Simulation of Venturi Tube Design for Column Flotation Using Computational Fluid Dynamics

Wan Wang

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
https://researchrepository.wvu.edu/etd/6909
Inclusion of 9 mm Firearm Type Using Quantitative Class Characteristics

Young Wang

Thesis submitted
to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science in
Forensic & Investigative Science

Keith Morris, Ph.D., Chair
James Carroll, M.Sc.
Casey Jelsema, Ph.D.

Department of Forensic & Investigative Science

Morgantown, West Virginia
2018

Keywords: Firearms, class characteristics, cartridge case, classification
Copyright 2018 Young Wang
ABSTRACT

Inclusion of 9 mm Firearm Type Using Quantitative Class Characteristics

Young Wang

The results of a data set of five models of 9 mm Luger caliber handguns suggest an algorithm such as the square of the Mahalanobis Distance as a step towards the creation of an electronic class characteristic database for cartridge cases to supplement the currently existing General Rifling Characteristic (GRC) database for bullets. The algorithm was validated using both hold-one out cross validation as well as on an entirely independent set of ground truth known cartridge cases which were blindly classified to mimic case work. A method for determining an objective threshold for inclusion onto a list for an investigator is proposed and the effects are illustrated using the blind set of hypothetical case work. The algorithm relied upon quantitative measurements of class characteristics. Three firearms per model and ten test fires per firearm were used to inform the algorithm of the mean and variance of the measurements taken. The test fires used to inform the algorithm were a combination of physical test fires and previous IBIS® entries where ground truth model was known. Measurements were taken of images which were retained in a digital cloud filing system where folders were used to organize test fires by the known donor model and firearm. Hold-one out cross validation was performed by withholding the measurements for a given test fire to serve as a questioned cartridge case, and computing the Mahalanobis Distance to each model. A threshold cut-off distance for inclusion onto a list which would be provided to an investigator was calculated based upon the results of the hold-one out cross validation and based upon the known-match Mahalanobis Distances following a central Chi-Square distribution. This threshold cut-off distance was used to guide decisions during the blind and independent classification of individual, physical cartridge cases. Each blind cartridge case was classified one at a time, independent of GRC bullet information, to mimic a crime scene where only one cartridge case is recovered. The blind set also included cartridge cases originating from models outside of the five considered by the algorithm or “database,” representative of real challenges experienced in case work.
## Contents

**Introduction**

1. Review of the Literature  
   1.1 Early Origins of Class Characteristic Use  
   1.2 Early Efforts to Classify Cartridge Cases  
   1.3 Early Toolmark Investigative Leads  
   1.4 Early Bullet Classification Efforts  
   1.5 Documentation and Identification Efforts  
   1.6 Early Warnings of False Exclusions  
   1.7 Early Origins of the Class Characteristic Database  
   1.8 Beginnings of Firearm Type Determination  
   1.9 Class Characteristic Use Since the 2000’s

2. Materials and Methods
   2.1 Materials
   2.2 Basis for measurements using ImageJ®
   2.3 Tests to determine adequate representation of variability present
   2.4 Construction of a model system of classification measurements

3. Results and Discussion
   3.1 Data Analysis
   3.2 Hold-one out cross validation
   3.3 Generation of Known Match and Known Non Match Density Plots
   3.4 Generation of Receiver-Operating-Characteristic Curve
   3.5 Calculation of AUC
   3.5 Potential use in case work via blind classification

4. Conclusion
   4.1 Practical limitations of this thesis work
   4.2 Directions for Future Research
   4.3 Final Conclusions
Introduction

In a shooting investigation, initially only cartridge cases and sometimes bullets are recovered. Very commonly, these cartridge cases are of the massively manufactured 9 mm Luger/Parabellum caliber. However, in the absence of a suspect firearm, no comparative identifications can be made. Thus, the initial investigation into a likely brand and perhaps model is a narrowing down process where certain types of firearms are placed on a candidate list for the investigator and others are not.

The greater detail a detective can describe in reconstructing the crime involving a specific type of firearm, the more likely a case will be solved quickly. Unfortunately, no presumptive test of this nature for recovered ammunition components is currently being in widespread use. The use of General Rifling Characteristic (GRC) databases for bullets and cartridge cases is not detailed enough to narrow down a brand of firearm. Similarly IBIS® systems connected to NIBIN will return many hits to cartridge cases which have previously been recovered from older crime scenes where the type of firearm is also unknown. If one were to search alternatively against previous test fired cartridge cases of previously seized known firearms, a similar problem exists if the firearm used was not already seized previously. The result is numerous hits of different makes and models are returned leaving the examiner with no ability to make sense of the IBIS® correlation ranking.

