Milam²

¹ Virginia Commonwealth University, Department of Forensic Science, 4609 Emmett Road,
Glen Allen, VA, 23060.

² Commonwealth of Virginia, Department of Forensic Sciences, 830 Southampton Avenue,
Suite 400, Norfolk, VA 23510.

³ John Jay College of Criminal Justice and the Graduate Center, City University of New York,
899 10th Avenue, New York, NY, 10019

Abstract: The validity and reliability of firearm and toolmark analysis has been debated in forensic science, often revolving around the subjectivity of the methods examiners use. This study attempts to confirm an examiner's subjective conclusions through objective computer analysis. Bullets, knowns and unknowns, fired through ten consecutively rifled barrels were used for the study. Unknown bullets were identified to the barrels from which they had been fired using traditional comparison techniques. Each land impression was photographed, and the distances of the prominent striae to one shoulder of the land engraved area (LEA) were measured using computer software. Two methods of selecting measurable striae were used. The data from these measurements was then converted into a barcode representative of the LEA from which it was taken. Barcodes were subjected to principle component analysis (PCA), and a support vector machine (SVM) was employed to evaluate the computer's ability to properly identify which barrel created a barcode, based on SVM analysis error rate (ideal error rate $\leq 5\%$). Optimal error rate varied based on selection technique, with 19.444% and 1.149% being the optimal values obtained by each method. The second result, generated by the majority of bullets analyzed, indicated the computer was able to adequately group barcodes according to their common origins, supporting the examiner's identifications. This research and described methodology may provide support for the reliability of firearm and toolmark analysis.

Key words: forensic science, firearm, toolmark, consecutively rifled barrels, barcoding, principle component analysis, support vector machine, pattern recognition.

Introduction

In its explanation of the Theory of Identification, AFTE acknowledges that, while based on training, expertise, and scientific theory, interpreting toolmarks and identifying them as being created from a common source is subjective in nature [1]. This acknowledgement has led to criticism of forensic firearm and toolmark analysis, and any conclusions that can be determined through its application, as being debatable or possibly invalid. These debates center on the questions of whether an examiner's experience, knowledge, and ability to draw conclusions without bias is sufficient for objective assessment of the evidence, or if it is necessary for conclusions to be based on an analysis method in which numerical data is gathered and generated, possibly without human participation [2].

These questions of subjectivity and reliability are of particular concern when evaluating firearm and toolmark analysis against the criteria for admissibility of scientific evidence in legal proceedings. In the instance of *Daubert* criteria, the basis of admissibility depends on, amongst other factors, the analysis employed having a known or potential error rate and general acceptance in the relevant scientific community [3]. It is under these two points that the subjective nature of toolmark identification becomes an issue. It is somewhat difficult to establish the error rate for the identification process, as proficiency testing is the only main method available to determine if an examiner's conclusions are valid [4, 5]. Of course, this may provide a measure of how well an examiner can perform, but still does not provide a fixed measurement of how likely it is that two toolmarks were made by the same tool. The second obstacle encountered when measuring toolmark identification against the *Daubert* criteria is its acceptance in the relevant scientific community. This becomes an issue depending on how the term "relevant" is defined. If the relevant community is only other firearm and toolmark

examiners, then the methods used in casework are more than likely generally accepted. However, if the relevant community includes other scientific disciplines, forensic or otherwise, than its reliability may be subject to scrutiny. A recent example of criticism against firearm and toolmark examination can be found in a report published by the National Academy of Sciences in 2009. While the report criticized the current practice of forensic science in general, it also claimed that there has not been enough research performed to determine how discriminatory the individual characteristics, like those used in firearms examination to establish sufficient agreement, are, and therefore the current scientific basis of this discipline remains unsatisfactory and subject to evaluation [6].

Literature Review

The question of this discipline's subjective nature is one that firearm and toolmark examiners have often tried to address in research studies. One of the most common methods used to answer this question is the analysis and comparison of toolmarks generated by tools that were consecutively manufactured. A common avenue of this type of research is the examination of consecutively rifled barrels and bullets fired from them to determine if they can be distinguished from one another [7-9]. One of the most noted of these studies is that reported by Brundage, in which sets of bullets fired from ten consecutively rifled pistol barrels, each set consisting of 20 knowns (two per barrel) and 15 unknowns, were sent to 30 different firearm and toolmark examiners in a blind study to determine if the examiners could attribute each unknown to its proper source [10]. This study was later expanded to include a total of 502 human examiners and five automated comparison systems [11]. In both studies, all unknowns were able to be identified with their appropriate source, with only seven inconclusive results being reported and no misidentifications being made. Additionally, other consecutively manufactured working

surfaces of firearms have been researched with similar goals as those of the consecutively rifled barrel studies. The working surfaces studied include bolt faces [12, 13] and extractors [14]. In these studies, the working surfaces were able to be successfully distinguished from one another, though the degree of similarity observed between sources varied based on working surface or manufacture.

Studies using consecutively manufactured tool surfaces and marks that are not produced by a firearm have also been conducted to examine the reproducibility of individual characteristics in these tools and the marks they produce. The variety of tools researched includes knives [15], tongue-and-groove pliers [16], and bunter tools [17]. Again, the results reported in these studies indicated that toolmarks made by consecutively manufactured tools could be identified to the tool that created them and that sufficient agreement was not demonstrated between marks made by different tools.

Another research avenue that has been historically utilized to support the conclusions that can be reached through firearm and toolmark analysis, as well as possibly providing a means to fulfill the *Daubert* error rate criterion, is the creation of statistical analyses of toolmarks of both common and uncommon origins. The foremost research in this area was performed by Biasotti when he compared the individual characteristics of 24 firearms with identical class characteristics. By comparing bullets fired from the same and different firearms, counting the number of total lines and matching lines in each comparison of land and groove impressions, and calculating the percent match for each comparison, he found the percent match for known matches was higher than known non-matches. However, Biasotti also notes that conclusions were much more discriminatory when the consecutiveness of the matching lines was considered. He calculated that the probability of observing more than four consecutively matched lines was

close to zero [18]. Studies conducted by Uchiyama also applied statistical tests to toolmark analysis, including Chi-squared and matching line probability of fired bullets [19] and analyses of computer generated striae patterns [20]. Each of these acknowledged the discriminatory power possible when measuring consecutively matching lines.

The notion of consecutively matching striae (CMS), suggested by Biasotti and Murdock as possibly having great benefit to firearm and toolmark examiners if used as the basis for a set of criteria in establishing identifications [21], has been the inspiration for much research related to firearm and toolmark examination [22-25]. Using this concept, a statistical study was performed using over 4,000 comparisons of marks created by various tools. An analysis of the data allowed the researchers to calculate and establish two conclusions: first, the probability of observing a specific number of CMS for known matches and known non-matches; and second, using likelihood ratios, the likelihood that a toolmark was created by the same tool rather than a different tool based on the number of CMS observed between the two marks [26, 27]. Application of statistics in this manner could potentially provide quantitative values for the discriminatory power of identifications.

The use of computer imaging techniques and statistical algorithms to perform toolmark comparisons, make identifications, or validate identifications is another area of study that has been researched and developed recently [28-35]. In one such study, researchers measured the characteristics of primer shear marks on cartridge cases through confocal microscopy. The software employed in the analysis allowed the subsequent data to be used to generate virtual 3-D images of the profiles, identify boundaries between significant lines in each profile, and convert the data into barcodes representative of each profile. The cartridge case profiles were then subjected to complex statistical analysis methods [36]. These methods included principal

component analysis (PCA), a procedure that reduces the number of variables, and thus the dimensions, of a data set into a smaller set of variables referred to as “principle components” [37], and support vector machines (SVM), learning programs for a computer system, which enables it to recognize pattern and make classifications [38]. Utilizing these and other analyses, it was determined the applied method was able to discriminate between the marks and group them in such a way as to suggest a common source [39].

The research described in this article is concerned with bullets fired from consecutively rifled barrels. The study is designed to compare the bullets using traditional toolmark examination methodology, identify which bullets were fired from the same barrels, and confirm those identifications through the employment of complex statistical analyses. The results of this study are of interest as it may provide objective validation of subjective observations made by the examiner, a generation of error rates for the identifications, and a reliable methodology that can easily be duplicated by most laboratories equipped for firearm and toolmark examination.

Materials & Methods

A set of full metal jacketed bullets fired from ten consecutively rifled Ruger P85 9mm barrels (Sturm, Ruger, & Co., Inc., Southport, CT) were selected for analysis. These bullets were obtained by the host laboratory as part of participation in the previously discussed Brundage study [10]. The set consisted of two known bullets for each barrel and 15 unknown bullets, a total of 35 samples. At least one unknown bullet was attributable to each barrel, providing a minimum of three replicate samples for each barrel.

A comparison and source identification analysis of the sample bullets was conducted through traditional comparison microscopy using a Leica UFM4 comparison microscope (Leica Microsystems Inc., Buffalo Grove, IL) with attached, position adjustable fluorescent light

sources. The examination was designed as a blind test to minimize examiner bias, though the examiner performing the analysis was aware that at least one unknown corresponded to each barrel. Index marks were created whenever the knowns were brought into proper phase or when an unknown bullet was identified as having been fired from a particular barrel. All identifications were recorded and then compared to a list of the proper barrel source for each unknown bullet after the traditional comparison had been completed.

The patterns of the most prominent striae present in each land impression were then examined and measured. This was accomplished by capturing the image of a land impression with a Leica DFC 420 digital camera (Leica Microsystems Inc., Buffalo Grove, IL), which was attached to the comparison microscope, and Leica Application Suite v. 3.6 (Leica Microsystems Inc., Buffalo Grove, IL) computer software, used in conjunction with the camera. The images were captured using the camera's grayscale setting in an attempt to maximize the contrast of the striae present. Images were captured under 15 times (15X) total magnification, this being the magnification at which the software had been calibrated. Each image was taken with the sample bullet on the left stage and a paper ruler on the right stage, which was used as a reference for both the size and proper orientation of the land impression. The light sources of the microscope were never consciously moved during the measurement taking process. A variety of tools employed by the software was then used to process and measure the image and the striae observed; all distance measurements were recorded in microns (μm). This was done by first using a line tool to draw a line across the bottom edge of the land impression, marking where the land impression started and establishing a baseline from which subsequent measurements would be taken. A measuring tool, which measures the distance between two points on the image and the angle at which the measurement was taken, was then used to draw a line on top of the

baseline to ensure it was flat (0°) and mark a point $850\mu\text{m}$ from the base end of the land impression. All other measurements were taken between this point and the base of the bullet.

