Performance Evaluation of a Cadre Forensics TopMatch-GS 3D System for Cartridge Case Comparisons

JOSEPH ALSDURF

Thesis submitted
to the Forensic Science Institute
at the University of Central Oklahoma

in partial fulfillment of the requirements for the degree of

Master of Science in
Forensic Science

Eric Law, Ph.D., Chair

James Creecy, Ph.D.

Stephanie Luehr, MSFS

Edmond, Oklahoma
2023

Keywords: *Cartridge Case Comparisons, Automated Comparisons, 3D Scanning*

Forensic Science Institute
University of Central Oklahoma
Edmond, Oklahoma

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of JOSEPH ALSDURF
has been approved by the
Thesis committee as satisfactory for the
Thesis required for the
M.S. degree in Forensic Science

[signature]	12/13/2023
Eric Law, Ph.D., Chair	Date
[signature]	12/13/2023
James Creecy, Ph.D.	Date
[signature]	12/13/2023
Stephanie Luehr, MSFS, External	Date

ABSTRACT

Performance Evaluation of a Cadre Forensics TopMatch-GS 3D System for Cartridge Case Comparisons

JOSEPH ALSDURF

This research presented a series of commonly used firearms and commonly used ammunition brands to evaluate the Cadre Forensics TopMatch-GS 3D system overall and to test for differences between same ammunition and different ammunition comparison scores. For this study, ten casework firearms were shot, producing 30 cartridge cases from each, except two firearms that only had 20 cartridge cases fired from each; for a total of 280 cartridge cases. The ammunition manufacturers used were Blazer, CCI, MaxxTech, Wolf, Remington, and Winchester. This study utilized match and nonmatch scatterplots and ROC curves to evaluate the overall performance of the Cadre Forensics TopMatch-GS 3D. All cartridge cases were inter-compared using TopMatch breech face impression and aperture sheer algorithms for 39,060 comparisons. This research aimed to validate the Cadre Forensics TopMatch-GS 3D, evaluate automated algorithm performance on various casework firearms, and examine possible differences in automated scores between ammunition manufacturers on system performance. The tested firearms and ammunition brands performed well using the Cadre Forensics algorithms. The AUC for all firearms in this study was 0.956 indicating the automated breech face algorithm can differentiate between same and different sources. The same source and same ammunition AUC was 0.988 while the same source and different ammunition comparisons performed slightly worse with an AUC of 0.952. A procedure for comparing AUCs based on the Mann-Whitney test was used to test for significance and the resulting p-value was less than the 0.05 threshold (or 95% level of confidence) indicating that same ammunition brands performed significantly better than different ammunition brands. MaxxTech and Wolf were brass primed and scored exceptionally well regardless of the firearm. Winchester and Remington resulted in lower scores when compared to other ammunition brands. Aperture shears under 800 microns in length provided misleading aperture shear results. Running the breech face algorithm over the combined score feature is recommended as the combined score defaults to the breech face score if the system believes the aperture shear values may be misleading. The data collected in this research will allow for further development and improvement to the Cadre Forensics comparison algorithm and open the door for future research opportunities using this system.

Contents

List of Figures

List of Tables

1. Introduction

In firearm identification, it is possible to identify fired cartridge cases to a specific firearm or to the exclusion of a firearm, through the observation of microscopic firearm imperfections. These imperfections are created during various manufacturing processes, that are then transferred to ammunition components during the firing process [5]. The components of a cartridge are illustrated in Figure 1.1. During the firing process, the firing pin and breech face leave impressions on the primer when the pressure of the propellant forces the case rearward for the ejection process. The breech face and firing pin impressions commonly used by firearms examiners are displayed in Figure 1.2. Some identifiable features on cartridge cases are common to certain firearm manufacturers, such as the elliptical firing pin impression a Glock handgun leaves on a cartridge case, as shown in Figure 1.3. This allows examiners to have a class characteristic that could help with firearm elimination. However, the appearance of individual characteristics left by a firearm can and will change with extended use caused by shot-to-shot variability [3]. This shot-to-shot variability calls for firearm examiners to be diligent in their examinations. Identifications are made using an examiner's training and experience using high-resolution images and identifying similar manufacturing features between the known and unknown cartridge cases and bullets. [6].

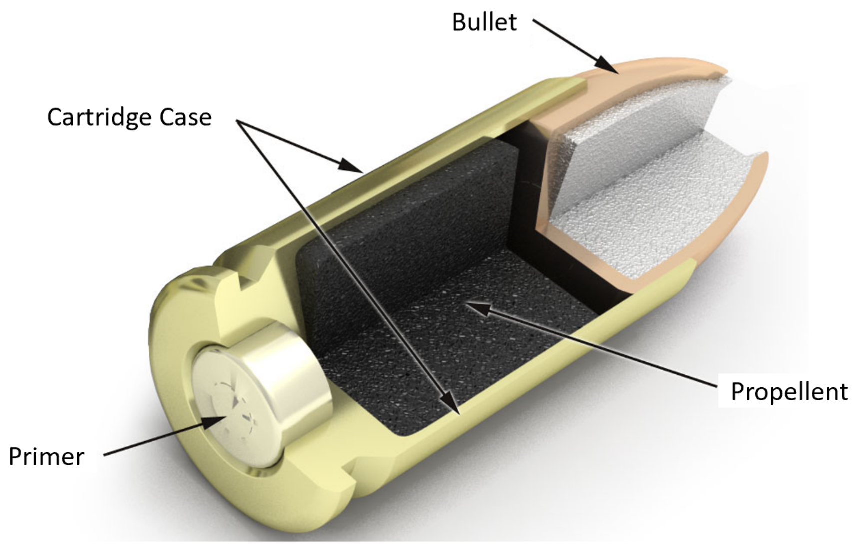


Figure 1.1: A cartridge with labeled ammunition components [1]

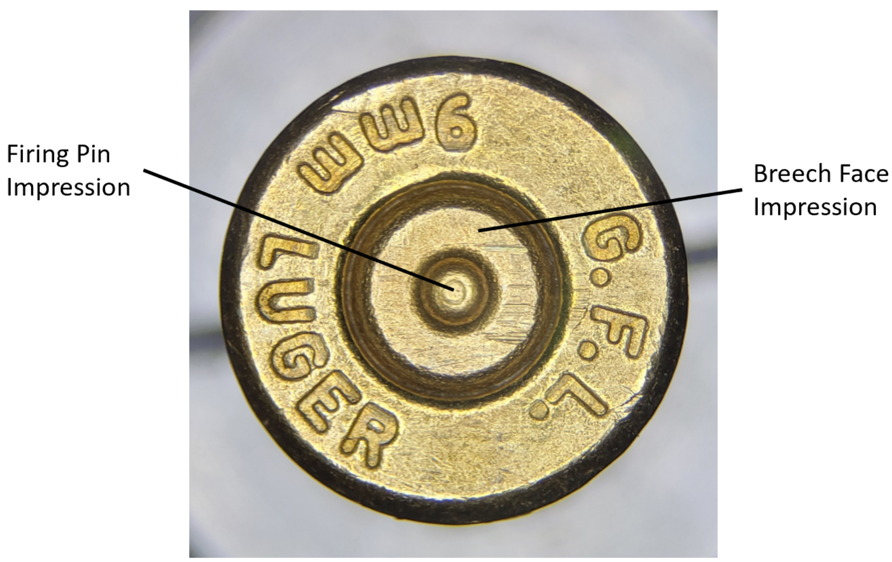


Figure 1.2: A fired cartridge case with labeled firing pin and breech face impression areas

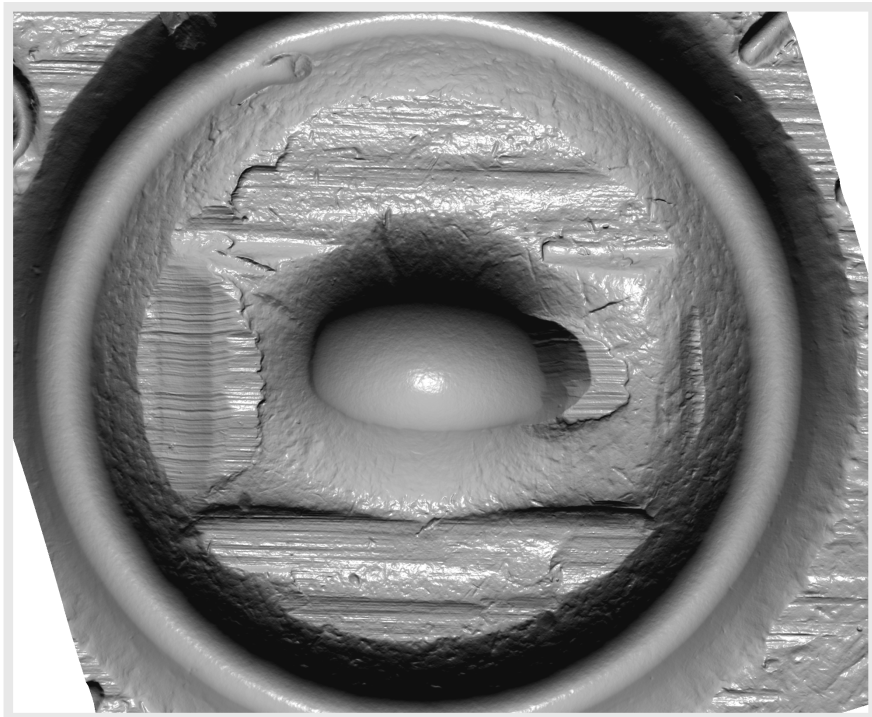


Figure 1.3: Example of a Glocks elliptical firing pin impression

Since the inception of firearm identification, examiners have been faced with the problem of only being able to see surface-level detail on cartridge cases and bullets which can change appearance based on lighting direction and intensity [7]. Surface features that are visible under certain lighting conditions may be virtually invisible with a small lighting change. Figure 1.4 shows the effect lighting position has on the same cartridge case. Over the years, there have been notable improvements to the way firearm evidence is examined. Now, examiners can visualize what is typically observed on a microscope stage on a screen, and use an automated comparison system such as the Integrated Ballistics Identification System (IBIS), which is part of the National Integrated Ballistics Information Network (NIBIN). However, all of this relies on the quality of the images that can be collected, the direction and intensity of the lighting used, and the quality of the comparison algorithms. Newer technologies use a 3D model of the cartridge case to allow the depth of marks to be considered. This system would ideally also eliminate the adverse effects that light variations cause. One such system is the Cadre Forensics TopMatch-GS 3D system which uses impression technology with a gel pad that conforms to the surface of interest at a micron-scale resolution allowing for very fine details, to be captured and scanned. The scan the system produces is of the gel-pad being deformed by the cartridge case being impressed into the gel. Scanning the gel rather than the cartridge case itself removes surface specularities [8].

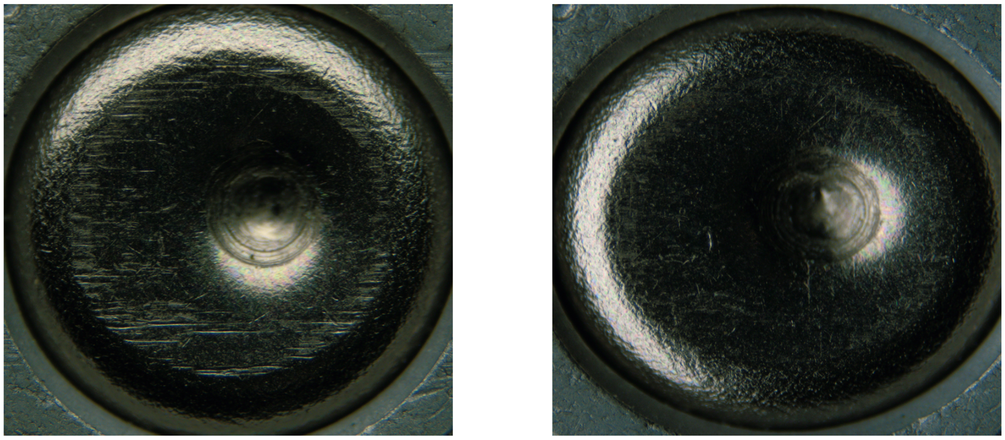


Figure 1.4: Two images from the same cartridge case under two different lighting conditions. In the left image the lighting was positioned at the 12 o'clock and in the right image the lighting was positioned at the 9 o'clock.

2. Literature Review

2.1 History

One of the first documented cases where a bullet was compared to a firearm dates as far back as 1835 in London, England. In this case, the examination involved marks from a bullet mold. In the 1800s, bullets were made by pouring molten lead into a mold in the shape of the bullet [7]. The imperfections in a bullet mold are transferred onto all bullets made by that mold. Although this method is not the same method as examining rifling marks, the examination sought to look at imperfections caused by the manufacturing process. One of the earliest studies on the examination of rifling marks on fired bullets was conducted in 1899 at the Medical Association of Central New York at Syracuse, and published in 1900. This article demonstrated the clear defacement of bullets caused by the rifling in the barrel [7]. These observations led to bullet and cartridge case examination emerging as a forensic science discipline.

For over 100 years, forensic firearm identification has used light microscopy, which relies on the examiner comparing individual characteristics on firearm evidence to make source conclusions. In traditional analysis, an examiner sets a known exemplar and a questioned cartridge case under a comparison microscope and conducts their examination visually. In 1996, the Integrated Ballistic Identification System (IBIS) was developed and tested by the Bureau of Alcohol, Tobacco, and Firearms (ATF). This system allows for these same comparisons to be done in mere minutes by complementing manual comparison with the introduction of a ballistics database and algorithm [9]. This algorithm automatically associates collected 2D images based on having similar class and individual characteristics. Once compared, the software produces a list of potential leads. Those potential leads are examined by a firearms examiner who performs a follow-up visual comparison to determine if any of those leads may have been fired by the same firearm [10].

2.2 Subjective Analysis

Firearm and toolmark identification conclusions have always been considered a subjective science. The AFTE Theory of Identification document acknowledges that firearm examiner conclusions are subjective [5]. Using 3D analysis as a tool in firearms and toolmark casework allows additional objective data to be collected and conclusions to be better supported in documentation. However, the result is still a subjective conclusion made by a firearm examiner. The 3D system scans and uses topography measurements to make its comparison rather than comparing visual characteristics. On the traditional side, the firearms examiner uses their expert judgment that cannot be quantified due to relying on an examiner's education, experience, and methods to reach conclusions such as identification or elimination. However, using only algorithm data to examine cartridge cases and bullets leaves the field of firearms and toolmarks with several issues. Algorithms generally compare the entire topography map and generate an overall score for the questioned bullet or cartridge case. This becomes problematic when firearms leave few reproducible marks on the cartridge case or bullet.

Microscopic comparisons rely on an examiner weighing the similarities and the dissimilarities and making conclusions based on their own experience [11]. Uncertainty is present in all sciences, especially when applying theories and techniques. This does not discount the area of science as being unreliable or unscientific [6]. To replace a physical microscopic comparison with an instrument using only algorithms is not in the best interest of the field of firearms and toolmarks; instruments and humans both have a possibility for errors. Instead, it would be wise to use 3D technology as a tool for the firearms examiner to support their physical examination under the comparison microscope. If an examiner is requested to testify in court based on an instrument's results, there would be no basis for the examiner to testify given that the algorithm would be the sole bearer of identification. The best course of action for tackling subjectivity in firearms analysis is to look for ways to prevent potential human error through different training methods and the use of automated systems that allow an examiner to compare and validate potential conclusions. Giving the examiner the final say in conclusions allows for expert testimony in cases that a one-size-fits-all program would fail to do.

2.3 Current Advancements

New technology in the firearm and toolmark field has tried to eliminate the effects of lighting on cartridge cases. One such system is the Evidence IQ Ballistics IQ, which uses five different lighting angles. Unfortunately, the image resolution is not always suitable for an accurate comparison of fine individual characteristics. The Ballistics IQ algorithm is a triage solution for sorting cartridge cases into firearm groups, after which an examiner would then have to use traditional microscopy methods for comparisons to verify the algorithm results. There is currently a "second-generation technology" [12] that has been developed allowing for cartridge cases and bullets to be captured in 3D scans. The Cadre Forensics 3D system, currently validated for casework use by the Federal Bureau of Investigation (FBI), still needs of additional research studies on the reliability of their algorithm in performing accurate comparisons.

Cadre Forensics conducted a study evaluating their automated algorithm over six years ago [8]. Since then, Cadre Forensics has made improvements to provide a better overall system. Some of those improvements have been updates to their automated algorithms for both breech face impressions and aperture shears, creating a 15 cartridge case holder to allow for more scans to be conducted at once, improvements to the gel pad used in the scanning process, and overall user interface improvements. However, there has been limited additional research on their system. The 2015 Cadre Forensics study used version 0.9 of the automated algorithm and Cadre has mentioned there has been additional tests on the algorithm. However, none of those additional internal validations have been published. The research conducted by Cadre Forensics in 2015 was focused on identifying true positives and true negatives. Research is needed on the effects of different firearms and ammunition types on the system to identify potential areas of system vulnerability.

The system measures the 3D surface topography of a cartridge case and compares both the breech face and aperture shear marks using their scoring algorithms. The Cadre Forensics system uses an acquisition method known as Photometric Stereo. Photometric Stereo is a technique to estimate an object's surface under different lighting conditions. The object being scanned is pressed into a gel made by the company GelSight [8]. Cartridge cases reflect light differently depending on the properties of the primer, and this gel acts as a medium to remove these influences of surface properties. This means that no matter the material, the scan becomes a function of only the surface geometry when scanned with the sensor by removing the influence of surface reflectivity on the measured topography. This contrasts significantly with confocal and focus-variation microscopy which are based on light reflected directly off of the surface of interest. There currently is a known problem of cartridge case materials playing a deciding factor in cartridge case comparison, as some materials do not reflect light as well as others. Comparing cartridge cases of different materials has often led to the wrong conclusions during analysis [13].

The Cadre Forensics system can measure at a micron scale [8]. This high resolution allows analysts to examine the smallest of marks needed to identify a cartridge case or bullet accurately. See Figure 2.1 for an example of a cartridge case scanned using the Cadre Forensics system next to a cartridge case using traditional light microscopy. Cadre's new 3D system eliminates light effects on the questioned cartridge case and captures an overall better quality image compared to older 2D instruments. When comparing traditional comparison microscopy versus the 3D scan system Cadre Forensics uses, the detail in the 3D scans already supersedes the traditional methods even by a simple glance.

Figure 2.1: An example of a cartridge case scanned using the Cadre Forensics system next to an image of a cartridge case under a comparison microscope.

Virtual comparison microscopy (VCM) is used to view fired ammunition components in a virtual environment, such as on a computer screen. In Cadre's 3D virtual environment, the user can adjust lighting and maneuver the 3D model. Cadre Forensics conducted a validation of VCM in the comparison of cartridge cases [14]. Cadre looked at 520 comparisons and found that with VCM, there were more additional positive identifications than with traditional light microscopy, where inconclusive conclusions were made more frequently [14]. Furthermore, when there were more complex identifications, VCM performed better than traditional light microscopy. There have also been studies on the potential error rates of examiners when using VCM [15]. In this study, there were 107 participants with 40 test sets. With a total of 693 comparisons there were only three false positives (.43 percent) and zero false negatives from 491 comparisons. Using VCM to visualize toolmarks in a 3D setting has the potential to lead to a lower false positive and false negative error rate, especially for complex comparisons.