While older efforts to create detailed class characteristic files using rolodex cards have had success, examiners today currently do not engage in this time consuming process and as a result, the older files have become very outdated as new firearms are constantly entering the market. Thus, there exists a need for an electronic database containing class characteristics in greater detail and is quantitative in nature to facilitate easier information sharing and updating amongst examiners. It is hoped this study will provide a method which is minimally time consuming to update which will provide a large return on time investment in terms of investigative leads. In addition, it is hoped the method will assist with case pre-assessment so examiners will not even need to do a test fire in certain circumstances as well as help to increase the number of correct eliminations which otherwise would have been classified as inconclusive. The following literature survey follows a chronological approach to paint a story of the progression and current state of class characteristic use by firearms and toolmark examiners.
1. Review of the Literature

1.1 Early Origins of Class Characteristic Use

One of the earliest references to the term “class characteristic” within the field of firearm and toolmark examination was a chart credited to Alfred Biasotti in 1969. One such class characteristic for a fired bullet would be the number of lands and grooves which were engraved on it when the bullet engaged the rifling of the barrel it was fired through. This class characteristic count of lands and grooves has also been referred to as a type of general rifling characteristic (GRC).

In a later 1970 study, Meyers found the GRC count of lands and grooves may be determined when examining a bullet’s entrance hole, specifically the appearance of the ring of residue the bullet wiped onto the surface. There are situations when bullets were not recovered or bullets were recovered outside the victims’ bodies due to over penetration. In the absence of a bullet corresponding to a victim’s entrance wound, inference of class characteristic number of lands and grooves from the bullet wipe could narrow down the type of firearm which may have been used to shoot the victim and to exclude firearm barrels having different count of lands and grooves.

This inference is particularly useful in shooting reconstructions involving multiple victims and a mixture of multiple firearms and bullets with different land and groove counts. The bullet wipe present on a victim could be a starting point for which gun was used to shoot which particular victim, even before any suspect guns are seized for comparative test fires.

1.2 Early Efforts to Classify Cartridge Cases

In addition to the bullet, the cartridge case also exhibits class characteristics which may be used to narrow down the type of firearm which fired the case. During firing, the head of the cartridge case (for rifles and handguns) or shotshell (for shotguns) is pressed against a surface inside the firearm called the breech before it is extracted and ejected out via other tool surfaces within the firearm.

Six types of shotguns were first successfully classified by McKay by breech face, extractor, and ejector marks these firearms left on fired shotshells in 1971. However, cartridge case class characteristics have also been able to mislead examiners if observed deviations from what is expected are not communicated and accounted for. For example, it was noted in 1973 by Sojat and Hart how a semi-automatic pistol such as the Heckler & Koch HK4 produces different and unusual class characteristics on cartridge cases when it is converted by the user from firing the original manufacturer’s caliber (.380 Automatic Colt Pistol/.380 ACP) to fire a different caliber (.22 Long Rifle/.22 LR) of cartridge.
While the default caliber chambered for the HK4 was either .380 or .32 ACP, the user could swap caliber-specific barrel assemblies and magazines to fire the HK4 using one of three different center-fire cartridges: .25, .32, or .380 ACP. These center-fire cartridges all rely on the gun’s firing pin hitting the central primer in the cartridge case to ignite the firing process. In contrast, the user could also change the position of the extractor faceplate and firing pin to accommodate the rim-fire cartridge .22 LR.

These changes caused additional breech face marks to be produced in both .380 ACP and .22 LR cartridge cases and additional flute markings (resulting from the expansion of a cartridge case in the chamber) for .22 LR cartridge cases by Celovsky in 1973. In response to Sojat and Hart’s earlier 1973 study, these additional changes were communicated in the form of descriptions and sketches in the Association of Firearms and Toolmark Examiner (AFTE) journal by Celovsky which demonstrated the medium of collaboration and information sharing for class characteristics during this time period.

