UNIVERSITY OF CENTRAL OKLAHOMA

GRADUATE COLLEGE

DEVELOPING A NEW DATASET AND METHOD FOR THE INTERPRETATION OF FIREARMS EVIDENCE

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF FORENSIC SCIENCE

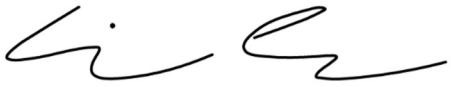
By

NATHAN KLAUS
Edmond, Oklahoma
2026

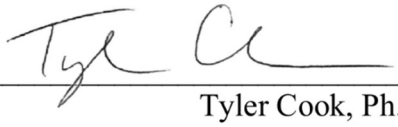
DEVELOPING A NEW DATASET AND METHOD FOR THE INTERPETATION OF FIREARMS EVIDENCE

A THESIS APPROVED FOR THE
W. ROGER WEBB FORENSIC SCIENCE INSTITUTE

BY



Eric Law, Ph.D., Chair



Tyler Cook, Ph.D.

Keith Morris Ph.D. (external)

Table of Contents

List of Figures

List of Tables

Abstract

Firearms evidence has long been utilized in investigations and in the courtroom. Emerging technologies have made the examination of firearms evidence, namely expended cartridge cases, an easier process. Recent technologies enable the use of computer algorithms to provide quantitative comparisons. However, many of the cartridge case datasets used to develop these algorithms are limited in number and scope, in particular insufficient accounting for variability expected to occur within markings produced by a single firearm. The goal of this research was to develop a new dataset consisting of 31 HiPoint C9 firearms with 17 cartridge cases fired from each. These cartridges were scanned using the Cadre Forensic TopMatch 3D instrument and assigned similarity scores by the TopMatch 3D algorithm. These similarity scores were then analyzed using both kernel density estimation and generalized linear mixed models. Kernel density estimation had difficulties due to the low distribution of scores among firearms tested. GLMM, however, offered a means by which a percentage probability of same- or different-source conclusion could be assigned to similarity scores of a model of firearm. This method has the potential to be expanded to other firearms and offer some quantification to the field of forensic firearms analysis.

Introduction

The use of forensic toolmark analysis for the investigation of crimes involving firearms is a longstanding practice. Firearm analysis relies on the same theory as toolmark analysis: when two hard objects collide, the harder object makes a characteristic impression on the softer object; in the case of firearms, this can refer to the barrel rifling interfacing with a fired bullet, or the firing pin and breech face contacting the cartridge case primer [1]. These markings can be

compared to determine if a particular firearm fired a particular cartridge case that is associated with a crime scene [1]. This practice dates back more than a century, with one of the first recorded instances being an 1835 case in London [1]. As technology has evolved, new tools became available for this practice; in the late 1800s, an early comparison microscope was developed, and by the 1930s it would become a commonplace tool for firearm analysis [1]. A view through a comparison microscope of a same-source comparison can be seen in Figure 1.

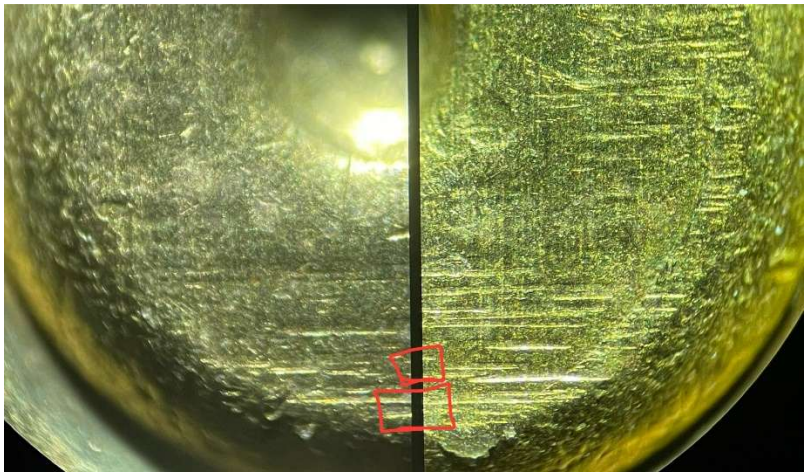


Figure 1: View from a Leica FS-M model of comparison microscope of a fired primer surface of a same-source 9mm Luger cartridge case. Areas of notable similarity have been marked in red. Image taken by Nathan Klaus.

Firearms analysis using comparison microscopes had been a qualitative practice: well-trained experts would make professional determinations based on what they observed. Beginning in the 1990s, firearms analysis entered the computer age: computerized microscope systems would capture 2D images of ammunition components (e.g. cases and bullets) and run comparisons between those images, producing a digital determination of whether the supplied materials came from the same firearm or different firearms [1]. These early comparison systems were far from perfect. Slight differences in lighting can create enough discrepancy to cause false negatives, and the proprietary nature of many of these older computer programs makes it difficult to understand the different quantitative factors the computer uses to make its determinations [1]. This created a

need for updated systems for comparing pieces of firearms evidence. Many proposed solutions, some proprietary and some not, were developed as a result. The development of these cartridge case comparison methods has been fueled by the creation of multiple experimental data sets, consisting of fired cases from a selection of firearms. However, all of these methods experience constant revision, and the creation of further data sets, distinct from those which have already been developed, is necessary to better understand factors affecting the outputs of various methods and enable better utilization of the methods available. One of these methods is known as the Congruent Matching Cells (CMC) method.

The Congruent Matching Cell Method

The Congruent Matching Cell, or CMC, method was first proposed in 2012 on recommendations from the National Institute of Standards and Technology (NIST) [2]. The basics of this method involve separation of the cartridge case surface into individual sectors, or cells, that the computer compares with the corresponding areas on another cartridge case surface [2]. An illustrated example of this separation of the surface by cells is shown in Figure 2. In any given cartridge case comparison, there exist two areas: valid and invalid [2].

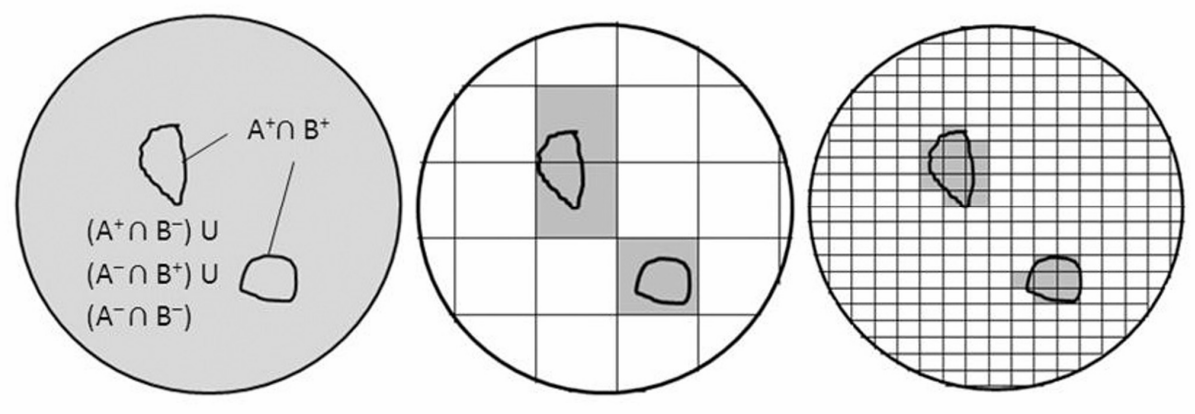


Figure 2: An example of dividing a surface into cells for analysis. Darkened regions represent cells containing the individual characteristics. Note how the overall size of the darkened areas decreases as more cells are used. Source: [2]

A valid region is one where individual characteristics are present; that is, a region with toolmarks specific to a particular firearm [2]. It would follow that invalid regions, for one reason or another, lack these individual characteristics, or at the very least do not possess useful characteristics [2]. Causes of invalid regions can include but are not limited to insufficient contact with the breech face [2]. The CMC method enables comparison systems to separate valid from invalid regions through use of the aforementioned cell separation technique, and within these cells the computer can assign a binary value of either congruent or non-congruent, representative of whether or not the compared cells have a significant degree of correlation with one another [2].

To calculate the similarity between two 3D topographical images of fired cartridge case primers, the CMC method utilizes three different parameters. One is the maximum cross correlation function, or CCF_{max}, which is a numerical descriptor of the maximum correlation between two surfaces, with the result expressed as a percentage [1]. For instance, two topographies in complete agreement with one another will exhibit a CCF value of 100% [1]. A cell is considered to be congruent if this CCF value exceeds 25% [3]. This is relatively low, but acceptable when considering the low likelihood of two distinct firearms, or even the same firearms, producing markings with even that level of similarity. The second parameter is the

registration angle, represented by θ [3]. This registration angle represents how much the cell on the second topography needs to be rotated to have the maximum correlation to the corresponding cell on the first topographical map [3]. If this angle differs by more than three degrees from the original cell, then the cell should not be counted as a congruent cell [3]. Similar to this is the final parameter, registration position. This parameter is represented by x and y values, typically in micrometers, corresponding to positions on a topographical image of a fired cartridge case primer [3]. Typically, if this position in a compared cell differs by more than 150 micrometers from the original cell's position on the topographical map from which the comparison is made, this cell is to be considered non-congruent, though this is dependent upon the sizes of the cells and the resolution of the topography images [3]. These particular values for the parameters are developed based upon experiments using sets of cartridge cases from the same firearms and different firearms [3]. However, given the sheer number of firearms and the different marks they may be prone to leaving on expended cartridge cases, these parameters should perhaps be regarded more as guidelines than absolutes and can be subject to adjustment. In order for two cartridge cases to be considered congruent, the number of contiguous congruent cells must be greater than or equal to six, according to recommendations from the NIST, adapted from the previous Consecutively Matching Striae (CMS) method [2].

Further Developments in the CMC Method

The CMC comparison method was not perfect at its introduction, and many different algorithms to improve it have been proposed and adopted since its creation. One of the first identified problems is the fact that evidence comparisons did not produce the same results if ran in reverse [4]. That is, comparing fired cartridge case A to fired cartridge case B using the CMC

method would not give the same output if cartridge case B were compared to cartridge case A. Usually, the difference between these outputs is not of any significance, but if a comparison run in one direction has more than six matching cells and the same comparison run in the opposite direction has fewer than six matching cells, this could be problematic. Therefore, some iterations of the CMC method will run both forward and backward comparisons in conjunction with using multiple registration angles to produce an output to reconcile both comparisons [4].

Another follow-up development to the CMC system has been the creation of the convergence algorithm. The convergence algorithm relies upon changes in cell position as the cell is rotated. As in the usual CMC method, the computer runs calculations to determine cell areas on the first marked surface that exhibit the greatest similarity to marked areas on the second surface [5]. However, convergence introduces the idea that these calculations of greatest similarity positions will change in response to the topographical map being rotated. When the image is rotated, the cell positions on cartridge case primers fired from the same gun will experience only a marginal change in position on the x and y axes [5]. However, when the image is rotated for a cartridge case from a different firearm, the x and y positions will experience drastic change [5]. This algorithm can, therefore, be utilized to further differentiate between evidence from the same firearm and evidence from two different firearms.

As in any statistically based algorithm, there are margins for error. Estimates for error in forensic science analyses are not a new process; probabilities for coincidental matches were considered for DNA analysis as early as the 1980s [6]. As stated previously, most data for these comparisons are constructed using examples of known matching and known nonmatching samples, and models for estimating error are no different [6]. In these instances, there are two kinds of identifiable error: a “false positive” would be identifying two pieces of firearms

evidence to the same firearm when they are in fact known to be nonmatching, and a "false negative" would be differentiation of two pieces of evidence that in fact do come from the same firearm [6]. The CMC method is useful for modeling the rates for these different kinds of errors due to its quantifying nature [6]. By running a dataset of known matches and known nonmatches through the CMC system and using its outputs to determine the rate of error, a model can be created to better understand and interpret the outputs of the system [6].