Unlike traditional comparison microscopy analysis, the Cadre Forensics system uses image-matching algorithms. The algorithm and its process are similar to that of Google Photos, which can categorize people and objects based on faces or features in the photograph. This type of sorting process used by the Cadre system claims to be more accurate because the scanner controls the cartridge case's scale and orientation [8]. The 3D scan generates height maps, allowing the automated system to avoid appearance variations like shadows from variations in lighting. The system produces a ranked list based on its scoring algorithm of how likely two cartridge cases are to have the same origin.

In addition to the FBI internally validating the system [16], Cadre Forensics has also validated its system [8]. Cadre's validation process consisted of real-world experiments using 290 firearms and over 700 cartridge cases totaling more than 100,000 comparisons [8]. The cartridge cases were then broken into five sets with each set having a unique test for the 3D system.

- Set 1 consisted of 47 firearms without preference for their ability to mark cartridge cases to allow for real-world conditions of toolmark quality and type. There were three test fires per firearm created using PMC brand ammunition. This set was used to evaluate system performance on firearms commonly seen in casework.

- Set 2 had a total of 101 firearms, 47 from the first test set and 64 additional firearms. There were seven ammunition types; Fiocci, PMC, RWS, Remington, Speer, and Winchester. There were two additional test fires created from the firearms in Set 1 using Remington ammunition. 26 of these firearms did not have a known match. The goal of this set was to test true positives and false positive rates.

- Set 3 was a series of 164 Glocks with two test fires from each. This test set was geared to testing the aperture shear algorithm.

- Set 4 was titled the Miami-Dade Study set. This set were well marked cartridge cases and had 10 pairs of matched known's and 15 unknown cartridge cases.

- Set 5 was titled the JFS set were there were 10 consecutively manufactured pistol slides, five test fires from each with a total of 50 cartridge cases. All of these test fires used Winchester brand ammunition and were strongly marked.

The results for well-marked cartridge case sets with a 0.6 threshold was 100 percent with no false positives [8]. The algorithm matched 80 percent of cartridge cases in sets one and three without identifying a false positive. The researchers noted that matching across ammunition brands is more difficult due to some brands marking poorly [8]. As in other research with two-dimensional analysis, this is a known problem for both automated and manual comparisons [13]. However, the extent of the different brands of ammunition and their effects on this system is not categorized, nor is the effect of shot-to-shot variability. Future research on 3D comparison systems should address these effects and how they influence the comparison. Laboratories using this new 3D technology need to be aware of which ammunition brands mark the best when compared to another brand recovered at a crime scene. Research is needed to look into the current reliability of the system using several ammunition brands and firearm types.

This study aimed to evaluate the Cadre comparison algorithm's accuracy using ten casework firearms and six ammunition brands. It was expected that same ammunition brands would score higher than cross-brand comparisons and that the same firearm cases would score higher than different firearm cases regardless of ammunition brands. The data partially supported this hypothesis with additional complexities that are detailed below.

3. Methods

3.1 Materials

In this study, there were a total of ten firearms chambered in 9mm Luger that were recovered through casework by the Oklahoma State Bureau of Investigation (OSBI). Of the ten firearms there were 30 cartridge cases fired except for two firearms that only had 20 cartridge cases fired from each, for a total of 280 cartridge cases. For each of the firearms there was a total of six different ammunition brands fired, except for the two firearms with only 20 cartridge cases (Table 3.1). The ammunition sources were Winchester, Remington, CCI, Blazer, Maxx Tech, and Wolf. Table 3.2 and Figure 3.1 illustrate each brand's primer and cartridge case components. These ammunition manufacturers were selected due to commonality and the variance in toolmarks they display. The purpose of using casework firearms for comparison is to allow for realistic and applicable results for the firearm and toolmark field.

Table 3.1: Firearm make and model with the corresponding ammunition used to make test fires for this study.

		Winchester	Remington	Maxx Tech	Wolf	CCI	Blazer
Beretta	92FS			x	x	x	x
Glock	19			x	x	x	x
Hi-Point	C9	x	x	x	x	x	x
Ruger	LC9	x	x	x	x	x	x
Ruger	P89	x	x	x	x	x	x
S&W	SD9VE	x	x	x	x	x	x
S&W	M&P9	x	x	x	x	x	x
SCCY	CPX	x	x	x	x	x	x
Springfield	XD9	x	x	x	x	x	x
Taurus	PT92	x	x	x	x	x	x

Table 3.2: Ammunition brands and their respective primer and cartridge case compositions that were used in this study.

	Primer	Cartridge Case
Winchester	Brass	Brass
Remington	Brass	Brass
Maxx Tech	Brass	Steel
Wolf	Brass	Steel
CCI	Nickle	Aluminum
Blazer	Nickle	Brass

Figure 3.1: Images of the ammunition head stamps of the ammunition used in this study.

3.2 Scan Acquisition

Each of the 280 cartridge cases were prepared by giving them a unique four-letter string that was engraved onto the side of the cartridge case and stored in a Microsoft Excel spreadsheet that identified the firearm from which each cartridge case was fired. All cartridge cases were cleaned using 91 percent Isopropyl alcohol and cotton swabs. A small air blower was used to dry the cases, and a sticky tack was used to remove any debris left on the cartridge case. Once prepped, the breech face was scanned using a Cadre Forensics TopMatch-GS 3D. This only included the breech face impression area of the primer. The firing pin impression and the headstamp were not scanned. There is currently no algorithm to compare firing pin impressions, leading to the breech face impression being the focus of this study.

A sinusoid reference standard provided by Cadre Forensics was used for quality control. A sinusoid is a calibrated micron scale reference that uses peak-to-peak distance, or wavelengths, that are known. These reference measurements have been tested and verified by the National Institute of Science and Technology (NIST). The software calculates the peak-to-peak distance and wave amplitude (as shown in Figure 3.2) and compares those measurements to those from NIST. If the measurements are similar, then that ensures the system is providing accurate 3D data. If the measurements are outside the NIST measurements then troubleshooting steps may be performed prior to continuing data collection. In this study, the sinusoid reference was used once before scanning was done on a new day and after every 30 cartridge cases following the first scan, establishing traceability in this study. The Cadre system showed that the scanner never went out of the ranges provided by the sinusoid reference and therefore, no corrective actions were needed during this study.

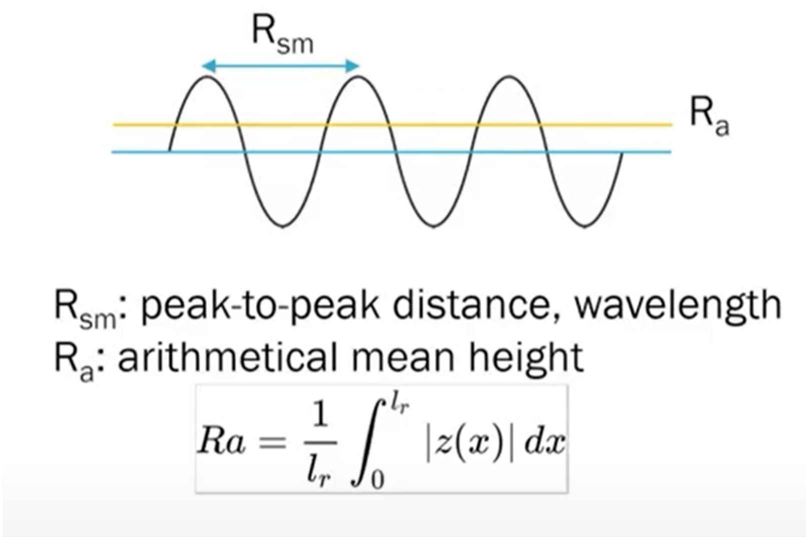


Figure 3.2: Image of a sinusoid wavelength [2] and formula the system uses to calculate the sinusoid reference measurements it compares to the measurements from NIST.

3.3 Automated Comparisons

The Cadre Forensics TopMatch software allows for the creation of masks that indicate the breech face and aperture shear areas of the scanned cartridge case for later comparison using the algorithm, including the breech face, aperture shear, and firing pin. Figure 3.3 shows an example of a masked cartridge case. A masked cartridge case is a cartridge case that has had the applicable breech face and aperture shear areas indicated using the Cadre Forensics TopMatch software. Using the mentioned masks, a user can identify the breech face area, firing pin area, and aperture shear in the software. The automated system only uses the masks created when comparing their respective areas. For example, a breech face compared to another breech face will use the breech face mask to generate an automated breech face score. This system has a proprietary algorithm that was developed by Cadre Forensics that produces a score from 0.0 to 1.0. A score closer to 0.0 indicates that the questioned case is from a different gun, while a score closer to 1.0 indicates that the questioned case came from the same gun. Once scanning was completed, each cartridge case was compared using TopMatch software to all other cartridge cases scanned as part of this study for a total of 39,060 comparisons. This produced a large data set of same versus different gun scores and same versus different ammunition scores.

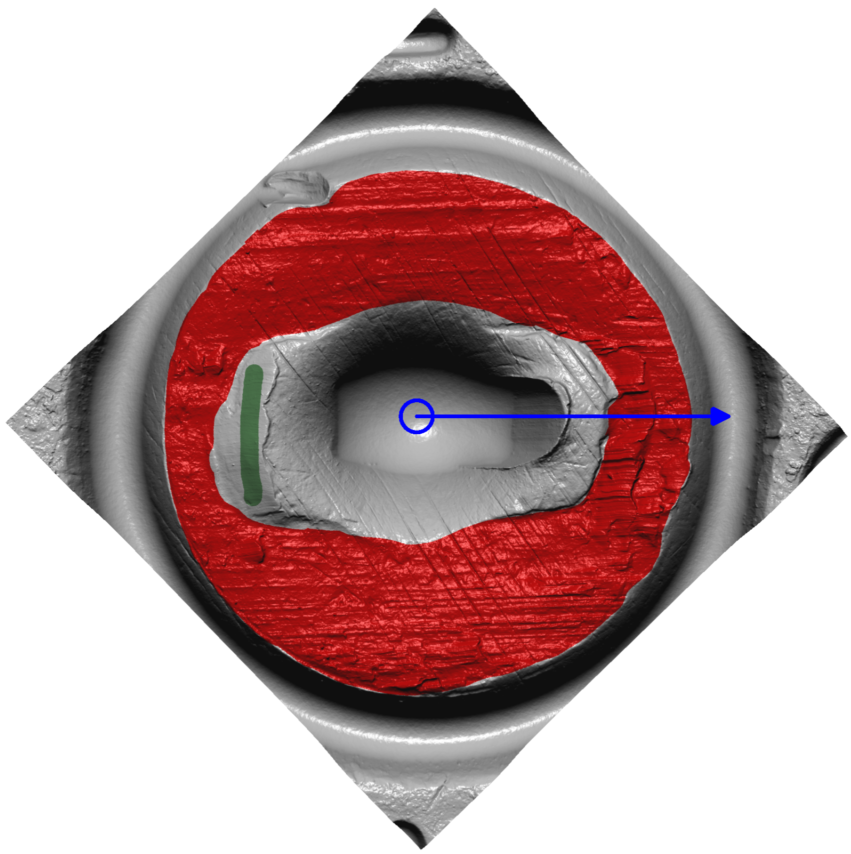


Figure 3.3: An example mask marking the comparable areas for the algorithm. The breech face impression is in red and the aperture shear is in green.

Figure 3.4 shows an ideal scenario where there is a complete separation of the match and nonmatch data [3]. The SD9 in this example leaves very distinctive, reproducible marks. The study that used this distribution used a different algorithm, so the x and y axis references slightly different similarity scores than what was used in the current study [3]. This example of a match and nonmatch distribution illustrates a definable gap between nonmatch and match scores for this particular firearm. In this case, the match data are clustered in the upper right corner of the plot area, while the nonmatch data are clustered near the origin. The degree of separation between these data points indicates the similarity between the cases from this firearm. However, we may observe the opposite, where scores of a nonmatch and a match overlap and have no identifiable gap.

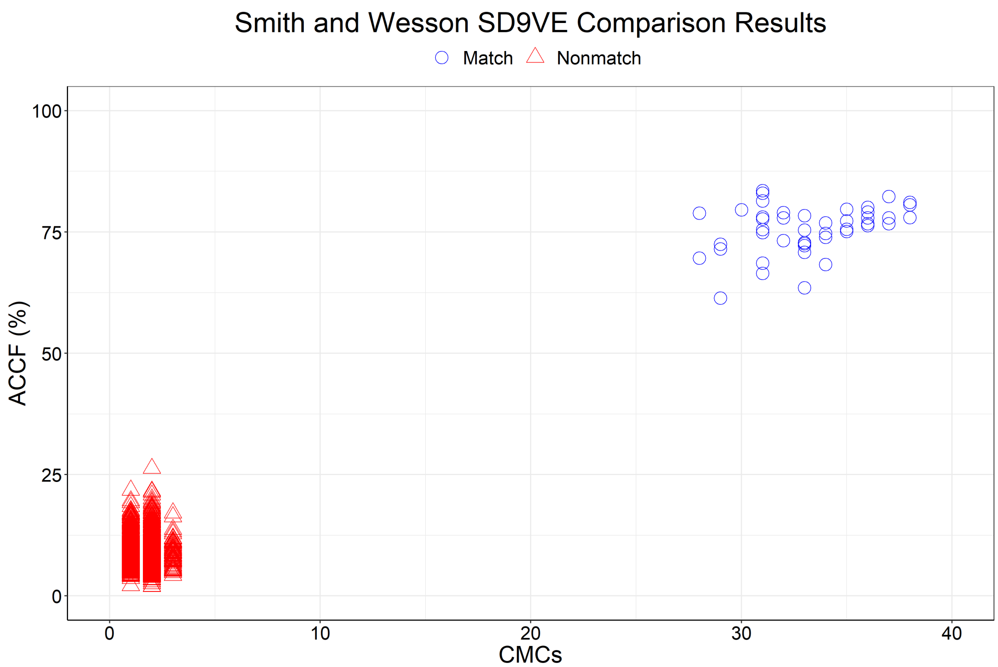
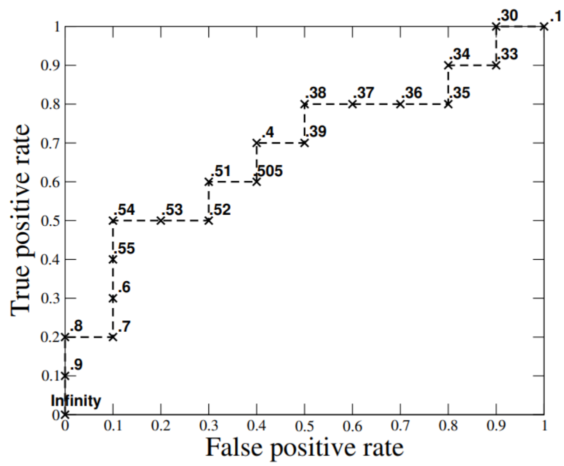


Figure 3.4: An example of a match and nonmatch distribution plot demonstrating complete separation between match and nonmatch data [3].

Receiver operating characteristic (ROC) curves were created using fully known match data sets for each firearm, and the resulting area under the ROC curve (AUC) values were calculated for each scatterplot. The AUC is defined as the probability that a randomly selected match will have a higher similarity score than a randomly selected nonmatch. It is a very interpretable statistic and describes the degree of separation observed on the match/nonmatch scatterplots. ROC curves were compared as described by DeLong et al.[17], and implemented using the algorithm developed by Sun and Xu [18] within the `pROC` R package [19]. These comparisons show which firearm and ammunition combinations perform best and worst in the Cadre Forensics comparison algorithm. Figure 3.5 shows a simple example of a ROC curve where the scores are organized by highest to lowest score [18]. For each positive (p), the curve goes up on the graph, and for each false positive (n), the curve goes horizontally. With a one-hundred percent true positive rate, the outcome would be a straight line going up the y-axis and a straight line along the x-axis, forming a right angle. This would mean that the highest scores represent matches, and the lowest scores represent nonmatches. The AUCs generated by this study will be compared using a procedure based on the Mann-Whitney test to determine significance between same ammunition and different ammunition comparisons. This procedure is located within the `pROC` R package [17], [18], [19].



Inst#	Class	Score	Inst#	Class	Score
1	**p**	.9	11	**p**	.4
2	**p**	.8	12	**n**	.39
3	**n**	.7	13	**p**	.38
4	**p**	.6	14	**n**	.37
5	**p**	.55	15	**n**	.36
6	**p**	.54	16	**n**	.35
7	**n**	.53	17	**p**	.34
8	**n**	.52	18	**n**	.33
9	**p**	.51	19	**p**	.30
10	**n**	.505	20	**n**	.1

Figure 3.5: An example ROC curve [4]. This figure illustrates how scoring data is used to create ROC curves.

4. Results and Discussion

4.1 Breech Face Similarity Scores

The same source and different source distribution plots were created to illustrate the numeric algorithm scores generated by the Cadre Forensics TopMatch-GS 3D. These distribution plots were created with all the data combined to evaluate the system as a whole. These data were also analyzed on a per firearm basis to assess differences in performance between different firearm models. Lastly, distribution plots were created to examine the impact of ammunition brands on the algorithm. The hypothesis in this study was that the same gun scores would be closer to a value of 1.0 while different gun scores would be closer to a value of 0.0. This separation between the same source and different source scores would allow for an expectation that when a score is plotted in a particular area, it would result in the correct ground truth conclusion.

The overall results showed that the same source same ammunition brand comparisons scored higher, with the same source different ammunition brand comparisons showing more variance between high and low scores. Figure 4.1 shows every score separated by same and different ammunition regardless of firearm model. The x-axis shows the source while on the y-axis is the breech face similarity score ranging from 0.0 to 1.0. Same source means that two cartridge cases were fired from the same gun and different source means two cartridge cases were fired by different firearms. During this study it was not expected that every firearm would perform with complete separation of same and different source due to the variability in the quantity and quality of the different marks firearms leave. It was expected in this study that same ammunition and same source scores would perform better than different ammunition scores from the same source, however, they would both have scores closer to 1.0 on the breech face similarity due to being from the same source firearm. The different source scores across both different and same ammunition scores are expected to have a breech face similarity score closer to 0.0. Overall, same ammunition and same source comparisons scored higher than different ammunition and same source comparisons. Different ammunition scores with the same source were more varied through the 1.0 to 0.0 similarity scale. Different ammunition and same source comparisons had Sixty-eight percent of comparisons scoring 0.5 on the similarity score or higher, while ninety percent of same ammunition same source comparisons scored a 0.5 similarity score or higher.

Figure 4.2 shows an overall distribution plot of scores split by firearm model. The Hi-Point C9 is the most notable, with every comparison from the same firearm having a 0.5 similarity score or above. It was expected that the Hi-Point C9 would score high due to its very prominent parallel marks. However, it was not expected to perform well across all ammunition brands.