Similarly, it was possible for examiners to write letters to the editor of the AFTE journal and make requests for information about certain class characteristics which are by nature relevant to all examiners who would likely have similar questions. One such example was a request for a reference file for .45 caliber cartridge case class characteristics and for help in resolving apparent confusion over discrepancies in the expected rifling count in a particular .22 caliber Ruger firearm by Cockherham in 1973. In response to Cockherham, Sojat published the class characteristic file cards being used at the Miami-Dade County Crime Lab which Sojat had already presented at the AFTE conference earlier in the year.

1.3 Early Toolmark Investigative Leads

The concept of using class characteristics for generating investigative leads was not yet utilized as it had been for marks generated by tools besides firearms. For example, in rural areas where hay, straw, or grass is an everyday item on a farm, baling wire was used as a common method to hand package the hay for ease of storage. In addition, being so readily available, it was also used like duct tape as a go-to item for repairing objects such as fences. Therefore, the forensic utility of examining the baling wire for both class characteristics as well as potential individualizing characteristics was explored by Boudreau and Smith in 1975.

In cases involving burglary and theft, tool marks generated on padlocks by pipe wrenches could likewise produce both class and individualizing characteristics. In a particular case study, care was advised to be cautious about using class characteristics to exclude a type of pipe wrench. Although the practice of exclusion works well for investigative leads where the simplest explanation being the most probable correct explanation, additional and more detailed comparative examinations were advised in searching for matching striations. In this particular 1976 case study by Dixon, the striations were found on the side of the padlock and were previously missed during the initial examination which had excluded the type of pipe wrench which was later identified as the source of these striations.
1.4 Early Bullet Classification Efforts

Another issue related to exclusion is, for bullets, the type of bullet determines whether or not class characteristics such as left twist versus right twist rifling can be reliably determined since certain types of bullets fragment or deform more than others. Of particular concern are the unjacketed lead bullets commonly used for target shooting. Fortunately, unjacketed lead bullets are becoming less commonly encountered in casework and methods have been devised to attempt to determine the inclination of rifling from fragments of frontally-impacted lead bullets.

These methods rely on the principle of when a bullet impacts a surface, while the surface stops the bullet’s tip from continuing to rotate, the base or rear portion of the bullet continues to spin at the cannelure groove due to inertia. Thus, in 1977, Molnar found small amounts of resulting gas scarfing on the bullet’s land fragments to aid in determining both direction of twist but also estimate the width of the lands expected on the base portion of the bullet.\textsuperscript{11}

One of the earliest references to a need for cooperation and communication via a standardized system for classifying bullets was discussed by Welch and Wilhelm in 1977.\textsuperscript{12} In a no-gun case, the authors argued the examiner could assist the agency’s investigative efforts by narrowing down the number of possible firearms using the previously measured widths of lands and grooves on bullets engraved by the barrels of known firearms. These measurements could be stored and used for future reference in case work where bullets but not firearms are recovered.\textsuperscript{12}

1.5 Documentation and Identification Efforts

A means for easier communication and documentation given the limited technological resources available during this time period was first termed the “evidence receipt/activity report” by Paholke in 1978.\textsuperscript{13} These receipts and reports were for general tool marks observed on collected evidence at the scene and subsequent laboratory activity for crimes of breaking and entering. The format provided allowed for information to be gathered from the tool marks left on security devices and locks to narrow down the class of tool which could have been used in the burglary.

The documentation guidelines facilitated a healthy communication between those working at the scene and those in the laboratory.\textsuperscript{13} Similarly, there were also efforts during this time period to classify the common types of evidence in burglaries such as safes. The classification of safes and their door types into more specific categories is another relevant variable when considering the toolmark substrate.\textsuperscript{14}

Another variable of consideration is the reproducibility of tool marks for identification purposes. In a comparison between crimping pliers (tool applied to a substrate) and seals on a bank bag (substrate) where the question was “can these pliers be identified as having sealed this seal,” knowledge of both the reproducibility of the tool and manufacturing process was needed.
Watson toured the factory in 1978 to obtain an understanding of the dies used to cast the shape of the pliers.\textsuperscript{15}

If class characteristics are consistent within a class of tool marks, than they may be used as an effective screening measure for comparisons of question and known samples. If class characteristics of a questioned mark and those produced by a known tool seized from a suspect are in agreement, an identification may or may not be possible depending upon individualizing characteristics. If individualizing characteristics are truly individual and not carried over during manufacture from tool to tool (subclass characteristics), than they may be used to make an identification.