Likelihood Ratios in Firearms Analysis

As much iteration as the CMC method has received, it is by no means the only method for comparing markings left on the primers of expended cartridge cases. Multiple other methods, often usable alongside or in place of the CMC method, have also come into being [1]. However, the interpretation of each algorithm's numerical scales can be challenging, leading to the use of likelihood ratios to express the weight of the evidence. The likelihood ratio is an expression of how likely it is that two compared cartridge cases have been fired from the same gun, expressed as a ratio [8]. Mathematically, this is derived from Bayes theorem, which can be expressed as the following equation:

$$\frac{P(H_P)}{P(H_D)} \times \frac{P(E|H_P)}{P(E|H_D)} = \frac{P(H_P|E)}{P(H_D|E)}$$

In this equation, $\frac{P(H_P)}{P(H_D)}$ represents the prior odds, $\frac{P(H_P|E)}{P(H_D|E)}$ represents the posterior odds, and $\frac{P(E|H_P)}{P(E|H_D)}$ represents the likelihood ratio, the probability of evidence supporting one hypothesis divided by the probability of the same evidence supporting the opposite hypothesis. In practice, a likelihood ratio of 100:1 suggests that it is 100 times more likely two cartridge cases were fired by the same gun, whereas a 3:1 ratio suggests it is only three times more likely that the cartridge cases

originate from the same firearm. The likelihood ratio itself is calculated by dividing the probability of a similarity score relating that both cartridge cases were fired by the same gun by the probability of that same score suggesting that both cartridge cases were fired by two distinct firearms [8]. The likelihood ratio can be utilized as a probability descriptor for any cartridge case comparison technology that produces a quantifiable result, proprietary or otherwise. This includes the CMC algorithm, and, in the case of the CMC method, the likelihood ratio calculation would involve the quantity of matching cells. These likelihood ratios serve to give any system further quantification of its results, assigning a numerical expression of probability to the comparison evidence.

The Use of Datasets in Firearms Analysis Technology

The development of the CMC method, likelihood ratios, and other, more proprietary means of firearm evidence comparison has been enabled by the creation and analysis of experimental datasets, consisting of lab-collected instances of firearm evidence. For example, Chen et. al.'s convergence developments utilized datasets of fired cartridge cases known as the Weller, Fadul, and Lightstone datasets [7, 9, 10]. The Fadul series, for instance, involved cases fired from 10 consecutively manufactured barrels and slides, with 40 rounds fired in total [9]. Weller, similarly, also had 10 consecutively manufactured pistol barrels and slides, with only nine rounds fired from each barrel [7]. The Fadul and Weller datasets, additionally, both utilized Ruger pistol slides [7, 9]. The Lightstone data set also utilizes 10 slides of consecutive manufacture, albeit with only 30 total rounds fired [10]. All of these series used approximately 10 consecutively manufactured slides for their experiments, and thus had consecutively manufactured breech faces producing markings on the cartridge cases. The importance of using

consecutively manufactured firearms is not to be understated; firearms with consecutive serial numbers are most likely to leave similar markings on primers as a consequence of similarity in manufacturing processes and the potential for subclass characteristics [7]. Subclass characteristics are characteristics shared between tools and firearms manufactured using the same tooling, e.g. the same broach being used to manufacture the breech face of multiple firearms, thus giving those breech faces similar markings [7]. However, the usage of consecutively manufactured firearms does pose a problem with regards to practicality. In a typical firearms case, the probability of the involved firearm and any firearm brought into evidence having been sequentially manufactured is small. As such, rigorous analysis of cartridge cases from consecutively manufactured firearms may be unnecessary. In contrast to these aforementioned data sets, another data set – the Baiker-Sørensen dataset – utilized 200 different firearms of multiple makes and models. Only two samples from each firearm were randomly selected for inclusion in this data set [11]. Also to be considered is the matter of which ammunition components to compare: the Baiker-Sørensen dataset, for instance, offers 19900 instances of non-matching samples alone [11]. Yet, that sample is limited to only two cartridge cases per firearm analyzed for a total of 400 rounds of ammunition, an amount that may be considered insufficient when considering the variability that can occur between shots. Meanwhile, the Fadul and Weller series offer only 10 firearms with a larger magnitude of rounds fired. These datasets, while instrumental in developing the necessary modifications to the CMC system and other algorithms, offer a relatively narrow range of methods for dataset creation. Based on the unfortunate amount of similarity between existing datasets and the propensity for datasets to exhibit the unrealistic expectation of consecutively manufactured firearms, additional

datasets are necessary to better understand the factors affecting computer comparison methods and their outputs.

Review Conclusion

In summary, the CMC system, over the course of its development, has undergone a series of changes fueled by the analyses of data sets. Unfortunately, current data collections all have very similar sample collection methods, resulting in a narrow understanding of the factors at play when utilizing cartridge case comparison algorithms and likelihood ratios. The creation of additional data sets will be necessary in order to further this understanding, as well as understanding the variation that can occur between shots and how much variation should be expected when performing comparisons. Additionally, a method is needed to calculate likelihood ratios that can take advantage of the variability present within the marks generated by a firearm to better account for this variability, producing more robust likelihood ratio estimates.

The use of likelihood ratios in testimony would be likely to change how firearms examination conclusions are delivered to the courtroom. The current AFTE range of conclusions includes Identification, Elimination, and Inconclusive outcomes [12]. It would not necessarily make sense to present a numerical likelihood ratio with this range of conclusions, as Identification or Elimination may be seen as absolute, whereas a likelihood ratio represents the strength of competing propositions. As such, it might be necessary to revise the way firearms expert opinions are delivered to accommodate this. A shift to likelihood ratios would also serve to address some of the criticisms of firearms examination. In 2016, the President's Council of Advisors on Science and Technology (PCAST) published a report on various forensic science disciplines that criticized the lack of quantification in forensic firearms examination [13].

Likelihood ratios in firearms examination would provide this requested quantification and bring firearms examination in line with some other forensic science disciplines that also utilize likelihood ratios.

Materials and Methods

This data set was constructed using a selection of fired cartridge cases. These cartridge cases were collected from Hi-Point C9 model pistols. All of the aforementioned cartridge cases are of 9mm Luger caliber, utilizing only 124-grain Federal brand ammunition for consistency. Some of these cases were provided for a related project, while others were provided for the purpose of this research.

Accounting for shot-to-shot variation was a necessary consideration for this project. Not all cartridge cases expended by one single firearm will have completely identical markings, a fact that must be considered when comparing cartridge cases, and this same fact is not showcased to an adequate degree by most data sets [14]. In order to get an acceptable level of shot variability, 17 cartridge cases from each firearm were utilized. This is based on data that suggests 15 cartridge cases is the minimum number required to exhibit an acceptable amount of shot-to-shot variation [14]. The extra two cartridge cases represent the two cartridges being compared to provide a single similarity score to calculate a likelihood ratio using the other 15 cartridge cases as data. These two unique cartridge cases were randomly selected from the 17 cartridge cases, with the similarity score being recorded for these two cartridge cases. To calculate the likelihood ratio, a pair of probability distributions was developed. One of these distributions was developed from similarity scores of the same-source (within-firearm) pairs within the 15 cartridge cases, whereas the other distribution was developed from similarity

scores of different-source pairs of the 15 cartridge cases, with the similarity scores along the x-axis and the probability density on the y-axis. From here, the similarity score of cartridge cases 16 and 17 was used to determine the y-value of both distribution curves. The same-source y-value would be divided by the different-source y-value to determine the likelihood ratio.

All of these cartridge cases – 17 cartridge cases from 31 firearms for a total of 527 cartridge cases – had both their breech face and firing pin impressions scanned using the Cadre Forensics TopMatch 3D scanning system, in order to generate a 3D topography of all of the cartridge cases so that they could be compared. To generate a 3D topography of the cartridge case, the TopMatch utilizes GelSight technology [15]. GelSight, as used in the TopMatch system is a gel disc, painted on one side; when an object is pressed into the clear side of the gel – in this circumstance, a cartridge case – a near-perfect topography of this surface pushes out on the opposing side of the gel, replicating even the most minute details in the cartridge case [15]. The TopMatch instrument uses photometric stereo, a method where pictures of this surface are taken with multiple lighting angles in order to generate a detailed 3D model of the primer surface of each cartridge case, with a lateral resolution of 1.8 micrometers per pixel and sub-micron depth resolution [16]. The use of the 3D model and multiple lighting angles minimizes any effect lighting or shadows might have on the comparison [15].

The TopMatch system not only generates these 3D topographies, but is also capable of calculating similarity between cartridge cases using its own algorithms. This method is proprietary, thus the exact algorithms it uses are not well understood, but it uses a method not too different from the CMC method for determining similarity. Specifically, the TopMatch method for comparison assesses specific features on the cartridge cases, attempting to identify features common between the both of them [15]. Regions of greater similarity will be highlighted, with

darker highlighting indicating areas of greater similarity, as illustrated in Figure 3, an example which possessed a similarity score of 0.9999. It is important to note that even among cartridge cases with such high similarity scores, it would not be expected for the entire surface to be highlighted, as the markings may be centered differently on the primer depending on where the firing pin struck the primer. Furthermore, only features that are individual characteristics contribute to the similarity of toolmarks, thus not the entire primer surface is of importance to the algorithm [2].

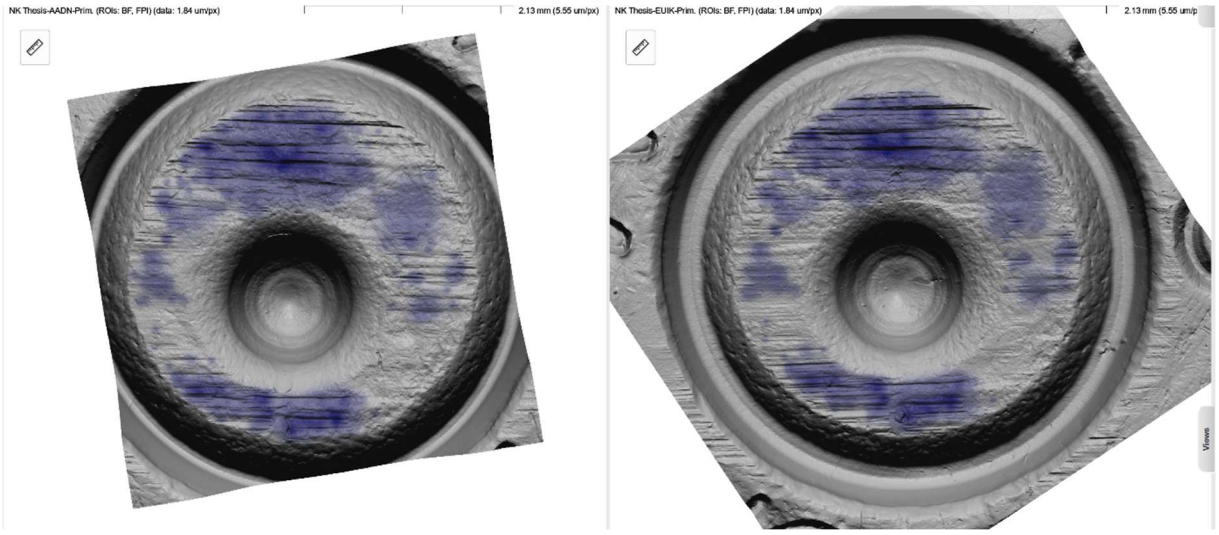


Figure 3: An example of a 3D scan of HiPoint C9 cartridge cases performed by the TopMatch system. Areas shaded in blue are areas of similarity according to TopMatch's own comparison algorithm. This example has a similarity score of 0.9999. Image captured by Nathan Klaus.

The TopMatch comparison algorithm will output a similarity score between 0 and 1, with lower values indicating lesser similarity and higher values indicating greater similarity [15]. This can be thought of as similar in concept to the number of congruent cells in a CMC method comparison, though it is important to recognize that the ratio of congruent cells to total cells in the CMC method should not be expected to be equivalent to this similarity score, as the systems use different processes to determine similarity. Furthermore, this similarity score is not a percentage indicating how close the two scans are to being identical and should not be treated as

such [15]. These comparison scores were generated for the breech face and firing pin impression areas, separately. However, as the firing pin comparison algorithm is still in its early stages, data collection and analysis focused on the breech face scores.