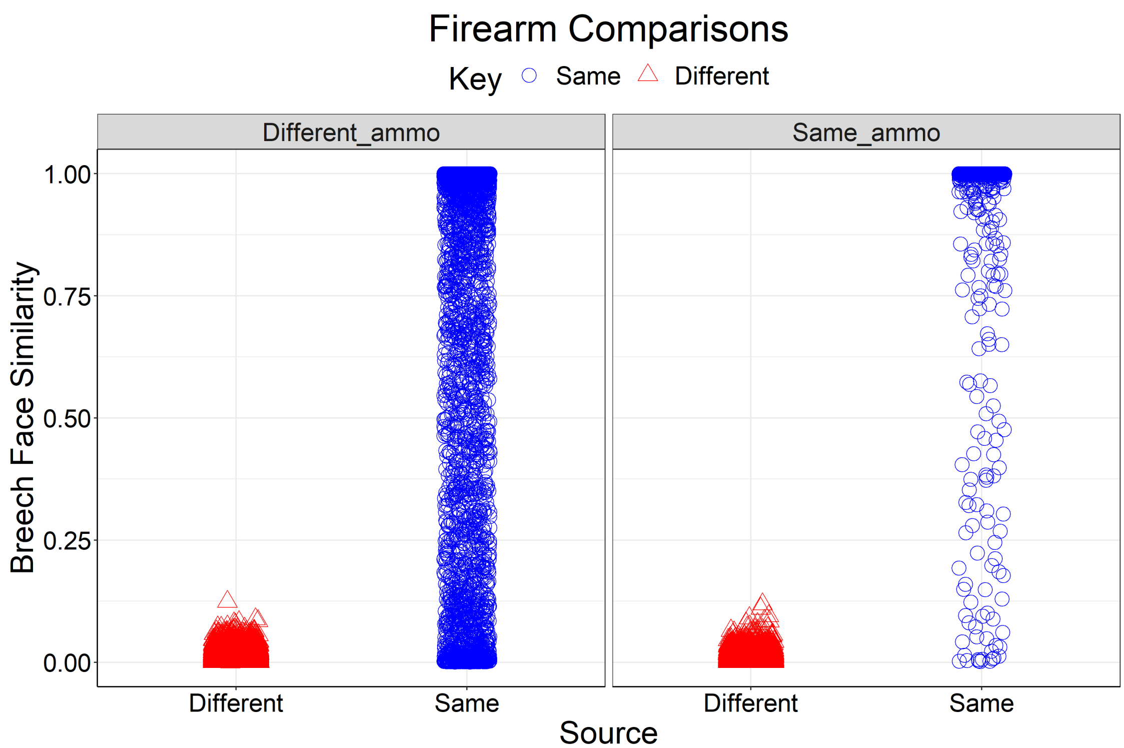


Figure 4.1: Scatterplot of every comparisons score split by same and different ammunition type.

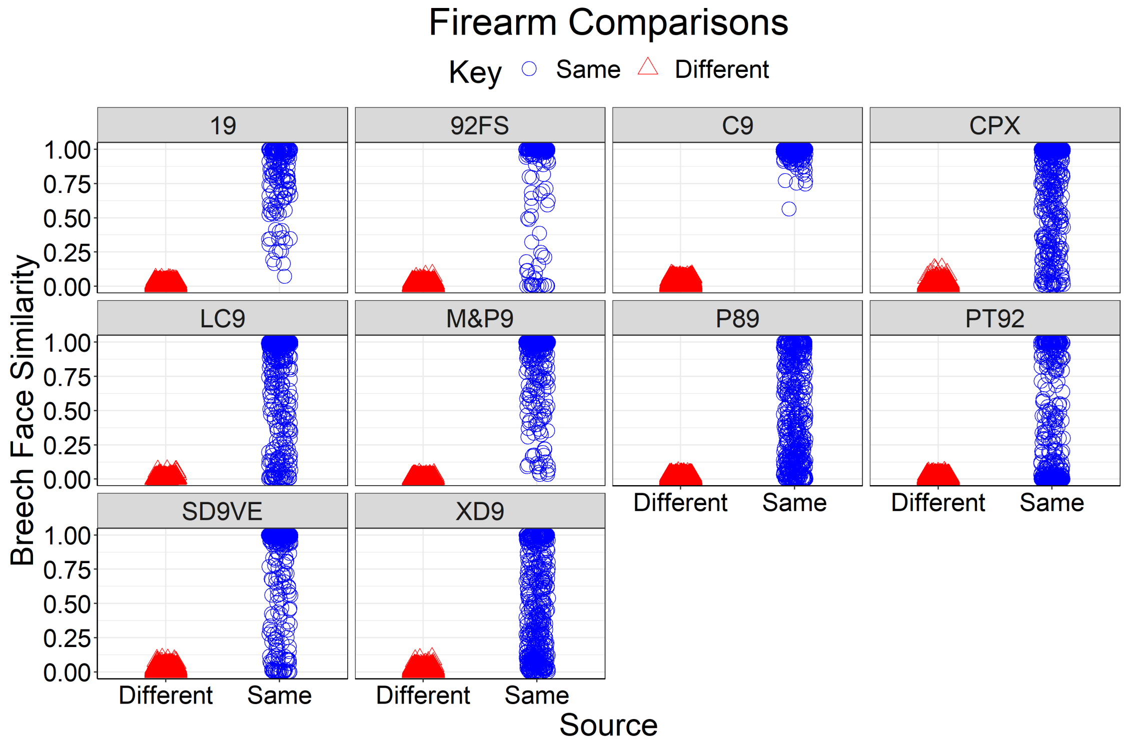


Figure 4.2: Overall scatter plot of comparison scores split by firearm model.

Figure 4.3 shows that regardless of the firearm model, the algorithm scored Maxx Tech and Wolf above a 0.9 between and within the two ammunition brands. Maxx Tech and Wolf scoring this high was not expected. These ammunition brands have a history of producing marks poorly due to their cartridge cases being made of steel [13]. However, in this study, the ammunition allowed for very definitive toolmarks. Figure 4.4 shows a comparison of Wolf to Maxx Tech cartridge cases from a Beretta 92FS.

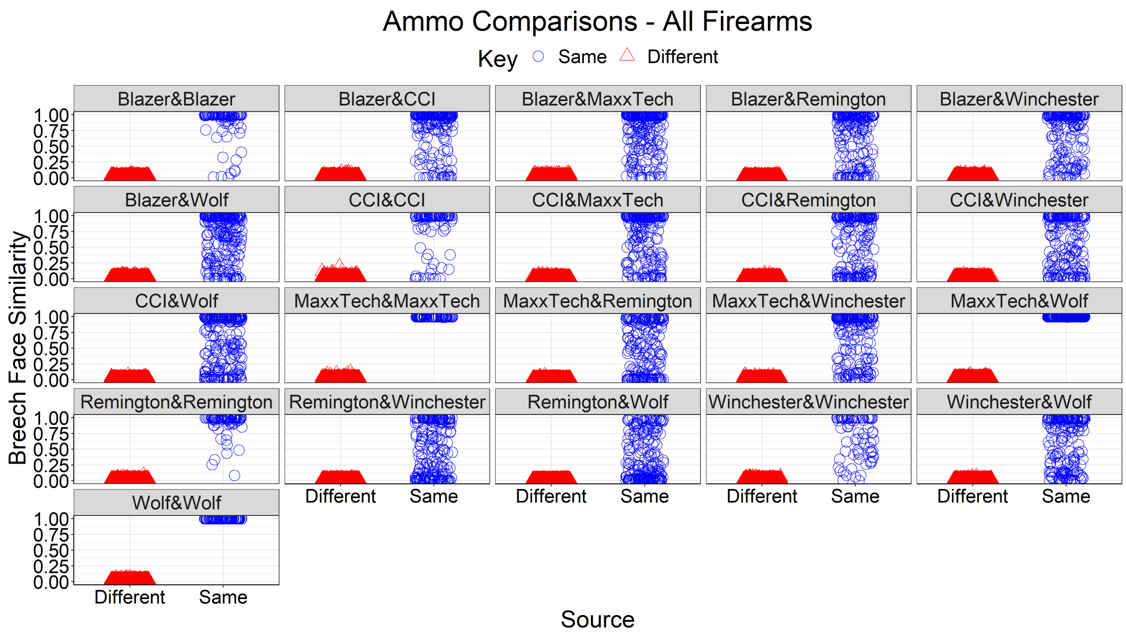


Figure 4.3: Overall scatterplot of comparison scores split by different ammunition combinations.

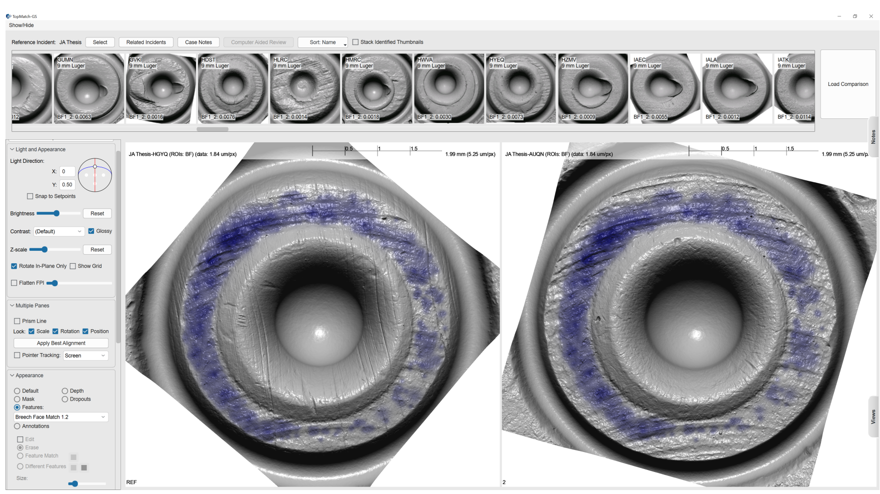


Figure 4.4: Cartridge case comparison of Wolf to Maxx Tech from a Beretta 92FS using Cadre software.

Distribution plots were created for each of the ten firearms to visualize the possible effects of ammunition brands better. The comparisons were then separated by the two ammunition brands used in the comparison. Figure 4.5 shows the same source and different source distribution plot for the Ruger P89 used in this study. Comparisons involving the ammunition brands Remington and Winchester scored low on the breech face similarity score. Remington compared to Blazer and Winchester caused a wide range of breech face similarity scores. However, when compared to CCI, Maxx Tech, and Wolf, most breech face similarity scores were below 0.5. One of the comparisons between Remington and Maxx Tech that scored a 0.005 on breech face similarity can be found in Figure 4.6. On the Maxx Tech (Left), there are parallel manufacturer marks. The Remington cartridge case (Right) appears to have had less breech face area for comparison. Figure 4.7 shows the comparison of two CCI cartridge cases fired from the Ruger P89 used in this study. The automated comparison gave these two cartridge cases a breech face similarity score of 0.32. The denoted areas of similarity were primarily on the top left and bottom right of both cartridge cases. On the left CCI cartridge case, concentric marks can be seen on the lower half. The algorithm appears to have had trouble identifying the similarity where those concentric marks were located.

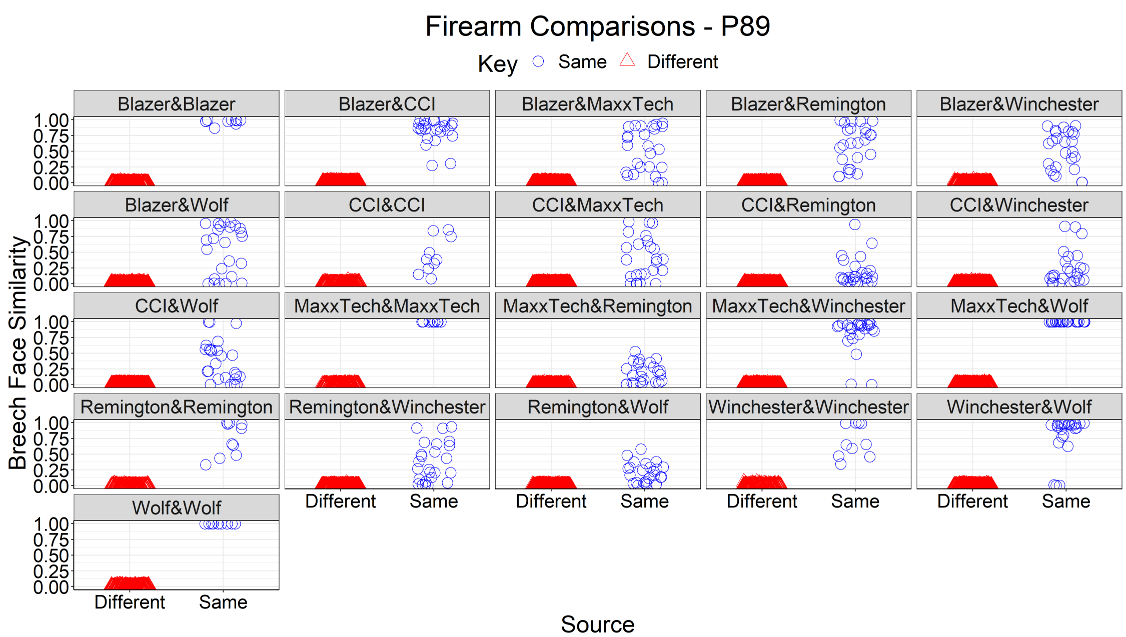


Figure 4.5: All same source and different source comparisons for the Ruger P89.

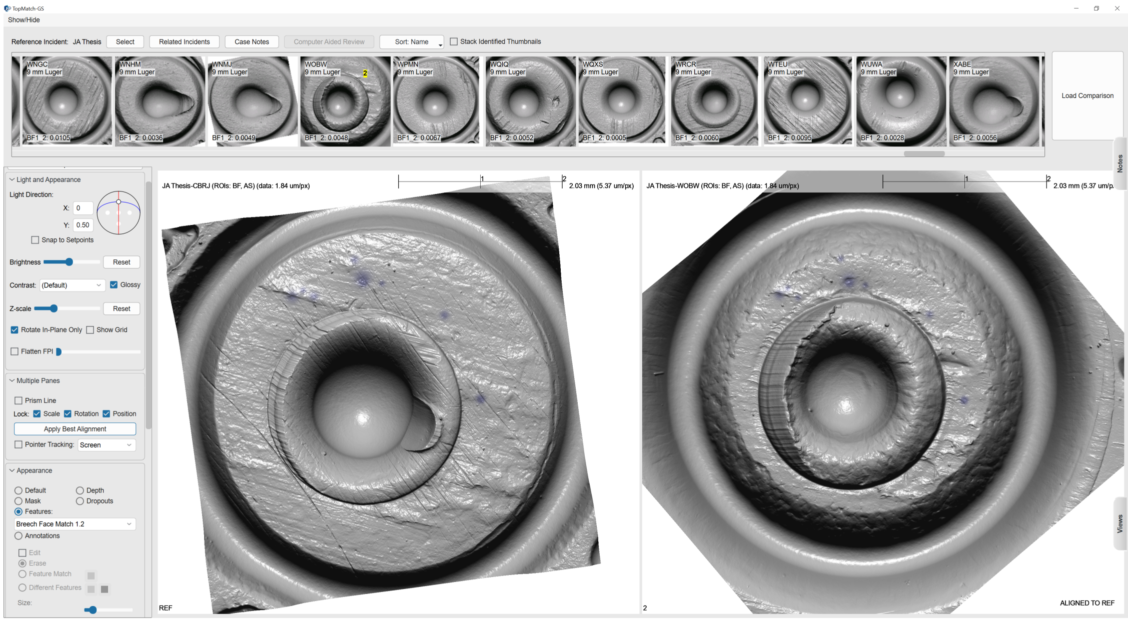


Figure 4.6: Comparison of a Ruger P89 Remington and Maxx Tech cartridge case that generated a score of 0.005.

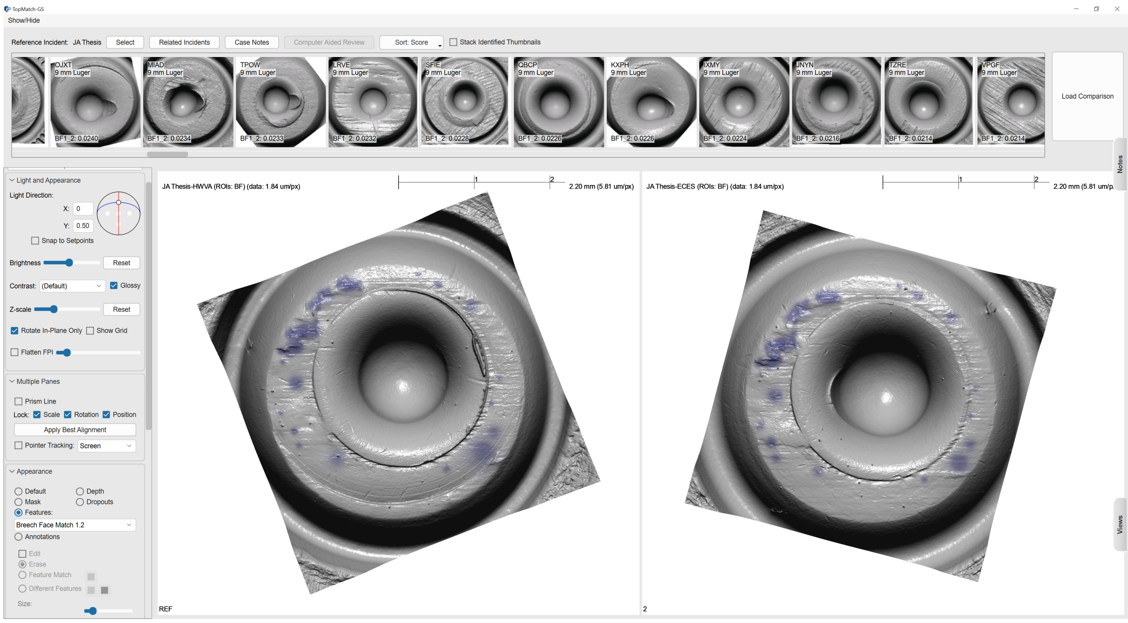


Figure 4.7: Comparison of two Ruger P89 CCI cartridge cases that generated a score of 0.32.

The Beretta 92FS same source and different source distribution plot for this study can be found in Figure 4.8. Comparisons involving CCI and Wolf had several low breech face similarity scores. A visual comparison between a CCI (Left) and a Wolf (Right) can be seen in Figure 4.9. On the CCI cartridge case, there was a larger flowback that took up most of the primer, leaving little usable breech face area for comparison. The algorithm was able to identify similarity despite the little breech face area available. That similarity was mainly found in the top portion of this cartridge case but none on the bottom. The algorithm gave this comparison a 0.06 breech face similarity score. This is likely due to the algorithm generating scores based on the overall mask similarity. The CCI cartridge case had high similarity when looking at the total masking area of the breech face, but the Wolf mask had significantly more area, and the similarity found was only a fraction of what area was available for comparison. This CCI cartridge case caused a couple of lower comparison scores when compared to both Wolf and Maxx Tech. There were several comparisons involving Blazer that generated a low score even when Blazer was compared to Blazer. Figure 4.10 compares a Blazer (Left) and a Maxx Tech (Right). This comparison generated a breech face similarity score of 0.003. This comparison is very similar to the previous comparison, where the Blazer cartridge case has flowback that covers most of the breech face area used for comparison. The algorithm was able to identify similarity on the top portion of the cartridge case, but there was no similarity found on the bottom portion due to there being such a small area for comparison purposes. Every same source cartridge case that was compared to this Blazer cartridge case generated below a 0.2 breech face similarity score; besides one comparison that was to another Blazer cartridge case, that comparison scored a 0.33. This Blazer cartridge case was the cause of almost every low same source score for the Beretta 92FS that involved Blazer.