The distances between the striae and the baseline were then measured using the measuring tool, all measurements being taken at a 90° angle. Both edges of the striae were measured to document its proper position in the land impression (Figure 1). For the purpose of this research, the striae being observed and measured were defined as the brighter, raised lines or ridges present in the land impressions. These specific marks were chosen because they could be easily distinguished under the grayscale setting. Striae which did not appear to be approximately parallel, with a consistent width, to the baseline were not selected for measurement. However, some striae which appeared parallel but seemed to widen or slant at a particular point were selected for measurement, with the point of measure being within the parallel portion of the striae. Also, in certain instances, multiple striae were measured as though they were one, due to either ill-defined boundaries between them or being a close, prominent group of striae which varied in width or angle through the land impression. At least ten and no more than 25 striae were identified and measured for each land impression. After a land impression had been analyzed, the measurement data was exported by the software into a Microsoft® Excel (Microsoft Corporation, Redmond, WA) file for storage and later analysis.

On each bullet measured, the index mark created during the traditional comparison was documented as land impression one (L1) for that bullet. The bullet was then rotated on the stage clockwise, in relation to the base, and the next land impression was documented as land impression two (L2), the next land impression three (L3), etc. This enabled all land impressions, a total of six per bullet, to be compared to the proper data of bullets having a common source. During examination, one bullet from each barrel was analyzed first. After the most prominent

striae were selected and measured, reference photos (similar in appearance to Figure 1B) were generated and used to measure the striae of subsequent bullets fired from the same barrel (known barrels). Striae were measured from the same point on the baseline as that of the reference whenever possible. However, the primary factor that affected from which point of the baseline measurements were taken was the appearance of the striae.

Two barrels, 17 and 19, possessing only three replicates each were selected to be examined without the use of reference photos (blind barrels). This was done to evaluate the recognition and reproducibility of prominent striae, as well as the ability to maintain consistent examination methodology. Bullets from these barrels were examined in an alternating order to prevent the examiner from being able to easily remember which striae were previously selected. Images of the measured land impressions and amount of striae selected were compared to the other replicates of the same barrel.

After review and compilation of the land impression striae measurements, the data was imported into the statistical program, R v. 2.13.1 (R Development Core Team, Vienna, Austria), for processing and analysis. The known and blind barrel data sets were processed separate from each other due to the difference in the method that was used to identify and measure the striae. Using R, the data sets of land impression measurements were converted into barcodes (The custom R codes written for this study are available upon request or downloadable through a user account at <http://toolmarkstatistics.no-ip.org/>). The following parameters were used to create the barcodes: land engraved area (LEA) width – 2000.00 μm ; resolution – 0.025 μm ; and tolerance – 10. These settings instructed the program where to begin and end each barcode, the minimum distance between striae, and how to distinguish between two striae in instances where one striation ended at the same point the next began. These barcodes were composed of numerous

variables and served as digital representations of what was observed and measured by the analyst. Each LEA was labeled with a number (1 – 48 for known barrels; 1 – 12 for blind barrels). The labels were the same for corresponding land impressions.

Once created, zero variance variables, i.e. variables that were the same for all barcodes, were identified and removed by the software. The barcodes were then subjected to PCA to reduce each one to a much smaller, more manageable number of variables. PCA measures information in a data set via variance. It is generally used to reduce redundant information in a data set by taking linear combinations of the original variables to form a new set of “derived variables”. Note that the entire set of “derived variables” (called principal components or just PCs) is equivalent to the original data set. The data in principal component form however has a special property. The PCs are naturally ordered by the amount of “information” (variance) they contain. If the first few PCs contain a majority of “information” in the data set (which is typically the case), then the remaining PCs can be deleted with a minimal amount of loss. The reason why a decision model should contain a minimal amount of variables is to help protect against over-fitting. PCA is an “automatic” and common way to reduce the number of variables used by the decision model.

As a matter of note, if only the first two or three PCs are retained then an approximate “picture” of the data can be plotted. The 2D or 3D PCA plot is usually only a very approximate “picture” of the data. They are pictures of the “math” none-the-less and we usually find them valuable in getting an overall understanding of what the data is trying to tell us. Thus a scatter plot (2-D PCA plot) and a 3-D plot of the PCA results were created using, respectively, the first two and first three principle components (dimensions) of the barcodes.

The correspondence of the barcodes to one another was then determined through employment of a support vector machine (SVM). With any computational identification task, a decision algorithm must be used. PCA by itself cannot make an association between a tool and a toolmark. Rather, it must be combined with a method to “learn” how to take the PCA data from a toolmark and associate it with a tool. SVM is a method to do this which has certain advantages from a forensic science point of view. The technique was developed to make reliable associations even if a dataset is small. Also, and perhaps most importantly SVM seeks to determine efficient classification rules for objects assuming nothing about the form of the underlying probability distribution generating the data [38]. This is a great advantage for application in forensic science. The fewer the decision algorithm’s underlying assumptions, the less vulnerable its conclusions are to attack in court.

The combined PCA-SVM analysis demonstrated the computer’s ability to distinguish between the numerically labeled barcodes using different numbers of dimensions per LEA. PCA-SVM decision models with 2 and up to 20 dimensions were analyzed. The computer’s identification error rate for each dimension level was generated and used to judge how well the barcodes could be differentiated, where the best separation between labeled groups occurred, and the lowest number of dimensions required to achieve acceptable separation. Error rates were estimated with the hold-one-out cross-validation (HOO-CV) technique. Given a number of PCs (2-20 for this study) HOO-CV computes the SVM decision rules using all but one of the toolmark patterns in the data set. The hold-one-out procedure is repeated for each toolmark pattern in the data set and the results are averaged to compute an estimated error rate. While simple in nature, HOO-CV is known to produce accurate error rate estimates and has a long published track record [39].

For the purposes of this research, acceptable separation was designated as an analysis with an HOO-CV identification error rate estimate less than or equal to 5%. The SVM results between the two data groups (known and blind barrels) were then compared to evaluate which measurement method provided better results.

Results

Traditional comparison of the known and unknown bullets demonstrated that all unknowns could be identified as having been fired from a particular barrel (Figure 3, Table 1). All barrels were found to have fired one unknown bullet, with the exception of barrels 12, 18, and 20. Comparison of the results to the reference list of barrel/unknown bullet identifications demonstrated that all identifications made were correct. Through the course of the study, 210 land impressions were examined. The known barrels had an average of 15.48 striae identified per LEA, while the blind barrels averaged 16.30 striae per LEA. In total, 3,281 striae were identified, selected, and measured in this study, resulting in a total of 6,562 measurements taken.

For the known barrels, suitable striae were able to be identified, selected, and measured for all initial reference bullets. Using reference photos from the initial bullets, all corresponding striae were able to be identified and measured in bullets originating from the same barrel. When the measurement data was uploaded into R, the barcodes created from the data demonstrated a similar appearance as the original measured images (Figure 2).

For the blind barrels, suitable striae were able to be identified, selected, and measured for all bullets analyzed. However, comparison of the resulting data sets from each measured LEA revealed the number of striae selected for measurement differed between land impressions (Table 2). An examination of the LEA images demonstrated that despite corresponding land impressions possessing similar patterns (Figure 3), different striae or different amounts of striae

in the corresponding LEAs had been selected and measured (Figure 4). Similarly, the barcodes created from the blind barrel data sets demonstrated similar patterns as the original measured images, but dissimilar patterns between corresponding LEAs (Figure 5).

Upon creation of the known barrel barcodes in R under the established LEA width, resolution, and tolerance parameters, it was found that each barcode was composed of 80,000 data points. After removal of the zero variance variables, this was reduced to 76,593 data points per barcode. Again, PCA reduces the dimension of the data space (here, barcode representations of the LEAs) and SVM “learns” to assign the resulting representations with the lands that generated them. For most application SVMs require many PCs to achieve reasonable error rates. The high dimensional spaces used in practice thus usually defy visual representation and only an ID error rate is available. It is non-the-less valuable to have a “mental picture” of the process the algorithm carries out to associate a toolmark with a tool. For this reason we will use a 2D PCA representation of the barcodes to show a “best-case scenario” and a “worst-case scenario” that the SVM can encounter when trying to pair a toolmark with a tool. (Remember though that this “mental picture” we are going to develop typically takes place in much higher dimensional space than 2D.)

Figure 6 shows PC-scores from all six LEAs on bullets from barrels 20 and 15. Most of the clusters of points are well separated, indicating that the barcodes from different LEAs are likely to be correctly identified by the algorithm. These LEAs comprising the clusters from which the circled points are drawn represent a “best-case scenario”. The points within each cluster are tightly spaced and the two different clusters themselves are well separated. The barcodes corresponding to the circled points appear at the bottom of Figure 6. A trained

examiner would be hard pressed to misidentify these known matches (KM) and known non-matches (KNM).

Figure 7 shows the same set of 2D PCA data as Figure 6; however a problem area is now indicated by the black square at the bottom center of the plot. This region represents a “worst-case scenario” for the decision algorithm. There are two different LEA clusters intermingled (both happen to be from barrel 15). The pink points are from LEA 6, and the green points are LEA 3. To the left in Figure 7, this area is zoomed in. The labeled points are close enough together such that the decision algorithm confuses their identities. However, when examining the bar codes that correspond to these points (immediately below the zoomed area), one can see that a trained human examiner should have little difficulty coming up with the correct IDs. This brings up a larger issue concerning the use of statistical pattern recognition and computers for toolmark analysis. Computers will not perform as well as trained examiners on pattern recognition tasks. The point to using computers is that they can establish ID error rate estimates in a scientifically reproducible/reviewable fashion. This figure highlights that, whatever the “true” error rate of the average human is, it is likely that it is not worse than the error rate of the average computer algorithm. The algorithm based error rates for this study were low, implying an examiner would have at least as low an error rate.