Kernel Density Estimation

To calculate likelihood ratios for each comparison, same source and different source distributions needed to be developed based on the similarity scores for each of the selected firearms. This involved the use of the programming language R, allowing for the generation of probability distributions and mathematical calculations with R Studio as the integrated development environment (IDE) [17]. Generating the likelihood ratios in RStudio required first generating graphs of TopMatch scores using both known matching and known nonmatching values, with one curve being generated for either known outcome. Two R scripts were used for this purpose. One R script organized the raw data from the TopMatch using a master list of all cartridge case identifiers as a reference, allowing them to be sorted by not only serial number but also firearm make, model, caliber, and ammunition manufacturer. With another script, probability density curves were generated from the data by utilizing kernel density estimation. Kernel density estimations generate applicable probability density functions from the data [18]. In general terms, a probability density is an expression of how probabilities are distributed over different x-values [18]. The greater the y-value at a given x-value, the more often that x-value will appear or is likely to appear in a data set. These functions were used to generate the two curves from which the likelihood ratios at a given similarity score could be determined, as illustrated in Figure 4, where the red curve represents different source probability densities and the blue curve represents the same source probability densities. The kernel of choice was the Gaussian kernel, which provides among the smoothest probability density functions and is

widely used. [18]. One of the variables in a kernel density estimation is bandwidth, which determines how much smoothing will be applied to the probability density function [18]. Bandwidth selection is of great importance, as having too high a bandwidth will result in the curve being oversimplified and not necessarily accurate to the data [18]. In this method's context, an oversmoothed function will not provide any useful information on whether a given similarity score will indicate a congruent or noncongruent set of markings on the cartridge case. However, a bandwidth that is too small could cause excessive small peaks within the curves [18]. Therefore, the optimal bandwidth was decided using the Sheather-Jones method [19]. The Sheather-Jones method is data-based and offers superior performance in practical application compared to other bandwidth selection methods, making it an effective choice for bandwidth determination for this data [19].

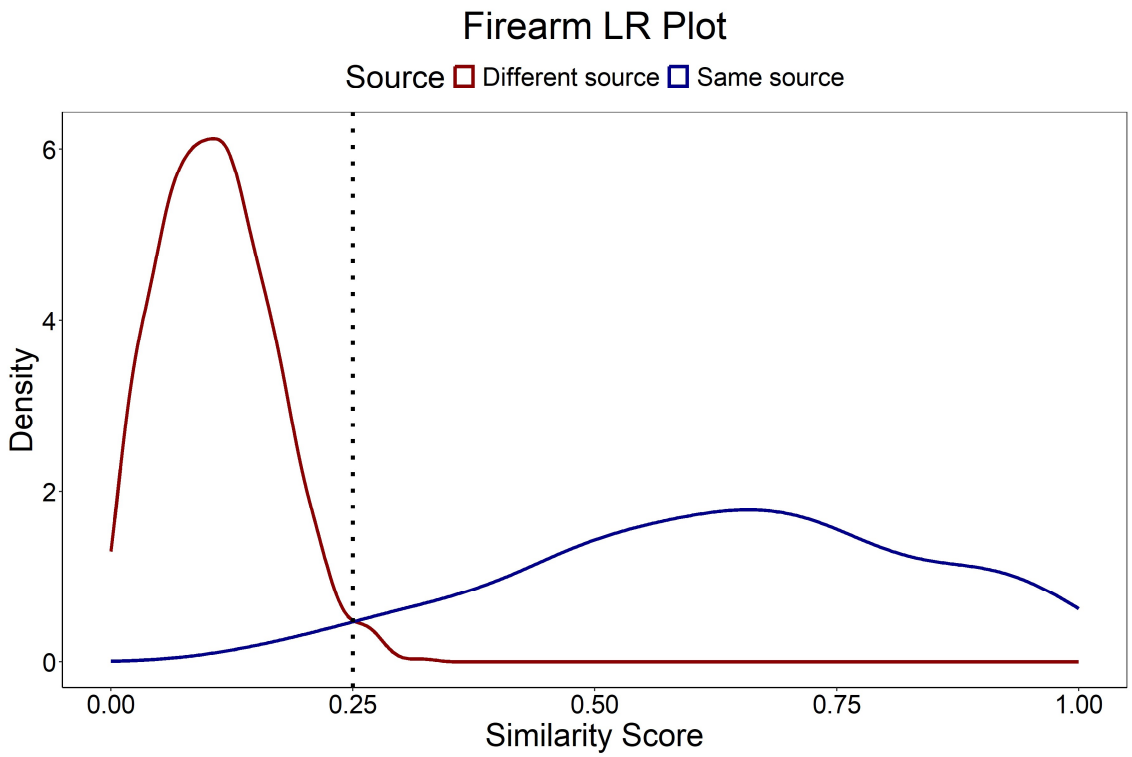


Figure 4: An example of a probability distribution, produced using random normal curves. The red and blue curves represent the probabilities of two cartridge cases having a different source and same source for a given similarity score, respectively. The green line is an example score of 0.25, which gives a likelihood ratio of 0.7607534 in this example curve. Image provided by Dr. Eric Law.

One of the aforementioned curves represents the probability of any given score being from two same source cartridge cases, while the second curve represents the probability of the compared cases being from different sources. These distributions were generated through the comparisons of the sets of 17 cartridge cases. From these distributions, a ratio for the likelihood that two cartridges were fired from the same firearm could be calculated, seeing as a likelihood ratio is the probability that one outcome fulfills one hypothesis divided by the probability of the same outcome fulfilling the opposing hypothesis. This method is capable of being expanded beyond the firearms present in this test set, with the end goal being that any two cartridge cases can be compared and given a likelihood ratio using this method. The primary drawback of this method is that a new same-source and different-source distribution would need to be generated for every criminal case involving a firearm. Repetition of this process would be time-consuming, a factor

that would be of great consideration to laboratories with a high firearms caseload. Multiple likelihood ratios were calculated, with a different pair of cartridge cases in each set serving as the tested cases in order to establish variation in likelihood ratios within a given model of firearm. These likelihood ratios could also be generated between one "case" cartridge case and numerous test cartridges. This method accounted for shot-to-shot variation when constructing likelihood ratios, something not covered by other methods and a shortcoming this method seeks to address, in particular with its 17-cartridge case sets [11].

Generalized Linear Mixed Model

The kernel density estimation and likelihood ratio encountered multiple issues during analysis, necessitating exploration of an alternative data analysis method. The result was the usage of a completely different analysis method: a generalized linear mixed model (GLMM). GLMMs offer a means of expressing likelihood among large datasets with binary outcomes, taking into account random effects [20]. Seeing as there are two possible conclusions for any cartridge case in the data set – same-source or different-source – the data for this project constitutes binary outcomes, making GLMMs an effective tool for analyzing this data. Normally, a logistic regression would be sufficient for this data, as it too supports binary outcomes [20]. However, the use of a logistic regression necessitates independence of all measurements [20]. This data set does exhibit dependence, as the ability for different firearms to create impressions in the cartridge cases varies between firearms, so the quality of the markings is dependent upon the firearm creating them. GLMMs do not require independence due to their inclusion of random effects in the model, making them a better choice for this data set than logistic regression [20]. Cartridge cases from the same firearm should be more similar to each other than to cartridge cases from different firearms and using the firearm itself as a random effect allows the model to

account for this grouping structure. GLMMs also incorporate fixed effects, and because most variables were controlled in this study the only fixed effect used to model likelihood ratios and match probabilities was the Cadre breech face similarity score. This GLMM method allows a given similarity score to be directly tied to a match probability, and likelihood ratios can be calculated from the output as well.

Results

Kernel Density Estimation

The data analysis encountered some unexpected difficulties in interpreting the similarity scores as probability densities. Many of the same-source comparisons yielded breech face similarity scores of 0.9999 for all comparisons within a single firearm. A data set consisting of all identical values cannot be represented by a probability distribution, as there is not a distribution of values to be represented. This also occurred for firearms where comparison scores, while not all 0.9999, had scores greater than 0.99, which also proved insufficient for creating a probability distribution. The firearms that did and did not experience non-distribution are highlighted in Table 1, with Figures 5, 6, and 7 showcasing the lowest-scoring comparison in each category. Figure 5 shows a same-source comparison with very few similar features, showcasing the level of variability that can occur in the "Sufficient Variability" category. Figures 6 and 7, however, show much similarity despite being the lowest-scoring in their category, showcasing a very low amount of similarity score variation.

Table 1: This table shows the serial numbers of the firearms that had necessary same-source similarity score distribution to generate probability density curves. "No Variability" had scores of 0.9999 throughout. "Insufficient Variability" had scores between roughly 0.9000 and 0.9999, which also proved insufficient for curve generation.

Sufficient Variability	Insufficient Variability	No Variability
804688	838795	I812038
004361	P019277	Obliterated S/N

P027307	P077656	P083640
P031079	P101611	P1224960
P062279	P1227506	P1312589
P1292278	P1265668	P1371068
P1427244	P144180	P1395494
P1441138	P1588948	P1726758
P1491856	P1700455	
P1672765	P1238527	
P1680774		
P193096		
P209027		

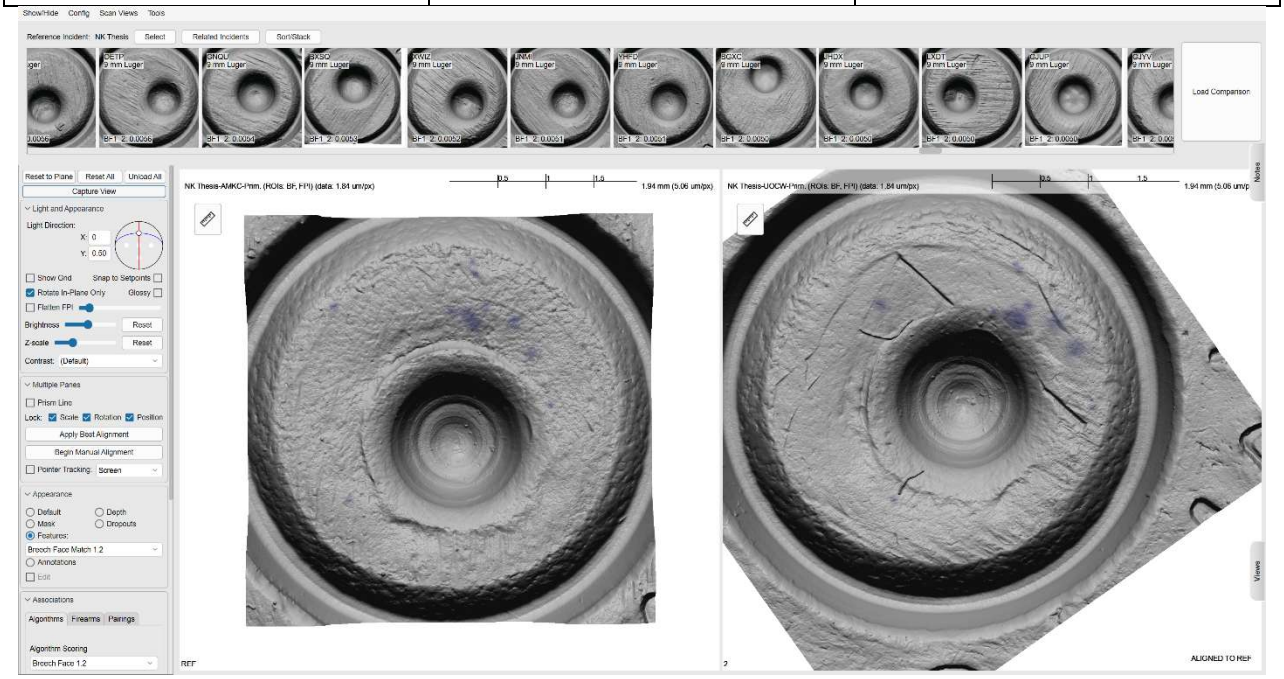


Figure 5: The lowest-scoring same-source comparison in the "Sufficient Variability" category. Note the lack of highlighting performed by the TopMatch algorithm. This comparison had a similarity score of only 0.004.