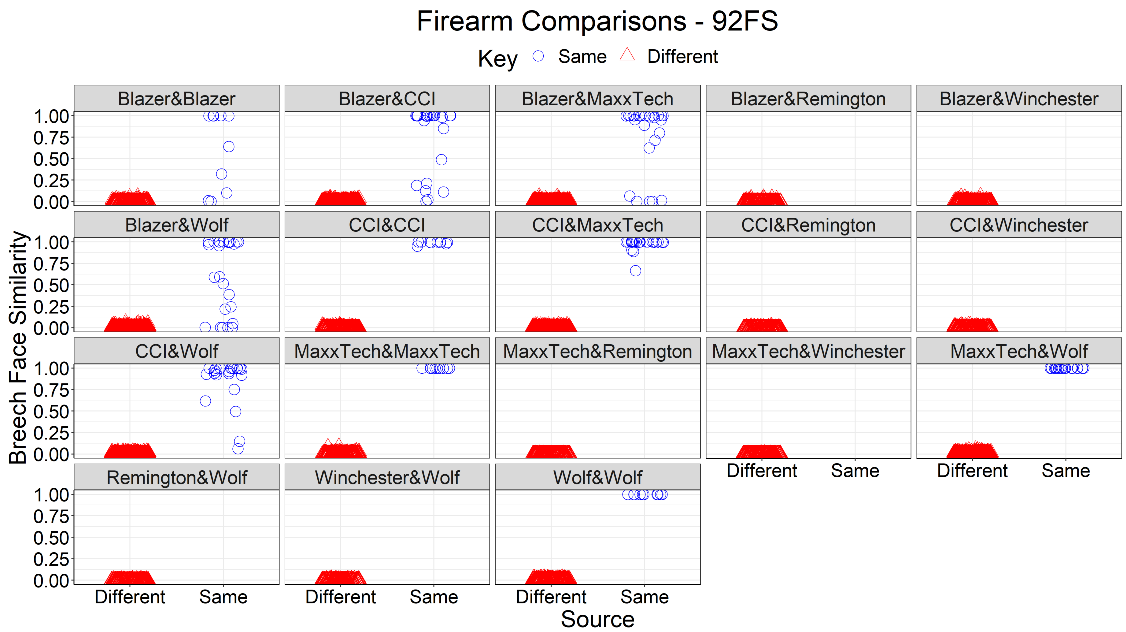


Figure 4.8: All same source and different source comparisons for the Berreta 92FS.

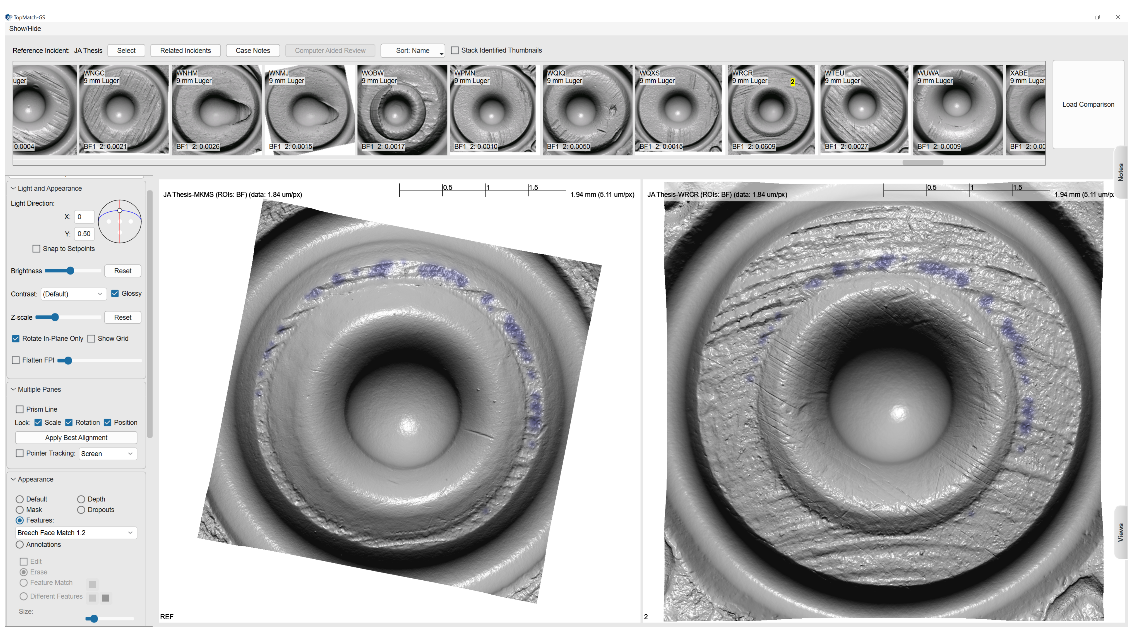


Figure 4.9: Comparison of a Beretta 92FS CCI and Wolf cartridge cases that generated a score of 0.06.

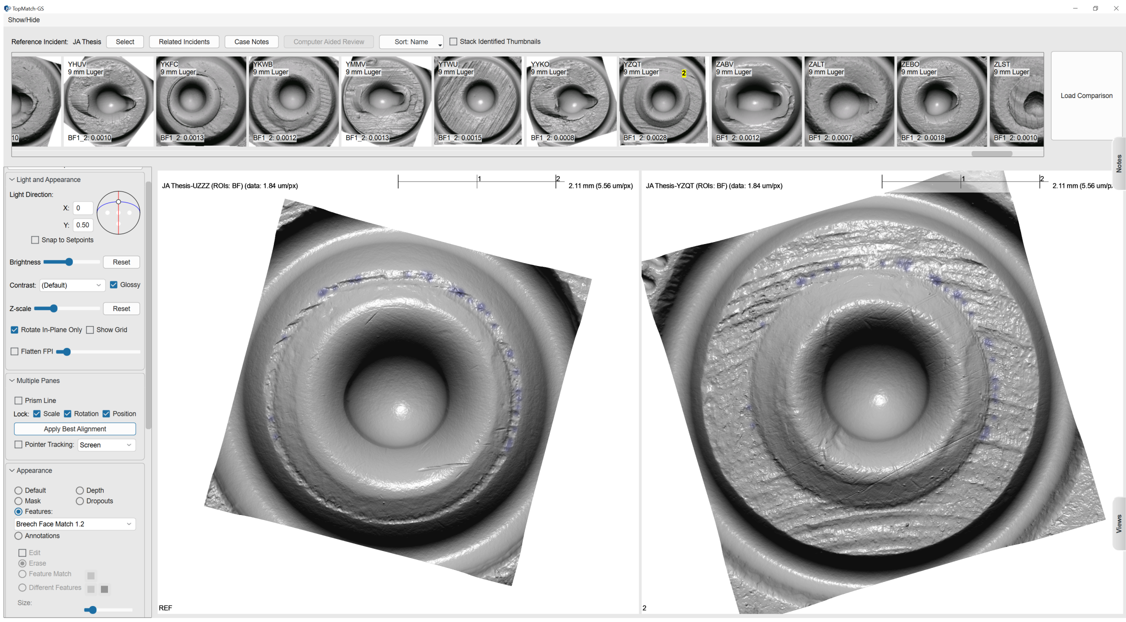


Figure 4.10: Comparison of a Beretta 92FS Blazer and Maxx Tech cartridge cases that generated a score of 0.003.

The distribution plot for the SCCY CPX can be found in Figure 4.11. Comparison results for this firearm are similar to previous firearms where Maxx Tech and Wolf scored high within and between. However, when Blazer, CCI, Wolf, and Maxx Tech are compared to Winchester and Remington there was a high variability of low and high scores with Winchester performing the poorest overall. Some comparisons gave interesting breech face similarity scores worth noting. Figure 4.12 compares a CCI (Left) and a Blazer (Right) cartridge case comparison that generated a 0.28 breech face similarity score. The areas of similarity identified by the algorithm are predominantly the shearing on both the right and left sides of both cartridge cases. The Blazer had more shearing on the right side than that of the CCI cartridge case. There was little to no similarity found anywhere else. Figure 4.13 compares two CCI cartridge cases with a breech face similarity score of 0.997. The algorithm found similarities in the same areas as the CCI and Blazer comparison. However, there were additional similarities between the two CCI cartridge cases found in some shearing on the right side. Visually there does not appear to be a significant difference in areas of comparison, but there was a difference in breech face similarity of roughly 0.70. The extra shearing locations and overall breech face area found in the Blazer cartridge case in Figure 4.12 likely contributed to a lower score. To help mitigate false positives, the algorithm has a tipping point where the similarity score will jump to a much higher score once it is reached. However, if the algorithm does not find the amount of similarity needed the comparison scores may appear low even when compared to another cartridge case that has similar marks. In Figure 4.13 the shearing locations are the same size and number. They also appear to have the same number of striations inside the shear. This means the algorithm saw these areas of similarity and weighted them more heavily when generating the breech face similarity score, leading to a much higher score.

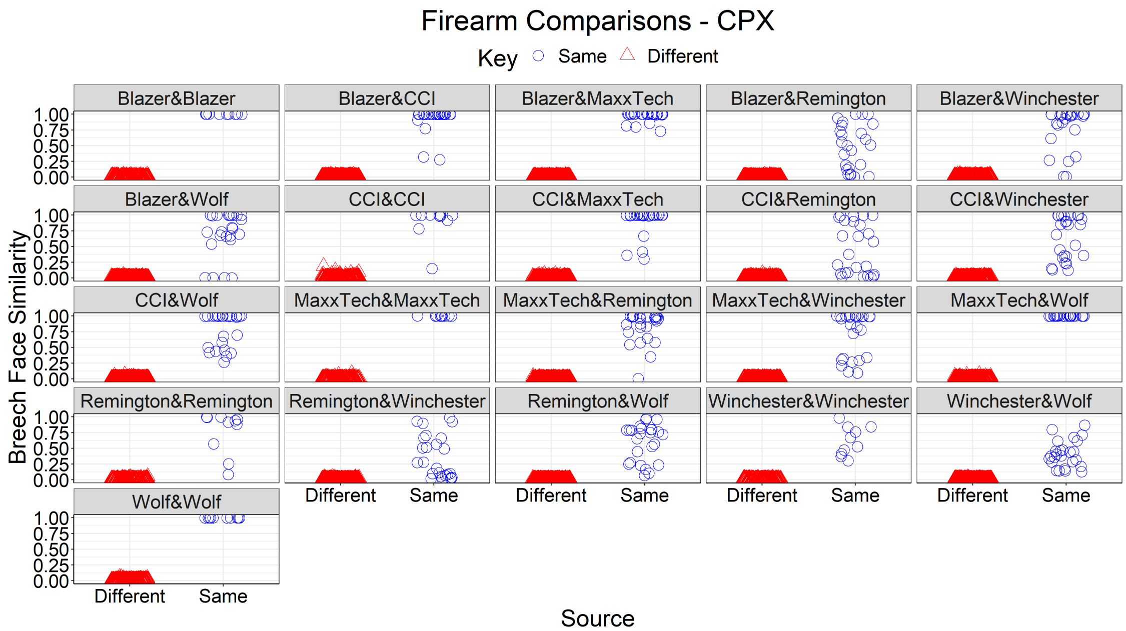


Figure 4.11: All same source and different source comparisons for the SCCY CPX.

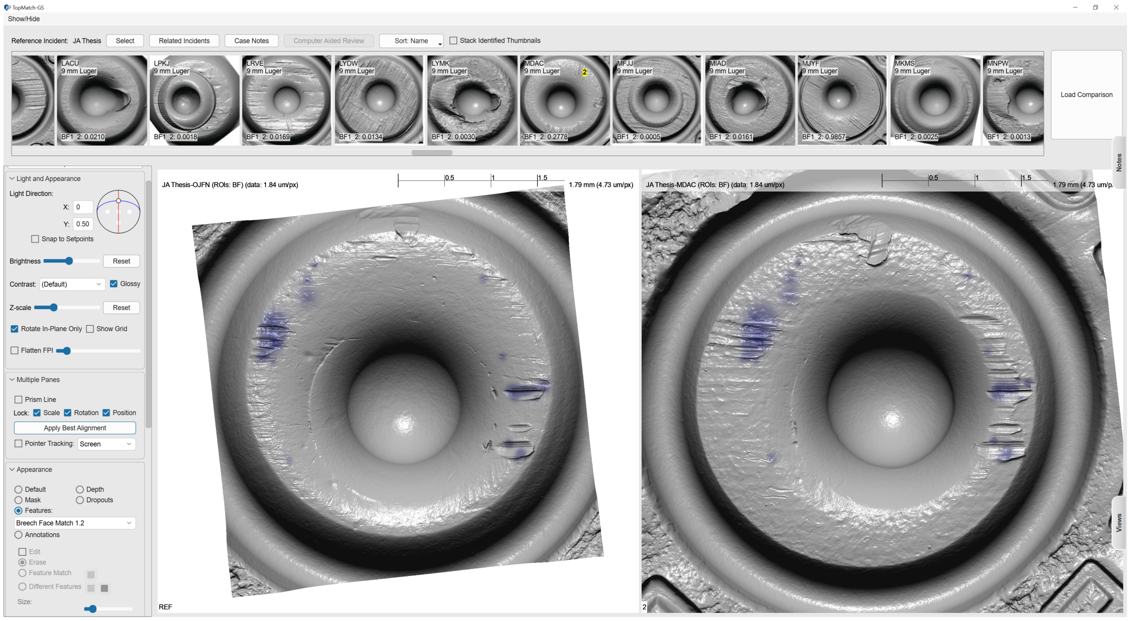


Figure 4.12: Comparison of a SCCY CPX CCI and Blazer cartridge cases that generated a score of 0.28.

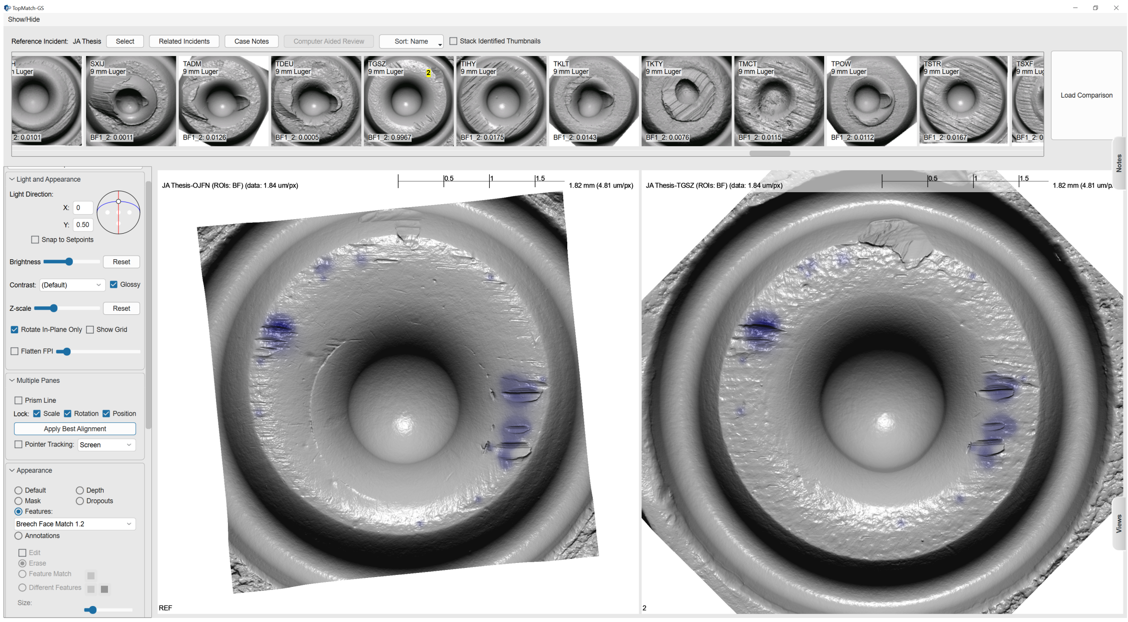


Figure 4.13: Comparison of two SCCY CPX CCI cartridge cases that generated a score of 0.997.

Figure 4.14 shows the same source and different source distribution plot for the Glock 19 used in this study. One of the Blazer to Blazer comparisons had a low breech face similarity score. Scores between the same ammunition brands and same firearm were expected to have a high breech face similarity score. That comparison can be found in Figure 4.15. Most of the similarity was found in the cartridge case's bottom right and top left. The breech face similarity score for this comparison was 0.25. Comparisons between CCI and Wolf were notably worse than other brands that CCI was compared to for the Glock 19. The four lowest scores were contributed to two cartridge cases. Figure 4.16 shows the comparison between a CCI and Wolf cartridge case that generated one of the low scores. On the CCI (Left) cartridge case you can see concentric marks on the top half of the cartridge case. On the Wolf (Right) cartridge case, there are parallel lines flowing through the firing pin impression onto the opposite side of the cartridge case. These parallel marks are likely marks transferred during the primer manufacturing process, not the firing process. The algorithm only identified marks to the bottom right of the cartridge case. The area above the rectangular aperture appears different. The parallel marks on the right case are not on the left case, and the concentric marks are all likely reasons why the breech face similarity score was low.

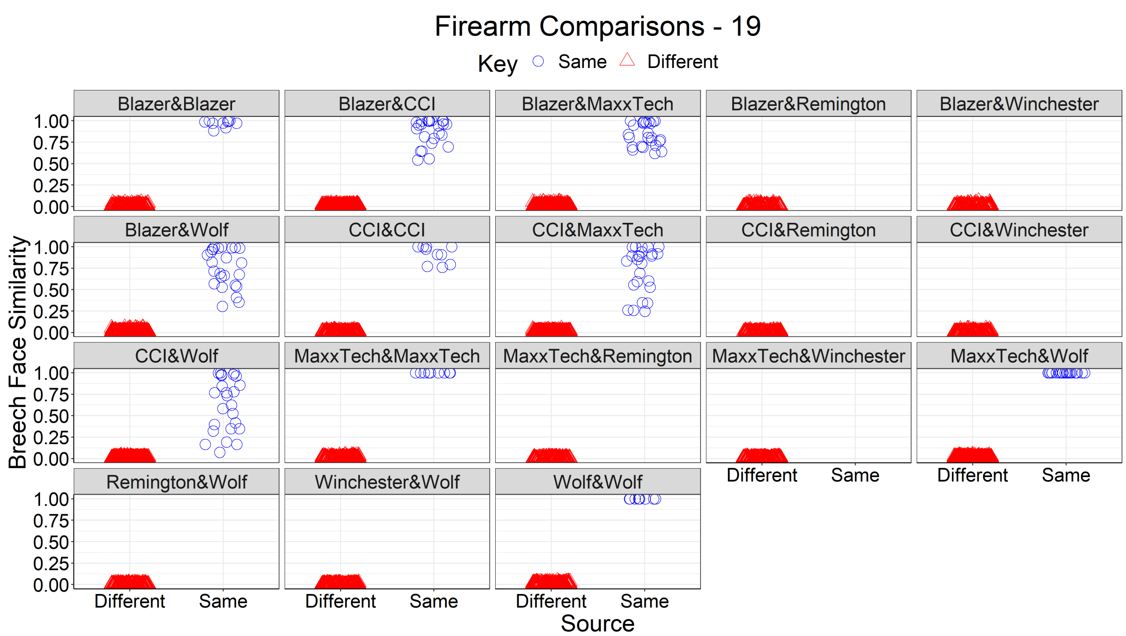


Figure 4.14: All same source and different source comparisons for the Glock 19.