The results of the PCA and subsequent classification of each barcode by their principle components demonstrated some separation of corresponding LEAs using two dimensions (Figure 8), and even better clustering and separation of these LEA groups when three dimensions were utilized (Figure 9). Analysis through SVM revealed the program was generally better able to distinguish between the LEA groups as the number of dimensions employed increased (Table 3, Figure 10). The optimal number of principle components, or dimensions, for analysis was

determined to be seven, due to the low error rate demonstrated at this level (1.1%) and the fact that the error rates did not largely fluctuate when more dimensions than that were used.

Creation of the blind barrel barcodes in R under the established LEA width, resolution, and tolerance parameters, found that, like the known barrel barcodes, each barcode was composed of 80,000 data points. However, after removal of the zero variance variables, this was reduced to 73,766 data points per barcode. When the results of the PCA were demonstrated by the classification of each barcode through their principle components in the 2-D plot (Figure 11) and 3-D plot (Figure 12), some apparent clustering and separation could be observed, though not to the degree that was observed in the known barrel plots. Analysis of the PCA data with SVM did not demonstrate the same trend in the program's ability to distinguish between LEA groups as was observed with the known barrel analysis (Table 4, Figure 13). The optimal number of dimensions was found to be 12, as it possessed the lowest error rate (19.4%) with the fewest number of variables. However, this value was still considerably larger than 5%, the separation significance level. Therefore none of the dimension levels were able to demonstrate acceptable separation. Based on the PCA plots, SVM error rates, and lack of acceptable separation, the methodology employed to select and measure striae in the known barrel barcodes was determined to be superior to that of the blind barrel barcodes.

Discussion

As previously noted, the most frequent objection firearm and toolmark examiners encounter is that the methods by which they reach their conclusions are too subjective [2, 6]. While it is acknowledged by examiners that the way in which they conclude two toolmarks were created by the same tool is subjective in nature [1], this does not necessarily insinuate that the identifications they report are invalid. Using the methodology outlined in this study, the

subjective conclusions of an examiner were confirmed by complex statistical analysis and computer pattern recognition. The selection of striae in each land impression, the characteristics of that toolmark that the examiner believed were significant, unique, or noteworthy, was indeed a subjective process. However, the barcodes created from the measurement data of those striae were a digital representation of those individual characteristics. Essentially, this method enables the examiner to instruct the computer as to how each LEA appears, and then have the computer confirm or reject his conclusion of sufficient agreement based on its ability to distinguish between the patterns with which it has been supplied. One way to test the validity of this method and its conclusions would be to have another examiner, or multiple examiners, repeat the procedure described and compare the examiner's results to those reported. It is very unlikely the second examiner would select all of the same striae as were chosen in this experiment. However, provided his selection and measurement methods are the same for each land impression analyzed, the results of the computer program should be similar. Because of the subjective nature of selecting striae, all measurements should ideally be taken by the same examiner. If these were to be taken by multiple examiners and used in one analysis, clear, unmistakable references to the original selected striae would need to be employed; the results of this method of striae measurement still could yield less conclusive results than those reported here.

One problem encountered during this research was difficulty in identifying striae in certain corresponding land impressions. Largely, this was due to a lack of striae in the initial bullet examined and used for selecting the significant striae for a particular LEA, specifically in the area 850 μ m from the base of the bullet. In these cases, the minimum amount of ten striae was able to be identified, though often not much more than that. The lack of striae in these areas of the land impression was likely due to that portion of the bullet not fully engaging the barrel's

rifling when it was fired, especially since, being test fires, the bullets had been properly packaged and were not damaged. When subsequent bullets from the same barrel were examined, many more striae were observed in the LEA, making the identification of the reference striae more difficult. In these situations, all striae were able to be properly identified and measured, though often use of the reference photos to do so was insufficient. Instead, the reference bullet was compared to the subsequent bullet under the comparison microscope, the striae corresponding to those selected on the reference bullet were identified on the subsequent bullet and its previously recorded image, and then those striae were measured. Fortunately, the opposite situation, more striae being present and selected on the reference bullet than what could be observed on the subsequent bullet, was not encountered during the research. Had this occurred, the subsequent barcode would have been lacking bars present on corresponding ones, and would likely have been misidentified by the computer's pattern recognition program. In any future studies related to this method, it would serve the examiner's interests well to first examine and compare all replicates of the same source before selecting which shall be used as the initial reference.

Analysis of the blind barrel bullets did not yield results that were as conclusive or favorable as that of the known barrel data sets. This is very likely due to the blind barrel bullets being examined and measured without the use of reference material. As stated, the number of striae and specific striae selected were often different between corresponding land impressions. This information is not necessarily surprising, as previous studies have noted that when examining a toolmark, certain striae are not deemed significant or even recognized until they are compared to those of another toolmark [19, 25]. Without a comparison, the significance of particular individual characteristics cannot be properly determined. Because of this, the striae

selection and measurement methods used for the analysis of the blind barrels is not considered to be reliable or recommended.

As successful and promising as the results of the pattern recognition analysis are, the methodology employed in this study is limited in certain ways. The primary limitation is the use of a comparison microscope to examine bullets and capture images of the land impressions. This instrument was chosen based on availability and its wide use amongst examiners. Unfortunately, as previously observed [30], in toolmark comparisons performed with a comparison microscope, lighting of the samples being examined is a critical factor that can affect the appearance of the toolmarks, the examiner's recognition of striae, and thus the conclusion of whether samples have a common origin or not. Every effort was made to maintain consistent lighting and positioning of the samples analyzed. However, it may be possible to obtain better or more accurate results if the light source was not a variable that needed to be considered.

The other issue with the employment of a comparison microscope is that when the image of the LEA is captured and measured, a three-dimensional surface is confined to a two-dimensional image. Therefore, the curved surface of the bullet is not being properly represented in the image from which the measurements are being taken, an issue that has been noted in other research [29]. Because of this two-dimensional distortion, the measurements of striae distances from the base of the LEA are likely not the actual distances. Again, each bullet was positioned on the stage as similar as possible, relative to the microscope's objective lens, so the difference between measured distances and actual distances should be comparable and consistent between all bullets. However, an interesting expansion to this study would be the employment of a confocal microscope, as this would allow the lighting and relative measurement of striae to be the same for all samples recorded.

The other main limitation of this study is the small sample size, particularly the small number of replicates that were employed for most barrels. These samples were selected based on availability and their previous use in studies with consecutively rifled barrels [10, 11]. However, it is always desirable to have as many samples and replicates per sample group as possible, as the results of an analysis are considered more reliable when more samples are analyzed.

Additionally, other statistical and pattern recognition analyses were employed by a previous study [36], which could have further validated the results produced by this research.

Unfortunately, the sample groups did not possess the amount of replicates required for the use of these analyses. Had they been employed, the results would not have been considered reliable enough to be of any valid use.

Conclusions

The research presented was performed with two specific objectives: 1) determine if, by microscopically examining land impressions, the examiner could properly identify the barrels from which the unknown bullets had been fired; and 2) if those identifications could be confirmed by statistical analysis and pattern recognition programs analyzing data the examiner had gathered from measuring striae that was deemed significant in a particular land impression. The first goal was fulfilled and easily confirmed by comparison of the examiner's conclusions with the known correlations. The attempt to validate the examiner's conclusions was also successful, based on the error rates obtained by SVM analysis. However, this is only true of the bullets in the known barrel group. The methods employed to select and measure striae in the blind barrel group yielded an inconsistent selection of significant striae, resulting in dissimilar patterns being produced for analysis, and thus high error rates. The selection and measurement of striae without comparisons to other toolmarks is not recommended for future study.

This research has been presented in an effort to add to the body of knowledge of firearm and toolmark examination, further expand previous studies conducted with bullets fired from consecutively rifled barrels (specifically those related to the original study from which the test samples were obtained), and provide a possible method that can be employed in future studies to help further establish the methods employed to reach conclusions regarding toolmark identification as being reliable and valid. Future studies related to this research should include investigating whether other examiners analyzing the same or similar sample sets can produce comparable results, a greater number of replicates per group, possible application of additional analysis methods that can be employed with a larger number of replicates, and the use of a confocal microscope for the examination and recording of sample surfaces and striae. Additionally, application of the described methodology to examine other types of toolmarks should be researched.

Acknowledgements

The authors would like to thank Dr. Marilyn T. Miller and Jessica Smith for all of their advice and suggestions throughout the course of conducting this research and preparing it for publication and presentation.

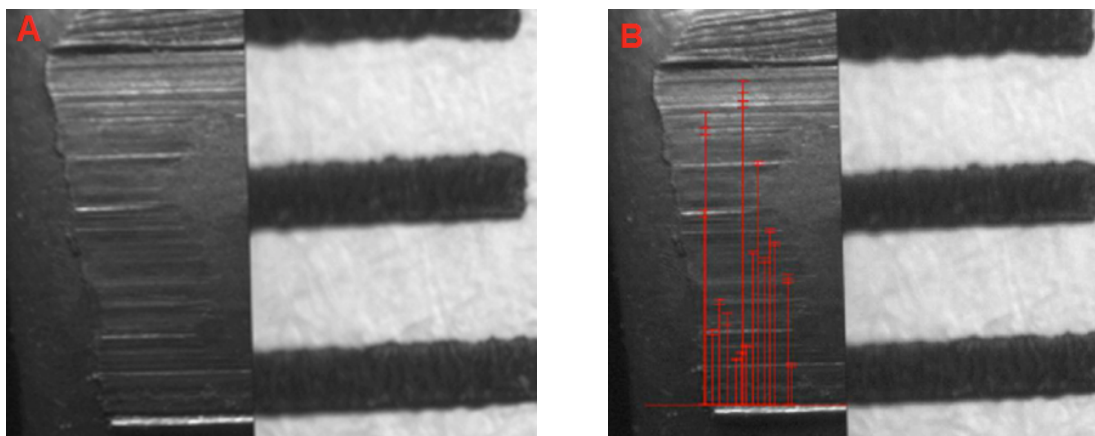


Figure 1. (A) Original captured image of Barrel 20, known bullet A, land impression 3, and (B) same image after striae have been selected and measured using computer software. Images were acquired at 15X magnification.