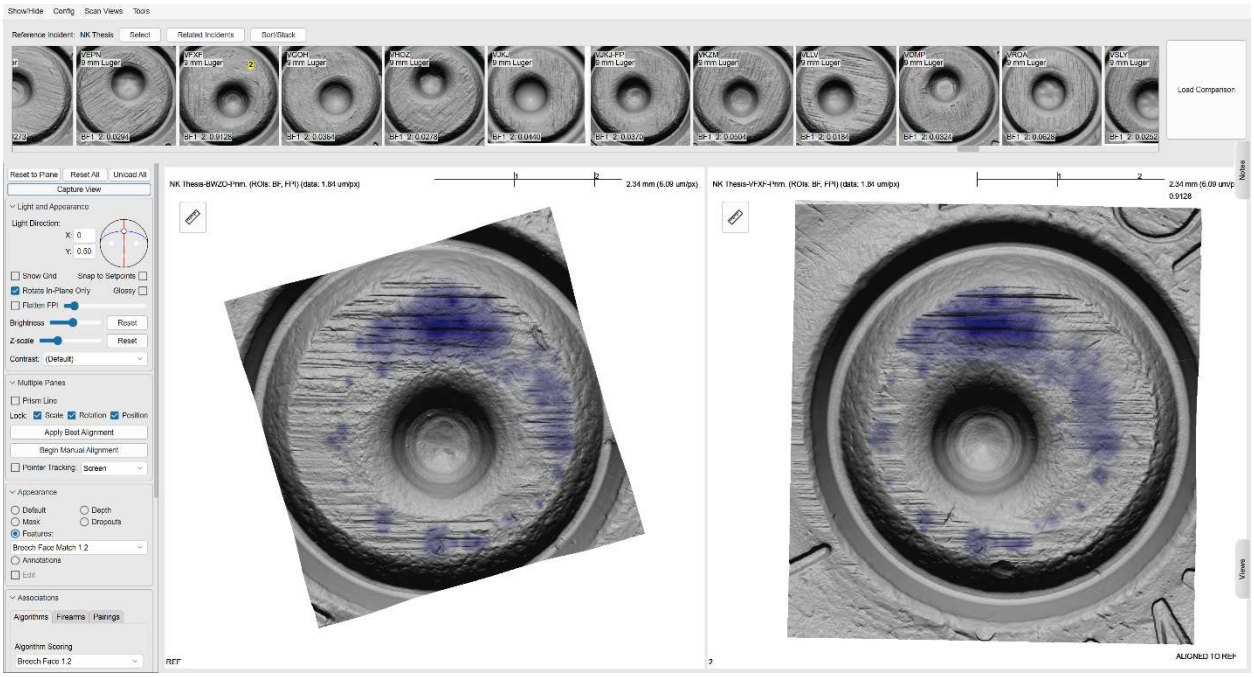


Figure 6: The lowest-scoring same-source comparison in the "Insufficient Variability" category. Similarity score of 0.9128.

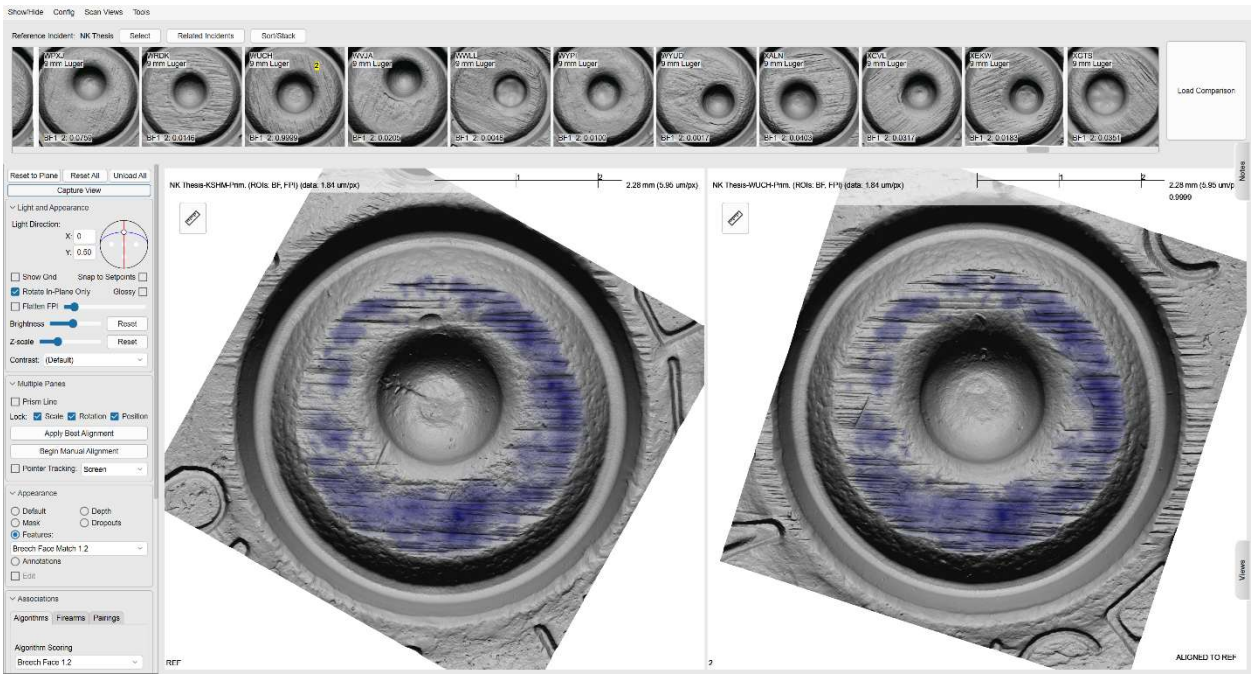


Figure 7: An example same-source comparison from the "No Variability" category. Similarity score of 0.9999.

This analytical curiosity was further compounded by near-zero probability densities at higher similarity scores among many different-source comparisons; since the different-source probability density would serve as a denominator for calculating likelihood ratios, many likelihood ratios came out as undefined where the computer tried to divide by zero. Essentially, there was little to no overlap between same-source and different-source similarity scores with which to calculate a likelihood ratio. This no-overlap phenomenon is illustrated effectively in Figure 8. This is not a failure of the TopMatch 3D; in fact, the considerable divide between same-source and different-source similarity scores actually speaks to the effectiveness of the TopMatch algorithm.

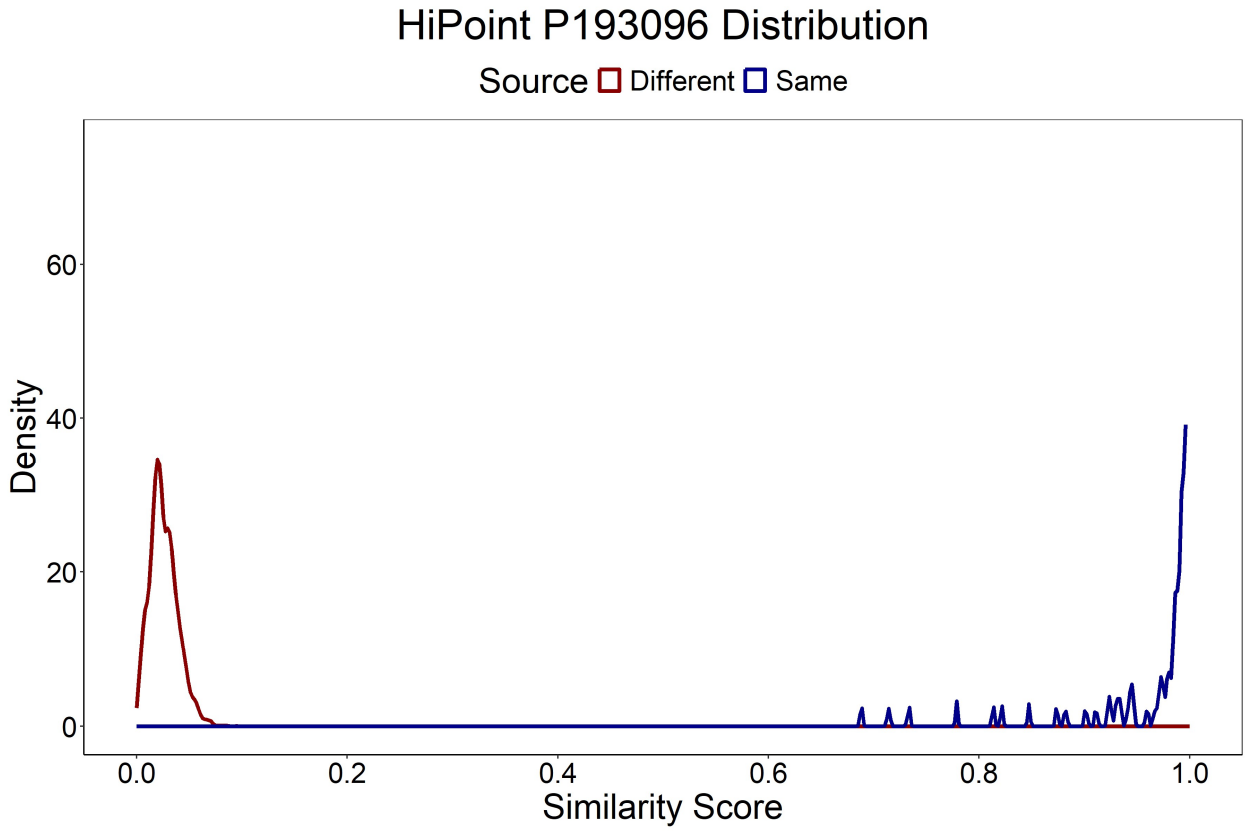


Figure 8: An actual set of probability density curves, generated for HiPoint P193096. This firearm exhibited a same-source and different-source distribution far narrower than expected, as well as no overlap between same-source and different-source scores

Another challenge occurred during data collection, wherein the TopMatch 3D would not scan to the depth necessary to create a topography of the bottom of the firing pin impression, which is where any identifying markings would be. Out of the 527 total cartridge cases, 68 cartridge cases suffered from this insufficient scan depth, with this phenomenon being concentrated among 4 serial numbers: P077656, P083640, P1371068, and P1312589. This, combined with the fact that the firing pin comparison algorithm is brand new and not yet widely available to forensic science laboratories at the time of this research, necessitated a focus on the breech face data.

Generalized Linear Mixed Model

Because the planned method utilizing likelihood ratios proved to be ineffective at analyzing the similarity score data obtained from the TopMatch 3D instrument, it became necessary to find another means of analyzing the data within RStudio. The Generalized Linear Mixed Model had been prepared for this purpose. Prior to building the GLMM, the model assumptions had to be verified, which included the following tests:

- Binary outcomes -> passed
- Independence of observations -> failed, thus usage of GLMM
- Linearity of independent variables -> test shown below
- Absence of strong multicollinearity -> passed, as only one variable is being used, that being the Cadre breech face score
- Absence of major outliers -> determined using Cook's test

The linearity of independent variables was tested using the Box-Tidwell test. This test determines whether or not the data meets the linearity requirements for using a GLMM by delivering a p-value [21]. The p-value for this data was calculated to be 0.294. This value is considered nonsignificant at a significance level of 0.05, resulting in a pass for the linearity test.

The absence of outliers was determined using the Cook's test. In the Cook's test for this data, the Cook score is representative of how much that breech face score would influence the model [22]. Put simply, the larger the Cook's distance, the further from the statistically expected value the comparison score. Large Cook's scores were prevalent among two specific serial numbers, 004361 and 804688, though the largest Cook's score belonged to P1238527. These scores were examined and found to be legitimate scores and implied no problems with the integrity of the data, thus these scores stayed in the data set. A chart indicating which comparisons among the cartridge cases had higher scores is shown in Figure 9. The comparisons between select cartridge cases that exhibited high Cook's distances are shown in Figures 10, 11, and 12.