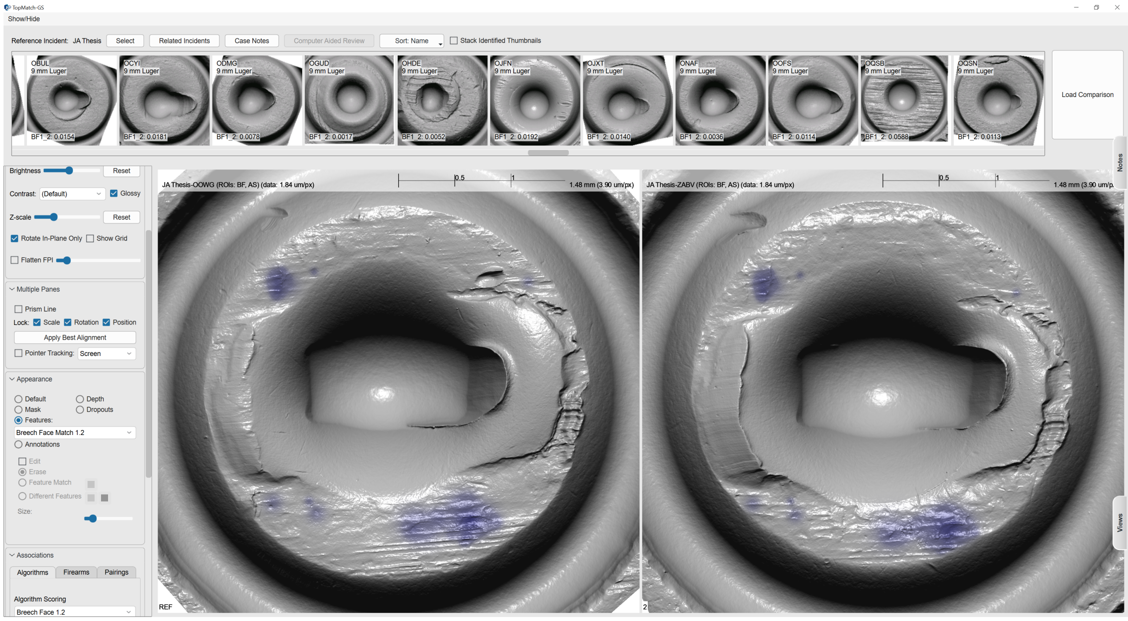


Figure 4.15: Comparison of two Glock 19 Blazer cartridge cases that generated a score of 0.25.

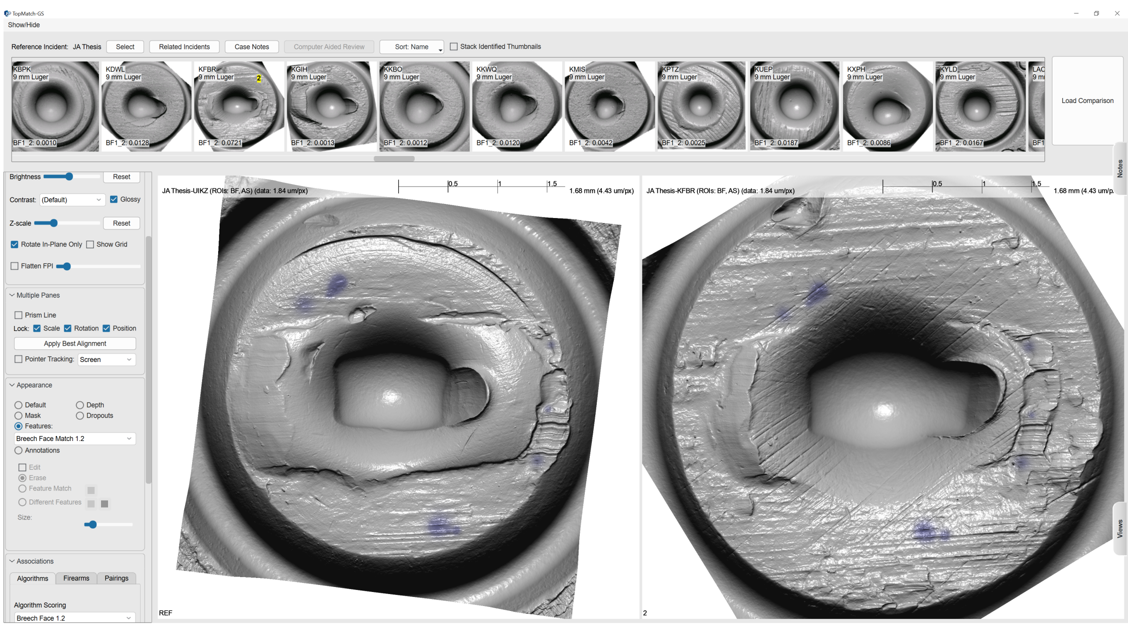


Figure 4.16: A Glock 19 comparison of a CCI and Wolf cartridge case that generated a score of 0.07.

The distribution plot for the Smith and Wesson M&P9 used in this study can be found in Figure 4.17. Overall, the Smith and Wesson M&P9 scored very high across most ammunition brands. However, the breech face similarity scores had more variability in scoring when the cartridge case brands Blazer and Remington were compared to Maxx Tech and Wolf. Figure 4.18compares a Blazer (Left) and a Wolf (Right) cartridge case that generated a breech face similarity score of 0.17. Minimal similarity was found between these two cartridge cases, with only a few areas above and below the aperture shear and center top and bottom of the breech face. The Blazer cartridge case appears to have had a smaller area for comparison than that of the Wolf cartridge case. The Wolf cartridge case also has parallel primer manufacturer marks that flow from top to bottom, with the marks flowing into the firing pin impression. The variability in marks between these two ammunition brands, coupled with Wolf's primer manufacturer marks and the overall surface area for comparison, has likely led to the lower scores when Blazer was compared to Maxx Tech and Wolf. Blazer scored high compared to all the other ammunition brands, even Remington, which also had low scores compared to Maxx Tech and Wolf. However, CCI and Winchester generated high scores when compared to Maxx Tech and Wolf.

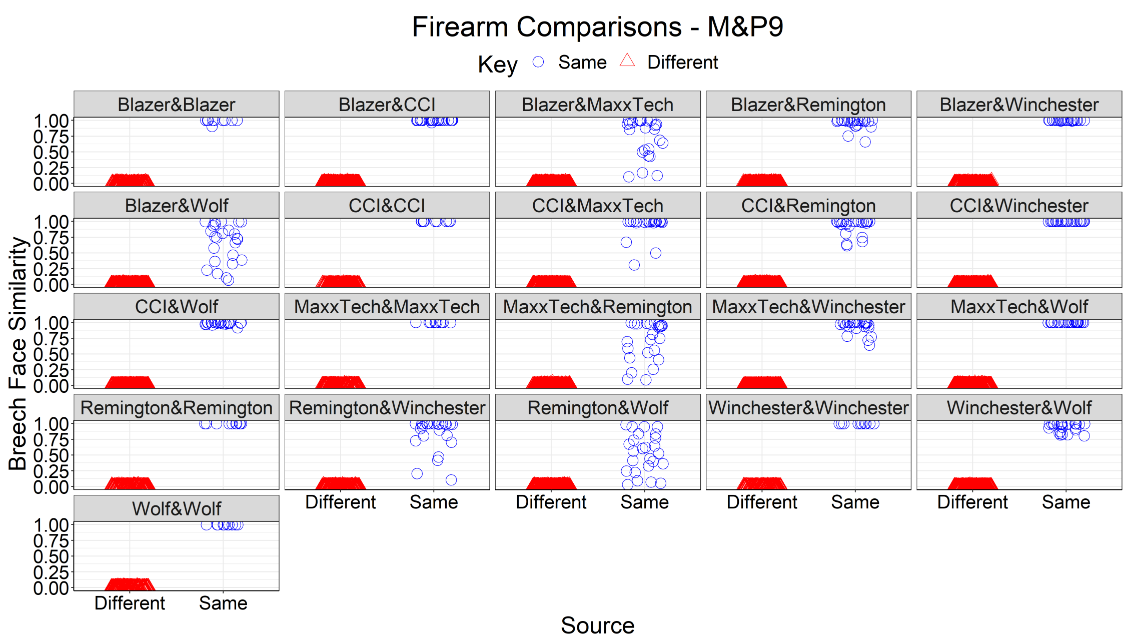


Figure 4.17: All same source and different source comparisons for the Smith and Wesson M&P9.

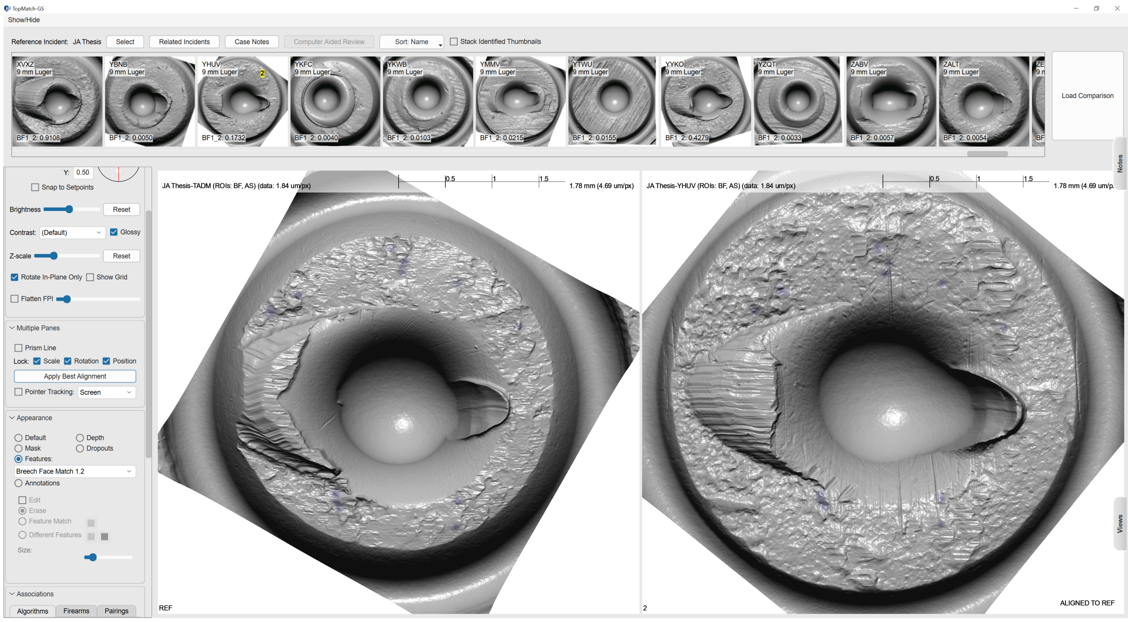


Figure 4.18: Comparison of a Smith and Wesson M&P9 Blazer and Wolf cartridge case that generated a score of 0.17.

The other firearm scatter plots showed similar results to the previously discussed firearms. Maxx Tech and Wolf scored exceptionally high when compared between and within. Where the breech face similarity score, on average, was above 0.85. Remington had a higher variability between breech face similarity scores compared to other ammunition brands, especially Maxx Tech and Wolf. Almost all brands scored high when compared to themselves. The exception to this was Winchester, where there was a higher variability between low and high scores. The exception to this is the Taurus PT92. Figure 4.19 shows all the different and same source comparisons for the Taurus PT92. This firearm scored exceptionally low across the board when different ammunition brands were compared to each other and themselves. CCI even had low scores when compared to itself. The exceptions are MaxxTech and Wolf, which again scored very high when compared to themselves and each other. The PT92 had more shot-to-shot variability than the other firearms used in this study, leading to the lower breech face similarity scores. Scatter plots for the remaining firearms can be found in Appendix A.

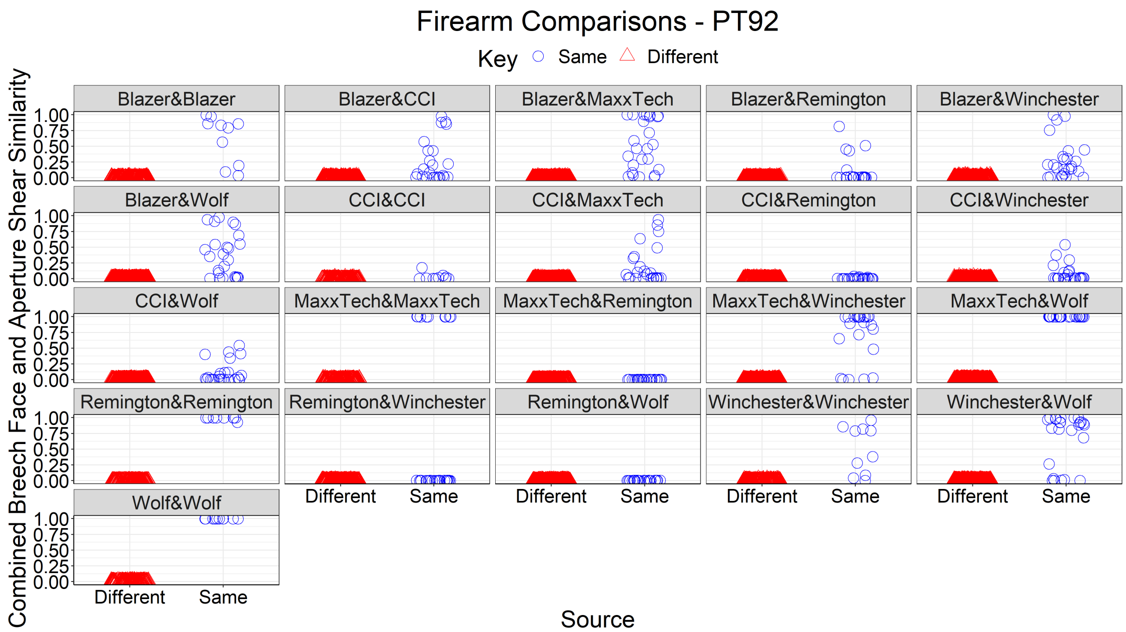


Figure 4.19: All same source and different source comparisons for the Taurus PT92.

4.2 Combined Breech Face and Aperture Shear Similarity Scores

At the time of this study Cadre Forensics released an addition to their current algorithm that takes the score of the aperture shear in conjunction with the breech face score and generates an overall comparison score. To illustrate these results a different distribution plot for each firearm was created. The x-axis is the same and different source, and the y-axis is the combined breech face and aperture shear similarity score. The combined algorithm runs both the breech face and aperture shear algorithm and displays the highest score as the combined breech face and aperture shear similarity score. In the following plots, different source comparisons are triangles, and the same source comparisons are circles. If the breech face scored highest, then the symbol will be blue. If the aperture shear scored the highest, the symbol will be orange, and if both scored the same, then the symbol will be red.

Figure 4.20 shows the Smith and Wesson SD9VE results when the combined breech face and aperture shear scores were first run. When first looking into this, it was noticed that several cartridge cases had low breech face scores and very high aperture shear scores for different source comparisons, indicating the aperture shear results were providing misleading values. Upon investigation, most of the high aperture shear scores were attributed to a handful of cartridge cases. Figure 4.21 shows a comparison between a Taurus PT92 and a Smith and Wesson M&P9 Remington cartridge cases. This comparison generated a breech face similarity score of 0.0025, an aperture shear score of 0.9951, and a combined score of 0.0025. This different source comparison generated a low combined score, which is to be expected. However, the aperture shear score being a 0.9951 for a different source comparison is a major concern. Looking at the aperture shear profile showed the algorithm was taking the shear on the Taurus and comparing it almost vertically to the M&P9 (see the green lines on the aperture shear in Figure 4.21). The system has trouble comparing shears that are less than 800 microns long, which can lead to high scores being generated by the aperture shear algorithm. When the algorithm detects something wrong with the shear it is comparing to, such as a length less than 800 microns, it will default to the breech face score even if there is a high aperture shear score. This helps prevent false positives from being reported by the algorithm. If only the aperture shear was used here, an examiner would quickly realize this is a nonsense result upon visual inspection of the cases. Several interesting aperture shear scores existed between a couple of Smith and Wesson M&P9 cartridge cases and a Springfield XD9 cartridge case. Figure 4.22 shows a comparison of a Springfield XD9 CCI cartridge case (Left) to a Smith and Wesson M&P9 Wolf cartridge case (Right) that generated a breech face similarity score of 0.0009, an aperture shear score of 0.8787, and a combined score of 0.0009. Again, if an examiner looked at these two cartridge cases, they could easily identify this as a nonsense result. However, unlike the almost vertical comparison between the PT92 and the Smith and Wesson M&P9, the aperture shear profiles are similar aperture shear profiles when comparing two different source cartridge cases. This PT92 CCI cartridge case led to a handful of high aperture shear scores in the Smith and Wesson M&P9 with the same and different source distribution plots, as seen in Figure 4.23.

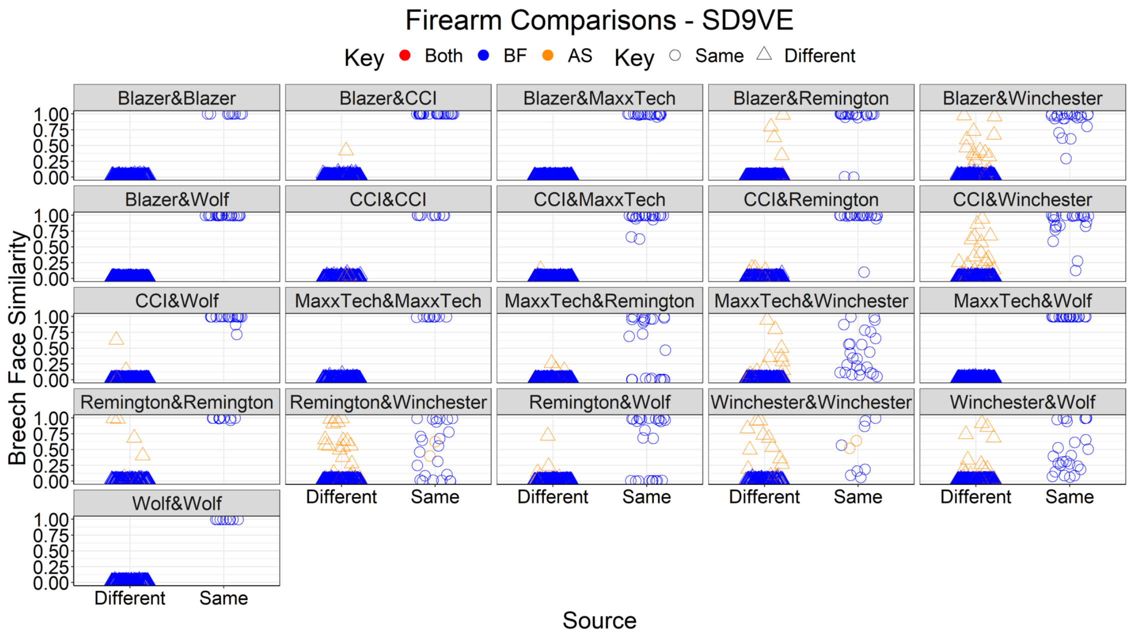


Figure 4.20: All same source and different source comparisons using the combined breech face and aperture shear algorithm for the Smith and Wesson SD9VE.