Barrel	Unknown(s) Identified with	Associated LEAs
11	Q2968	Knowns 1-6
12	Q2427, Q2500	Knowns 7-12
13	Q2407	Knowns 13-18
14	Q2299	Knowns 19-24
15	Q2743	Knowns 25-30
16	Q2737	Knowns 31-36
17	Q2762	Blinds 1-6
18	Q2289, Q2415	Knowns 37-42
19	Q2571	Blinds 7-12
20	Q2683, Q2849, Q2944, Q3075	Knowns 43-48

Table 1. List of unknown bullets identified as having been fired from specific barrels and LEAs associated with each barrel.

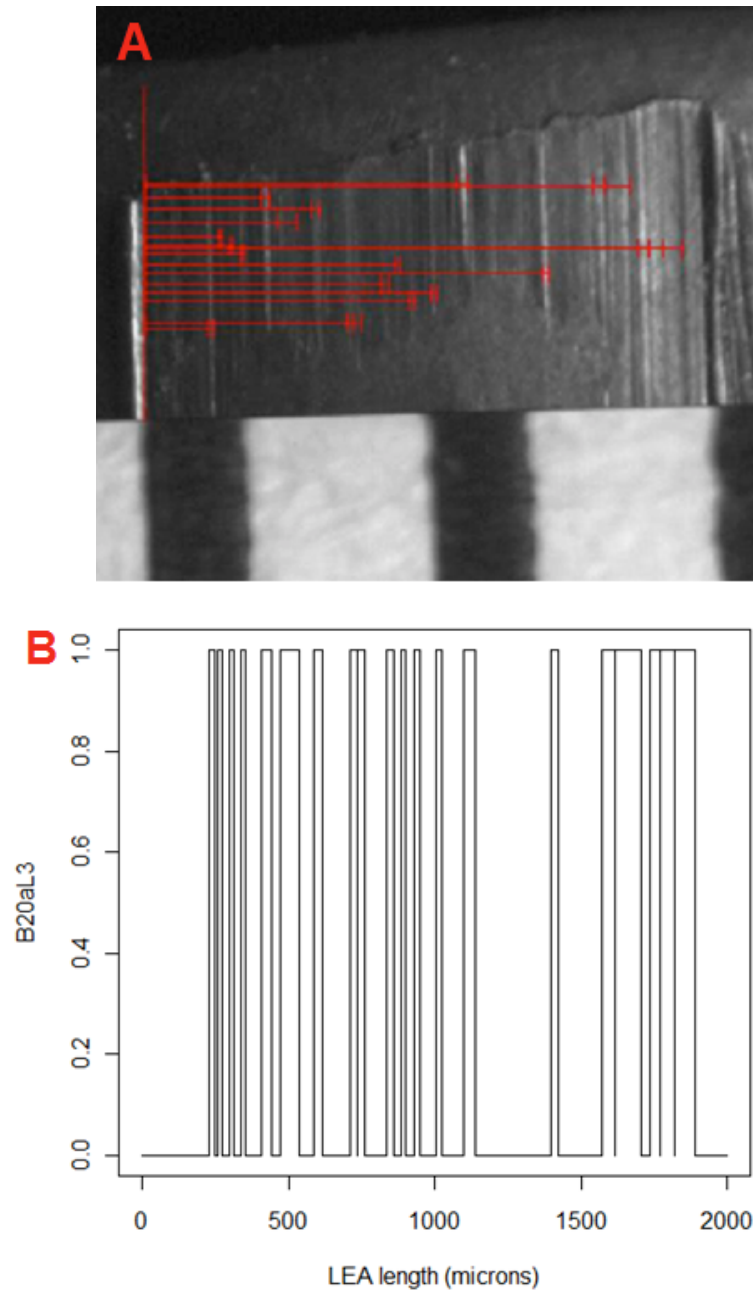


Figure 2. (A) Measured image of Barrel 20, bullet A, land impression 3, and (B) barcode created from those measurements. The pattern of the selected striae is replicated in the barcode. Image A acquired at 15X magnification.

Land Engraved Area	Bullet Measured				
	Barrel 17A	Barrel 17B	Q2762	Barrel 19A	Barrel 19B Q2571
LEA 1	18	17	17	14	20 18
LEA 2	17	16	16	20	21 19
LEA 3	17	15	16	16	13 16
LEA 4	18	18	19	15	18 19
LEA 5	14	13	14	14	15 13
LEA 6	13	13	13	15	15 15

Table 2. Frequency of striae identified in land impressions of blind barrel bullets. The number of striae identified and measured varies between corresponding land impressions.

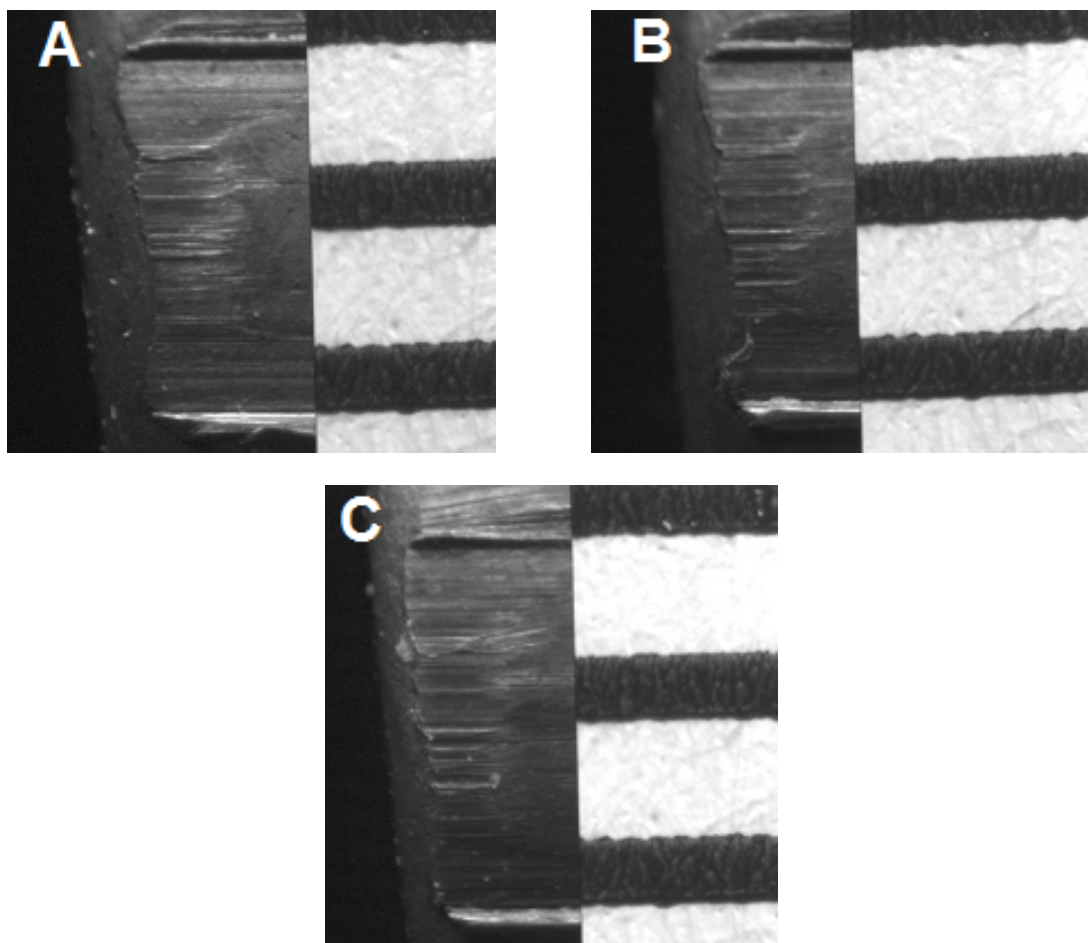


Figure 3. Images of the first land impressions of bullets: (A) Known 19A, (B) Known 19B, and (C) Q2571. All bullets were fired from Barrel 19. Correlation between the striae patterns can be observed. All images acquired at 15X magnification.