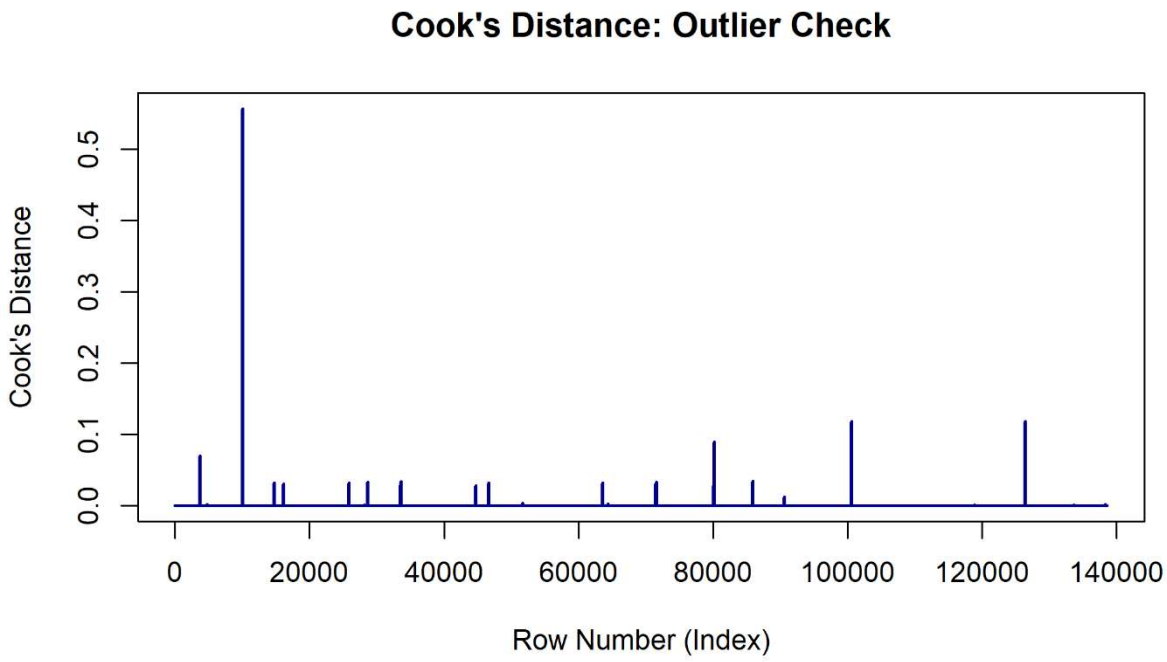


Figure 9: A chart indicating Cook's distances among the many thousands of comparisons. The Row Number simply indicates the position in the data spreadsheet where the comparison lies for that particular score.

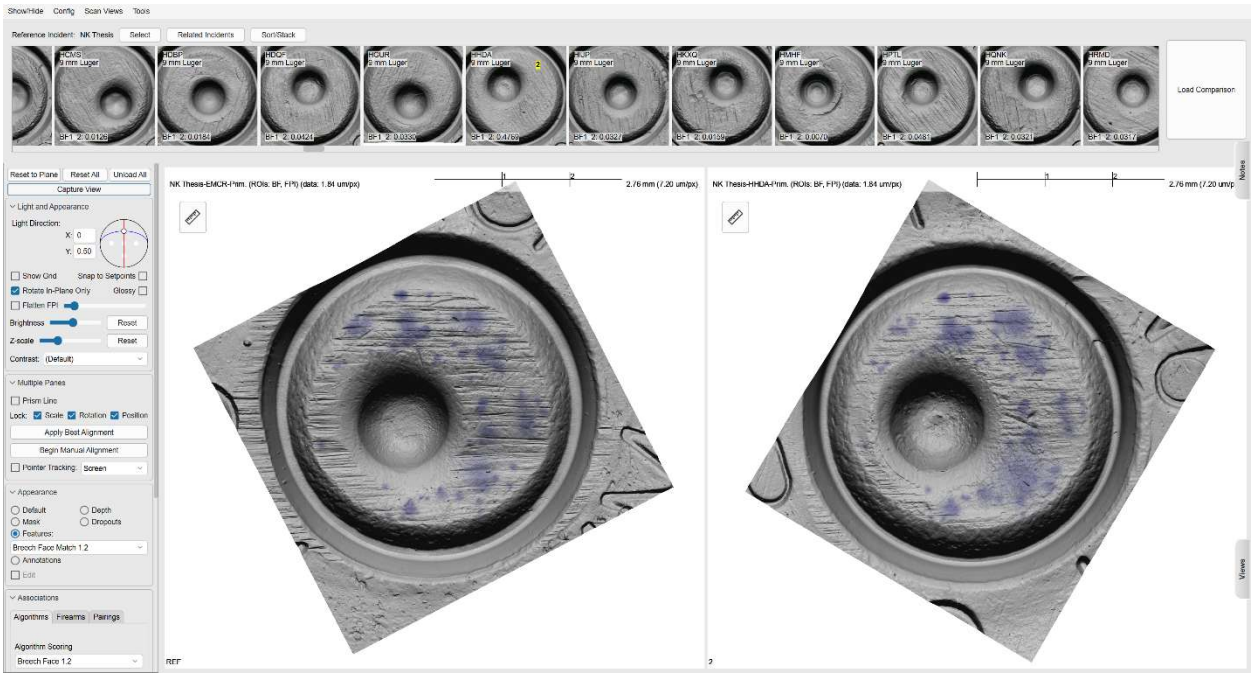


Figure 10: The comparison with the highest Cook's score. In this case, the cartridges had an unusually high similarity score of 0.4769 despite hailing from different serial numbers. A value this high might be expected to exert an influence on the generalized linear mixed model.

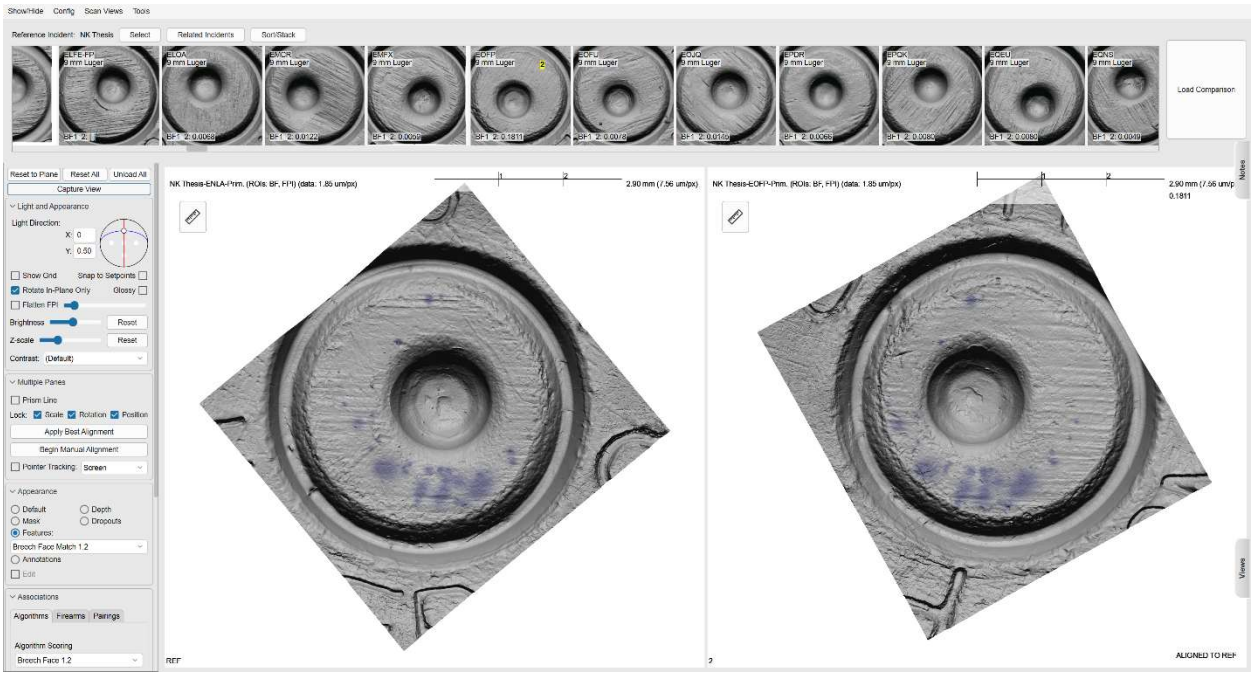


Figure 11: A high Cook's score comparison from serial number 004361, which exhibited multiple high Cook's scores compared to other serial numbers. Here, the similarity score was 0.1811 despite being a same-source comparison.

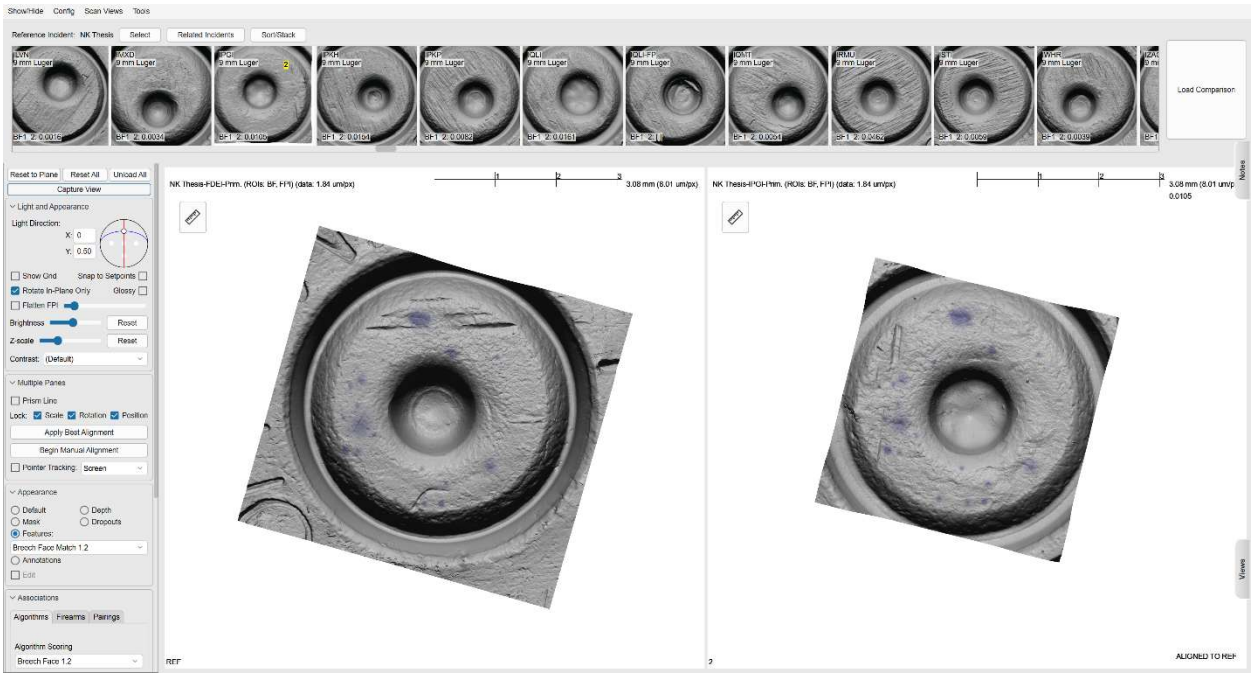


Figure 12: A high Cook's score comparison for serial number 804688, a firearm that exhibited multiple higher Cook's scores. This example, despite being same-source, had a similarity score of only 0.0105.

A leave-one-out cross validation (LOOCV) was conducted to verify prediction validity of the model using ground truth data. Because this is a mixed effects model with the firearm as a random effect, the cross validation was tested by removing all results for a single firearm, building the model on the remaining firearms, and testing the model on the held-out firearm. This was repeated for each of the 31 guns, with the results plotted in a receiver operating characteristic (ROC) curve. A ROC curve plots true positives on the y-axis and false positives on the x-axis [23]. An ideal ROC curve would possess an area under the curve (AUC), calculated by integration, of 1.0, indicating no false positives [23]. The ROC curve for this data set had an AUC of approximately 0.997, very close to the perfect value, serving to validate the GLMM. This validation was also practical in nature: any casework firearm fitting the parameters of the model (HiPoint C9, Federal ammunition) would not itself be part of the model, a circumstance well replicated by leaving one of the firearms out.

Figure 13 shows the results of the logistic GLMM: each curve represents the relationship between similarity score and the probability of that score indicating a same-source ground truth. The gray curves are for individual firearms, whereas the blue curve near the center of the figure is the average for all firearms. The gray curves do well to demonstrate the variability that can occur within a single model of firearm. The gray curves to the left of the blue curve generated more consistent markings, as a lower similarity score is required to achieve a high probability of same-source conclusion. The curves to the right of the blue curve show the opposite, where firearms produced less consistent markings and thus have a lower same-source probability with even above-average similarity scores. The average from this chart suggests that a similarity score as low as 0.50 would have, in an average comparison, a greater than 60% chance of being a same-source comparison, with this probability increasing sharply with only miniscule increases in similarity score value. This sort of data could be directly presentable in court for a firearm of this specific type.