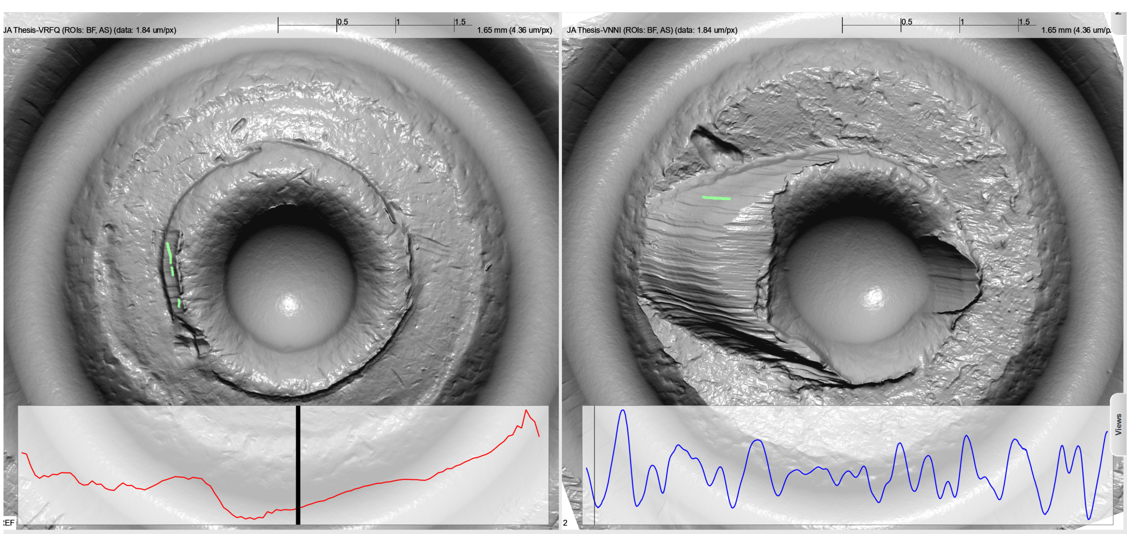


Figure 4.21: Comparison between a Taurus PT92 and a Smith and Wesson M&P9 Remington cartridge cases. This comparison generated a breech face similarity score of 0.0025, a aperture shear score of 0.9951, and a combined score of 0.0025.

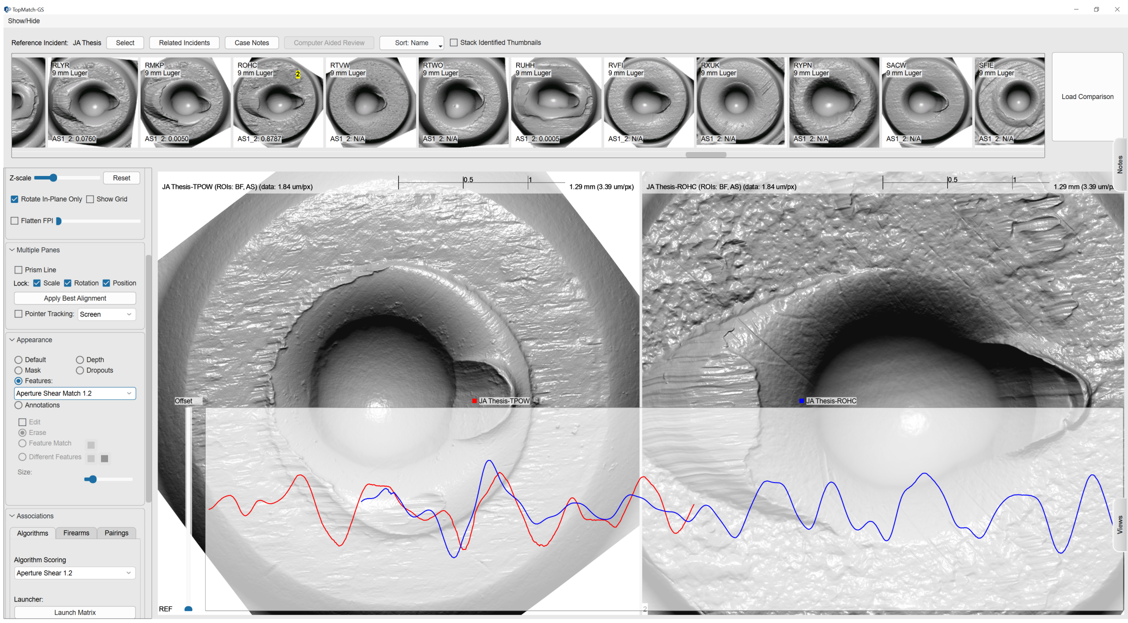


Figure 4.22: Comparison of a Springfield XD9 CCI cartridge case (Left) to a Smith and Wesson M&P9 Wolf cartridge case (Right) that generated a breech face similarity score of 0.0009, a aperture shear score of 0.8787, and a combined score of 0.0009.

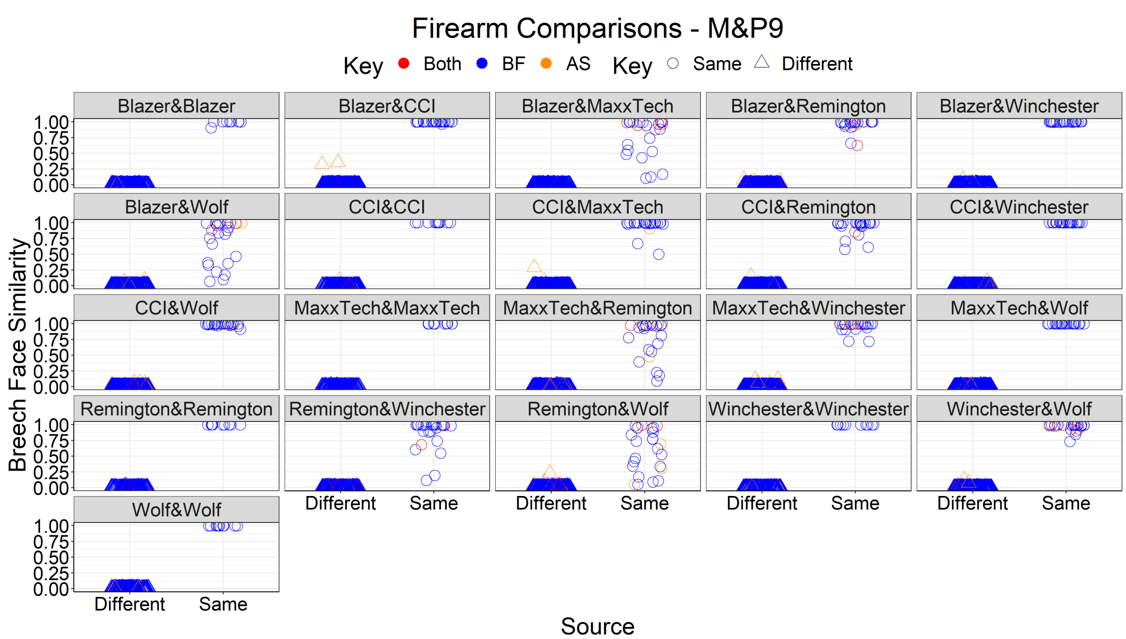


Figure 4.23: All the same source and different source comparisons distribution plot with the combined breech face similarity and aperture shear scores for the Glock 19.

Figure 4.14 shows the same source and different source distribution plot for the Glock 19 used in this study when using the breech face algorithm. Figure 4.24 shows the same source and different source distribution plot with the combined breech face similarity and aperture shear scores for this firearm. For the Glock 19, there were no noticeable changes between the breech face similarity scores and the combined breech face and aperture shear similarity scores. The assumption was that since the Glock has very prominent shearing scores would increase, but most of the breech face scores were large enough that the aperture shear was not used in computing the combined score. There were a couple of aperture shears that were higher than expected, all less than 0.28, from different source comparisons. Figure 4.25 shows a comparison of a Glock 19 Blazer (Left) and a SCCY CPX Remington cartridge case that gave a breech face similarity score of 0.0049, an aperture shear score of 0.2622, and a combined score of 0.0049.

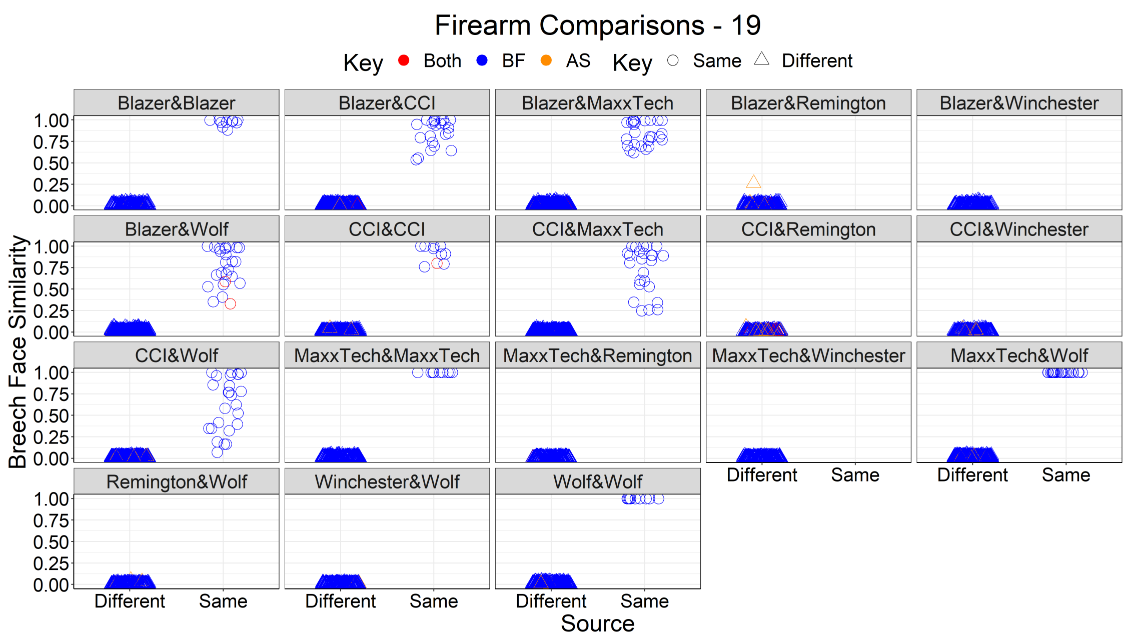


Figure 4.24: All the same source and different source comparisons distribution plot with the combined breech face similarity and aperture shear scores for the Glock 19.

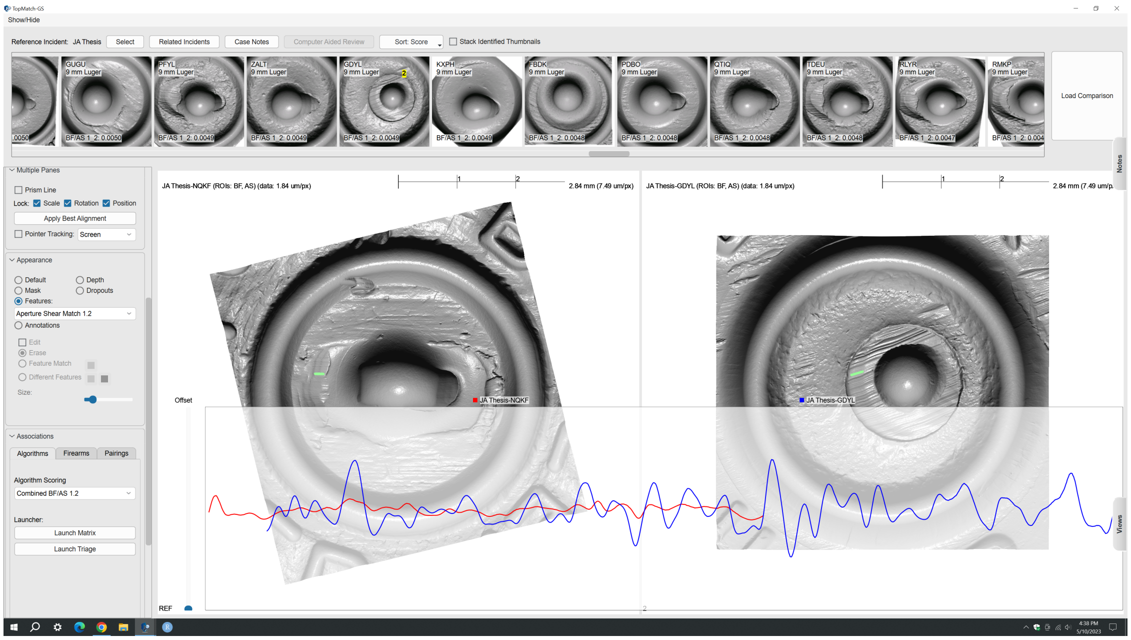


Figure 4.25: A comparison from a Glock 19 Blazer (Left) and a SCCY CPX Remington cartridge case that gave a breech face similarity score of 0.0049, a aperture shear score of 0.2622, and a combined score of 0.0049.

Figure 4.17 shows the same source and different source distribution plot for the Smith and Wesson M&P9 used in this study when using the breech face algorithm. Figure 4.23 shows the same source and different source distribution plot with the combined breech face similarity and aperture shear scores for this firearm. The Smith and Wesson M&P9 had several comparisons that scored higher due to their aperture shear scores. Figure 4.26 compares a Remington (Left) and a Wolf (Right) cartridge case. The breech face similarity score was 0.436, the aperture shear score was 0.961 and the combined score was 0.436. The breech face similarity scored lower for the same firearm comparison. However, the algorithm was able to identify high similarity in the aperture shear. When the algorithm combined these scores, the score was still 0.436 even with the high aperture shear similarity found.

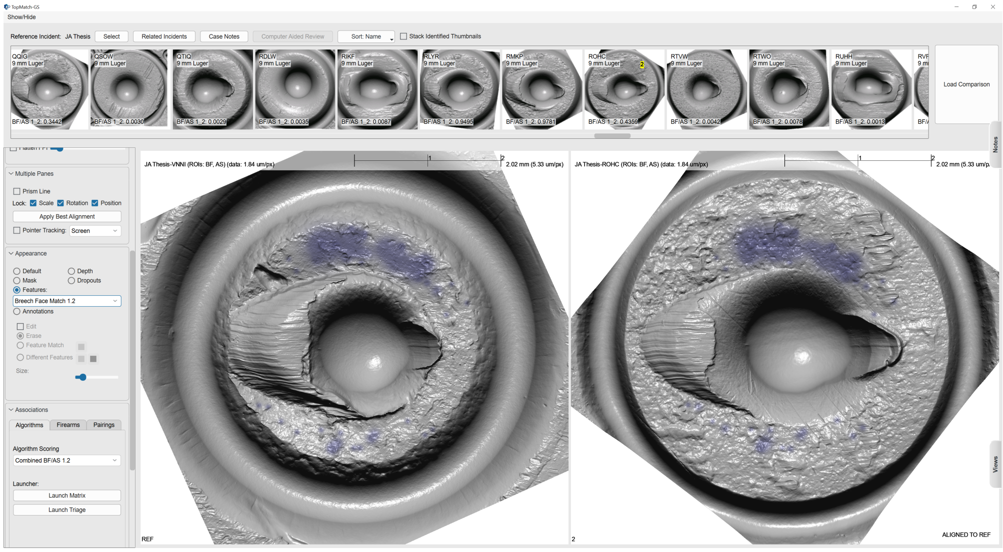


Figure 4.26: Comparison of a Smith and Wesson M&P9 Remington and Wolf cartridge case that generated a breech face similarity score of 0.436, a aperture shear score of 0.899, and a combined score of 0.436.

In the current state of both algorithms, the breech face similarity score algorithm is the better of the two scoring methods. The combined breech face and aperture shear similarity scores do not appear to be combining scores. Instead, one of the scores is being used as the displayed score. The combined breech face and aperture shear similarity scores have generated high scores for different source comparisons. It is recommended that the breech face similarity score be used because it appears to be the more reliable metric.

4.3 ROC Curves

Figure 4.27 shows ROC curve data for all firearms and ammunition types. The AUC for the ROC curve is printed in the middle. The AUC is the probability that a randomly selected same source comparison will have a higher similarity score than a randomly selected different source comparison. The overall AUC was found to be 0.956. This means that there is a 95.6 percent probability that a randomly selected same source comparison will have a higher similarity score than a randomly selected different source comparison. Figure 4.28 shows the ROC curve for all firearm models where all comparisons were the same ammunition brand. In contrast, Figure 4.29 displays a ROC curve for all firearm models where all comparisons were different ammunition brands. The different ammunition AUC of 0.952 is slightly lower than the same ammunition curve where the AUC was 0.988. This difference was expected since better results were seen with the same ammunition brand comparisons.

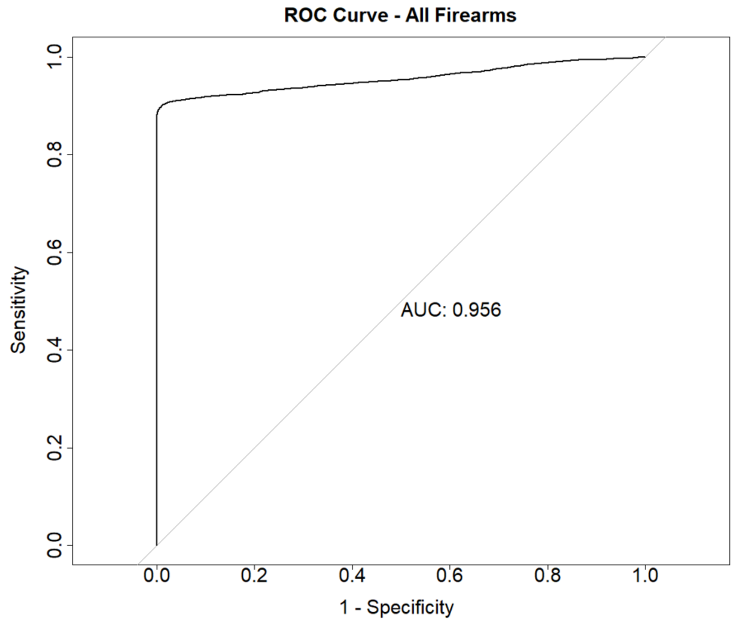


Figure 4.27: ROC curve for all data, combining all firearms and ammunition.