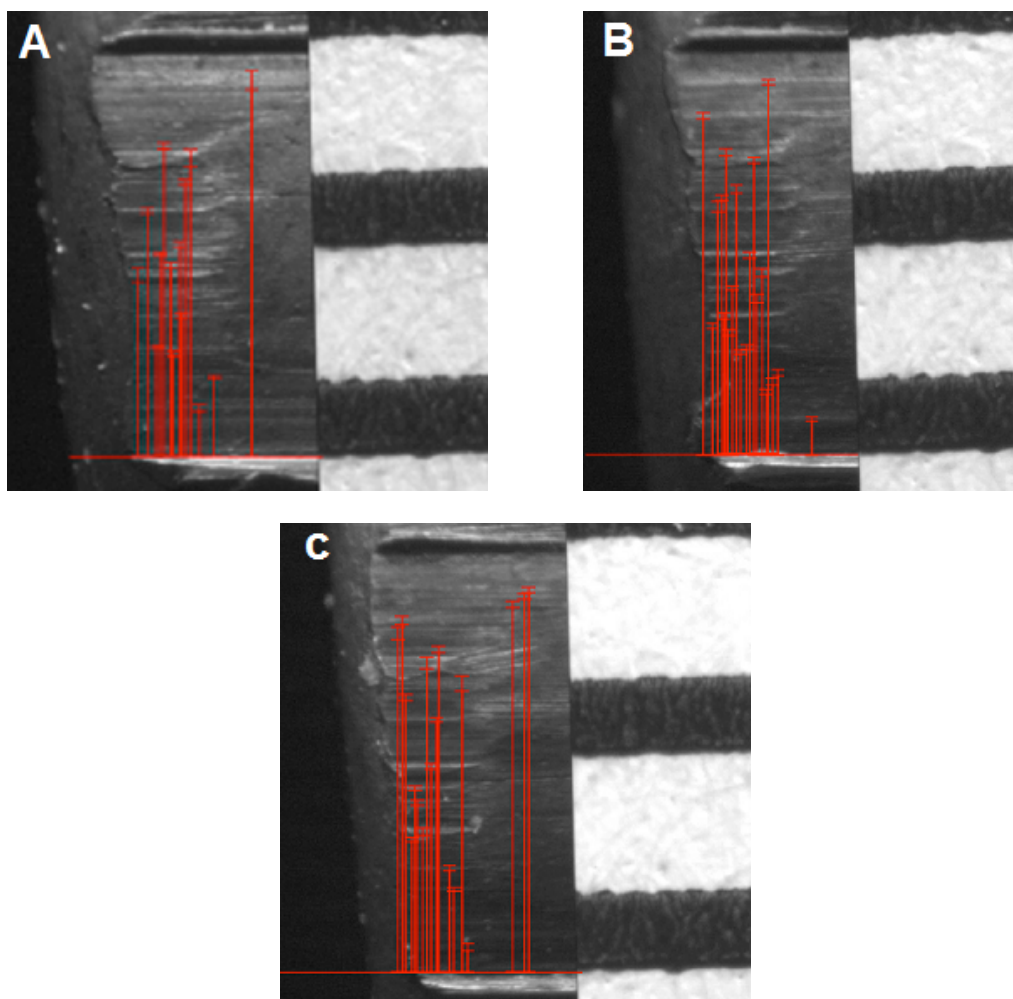


Figure 4. Measured images of the first land impressions of bullets: (A) Known 19A, (B) Known 19B, and (C) Q2571. All bullets were fired from Barrel 19. Selected striae and amounts of striae varied between the corresponding LEAs. All images acquired at 15X magnification.

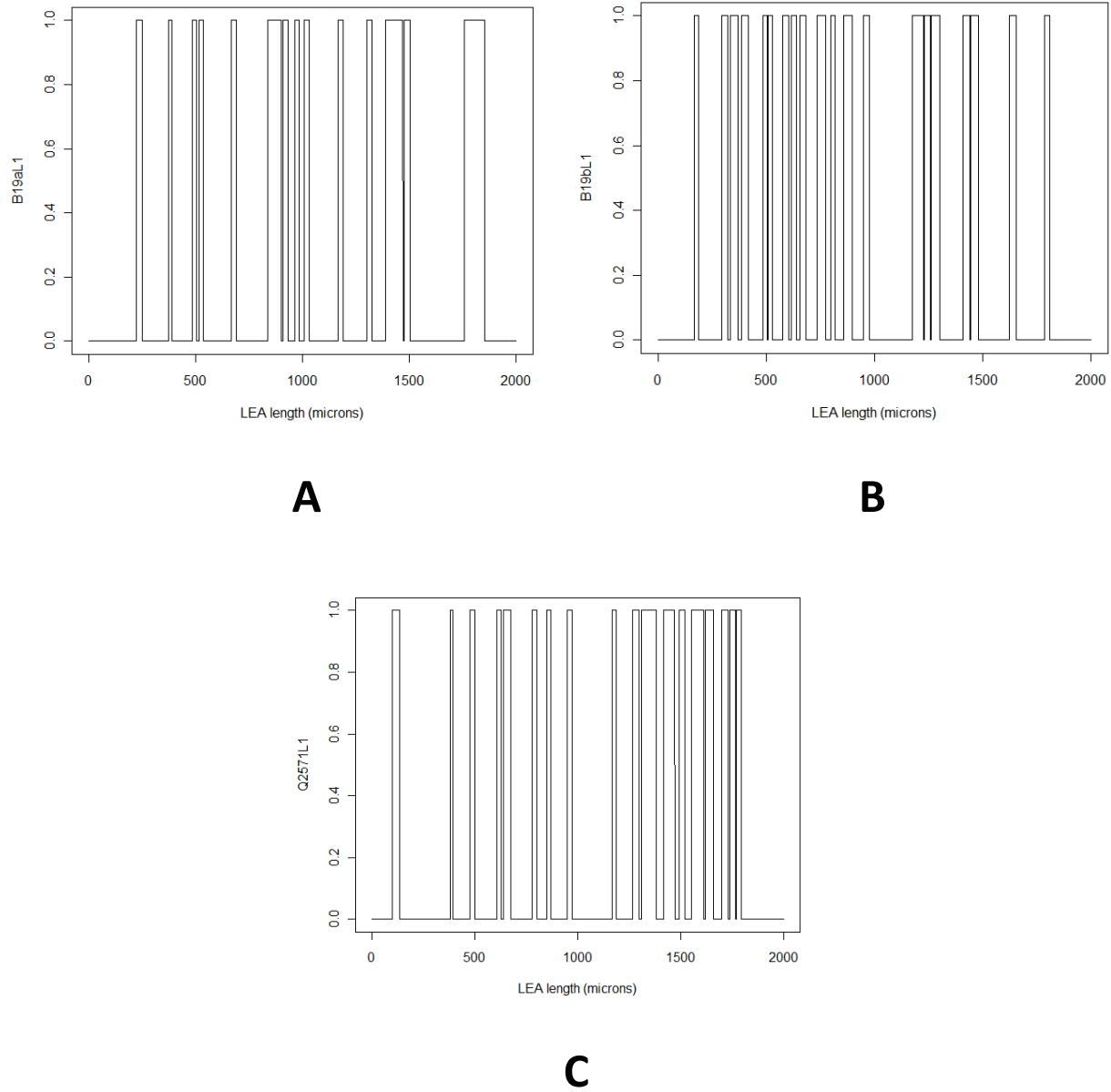


Figure 5. Barcodes of the first land impressions of bullets: (A) Known 19A, (B) Known 19B, and (C) Q2571. All bullets were fired from Barrel 19. Barcodes were created with the measurement data gathered from the images in Figure 6. Despite the bullets being fired from the same barrel, different striae being selected created different signature profiles for each LEA.

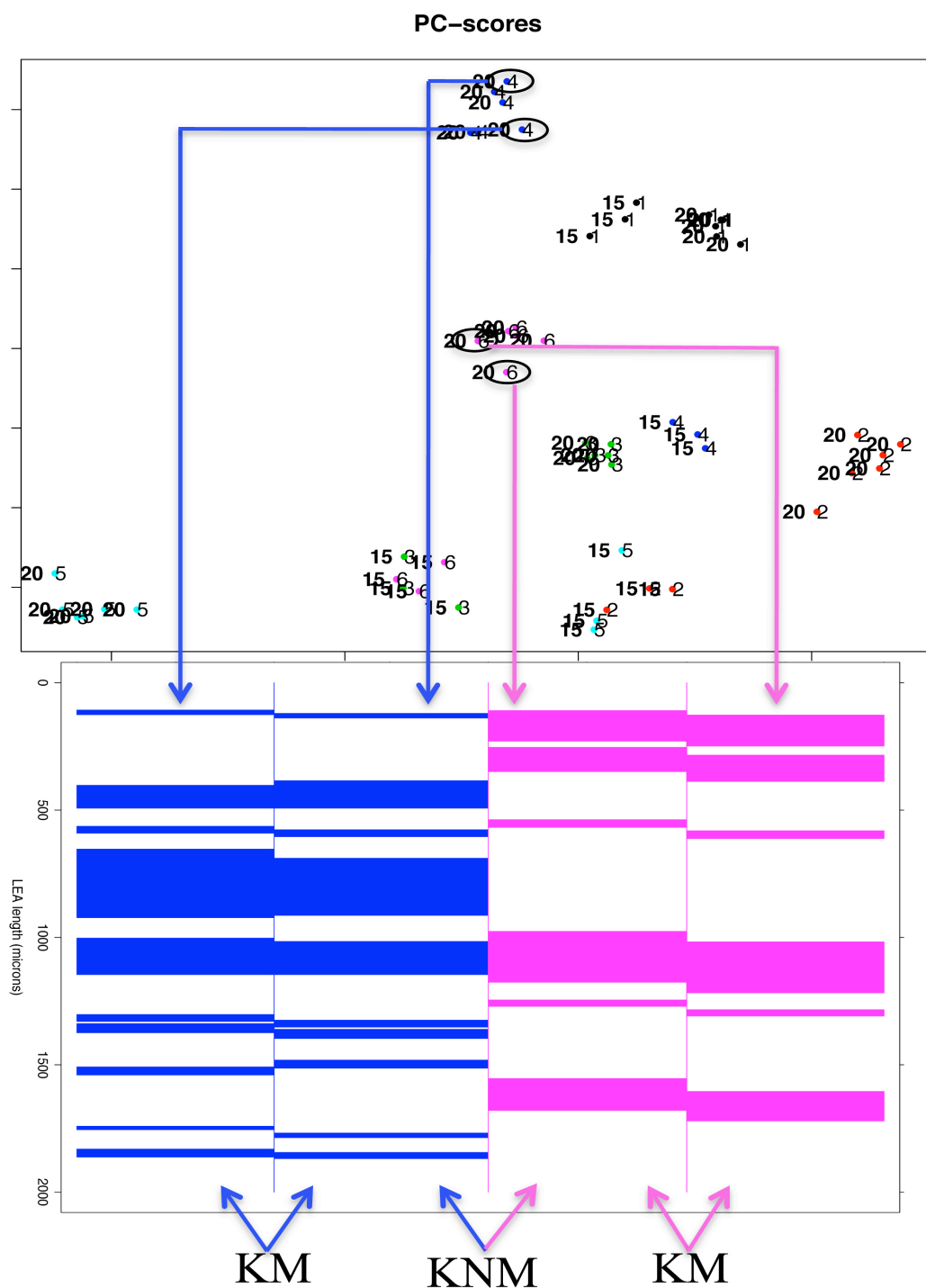


Figure 6. A 2D “mental picture” of the best-case scenario a decision algorithm can encounter when attempting to pair a toolmark with the tool that generated it. Here a computer and a trained human examiner with always make a correct identification.