Since the standard for identification is in relation to the most similar tool mark produced by a different tool or best known non-match (BKNM), Watson considered two consecutively-manufactured sets of pliers. He noted reproducibility of both class and individualizing characteristics (with no sub-class carryover) amongst the pliers. In other words, Watson concluded it was possible to identify a seal as having been created by a specific pair of pliers.\textsuperscript{15}

1.6 Early Warnings of False Exclusions

A lack of reproducibility can be problematic with respect to false exclusions. For firing pin impressions on 22LR cartridge cases, it was found by Cochrane in 1981 that the dimensions of the firing pin impression can be quite different yet they could be identified as having been discharged by the same firearm using other markings observed upon more detailed examination.\textsuperscript{16} It is thought this may be due to the firing pin striking the rim of the cartridge with varying degrees of contact due to the potential looseness of the firing pin.

Another variable which was thought to be significant with regards to false exclusion via disagreements in class characteristics was pressure of the cartridge. For cartridges overloaded with powder and using bullets produced from a casting process, these bullets were found by Jordan and Looney in 1981 to be incorrectly eliminated as having been fired by the same gun.\textsuperscript{17}

This variability may be less of a problem if the goal is to separate classes, so long as the overlap between classes is either small or can be resolved through consideration of other markings. Unfortunately, for .22 LR, the low pressure of the cartridge results in an absence of many other markings which could be used to discriminate between classes.

Regardless of the caliber of the cartridge, type determination via class characteristics can be analogous to a presumptive test. Although useful for investigative purposes, exclusions or negative results should never truly be excluded if additional confirmatory tests are available. Thus, there exists a need to quantify the uncertainty with regards to false or incorrect exclusion or inclusion into a class.
1.7 Early Origins of the Class Characteristic Database

When starting a class characteristic file or database, it is of importance from a return on time investment standpoint to consider starting with guns most frequently encountered in casework which may be specific for a given area. This consideration was first highlighted by Miami-Dade County Crime Laboratory examiner Robert Hart in the 1981 publication of the laboratory’s GRC data for the firearms they most commonly encountered.\(^\text{18}\)

Distinguishing between class and individualizing characteristics can be difficult without an understanding of the manufacturing process of the tools responsible for creating the marks. Once again, three sets consecutively broached Craftsman tongue and groove pliers (commonly used to break off door knobs in burglaries) were studied by Cassidy in 1982 and it was found that pliers leave both individualizing marks distinct from class characteristic marks.\(^\text{19}\)

For cartridge case class characteristics, .22 LR cartridges were again re-visited for their firing pin impressions and a list of common firing pin styles along with example firearm models using each type of firing pin was published by Fraser in 1983.\(^\text{20}\) In addition, the position of ejector and extractor markings relative to the 12 o’clock position of the firearm and the types of extractors and ejectors were documented.\(^\text{20}\)

In 1984, Starrs argued one of the most important abilities of a firearms and toolmarks examiner is the ability to testify in a heated courtroom environment where one would need to explain what class characteristics are to a judge or jury.\(^\text{21}\) An understanding of the fundamentals of courtroom testimony is valuable since a system of type determination may be scrutinized in court. Thus, the ability to articulate the basis for the generated types must be considered from the initial development phases of the system.

However, in this thesis study, it is hypothesized it may be easier to convince a skeptic to accept the scientific soundness of determining a possible class of firearm in a no gun case than the soundness of individualization to the exclusion of all other firearms in a test fired gun case. This is because the concept behind databases is better understood.

The limitation of using class characteristics lies in the ability for the database component of the model to be updated with conditions as there will always be strange firearms which either deviate from expected class characteristics or produce class characteristics which were previously not accounted for in the database. In 1984, Cassidy discussed a case of a submitted .25 ACP Norton TP-70 pistol and cartridge cases where the primer displayed an unusual breech face impression.\(^\text{22}\) Subsequent test fired cartridge cases from the TP-70 demonstrated reproducibility of this unusual breech face impression and Cassidy discussed the cause for these impressions which may need to be considered in the construction of a class characteristic classification model.\(^\text{22}\)

However, even considering the limitations and practical difficulties of creating a class characteristic model or database, the author believes there is potential since the basis for classification is more easily understood by the layperson or non-toolmark examiner than the basis for individualization or identification, since the concept behind databases is already
understood. By 1984, Biasotti and Murdock, and the Association of Firearms and Toolmark Examiners (AFTE) as a whole, had already been trying for decades to explain the foundational validity of using a combination of randomly acquired characteristics on top of class characteristics to reach the conclusion of identification to the reasonable exclusion of all other tools for a given tool mark. Despite empirical and theoretical arguments, there still exists a portion of the scientific community who regards their logic as circular reasoning or as more of an art form than a science. In contrast, the use of databases to determine a possible class origin for investigative leads such as for Original Equipment Manufacturer (OEM) paints using the Paint Data Query (PDQ) database meets relatively little opposition, since analytical chemists have been better able to articulate the basis for manufacturing differences as it relates to conclusions made.