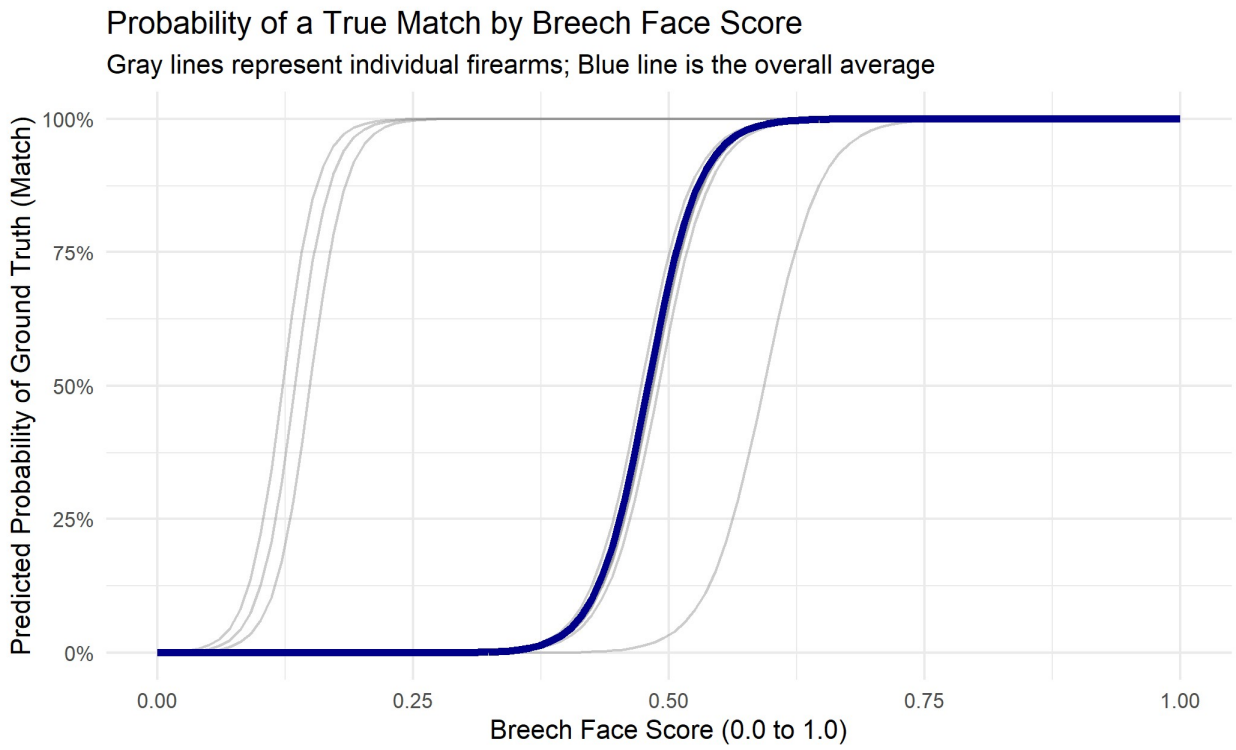


Figure 13: A chart showing a relationship between similarity score and the probability of that score indicating a same-source comparison.

Through rearrangement of Bayes' Theorem, likelihood ratios were able to be calculated from the GLMM output. The ratio of same-source to different-source comparisons within the overall dataset served as the prior odds, approximately 0.03137, naturally weighed heavily towards different-source due to the far larger number of different-source comparisons. This larger number of different-source comparisons is in part why the LR values increase so sharply beyond a breech face score of 0.5. This is because the evidence would first need to be strong enough to overcome this large number of different-source comparisons, after which point there is no other significant statistical hurdle. The probabilities provided by the GLMM were converted to the posterior odds, enabling a likelihood ratio to be calculated by using Bayes' Theorem, dividing the posterior odds by the prior odds. Table 2 shows the likelihood ratio at different breech face test scores in the GLMM.

Table 2: A table showing the likelihood ratios calculated from different breech face scores in the GLMM.

Breech Face Test Score	Likelihood Ratio
0.1	8.28725e-06
0.25	0.003332357
0.5	72.99951
0.75	1599147
0.9	643027419

The expansion of this method is expected to have an important impact on the field of forensic firearms examination. However, this execution of this method is limited in scope to strictly Hi-Point C9 handguns using Federal brand 9mm ammunition. Traditionally, firearms examination had been a qualitative process, without many numerical outputs to assist in determinations. This is among the biggest faults to have been mentioned in the 2016 PCAST report, which claimed such qualitative methods offered only the possibility for subjective conclusions [13]. A method such as what has been outlined above would provide some much-needed quantification and objectivity to the comparison of firearms evidence. This quantification will allow presentation of firearms examination findings to have more concrete figures in a format that is potentially more presentable to a courtroom setting.

While this method began with KDE as a means of obtaining likelihood ratios before pivoting to a GLMM, it is important to note that the utilizations of these methods have distinctions between one another. The KDE method would require 15 test firings from a case firearm and require all of the resulting cartridge cases to be compared to similar or same firearms, generating a large nonmatch distribution, in order to calculate the likelihood ratio for the evidence cartridge case(s). Alternatively, the GLMM is directly applicable to any new

Federal brand cartridge cases fired by HiPoint C9 pistols without having to develop case-specific match and nonmatch distributions due to the specific firearms being included in the model as a random effect.

Conclusions

The model developed in this paper demonstrates an alternative approach to more traditional kernel density estimation for calculating likelihood ratios in firearm evidence comparisons and further suggests that the Cadre TopMatch 3D instrument and its built-in breech face comparison algorithms are reliable for the identification and differentiation of firearms evidence, particularly in Hi-Point C9 model firearms. It is worth noting, however, that Hi-Point C9 handguns produce pronounced, consistent parallel impressions in their breech face marks that allows for generally easy comparison even without the TopMatch 3D algorithm. This is not to understate the importance of the algorithm's outputs, however, as even with these easily comparable marks in C9 model pistols, these numerical outputs can help to bolster an examiner's conclusions when delivered as expert testimony. More varied models of firearm would need to undergo similar testing to achieve a thorough understanding of the TopMatch algorithm's efficacy.

Further Research

This data set serves to represent the variability of only HiPoint C9 model of firearms and is thus only currently practical for applications pertaining to that firearm. Additional research will need to focus on applying this methodology to other firearms. Given that Hi-Point C9 pistols are known for leaving pronounced and generally consistent impressed markings within a single firearm, a good proving of this method would be to use the method with other 9mm Luger

firearms that do not leave such distinguishable markings or possess different kinds of markings altogether (e.g. granular patterns). Other firearms of different calibers and brands of ammunition would need to be subjected to similar GLMM construction so that should those firearms show up in casework, an existing model could be used to analyze the fired cartridges. A fully expanded model would allow a user to input firearm make, model, chambering, and brand of ammunition to provide examiners likelihood ratios applicable to a specific case firearm. Obtaining an understanding of how the TopMatch system tends to score same-source and different-source comparisons of various firearm models, calibers, and ammunition brands would allow examiners to use the TopMatch algorithm effectively in casework.

References

1. Vorburger, T. V., Song, J., & Petraco, N. (2015). Topography measurements and applications in ballistics and tool mark identifications. *Surface Topography: Metrology and Properties*, *4*(1), 013002.
2. Song, J. (2013). Proposed NIST ballistics identification system (NBIS) based on 3D topography measurements on correlation cells. *AFTE Journal*, *45*(2), 184-194.
3. Tong, M., Song, J., Chu, W., & Thompson, R. M. (2014). Fired cartridge case identification using optical images and the congruent matching cells (CMC) method. *Journal of Research of the National Institute of Standards and Technology*, *119*, 575.
4. Tong, M., Song, J., & Chu, W. (2015). An improved algorithm of congruent matching cells (CMC) method for firearm evidence identifications. *Journal of Research of the National Institute of Standards and Technology*, *120*, 102.
5. Chen, Z., Song, J., Chu, W., Soons, J. A., & Zhao, X. (2017). A convergence algorithm for correlation of breech face images based on the congruent matching cells (CMC) method. *Forensic Science International*, *280*, 213-223.
6. Song, J., Vorburger, T. V., Chu, W., Yen, J., Soons, J. A., Ott, D. B., & Zhang, N. F. (2018). Estimating error rates for firearm evidence identifications in forensic science. *Forensic Science International*, *284*, 15-32.
7. Weller, T. J., Zheng, A., Thompson, R., & Tulleners, F. (2012). Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides. *Journal of Forensic Sciences*, *57*(4), 912-917.

8. Riva, F., & Champod, C. (2014). Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases. *Journal of Forensic Sciences*, *59*(3), 637-647.
9. Fadul, T. G., Hernandez, G. A., Wilson, E., Stoiloff, S., & Gulati, S. (2013). An empirical study to improve the scientific foundation of forensic firearm and tool mark identification utilizing consecutively manufactured Glock EBIS barrels with the same EBIS pattern. *US Department of Justice Report*.
10. Lightstone, L. (2010). The potential for and persistence of subclass characteristics on the breech faces of SW40VE Smith & Wesson Sigma pistols. *AFTE Journal*, *42*(4), 308-322.
11. Baiker-Sørensen, M., Alberink, I., Granell, L. B., van der Ham, L., Mattijssen, E. J., Smith, E. D., ... & Zheng, X. A. (2023). Automated interpretation of comparison scores for firearm toolmarks on cartridge case primers. *Forensic Science International*, *353*, 111858.
12. Hofmann, H., Carriquiry, A., & Vanderplas, S. (2020). Treatment of inconclusives in the AFTE range of conclusions. *Law, Probability and Risk*, *19*(3-4), 317-364.
13. President's Council of Advisors on Science and Technology (PCAST). (2016). Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. Report to the President.
14. Law, E. F., Morris, K. B., & Jelsema, C. M. (2017). Determining the number of test fires needed to represent the variability present within 9 mm Luger firearms. *Forensic Science International*, *276*, 126-133.
15. Weller, T., Brubaker, M., Duez, P., & Lilien, R. (2015). Introduction and initial evaluation of a novel three-dimensional imaging and analysis system for firearm forensics. *AFTE Journal*, *47*(4), 198-208.

16. Cadre Forensics (2022). TopMatch-3D High-Capacity Desktop Scanner. *Technical Specifications*. Cadre Forensics. Retrieved from https://www.cadreforensics.com/techspecs.html
17. R Core Team (1993). *R* (Version 4.6.1) [Coding language]. r-project.org.
18. Węglarczyk, S. (2018). Kernel density estimation and its application. *ITM Web of Conferences* (Vol. 23, p. 00037). EDP Sciences.
19. Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(3), 683-690.
20. Bolker, B. M. (2015). Linear and generalized linear mixed models. *Ecological Statistics: Contemporary Theory and Application*, *2015*, 309-333.
21. Heit, D. R., Ortiz-Calo, W., Poisson, M. K., Butler, A. R., & Moll, R. J. (2024). Generalized nonlinearity in animal ecology: Research, review, and recommendations. *Ecology and Evolution*, *14*(7), e11387.
22. Cook, R. D., & Weisberc, S. Influence and Outliers. *An Introduction to Regression Graphics*, *203*.
23. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874.

Software Used

1. R Core Team (1993). *R* (Version 4.6.1) [Coding language]. r-project.org.
2. Posit PBC (2011). *RStudio* (Version 4.5.1) [Computer software]. https://posit.co.