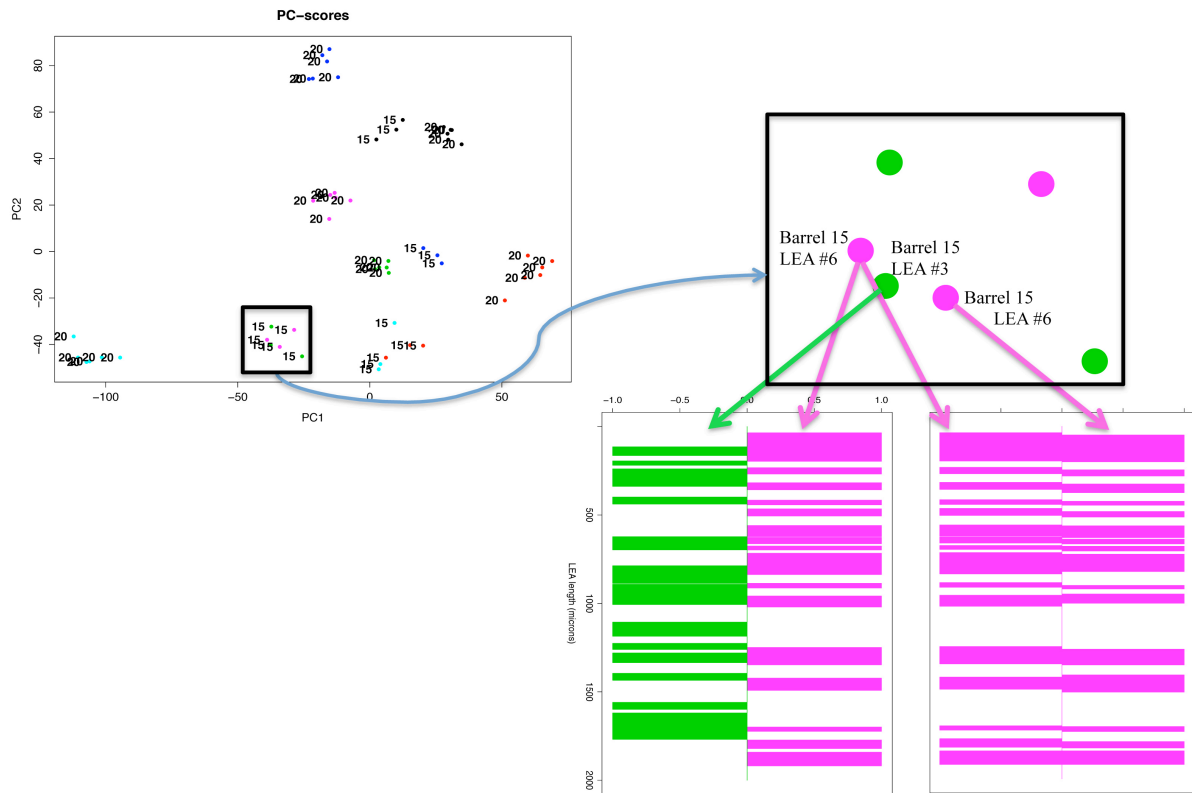


Figure 7. A 2D “mental picture” of the worst-case scenario a decision algorithm can encounter when attempting to pair a toolmark with the tool that generated it. Here a computer will make a wrong identification of some of the points in the box.

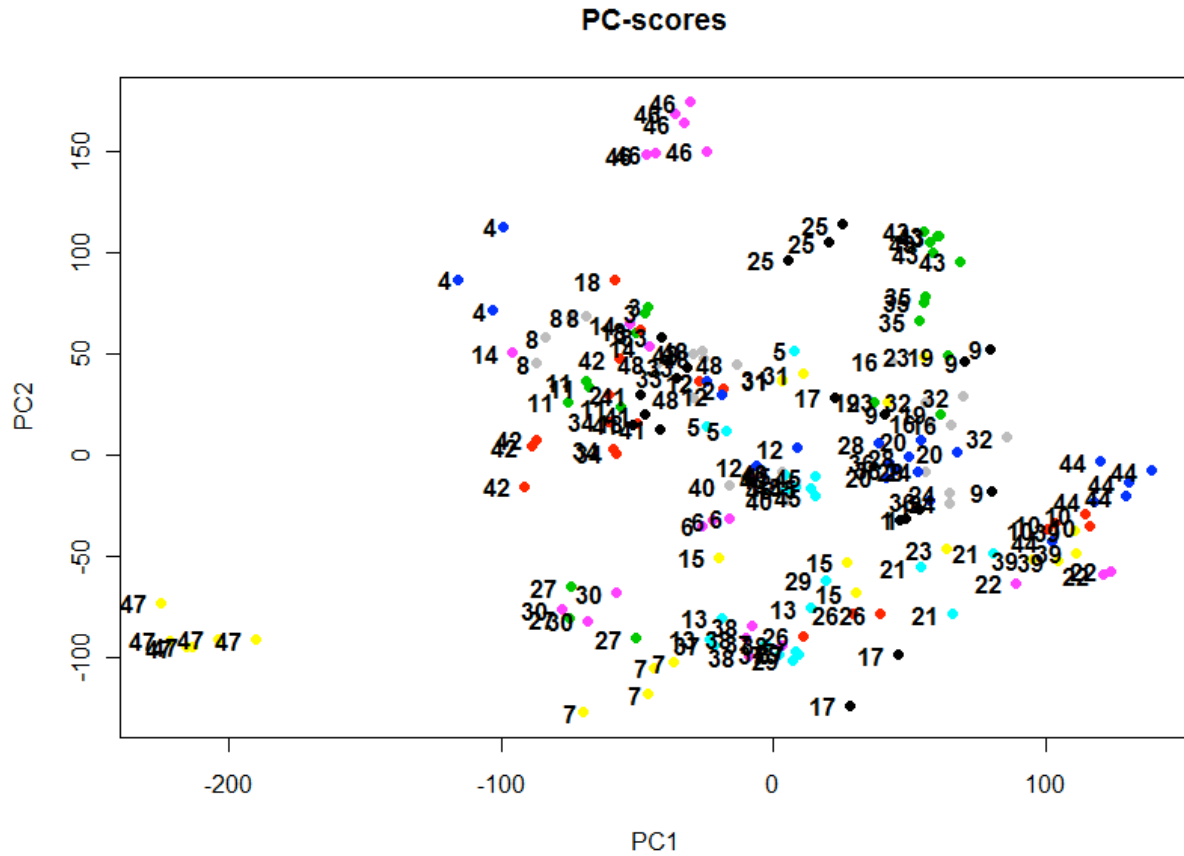


Figure 8. Scatter-plot created with the first two principle components of the known barrel data sets generated through PCA. Number labels (1-48) represent corresponding LEAs of bullets fired from the same barrel. Same groups appear to cluster together, some being separated from the rest of the groups.

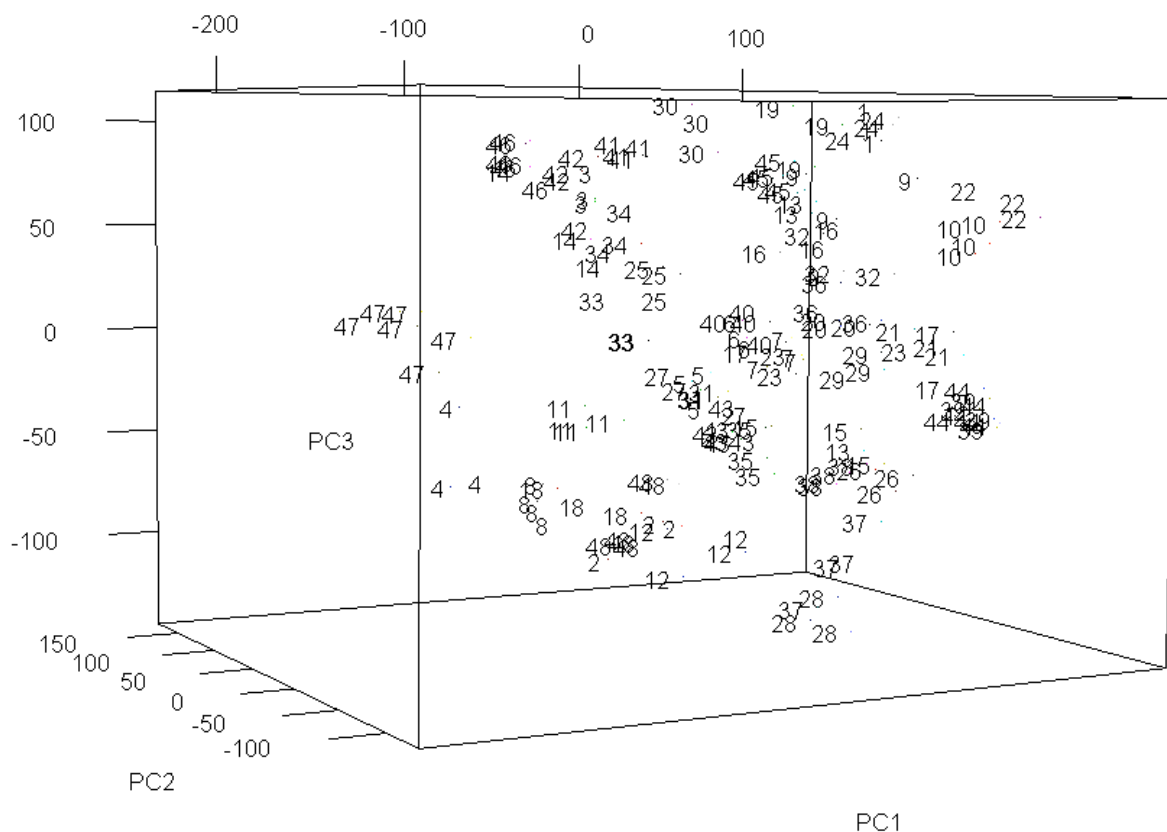


Figure 9. Three-dimensional plot created with the first three principle components of the known barrel data sets generated through PCA. Number labels (1-48) represent corresponding LEAs of bullets fired from the same barrel. Better clustering and separation of groups can be observed, compared to the scatter-plot.