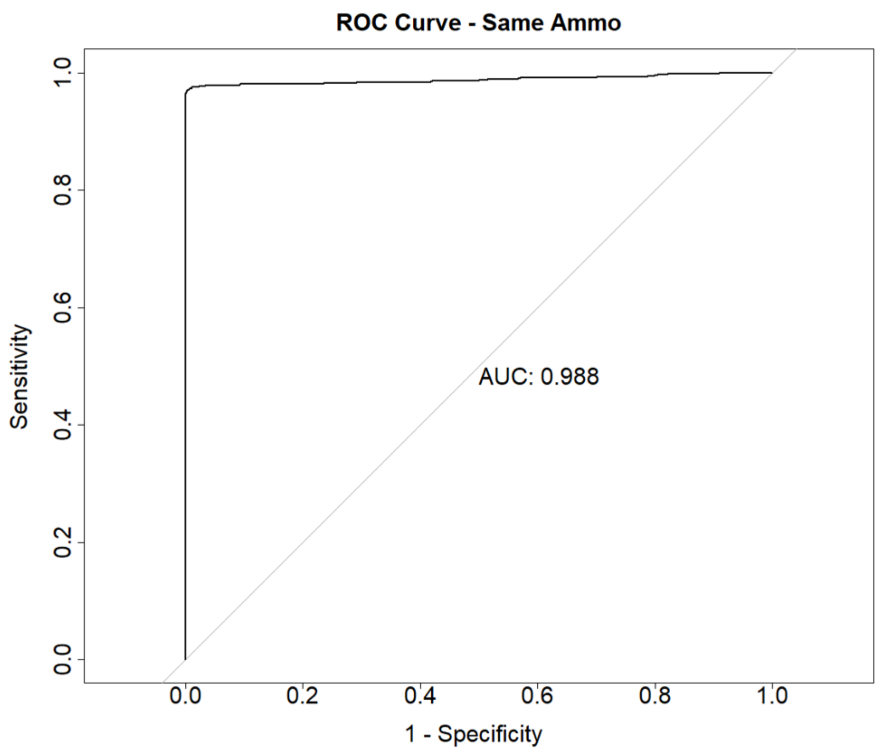


Figure 4.28: ROC curve for all comparisons between same ammunition type

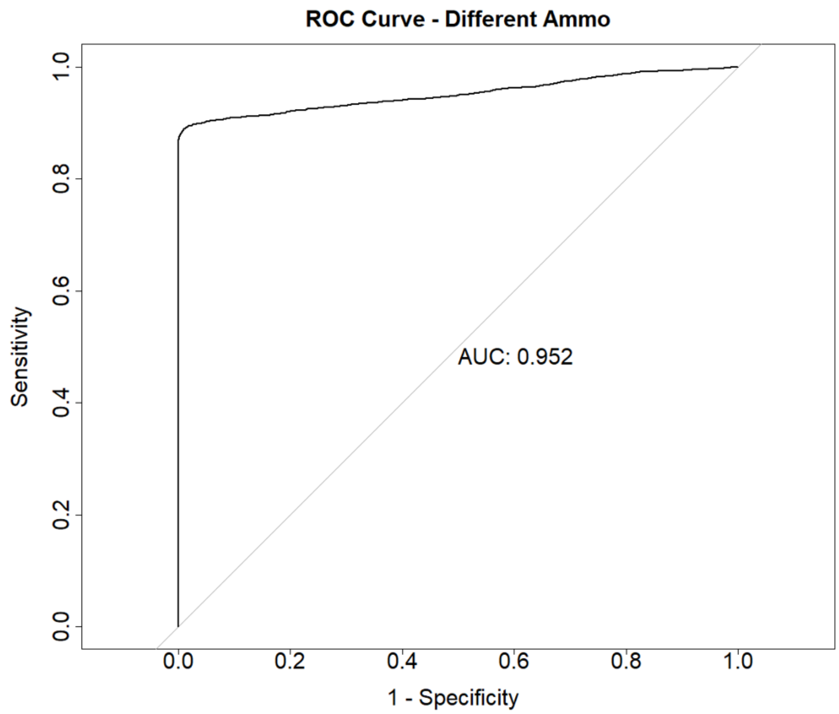


Figure 4.29: ROC curve for all comparisons where ammunition type was different.

A procedure for comparing AUCs based on the Mann-Whitney test was used to determine if there was a significant difference in the performance of the system between the same ammunition and different ammunition comparisons [17], [18], [19]. The result of this test gave a p-value of 9.597e-15. Because the value was less than the 0.05 threshold, there was evidence that same ammunition brands performed significantly better than different ammunition brands. However, a 0.952 AUC does not indicate poor performance. The system performs well with different ammunition, just not as good as with the same ammunition.

Figure 4.30 shows ROC curves regardless of ammunition type separated by firearm model. Every firearm had an AUC above 0.96, with the exception of the PT92, which had an AUC of 0.802. The M&P9, Glock 19, and the C9 had a 1.000 AUC, meaning that all higher similarity scores were the same source comparisons.

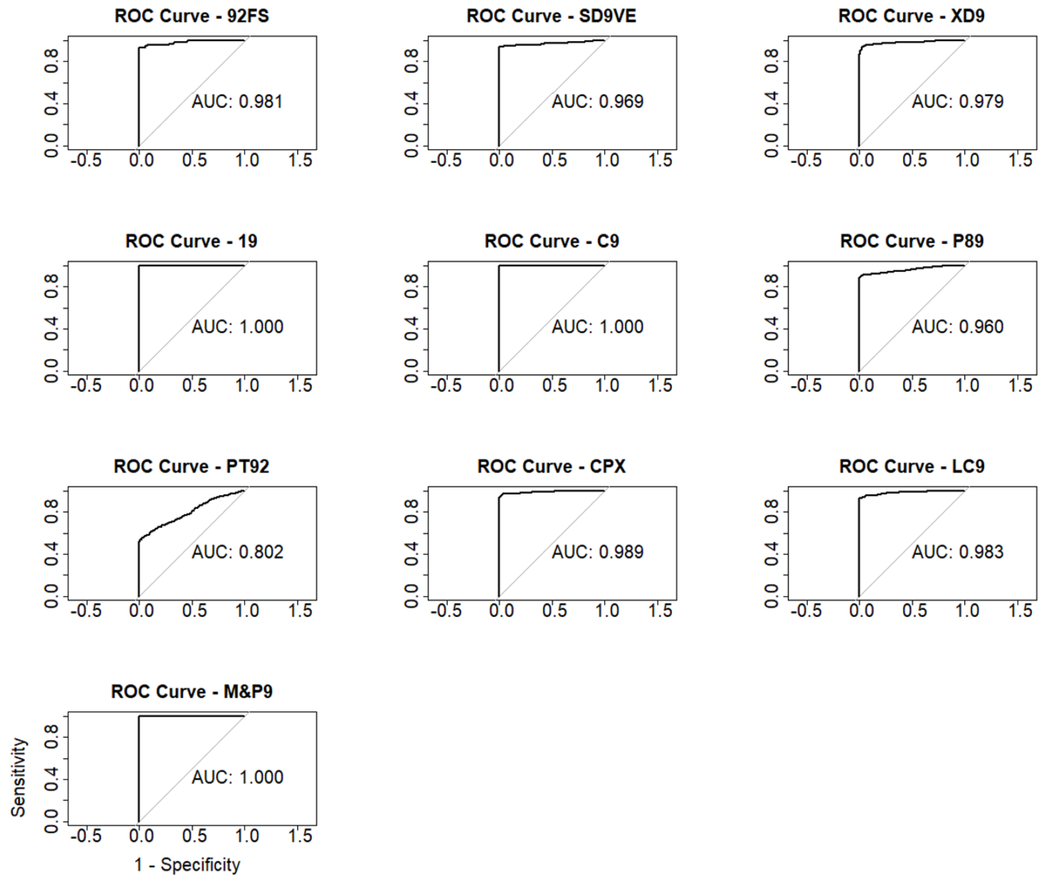


Figure 4.30: ROC curves regardless of ammunition type separated by firearm model.

Figure 4.31 shows ROC curves for different ammunition comparisons broken down by model, while Figure 4.32 displays ROC curves for the same ammunition comparisons broken down by model. AUCs for different ammunition scores performed slightly worse when compared to the same ammunition scores. This shows that using the same ammunition leads to a higher AUC and better system performance. The firearm model that was impacted the most when comparing different and same ammunition scores was the PT92. The PT92 had an AUC of 0.786 for different ammunition scores, while the AUC for the same ammunition scores was 0.936. This displays the AUC being impacted due to the different types of ammunition brands used in this particular firearm. The AUC for different ammunition scores being lower for the PT92 when compared to the other firearms is likely due to shot-to-shot variability.

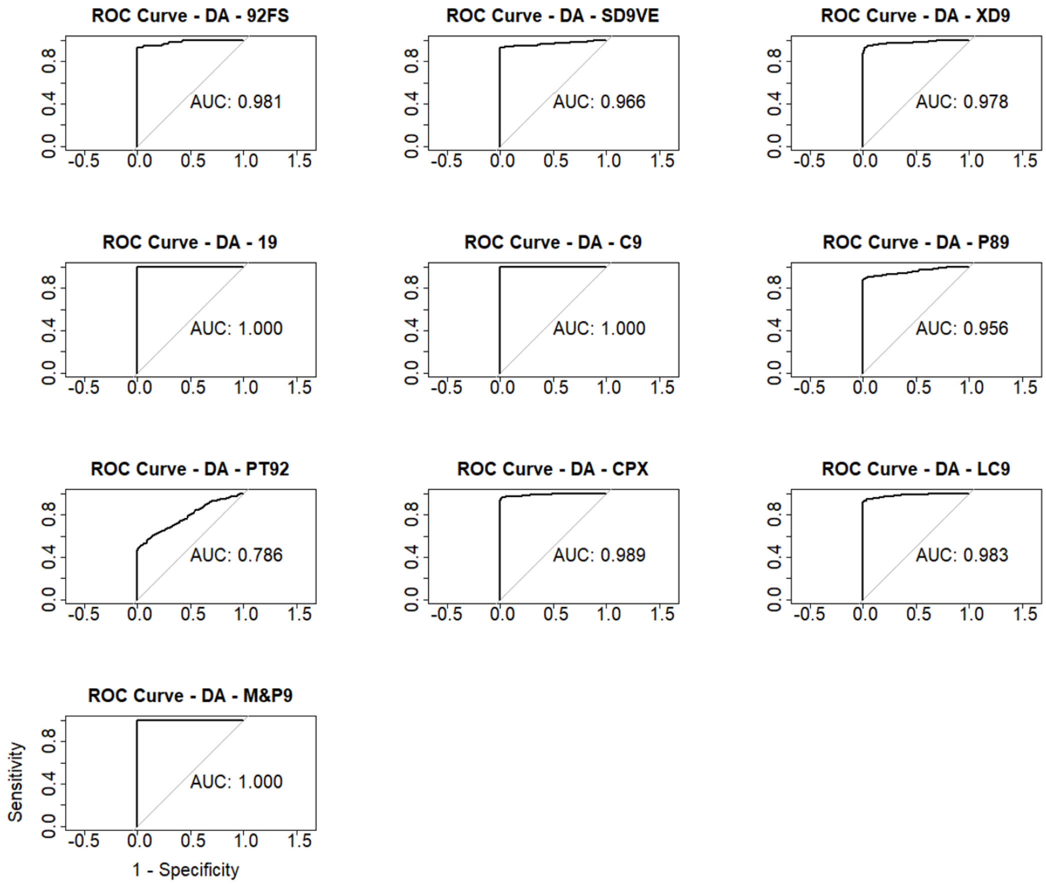


Figure 4.31: ROC curves for different ammunition comparisons broken down by model.

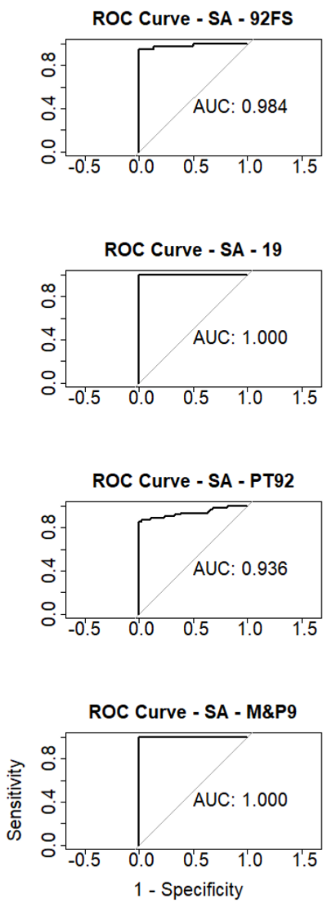


Figure 4.32: ROC curves for same ammunition comparisons broken down by model.

4.4 Shot-to-Shot Variation

When there are numerous same source comparisons, it is not expected every single comparison in the same and different brands would be perfect. This means that not all pairs will have high similarity to each other. If there is one comparison between two specific cartridge cases that are not scoring well due to the marks present or absent, you may still be able to link the cartridge cases to a firearm through other cartridge case comparisons. There were several examples in the distribution plots discussed earlier of a handful of low comparison scores when most of the other scores were much higher. To illustrate this the triage feature in the Top Match virtual comparison microscopy workspace was used. Triage is defined as the preliminary sorting of cases by firearm.

Figure 4.33 shows the triage matrix for all P89 cartridge case comparisons between Blazer and MaxxTech with a threshold of 0.5. The threshold set to 0.5 means that every cartridge case comparison scored 0.5 and above will have a connecting line to the cartridge case that generated that score. There are a couple of connecting lines that could appear; a green line indicates a high similarity, while the black line simply indicates the comparison matches the threshold selected. The threshold can be changed between 0.0 to 1.0. This could be used to illustrate the lowest score that still allows all cartridge cases to connect to each other. Comparisons between the P89 Blazer and MaxxTech were selected due to the high variability observed between scores in the distribution plot seen in Figure 4.5. Cartridge case FXLS does not have a score above 0.5 when compared to EVUE but can connect to EVUE through cartridge cases MZOP and FUTN. This is illustrated by the blue line in Figure 4.33. As the threshold increases, the number of connected pairs decreases. In Figure 4.34 it shows all connected pairs above a 0.9 threshold. This Figure shows two groups of cartridge cases that scored above a 0.9 to each other with a connecting cartridge case, GUGU, which is the only cartridge case connecting the two groups. This linking cartridge case had toolmarks found in both groups. Without this linking cartridge case, it may become difficult to identify the firearm if we only had one cartridge case from either group. Figure 4.35 shows images of the scanned cartridge cases where the connecting cartridge case GUGU is in the center. On the left and right of GUGU are representatives of the grouped pairs in Figure 4.34.

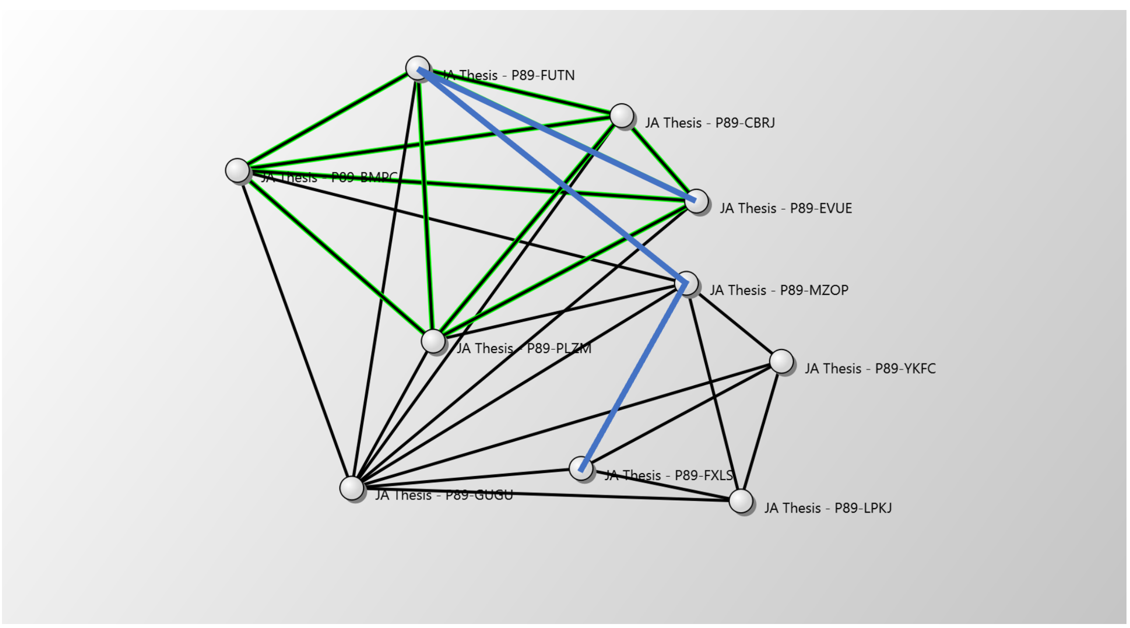


Figure 4.33: The triage matrix for all P89 Blazer to MaxxTech comparisons above a 0.5 breech face similarity score.

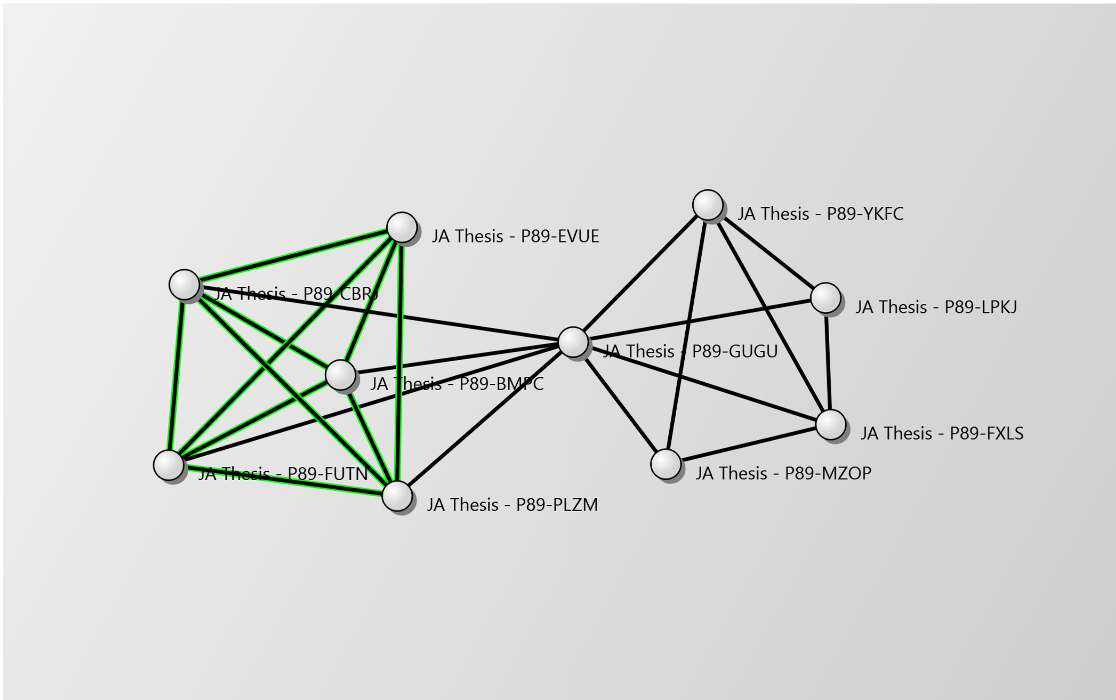


Figure 4.34: The triage matrix for all P89 Blazer to MaxxTech comparisons above a 0.9 breech face similiarity score.

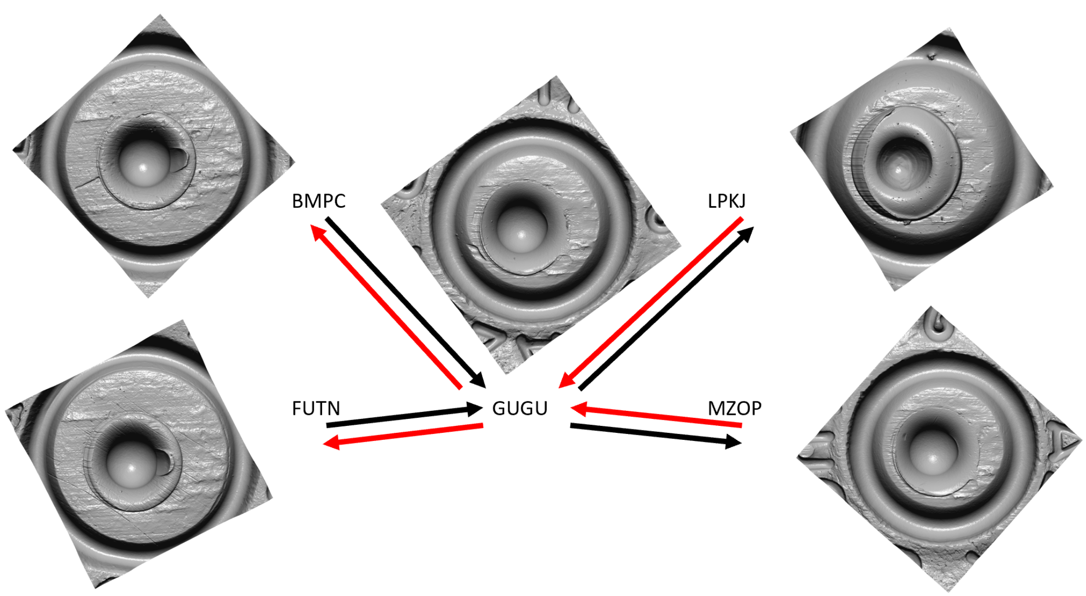


Figure 4.35: A diagram illustrating the grouped pairs seen in the P89 triage matrix.