3. Wickham, H. & Chang, W. (2007). *ggplot2* (Version 4.0.3) [Computer software]. ggplot2.tidyverse.org.
4. Wickham, H., François R., Henry L., Kirill Müller K., & Vaughan D (2014). *dplyr* (Version 1.2.1) [Computer Software]. dplyr.tidyverse.org.
5. Bates D., Mächler M., Bolker B., & Walker S. (2015). *lme4* (Version 2.0-1) [Computer Software]. https://cran.r-project.org/web/packages/lme4/index.html
6. Robin X., Turck N., Hainard A., Tiberti N., Lisacek F., Sanchez J., & Müller M. (2011). *pROC* (Version 1.19.0.1) [Computer Software]. https://cran.r-project.org/web/packages/pROC/index.html

Appendix

Cadre Comparison File Cleaning R Script

```
library(dplyr)

master_file <- read.csv("C:/Users/natha/OneDrive/Desktop/ThesisData/Cartridge Case Identifier Master List.csv", header = TRUE)
#master_file <- master_file[master_file$Notes == "NIJ",]

raw_data <- read.csv("C:/Users/natha/OneDrive/Desktop/ThesisData/FullResultsFiringPin_PW.csv", header = TRUE)
raw_data <- raw_data[-c(4)]
raw_data$Case_ref <- sub(pattern = "NK Thesis-", replacement = "", raw_data$Case_ref)
raw_data$Case_comp <- sub(pattern = "NK.Thesis.", replacement = "", raw_data$Case_comp)
colnames(raw_data) <- c("Case_ref", "Case_comp", "FP_score")

raw_data2 <- read.csv("C:/Users/natha/OneDrive/Desktop/ThesisData/FullResultsBreechFace_PW.csv", header = TRUE)
raw_data2 <- raw_data2[-c(4)]
raw_data2$Case_ref <- sub(pattern = "NK Thesis-", replacement = "", raw_data2$Case_ref)
raw_data2$Case_comp <- sub(pattern = "NK.Thesis.", replacement = "", raw_data2$Case_comp)
colnames(raw_data2) <- c("Case_ref", "Case_comp", "BF_score")
```

```
raw_data3 <- left_join(raw_data, raw_data2, by = c("Case_ref", "Case_comp"))

###############
# Test Section

#test_raw_data <- raw_data[1:5,]
#which(master_file$Case.ID == "TXZY", arr.ind=TRUE) # find row number for given Case.ID
#test_master_file <- master_file[c(87,89,64,20,63,61),]

###############

firearm_data_temp <- merge(raw_data3, master_file, by.x="Case_ref", by.y="Case.ID")
firearm_data_temp <- firearm_data_temp[-c(10:12)] # remove notes column and extra blank
columns
colnames(firearm_data_temp) <- c("Case_ref", "Case_comp", "FP_score", "BF_score",
"Make_ref", "Model_ref",
                          "Serial_num_ref", "Ammunition_ref", "Caliber_ref")

firearm_data_temp2 <- merge(firearm_data_temp, master_file, by.x="Case_comp",
by.y="Case.ID")
firearm_data_temp2 <- firearm_data_temp2[-c(15:17)] # remove notes column and extra blank
columns
colnames(firearm_data_temp2) <- c("Case_comp", "Case_ref", "FP_score", "BF_score",
"Make_ref", "Model_ref",
                           "Serial_num_ref", "Ammunition_ref", "Caliber_ref", "Make_comp",
                           "Model_comp", "Serial_num_comp", "Ammunition_comp",
"Caliber_comp")
col_order <- c("Case_ref", "Case_comp", "FP_score", "BF_score", "Make_ref", "Model_ref",
           "Serial_num_ref", "Ammunition_ref", "Caliber_ref", "Make_comp",
           "Model_comp", "Serial_num_comp", "Ammunition_comp", "Caliber_comp")
firearm_data_temp2 <- firearm_data_temp2[,col_order]
firearm_data_temp2$Ammunition_ref <- sub(pattern = " ", replacement = "",
firearm_data_temp2$Ammunition_ref)
firearm_data_temp2$Ammunition_comp <- sub(pattern = " ", replacement = "",
firearm_data_temp2$Ammunition_comp)
firearm_data_temp3 <- firearm_data_temp2

# add in source column based on serial number
Source <- c()

for(ii in 1:nrow(firearm_data_temp3)){

  if(firearm_data_temp3$Serial_num_ref[ii] == firearm_data_temp3$Serial_num_comp[ii]){
```

```
    Source[ii] <- "Same"
  } else if(firearm_data_temp3$Serial_num_ref[ii] !=
firearm_data_temp3$Serial_num_comp[ii]){
    Source[ii] <- "Different"
  } else{
    Source[ii] <- "Error"
  }

}

firearm_data_temp3 <- cbind(firearm_data_temp3, Source)
col_order <- c("Case_ref", "Case_comp", "FP_score", "BF_score", "Source", "Make_ref",
"Model_ref",
               "Serial_num_ref", "Ammunition_ref", "Caliber_ref", "Make_comp",
               "Model_comp", "Serial_num_comp", "Ammunition_comp", "Caliber_comp")
firearm_data_temp3 <- firearm_data_temp3[,col_order]

# add in column about whether the reference and comparison ammo brands are the same and
# for handling reciprocal comparisons:
# ex: MaxxTech&Blazer == Blazer&MaxxTech (sort based on alphabetical order)
Ammo_equal <- c()

for(ii in 1:nrow(firearm_data_temp3)){

  if(firearm_data_temp3$Ammunition_ref[ii] == firearm_data_temp3$Ammunition_comp[ii]){
    Ammo_equal[ii] <- "Same_ammo"
  } else if(firearm_data_temp3$Ammunition_ref[ii] !=
firearm_data_temp3$Ammunition_comp[ii]){
    Ammo_equal[ii] <- "Different_ammo"
  } else{
    Ammo_equal[ii] <- "Error"
  }

}

firearm_data_temp3 <- cbind(firearm_data_temp3, Ammo_equal)
col_order <- c("Case_ref", "Case_comp", "FP_score", "BF_score", "Source", "Make_ref",
"Model_ref",
               "Serial_num_ref", "Ammunition_ref", "Caliber_ref", "Make_comp",
               "Model_comp", "Serial_num_comp", "Ammunition_comp", "Caliber_comp",
"Ammo_equal")
firearm_data_temp3 <- firearm_data_temp3[,col_order]
```

```
write.table(firearm_data_temp3,
"C:/Users/natha/OneDrive/Desktop/ThesisData/FullResultsBFFP_Clean.csv",
          row.names=FALSE, col.names=TRUE, sep=",")
```

Kernel Density Estimation R Script

```
library(ggplot2)
library(dplyr)
  # these load the libraries to perform necessary functions

# read and organize data

dat01 <- read.csv("C:/Users/natha/OneDrive/Desktop/ThesisData/FullResultsBFFP_Clean.csv",
header = TRUE)
#serial_nums <- unique(dat01$Serial_num_ref)
  # This should target all unique serial numbers in the data
serial_nums <- c("804688", "004361", "P027307", "P031079", "P062279", "P1292278",
"P1427244", "P1441138", "P1491856", "P1672765", "P1680774", "P193096", "P209027")
  # This targets only defined serial numbers

# create save directory for plots
save.dir <- "C:/Users/natha/OneDrive/Desktop/ThesisData/Data Plots/SJ"

sample_level <- c(2) # this is the number of cartridge cases being removed for LR calculation
                  # we probably only need 2, but this is a good placeholder in case we need to do
other tests
LR_results_all <- c() # create and empty string to store all results

#ii <- 1

# The following begins the loop:
for(ii in 1:length(serial_nums)){

  dat01_firearm <- dat01[dat01$Serial_num_ref == serial_nums[ii],]
  dat01_firearm_ss <- dat01[dat01$Serial_num_ref == serial_nums[ii] & dat01$Source ==
"Same",]
    # This narrows down dat01 to a single firearm, and to same source comparisons

  # create plot filename using the save.dir directory
  filename1 <- file.path(save.dir, paste(serial_nums[ii], "Test HiPoint.png", sep=" "))

  ggplot(dat01_firearm, aes(x=BF_score, color = Source)) +
    geom_density(linewidth = 1.25, bw="sj") +
    scale_color_manual(values = c("darkred", "darkblue")) +
    scale_x_continuous(limits = c(0,1), breaks = seq(0, 1, 0.2)) +
    scale_y_continuous(limits = c(0,75)) +
    labs(title = paste("HiPoint", serial_nums[ii], "Distribution", sep=" "),
```

```
        x = "Similarity Score",
        y = "Density")+
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5), panel.background = element_blank(),
          axis.line = element_line(colour = "black"), text = element_text(size=24),
          legend.position = "top", axis.text.x = element_text(colour = "black"),
          axis.text.y = element_text(colour = "black"),panel.grid.major = element_blank(),
          panel.grid.minor = element_blank())
    # This creates the plot.

  #save plot
  ggsave(filename1, width = 12, height = 8, dpi = 300, bg = "transparent")

  # how am I going to get likelihood ratios????
  cases <- as.data.frame(unique(dat01_firearm_ss$Case_ref))
  colnames(cases) <- "Cases"

  #jj <- 1
  for(jj in 1:length(sample_level)){

    # create all unique combinations of 2 cases
    cases_sampled <- combn(cases$Cases, sample_level[jj])

    LR_results_gun <- c() #create empty string to store results from next loop

    for(kk in 1:ncol(cases_sampled)){

      #pull out first unique pair (on first loop run, then do all subsequent pairs)
      cases_to_rem <- cases_sampled[,kk]

      # remove sampled cases to build match/nonmatch distribution with everything else
      sampled_dist <- dat01_firearm[dat01_firearm$Case_ref != cases_to_rem[1] &
dat01_firearm$Case_ref != cases_to_rem[2] &
                                dat01_firearm$Case_comp != cases_to_rem[1] &
dat01_firearm$Case_comp != cases_to_rem[1], ]

      match_dist <- sampled_dist[sampled_dist$Source=="Same",]
      match_dens <- density(match_dist$BF_score, bw="sj")

      nm_dist <- sampled_dist[sampled_dist$Source=="Different",]
      nm_dens <- density(nm_dist$BF_score, bw="sj")

      #pull out BF score for the pair of cases removed
      Test_val <- dat01_firearm[dat01_firearm$Case_ref == cases_to_rem[1] &
dat01_firearm$Case_comp == cases_to_rem[2],]$BF_score
```

```
    if(length(Test_val) == 0){ #this logical chunk tests the reverse of the previous line depending
on which column each identifier is in
      Test_val <- dat01_firearm[dat01_firearm$Case_ref == cases_to_rem[2] &
dat01_firearm$Case_comp == cases_to_rem[1],]$BF_score
    }

    #calculate LR by getting the density from the match and nonmatch distributions at the unique
pair BF score
    Test_match_dens <- density(match_dist$BF_score, from=Test_val, to=Test_val, n=1,
bw="sj")$y
    Test_nm_dens <- density(nm_dist$BF_score, from=Test_val, to=Test_val, n=1, bw="sj")$y
    LR <- Test_match_dens/Test_nm_dens #match/nonmatch
    LR <- round(LR,2)
    logLR <- log10(LR) #positive supports prosecution, negative supports defense
    #try rounding to 2 decimals! round(LR) AND round(logLR)
    logLR <- round(logLR, 2)

    output <- c(serial_nums[ii], cases_to_rem[1], cases_to_rem[2], Test_val, Test_match_dens,
Test_nm_dens, LR, logLR)

    LR_results_gun <- rbind(LR_results_gun, output)
    rownames(LR_results_gun) <- NULL #remove row names
    colnames(LR_results_gun) <-
c("Serial_num","Case_ref","Case_comp","BF_score","Match_dens",
                          "NM_dens","LR","logLR")

    cat(cases_sampled[,kk],"\n")

    } #closes cases_sampled loop

  } #closes sample_level loop

  # row 1 in ss spreadsheet - VBIL and CQKF and remove,
  #filter(df, Case_ref != VBIL | Case_ref != CQKF | Case_comp != VBIL | Case_comp != CQKF)
  # create densities with remaining scores
  # Pull score from VBIL and CQKF for LR
  # Code for different folders???
    #dir.create()
  # cat()

  LR_results_all <- rbind(LR_results_all, LR_results_gun)

  cat(paste(serial_nums[ii], "Complete", sep=" "),"\n")

} #closes serial_nums loop
```

```
# Creates a new directory; useful for separating plots by serial number

#for(ll in 1:length(serial_nums)){
  #dir.create("C:/Users/natha/OneDrive/Desktop/ThesisData/Data Plots/Example Directory")
#}
write.csv(LR_results_all, file="C:/Users/natha/OneDrive/Desktop/ThesisData/Data
Plots/FullResults_LR_results_all.csv")

#use sort command? google this!

ggplot(LR_results_all, aes(x=Serial_num, y=LR, group=Serial_num)) +
  geom_boxplot() +
  labs(title = "Likelihood Ratio Box Plot", x = "Serial Number", y = "Likelihood Ratio") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5), panel.background = element_blank(),
        axis.line = element_line(colour = "black"), text = element_text(size=16),
        legend.position = "top", axis.text.x = element_text(colour = "black"),
        axis.text.y = element_text(colour = "black"),panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())

ggplot(LR_results_all, aes(x=Serial_num, y=logLR, group=Serial_num)) +
  geom_boxplot() +
  labs(title = "Logarithm of LR Box Plot", x = "Serial Number", y = "log(LR)") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5), panel.background = element_blank(),
        axis.line = element_line(colour = "black"), text = element_text(size=16),
        legend.position = "top", axis.text.x = element_text(colour = "black"),
        axis.text.y = element_text(colour = "black"),panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
```