Number of Dimensions	Analysis Error Rate (%)
2	53.4
3	23.0
4	14.4
5	7.5
6	2.9
7	1.1
8	1.1
9	1.7
10	1.1
11	1.1
12	1.1
13	1.1
14	2.3
15	1.7
16	1.7
17	1.1
18	1.1
19	1.1
20	1.1

Table 3. Error rates of known barrel SVM analysis, based on number of principle components (dimensions) used.

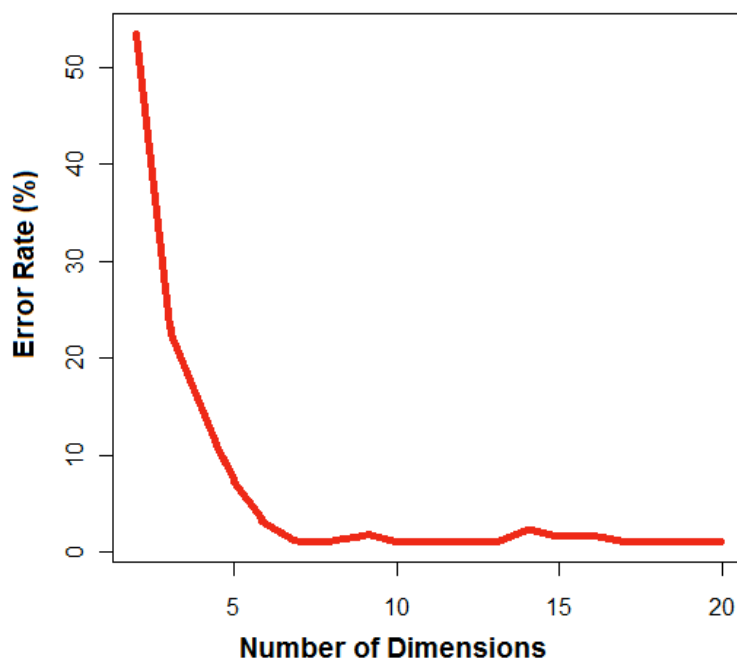


Figure 10. Plot of known barrel SVM analysis error rates and number of dimensions used.

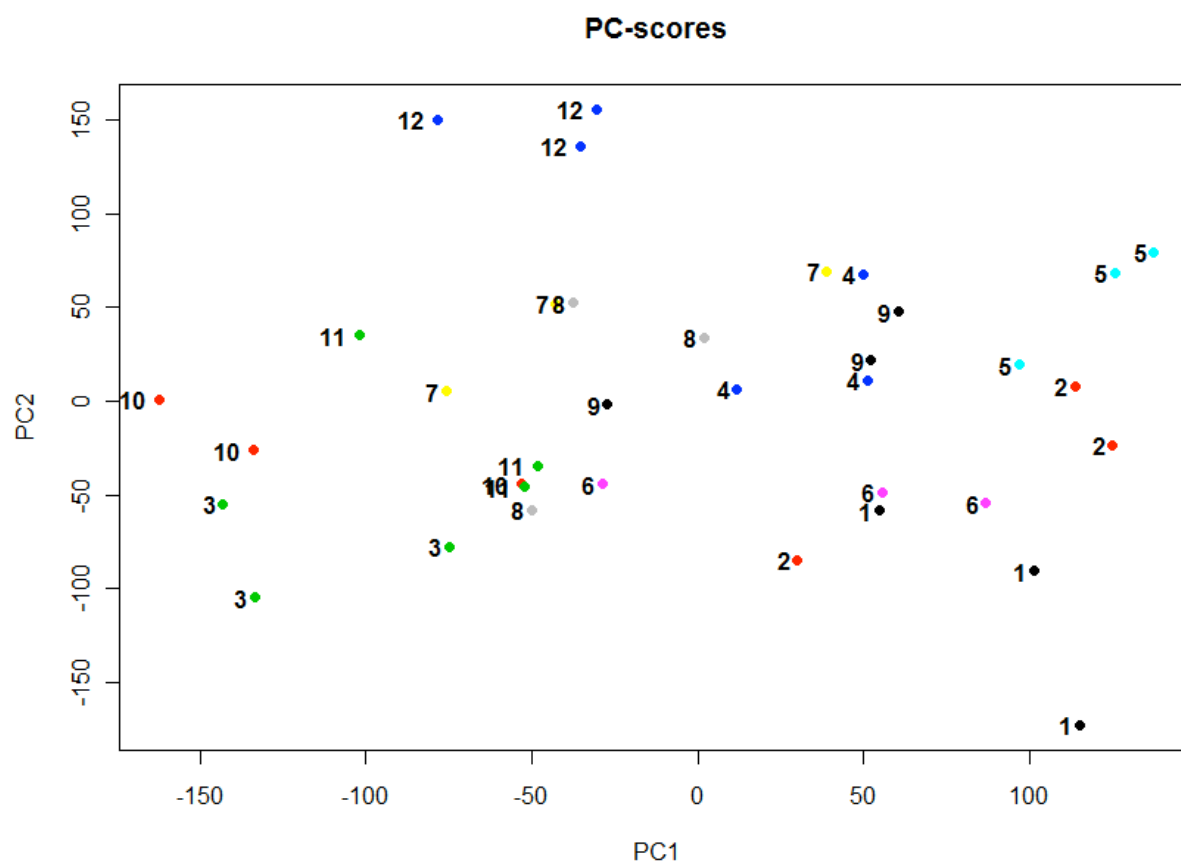


Figure 11. Scatter-plot created with the first two principle components of the blind barrel data sets generated through PCA. Number labels (1-12) represent corresponding LEAs of bullets fired from the same barrel. Some groups appear to cluster together or separate from the rest of the groups, however not as well as those of the known barrels.

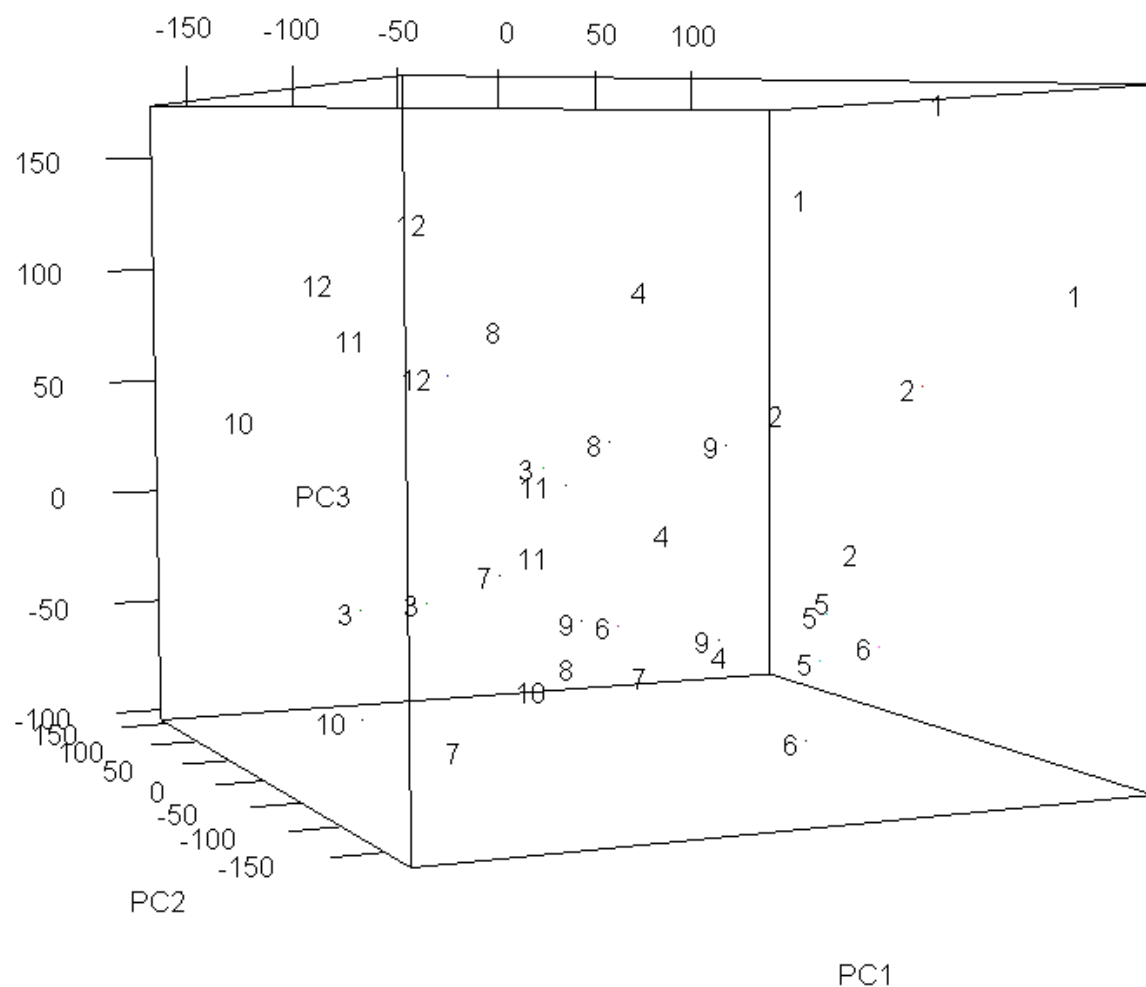


Figure 12. Three-dimensional plot created with the first three principle components of the blind barrel data sets generated through PCA. Number labels (1-12) represent corresponding LEAs of bullets fired from the same barrel. Again, some clustering and separation can be observed, but not as distinctive as that of the known barrels.