Figure 4.36 shows the triage matrix for all CPX cartridge case comparisons between Winchester and Remington with a threshold of 0.5. The CPX had a high variability between scores, which can be observed in the breech face similarity plot found in Figure 4.11. The triage created for the CPX had fewer high-similarity cartridge case comparisons compared to the P89 triage. There were also fewer connected pairs. This means that the CPX Winchester and Remington comparisons did not have as many comparison scores over the 0.5 threshold when compared to the P89 Blazer to MaxxTech comparison triage. The CPX cartridge case, SGDN, only has three connected but can connect to every cartridge case through other connected cartridge cases. The blue lines in 4.36 illustrate this. This triage feature visualizes the fact that no matter the ammunition brand used in the comparison, even some of the lowest comparisons can still be linked to a firearm through other cartridge cases fired by the same source.

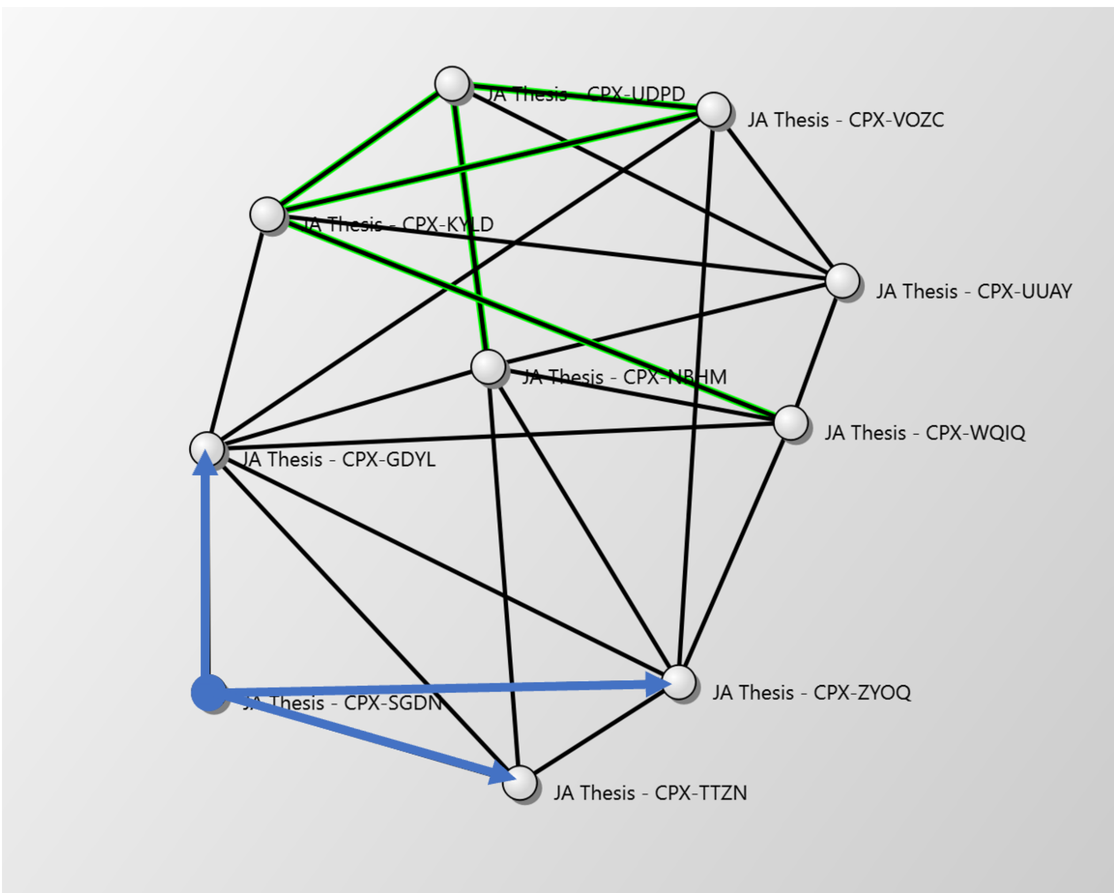


Figure 4.36: The triage matrix for all CPX Winchester to Remington comparisons above a 0.5 breech face similarity score.

5. Conclusions

The algorithms overall performance with the tested firearms and ammunition was excellent, as illustrated by the ROC AUC of 0.956. The automated breech face algorithm was able to differentiate between the same and different sources, with the same source and same ammunition types having the highest scores, as was expected. Brass-primed MaxxTech and Wolf scored exceptionally well regardless of the firearm. Winchester and Remington did not compare very well against other ammunition brands and, often, themselves. The aperture shear algorithm does well when the shear is 800 microns long or greater. However, with smaller shears, it is recommended that an examiner use caution when masking, as the algorithm may report a higher than expected score for different source comparisons. The combined score defaults to the breech face score when the system believes the aperture shear values may be misleading. This means that running the breech face algorithm is the most reliable in most situations. The combined score feature came out relatively recently, and Cadre Forensics plans to update the algorithm to fix some of the inconsistencies observed in this study.

5.1 Future Research

Further research into the Cadre Forensics TopMatch-GS 3D system should use the scans created for this research project in future updated algorithm versions. The large and already created dataset with numerous ammunition brands and firearms would allow for a great way to illustrate the overall improvements Cadre Forensics makes to its algorithm scoring. The dataset used in this study could be expanded to include additional firearms and ammunition types and their effects on the system. Pistols chambered in 9mm Luger are a great starting point for overall performance, but a variety of other calibers, specifically rifle calibers, should be evaluated. Historically, rifles do not mark as well as pistols and can prove troublesome in casework. AR platformed rifles could be evaluated by seeing differences, if any, in scoring between calibers 5.56x45mm and .223 Remington and various ammunition brands. Certain ammunition brands have different primer sizes and can lead to lower results when the overall area of a primer of one cartridge case is larger than that of the cartridge case it is comparing to. A future study could examine the effects of different primer sizes on algorithm scores. Another area that needs research is to look into the triage feature, evaluate connected pairs, and provide recommendations on how many test fires are needed to represent the variability within a firearm.

6. Appendix A

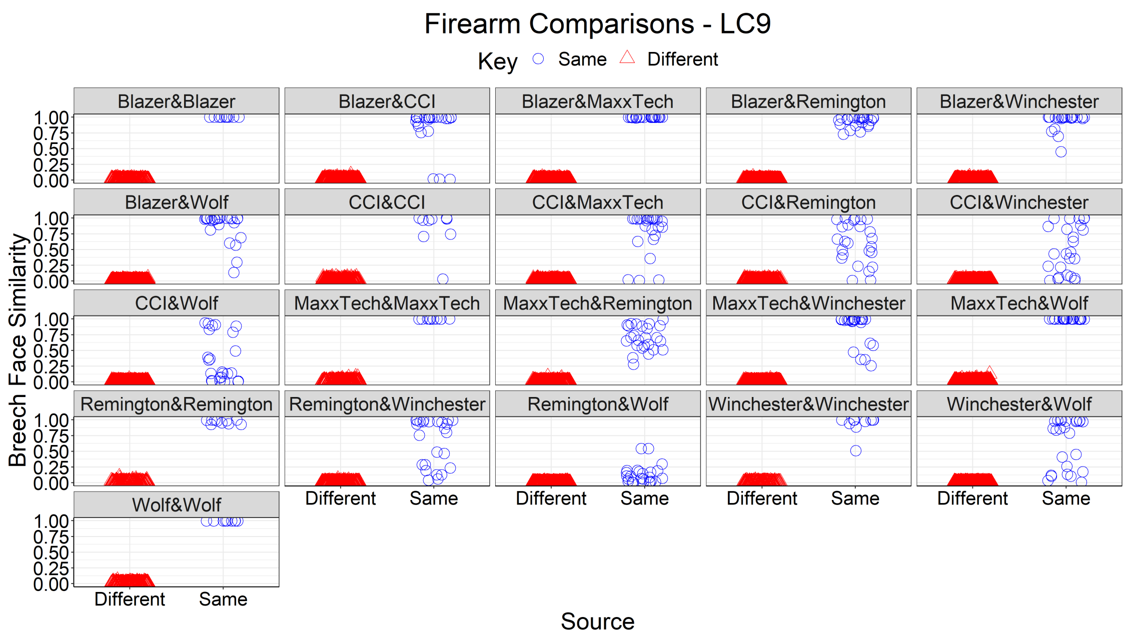


Figure 6.1: All same source and different source comparisons for the Ruger LC9.

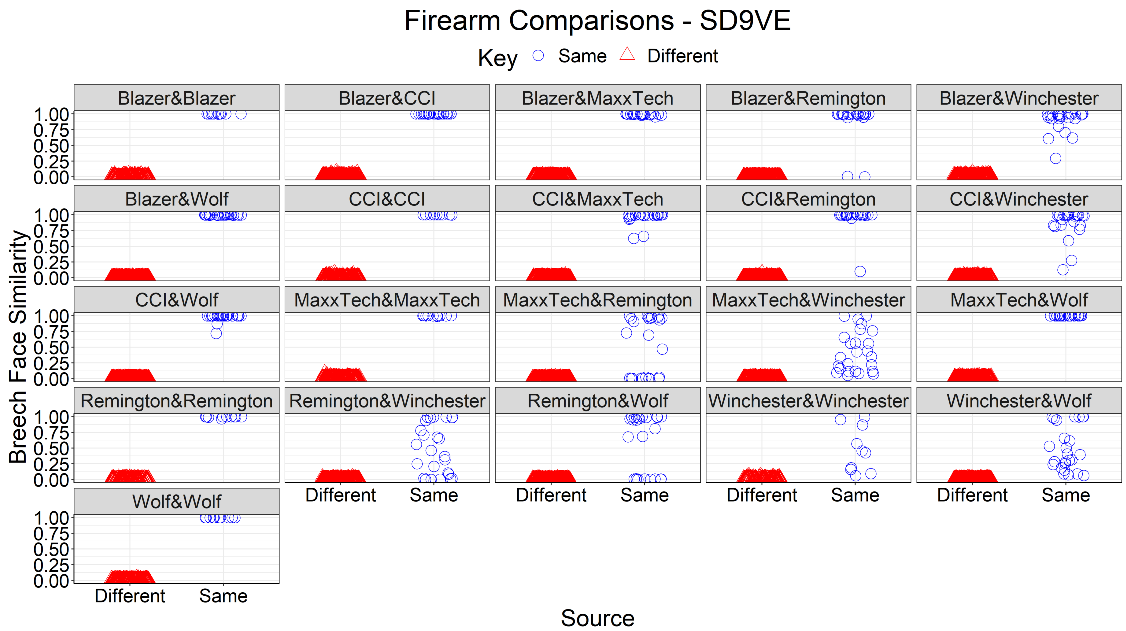


Figure 6.2: All same source and different source comparisons for the Smith and Wesson SD9VE.

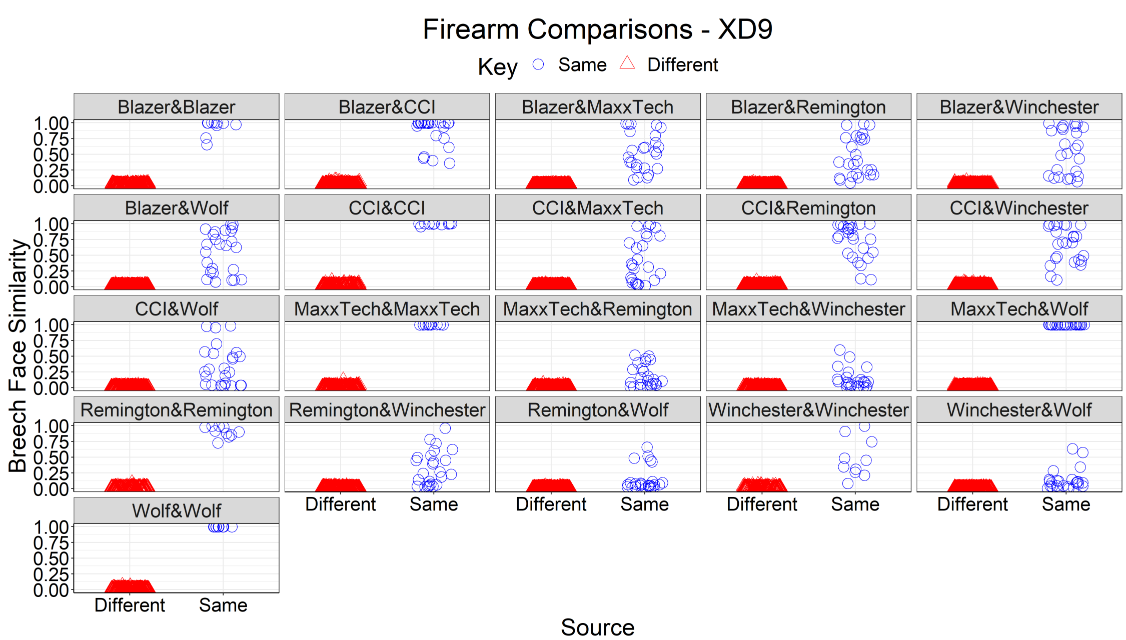


Figure 6.3: All same source and different source comparisons for the Springfield Armory XD9.

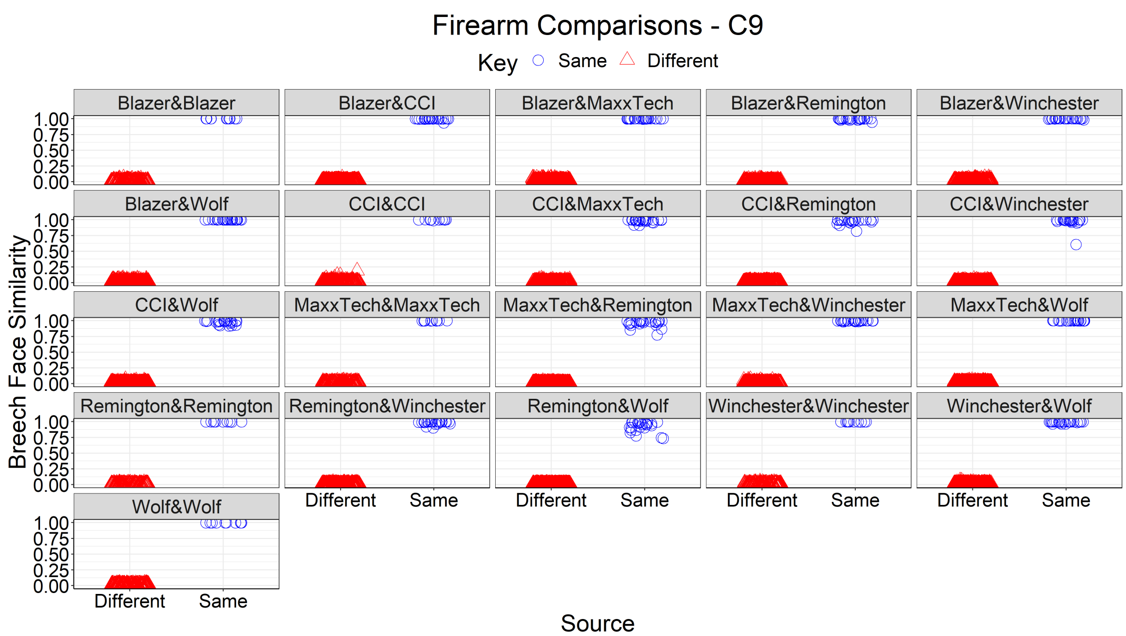


Figure 6.4: All same source and different source comparisons for the Hi-Point C9.

7. References

[1] Guns & Ammo. Guns & Ammo 101 The Bullet Cartridge. Retrieved from http://www.gunsandammo.info/ammo/ammo-101. *Guns & Ammo*, 2015.

[2] R Lilien. Presentation over quality control on the topmatch-gs 3d sytem. *Topmatch*, 2022.

[3] E.F. Law and K.B Morris. Evaluating firearm examiner conclusion variability using cartridge case reproductions. *Journal of Forensic Sciences*, 2021:1–17, February 2021.

[4] T Fawcett. Roc graphs: Notes and practical considerations for data mining researchers. *Intelligent Enterprise Technologies Laboratory*, page 4, January 2003.

[5] AFTE. AFTE Theory of Identification as it Relates to Toolmarks. Retrieved from https://www.afte.org. *AFTE Journal*, 2022.

[6] S Bunch. Is a match really a match? *FBI*, March 2011.

[7] C Monturo. Forensic firearm examination. *Science Direct*, 2010.

[8] T Weller, M Brubaker, P Duez, and R Lilien. Introduction and initial evaluation of a novel three-dimensional imaging and analysis for firearm forensics. *AFTE Journal*, 47:198–208, 2015.

[9] Robert Thompson. Automated firearms evidence comparison using the integrated ballistic identification system (IBIS). *Investigation and Forensic Science Technologies*, 42, 1999.

[10] J DeKinder, F Tulleners, and H Thiebaut. Reference ballistic imaging database performance. *Forensic Science International*, 140:207–215, 2004.

[11] R Nichols. Defending the scientific foundations of the firearms and tool mark identification discipline: Responding to recent challenges. *J Forensic Science*, 52, May 2007.

[12] Pierre Duez. Development and validation of a virtual examination tool for firearm forensics. *Journal of Forensic Sciences*, 63:1069–1084, 2017.

[13] J Davis. Primer cup properties and how they affect identification. *AFTE Journal*, 42:3–29, 2010.

[14] L Knowles, D Hockey, and J Marshall. The validation of 3D virtual comparison microscopy (VCM) in the comparison of expended cartridge cases. *Journal of Forensic Sciences*, 2021.

[15] C Chapnick, T Weller, P Duez, E Meschke, J Marshall, and R Lilien. Results of the 3D Virtual Comparison Microscopy Error Rate (VCMER) Study for firearm forensics. *Journal of Forensic Sciences*, 2020.

[16] Cadre Forensics. Cadre Forensics Announcements. Retrieved from https://www.cadreforensics.com/pdf/FBIFTU_Announce.pdf. *Cadre Forensics Announcements*, 2017.

[17] E.R. DeLong, D.M. DeLong, and D.L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characterisitc curves: a nonparametric. *Biometrics*, 44:837–845, 1988.

[18] X. Sun and W. Xu. Fast implementation of delong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEE Signal Process*, Lett. 21:1389–1393, 2014.

[19] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, and M. Muller. proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinform*, 12:77, 2011.