The concept of a database for determining the potential class of tool was first alluded to as a collection of photographs for general tool marks created by cutting tools on various surfaces composed of lead, aluminum, and brass by Cochrane in 1985. The photographs were to serve as a reference in cases where no tool was recovered or if the examiner’s initial instinct was the tool recovered was unlikely to have produced the mark in question. The ability to reference photographs of previous marks known to have been produced by a certain class of tool may serve as a form of case pre-assessment, perhaps by an independent supervisor, as to whether time is even worth spending to use the submitted tool to create a new mark for comparison purposes.

In addition to considering the material of the surface which a mark was left on, one also needs to be aware of marks produced by tools during manufacture, specifically defects, which were not created by the evidence tool in question. An example for consideration are the surfaces of plastic polyester and polyethylene bags which have been sealed by a heat sealer tool. In 1986, Stone reported knowledge of the plastic extrusion manufacturing process and the marks produced as critical for comparison purposes when considering the class characteristic marks left by the heat sealers themselves.

1.8 Beginnings of Firearm Type Determination

If the tool in consideration is a firearm, one must consider the many tools within a firearm which are the source of marks on the surface of a bullet and cartridge case. Evans, in 1986, considered one particular firearm, the Steyr Armee-Universal-Gewehr (AUG) or “universal army rifle” in terms of its gas operated, rotating bolt action as the mechanism for 5.56 x 45 mm caliber ammunition cartridges to be loaded, discharged, and ejected. These processes generated class characteristics on the discharged ammunition components corresponding to the mechanisms utilized by the gas piston-driven AUG which include a plastic hammer, claw extractor, and spring-loaded ejector.

In 1986, the beginnings of true type determination of a firearm manufacturer from multiple candidate manufacturers using class characteristics originated from Nennstiel, of the German
Federal Police Crime laboratory, and built upon earlier, more obscure efforts by a 1930 publication of cartridge case photographs by Mezger, Hess, and Haslacher for various firearms used around the time of World War I. Nennstiel presented his reference collection to AFTE and demonstrated the reproducibility of markings on both cartridge cases and bullets which corresponded to a particular class or type of firearm. Nennstiel argued for the merits of using computerized methods to assist the firearms and toolmark examiner but emphasized the method revolved around the experience and knowledge of the examiner and how the method would not be fully automatic as it is still up to the examiner to make judgment calls on the presence or absence of features.

One should note class characteristics are not necessarily intended by the manufacturer in the sense the end resulting marks or features are not necessarily being considered when the manufacturer designs the tool. For example, for cartridge cases, the number of case cannelures or grooves had been found to vary within one batch or box of Federal brand .22 caliber cartridges. Although the cartridges within the box of ammunition could have been reasonably assumed to be manufactured using the same machinery, some cartridge cases ended up having two cannelures while others ended up having three, as reported by Haag in 1987. Thus, in a situation when Manufacturer A is known to use two cannelures consistently and Manufacturer B uses two or three, the observation of three cannelures suggests A is not the source.

In contrast, there are class characteristics which the manufacturer intended to include on a particular surface, whether for aesthetic design or technical performance purposes. For example, the knurling process used by Hornandy on certain lead pistol caliber bullets may be used to distinguish these bullets from other bullets of the same broader class characteristic caliber, as reported by Haag in 1990. Microscopically, the knurled surface and coating is a unique form of trace evidence when transferred but also a unique pattern of staggered rows of tiny, uniformly distributed diamond “impressions” which can persist on the bullet even post-impact.