Number of Dimensions	Analysis Error Rate (%)
2	66.7
3	63.9
4	47.2
5	36.1
6	38.9
7	33.3
8	25.0
9	25.0
10	25.0
11	22.2
12	19.4
13	19.4
14	22.2
15	22.2
16	30.6
17	30.6
18	25.0
19	25.0
20	30.6

Table 4. Error rates of blind barrel SVM analysis, based on number of principle components (dimensions) used.

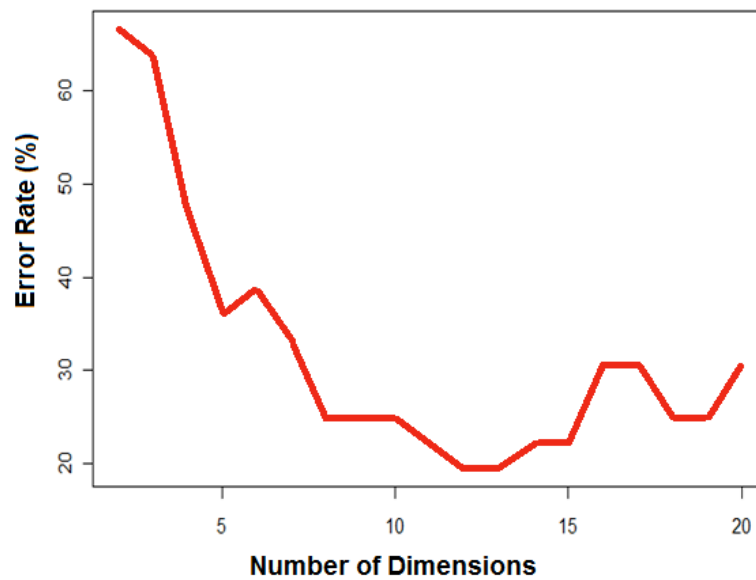


Figure 13. Plot of blind barrel SVM analysis error rates and number of dimensions used.

References

1. AFTE, "Theory of Identification as it Relates to Toolmarks," AFTE Journal, Vol. 30, No. 1, Winter 1998, pp. 86.
2. Meyers, C.R., "The Objective v. Subjective Boondoggle," AFTE Journal, Vol. 19, No. 1, Jan. 1987, pp. 24-30.
3. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).
4. Grzybowski, R.A. and Murdock, J.E., "Firearm and Toolmark Identification – Meeting the *Daubert* Challenge," AFTE Journal, Vol. 30, No. 1, Winter 1998, pp. 3-14.
5. Rosenberry, J.L., "Firearm/Toolmark Examination and the *Daubert* Criteria," AFTE Journal, Vol. 35, No. 1, Winter 2003, pp. 38-48.
6. National Research Council, National Academy of Sciences, Strengthening Forensic Science in the United States: A Path Forward, The National Academic Press, Washington, D.C., 2009.
7. Murdock, J.E., "A General Discussion of Gun Barrel Individuality and an Empirical Assessment of the Individuality of Consecutively Button Rifled .22 Caliber Rifle Barrels," AFTE Journal, Vol. 13, No. 3, July 1981, pp. 84-111.
8. Hall, E.E., "Bullet Markings from Consecutively Rifled Shilen DGA Barrels," AFTE Journal, Vol. 15, No. 1, Jan. 1983, pp. 33-47.
9. Matty, W., "A Comparison of Three Individual Barrels Produced from One Button Rifled Barrel Blank," AFTE Journal, Vol. 17, No. 3, July 1985, pp. 64-9.
10. Brundage, D.J., "The Identification of Consecutively Rifled Gun Barrels," AFTE Journal, Vol. 30, No. 3, Summer 1998, pp. 438-44.

11. Hamby, J.E., Brundage, D.J., and Thorpe, J.W., "The Identification of Bullets Fired from 10 Consecutively Rifled 9mm Ruger Pistol Barrels: A Research Project Involving 507 Participants from 20 Countries," *AFTE Journal*, Vol. 41, No. 2, Spring 2009, pp. 99-110.
12. Matty, W., "Raven .25 Automatic Pistol Breech Face Tool Marks," *AFTE Journal*, Vol. 16, No. 3, July 1984, pp. 57-60.
13. Lopez, L.L. and Grew, S., "Consecutively Machined Ruger Bolt Faces," *AFTE Journal*, Vol. 31, No. 1, Winter 2000, pp. 19-24.
14. Nichols, R.G., "Firearm and Tool Mark Identification: the Scientific Reliability and Validity of the AFTE Theory of Identification Discussed Within the Framework of a Study of Ten Consecutively Manufactured Extractors," *AFTE Journal*, Vol. 36, No. 1, Winter 2004, pp. 67-88.
15. Watson, D.J., "The Identification of Tool Marks Produced from Consecutively Manufactured Knife Blades in Soft Plastics," *AFTE Journal*, Vol. 10, No. 3, 1979, pp. 43-5.
16. Cassidy, F.H., "Examination of Toolmarks from Sequentially Manufactured Tongue-and-Groove Pliers," *J Forensic Sci*, Vol. 25, No. 4, Oct. 1980, pp. 796-809.
17. Rosati, C.J., "Examination of Four Consecutively Manufactured Bunter Tools," *AFTE Journal*, Vol. 32, No. 1, Winter 2000, pp. 49-50.
18. Biasotti, A.A., "A Statistical Study of the Individual Characteristics of Fired Bullets," *J Forensic Sci*, Vol. 4, No. 1, Jan. 1959, pp. 34-50.
19. Uchiyama, T., "A Criterion for Land Mark Identification," *AFTE Journal*, Vol. 20, No. 3, July 1988, pp. 236-51.

20. Uchiyama, T., "The Probability of Corresponding Striae in Toolmarks," *AFTE Journal*, Vol. 24, No. 3, July 1992, pp. 273-90.
21. Biasotti, A.A. and Murdock, J., "Criteria for Identification or State of the Art of Firearm and Toolmark Identification," *AFTE Journal*, Vol. 16, No. 4, Oct. 1984, pp. 16-34.
22. Miller, J. and McLean, M., "Criteria for Identification of Toolmarks," *AFTE Journal*, Vol. 30, No. 1, Winter 1998, pp. 15-61.
23. Miller, J., "Criteria for Identification of Toolmarks. Part II: Single Land Impression Comparisons," *AFTE Journal*, Vol. 32, No. 2, Spring 2000, pp. 116-31.
24. Miller, J., "An Examination of the Application of the Conservative Criteria for Identification of Striated Toolmarks Using Bullets Fired from Ten Consecutively Rifled Barrels," *AFTE Journal*, Vol. 33, No. 2, Spring 2001, pp. 125-32.
25. Miller, J. and Neel, M., "Criteria for Identification of Toolmarks. Part III: Supporting the Conclusion," *AFTE Journal*, Vol. 36, No. 1, Winter 2004, pp. 7-38.
26. Neel, M. and Wells, M., "A Comprehensive Statistical Analysis of Striated Tool Mark Examinations. Part 1: Comparing Known Matches and Known Non-Matches," *AFTE Journal*, Vol. 39, No. 3, Summer 2007, pp. 176-98.
27. Wevers, G., Neel, M., and Buckleton, J., "A Comprehensive Statistical Analysis of Striated Tool Mark Examinations. Part 2: Comparing Known Matches and Known Non-Matches Using Likelihood Ratios," *AFTE Journal*, Vol. 43, No. 2, Spring 2011, pp. 137-45.
28. DeKinder, J.D. and Bonfanti, M., "Automated Comparisons of Bullet Striations Based on 3D Topography," *Forensic Science International*, Vol. 101, 1999, pp. 85-93.

29. Bachrach, B., "Development of a 3D-based Automated Firearms Evidence Comparison System," J Forensic Sci, Vol. 47, No. 6, Nov. 2002, pp. 1253-64.
30. Banno, A., Masuda, T., and Katsushi, I., "Three Dimensional Visualization and Comparison of Impressions on Fired Bullets," Forensic Science International, Vol. 140, 2004, pp. 233-40.
31. Leon, F.P., "Automated Comparison of Firearm Bullets," Forensic Science International, Vol. 156, 2006, pp. 40-50.
32. Faden, D., Kidd, J., Craft, J., Chumbley, L.S., Morris, M., Genalo, L., et al., "Statistical Confirmation of Empirical Observations Concerning Tool Mark Striae," AFTE Journal, Vol. 39, No. 3, Summer 2007, pp. 205-14.
33. Chumbley, L.S., Morris, M.D., Kreiser, M.J., Fisher, C., Craft, J., Genalo, L.J., et al., "Validation of Tool Mark Comparisons Obtained Using a Quantitative, Comparative, Statistical Algorithm," J Forensic Sci, Vol. 55, No. 4, July 2010, pp. 953-61.
34. Chu, W., Song, J., Vorburger, T., Yen, J., Ballou, S., and Bachrach, B., "Pilot Study of Automated Bullet Signature Identification Based on Topography Measurements and Correlations," J Forensic Sci, Vol. 55, No. 2, Mar. 2010, pp. 341-7.
35. Chu, W., Song, J., Vorburger, T., and Ballou, S., "Striation Density for Predicting the Identifiability of Fired Bullets with Automated Inspection Systems," J Forensic Sci, Vol. 55, No. 5, Sep. 2010, pp. 1222-6.
36. Gambino, C., McLaughlin, P., Kuo, L., Diaczuk, P., Petillo, G., Kammerman, F., et al., "Confocal Microscopy and Striated Tool Marks: A Statistical Study and Potential Software Tools for Practitioners", Presented at the 42nd AFTE Training Seminar, Chicago, IL, May, 2011.

37. Joliffe, I.T., Principal Component Analysis, 2nd ed., Springer, New York, 2004.
38. Cortes, C. and Vapnik, V., "Support-Vector Networks," Machine Learning, Vol. 20, 1995, pp. 273-97.
39. Efron, B. and Tibshirani, R.J., An introduction to the bootstrap, Chapman & Hall, London, 1993.