GLMM Regression R Script

```
library(ggplot2)
library(dplyr)
library(lme4)
library(pROC)

dat01 <- read.csv("C:/Users/natha/OneDrive/Desktop/ThesisData/FullResultsBFFP_Clean.csv",
                  header = TRUE)

# the below two rows just show data file organization and unique values per variable
head(dat01)
sapply(dat01, function(x) length(unique(x)))

#####################################
```

```
# Build the GLMM
# Also measures system time -> lots of data

start_time <- Sys.time()

# use a random intercept and random slope
# --- RUN THE GLMM ---
model_glmm <- glmer(Ground_truth ~ BF_score + (BF_score | Serial_num_ref),
                    data = dat01,
                    family = binomial(link = "logit"),
                    control = glmerControl(optimizer = "bobyqa"))

end_time <- Sys.time()
time_taken <- end_time - start_time

print(time_taken)
print(summary(model_glmm))

#####################################

# Assumption 1 - Test for linearity in logit

dat01$score_log_score <- dat01$BF_score * log(dat01$BF_score)
box_tidwell_model <- glm(Ground_truth ~ BF_score + score_log_score,
                         family = binomial, data = dat01)
summary(box_tidwell_model) # Look at the p-value for score_log_score

#####################################

# Assumption 2 - outliers
# Cook's Distance (rule of thumb >1.0 is strong outlier)
# Run a standard GLM just for the outlier check
outlier_model <- glm(Ground_truth ~ BF_score,
                     data = dat01,
                     family = binomial(link = "logit"))

# Extract the Cook's Distance scores
cooks_d <- cooks.distance(outlier_model)

plot(cooks_d,
    type = "h",
    main = "Cook's Distance: Outlier Check",
```

```
    ylab = "Cook's Distance",
    xlab = "Row Number (Index)",
    lwd = 2,
    col = "darkblue")

# Have R find the row numbers where Cook's Distance is greater than the threshold
outlier_rows <- which(cooks_d > 0.02)

# Extract the actual Cook's Distance scores for just those specific rows
outlier_scores <- cooks_d[outlier_rows]

# Put them into a neat dataframe so you can actually read them
outlier_table <- data.frame(
  Row_Number = outlier_rows,
  Cooks_Distance = outlier_scores
)

# Sort the table from the worst offender (highest score) down to the lowest
outlier_table <- outlier_table[order(-outlier_table$Cooks_Distance), ]
outlier_table$Serial_num_ref <- dat01$Serial_num_ref[outlier_table$Row_Number]

# Rearrange the columns so the Serial Number is easy to read
# This just moves it to the second column, right after the Row Number
outlier_table <- outlier_table[, c("Row_Number", "Serial_num_ref", "Cooks_Distance")]

print(outlier_table)

#####################################

# Get the odds ratio
# This converts the fixed effects estimates from the model summary to actual odds

exp(fixef(model_glmm))

#####################################

# We want to predict the probability for scores ranging from 0 to 1
# for every single one of the 30 guns
plot_data <- expand.grid(
  Serial_num_ref = unique(dat01$Serial_num_ref),                    # All 30 firearms
  BF_score = seq(0, 1, length.out = 100) # 100 points between 0.0 and 1.0
)
```

```
# Calculate predicted probabilities
# Get the predictions for EACH SPECIFIC GUN (incorporates random effects)
# type = "response" converts the log-odds back into plain percentages (0.0 to 1.0)
plot_data$gun_specific_prob <- predict(model_glmm,
                                       newdata = plot_data,
                                       type = "response")

# Get the OVERALL AVERAGE prediction (fixed effects only)
# re.form = NA tells the model to ignore the gun-specific differences
plot_data$overall_prob <- predict(model_glmm,
                                  newdata = plot_data,
                                  type = "response",
                                  re.form = NA)

# Create ggplot to visualize results
ggplot(data = plot_data, aes(x = BF_score)) +

  # Plot the 30 individual guns as thin, transparent lines
  geom_line(aes(y = gun_specific_prob, group = Serial_num_ref),
            color = "gray50", alpha = 0.4, linewidth = 0.5) +

  # Plot the overall average as a thick, solid blue line
  geom_line(aes(y = overall_prob),
            color = "darkblue", linewidth = 1.5) +

  # Formatting and Labels
  theme_minimal() +
  labs(
    title = "Probability of a True Match by Breech Face Score",
    subtitle = "Gray lines represent individual firearms; Blue line is the overall average",
    x = "Breech Face Score (0.0 to 1.0)",
    y = "Predicted Probability of Ground Truth (Match)"
  ) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1)) # Formats y-axis as %

#####################################

# Find optimal cutoff to best balance sensitivity, specificity, FPs, and FNs
# We compare your actual ground truth to the raw continuous algorithm scores
roc_breech <- roc(response = dat01$Ground_truth,
                  predictor = dat01$BF_score,
                  levels = c(0, 1), # Explicitly define 0 as negative, 1 as positive
                  direction = "<")  # Assumes higher scores = more likely to be a match
```

```
# Find optimal score threshold
# "best" automatically uses Youden's J statistic to find the optimal balance
optimal_cutoff <- coords(roc_breech,
                         x = "best",
                         best.method = "youden",
                         ret = c("threshold", "accuracy", "sensitivity", "specificity"))

print(optimal_cutoff)

#####################################

# calculate accuracy, sensitivity, specificity, false positives, false negatives
# can use any score - start with 50% match probability (0.479)
custom_score_threshold <- 0.479

# 1. Define the custom function
calculate_metrics <- function(actual_labels, predicted_scores, threshold) {

  # Classify scores based on your custom threshold
  # If the score is >= threshold, it gets a 1. Otherwise, a 0.
  predicted_classes <- ifelse(predicted_scores >= threshold, 1, 0)

  # Calculate the raw Confusion Matrix counts
  True_Positives  <- sum(actual_labels == 1 & predicted_classes == 1, na.rm = TRUE)
  True_Negatives  <- sum(actual_labels == 0 & predicted_classes == 0, na.rm = TRUE)
  False_Positives <- sum(actual_labels == 0 & predicted_classes == 1, na.rm = TRUE)
  False_Negatives <- sum(actual_labels == 1 & predicted_classes == 0, na.rm = TRUE)

  # Calculate the percentages
  Total <- length(actual_labels)
  Accuracy    <- (True_Positives + True_Negatives) / Total
  Sensitivity <- True_Positives / (True_Positives + False_Negatives)
  Specificity <- True_Negatives / (True_Negatives + False_Positives)

  # Package it all up into a neat, readable table
  results <- data.frame(
    Threshold = threshold,
    Accuracy = round(Accuracy, 4),
    Sensitivity = round(Sensitivity, 4),
    Specificity = round(Specificity, 4),
    TP = True_Positives,
    TN = True_Negatives,
```

```
   FP = False_Positives,
   FN = False_Negatives
  )

  return(results)
}

# Just pass your ground truth column, your score column, and any threshold you want!

my_results <- calculate_metrics(actual_labels = dat01$Ground_truth,
                                predicted_scores = dat01$BF_score,
                                threshold = custom_score_threshold)

print(my_results)

#####################################

# Match probability at a given score

breech_face_score <- 0.55

Match_prob <- 1 / (1 + exp(-(-19.151 + (39.955*breech_face_score))))
cat(paste(round(Match_prob*100, 2), "% match probability", sep="")
```

GLMM LR Calculation R Script

```
library(lme4)

dat01 <- read.csv("C:/Users/natha/OneDrive/Desktop/ThesisData/FullResultsBFFP_Clean.csv",
                  header = TRUE)

# Build the GLMM
# Also measures system time -> lots of data

start_time <- Sys.time()

# use a random intercept and random slope
# --- RUN THE GLMM ---
model_glmm <- glmer(Ground_truth ~ BF_score + (BF_score | Serial_num_ref),
                    data = dat01,
                    family = binomial(link = "logit"),
                    control = glmerControl(optimizer = "bobyqa"))
```

```
end_time <- Sys.time()
time_taken <- end_time - start_time
print(time_taken)

#####################################
# Likelihood Ratio predictions

# calculate prior odds
prior_odds <- (sum(dat01$Ground_truth == 1) / nrow(dat01)) / (1 - sum(dat01$Ground_truth ==
1) / nrow(dat01))

# pick a breech face score at which to calculate an LR
test_BF_score <- 0.9

# Define your new casework scenario
new_case <- data.frame(
  BF_score = test_BF_score # The raw score from the algorithm
)

# Get the model's absolute probability (Posterior Probability)
new_case$predicted_prob <- predict(model_glmm,
                                   newdata = new_case,
                                   type = "response",
                                   re.form = NA)

# Convert Posterior Probability to Posterior Odds
new_case$posterior_odds <- new_case$predicted_prob / (1 - new_case$predicted_prob)

# Divide by the Dataset Prior Odds to get the Likelihood Ratio
new_case$Likelihood_Ratio <- new_case$posterior_odds / prior_odds

# 5. Print the final report
print(new_case[, c("BF_score", "Likelihood_Ratio")])
```

GLMM Validation R Script

```
library(lme4)
library(pROC)
library(ggplot2)

dat01 <- read.csv("C:/Users/natha/OneDrive/Desktop/ThesisData/FullResultsBFFP_Clean.csv",
                  header = TRUE)
```

```
# 1. Create an empty column in your dataset to hold the cross-validated predictions
dat01$cv_predicted_prob <- NA

# 2. Get the unique list of your 31 guns
gun_list <- unique(dat01$Serial_num_ref)

start_time <- Sys.time()

# 3. Start the Leave-One-Gun-Out loop
for(current_gun in gun_list) {

  # Print progress so you know it hasn't frozen
  cat("Training model without gun:", current_gun, "\n")

  # Split the data into Training (30 guns) and Testing (1 left-out gun)
  train_data <- dat01[dat01$Serial_num_ref != current_gun &
                      dat01$Serial_num_comp != current_gun, ]
  test_data  <- dat01[dat01$Serial_num_ref == current_gun, ]

  # Train the model on the 30 guns
  # (Make sure this formula matches your actual model exactly)
  cv_model <- glmer(Ground_truth ~ BF_score + (BF_score | Serial_num_ref),
                    data = train_data,
                    family = binomial(link = "logit"),
                    control = glmerControl(optimizer = "bobyqa"))

  # Predict on the 1 left-out gun.
  # CRITICAL: re.form = NA tells the model "You have never seen this gun, just use the
baseline"
  dat01$cv_predicted_prob[dat01$Serial_num_ref == current_gun] <- predict(cv_model,
                                                          newdata = test_data,
                                                          type = "response",
                                                          re.form = NA)
}

cat("Cross-validation complete!\n")

end_time <- Sys.time()
time_taken <- end_time - start_time
print(time_taken)

#####################################
#####################################
```

```
# ROC Curve to evaluate model prediction performance
# This calculates the Area Under the Curve (AUC) for your out-of-sample data

cv_roc <- roc(response = dat01$Ground_truth,
              predictor = dat01$cv_predicted_prob)

as.numeric(auc(cv_roc))

# Plot the ROC curve
ggroc(cv_roc, color = "darkblue", linewidth = 1, legacy.axes = TRUE) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "gray") +
  labs(
    title = "HiPoint Data Cross Validation ROC Curve",
    x = "Specificity (False Positive Rate)",
    y = "Sensitivity (True Positive Rate)"
  ) +
  theme_minimal()

save.dir <- "C:/Users/natha/OneDrive/Desktop/ThesisData/Data Plots"
filename <- file.path(save.dir, paste("HiPoint GLMM Cross Validation ROC.png", sep=""))
ggsave(filename, width = 12, height = 8, dpi = 500, bg = "transparent")
```