In 1990, the AFTE glossary was updated by Murdock and colleagues to distinguish class characteristics from “subclass” characteristics. Subclass characteristics are a more restricted form of class characteristics which are randomly acquired during manufacture but are not unique to one particular firearm, but many firearms within a batch. More detailed classification using subclass characteristics does have challenges. For example, ejector class characteristic shapes of wedge as opposed to arc or rectangular may be used for eliminations particularly upon examining the actual ejector in a seized firearm. However, to measure class shapes in terms of quantitative angles and dimensions could be an example of subclass characteristics which carry over within a batch of firearms. Using subclass characteristics to make eliminations would require knowledge of manufacturing tolerances, part sourcing, and other manufacturing processes.

Unlike bullets, it has been noted that identification of a particular class of shotguns from the recovered projectile wads which contained the shot/pellets or slug is generally not feasible except under certain circumstances. One of these special circumstances studied in the literature
was the C-Lect-Choke designed to constrict the spread of pellets in the barrel of the 20 gauge Mossberg 385K shotgun by Heuske in 1990.\textsuperscript{31} It was found the surfaces and edges of the Winchester branded wads, also known as shotcups, fired in the Mossberg 385K were marked by both the choke and the ports of the muzzle.

Thus, in a case where the suspect firearm was the Mossberg 385K and three wads were recovered from the victim’s body which lacked these class characteristic markings, an elimination conclusion was reached which otherwise would have been an inconclusive had the more-detailed class characteristic features not been considered.\textsuperscript{31}

Keeping current with more detailed class characteristics specific to a particular model and make of firearm as it entered the market was clearly evident in the early 1990s. Nordhoff published an article dedicated to a new model of .25 ACP Lorcin L25 pistol.\textsuperscript{32} Lorcin had been established less than a year prior and was marketing inexpensive firearms. The designs of the firearms were similar to corresponding models made earlier by another earlier manufacturer, Raven Arms. Specifically, the Lorcin L25 and the Raven MP-25 were both 7-shot pocket pistols designed to be easily concealed. Nordhoff published the similarities and important differences in the class characteristics of these similar firearms which would likely become so called “Saturday night special” firearms distributed through pawn shops and eventually used in the commission of a crime.\textsuperscript{32}

In response to Nordhoff\textsuperscript{32}, Thompson published a similar article in 1991 covering class characteristics, for the 10 mm Wyoming Arms Parker.\textsuperscript{33} One reason for a 10 mm handgun to make its way in the commission of a crime can be seen given the historic move of the FBI towards .40 Smith & Wesson (S&W) in 1990. The .40 S&W as the replacement for the higher recoil 10 mm cartridge and delivered the desired terminal ballistics in a cartridge which could be chambered in smaller frame handguns.

Later in 1991, Gieszl argued the need for examiners to update their class characteristic files for the new Model 83 and 85 handguns chambered in .380 ACP manufactured by Argentine arms maker Bersa.\textsuperscript{34} Once again, the 83 and 85 were examples of affordable and concealable firearms which would be popular and potentially end up being used in crimes. The significance of selecting firearms based on market trends and how often they likely will be encountered in casework provides the highest return on investment for an examiner. Updating the laboratory’s reference files of the detailed class characteristics produced by the firearm allows for easy referencing in future case work involving no-gun cases.

In 1991, Shem and Rogers noted another use of class characteristics in no-gun cases is the study of the unique gun powder residue pattern produced and found on a victim shot by a firearm which had its muzzle ported to reduce the shooter’s felt recoil.\textsuperscript{35} It was found in this particular case study the suspect’s firearm was modified with this muzzle break by a company named Mag-na-port\textsuperscript{®}.\textsuperscript{36} Upon consulting the pathologist who examined the homicide victim, it was found the victim had a gun powder residue pattern of the class characteristic pattern produced by the Mag-na-port\textsuperscript{®} muzzle break. The article assists investigators, pathologists, and
examiners by drawing awareness of this class characteristic gun powder residue pattern as it indicates the gun which is to be sought after must have a ported barrel.

In 1992, Carr of the Miami-Dade County Crime Laboratory re-examined the class characteristics of the Lorcin L25 for the unique way it marked the shoulder of bullets due to the leade, the gap between the firearm’s chamber and the barrel.\(^\text{37}\) Again, in a no-gun case, knowledge of these class characteristics would allow the examiner to narrow down the make and model of firearm.

Perhaps the greatest advancement for class characteristic prediction of firearms during this time was Kennington’s “The Matrix; 9 mm Parabellum: An Empirical Study of Type Determination.” Kennington, also of the Miami-Dade County Crime Laboratory, and his laboratory had been using physical cards for class characteristics since the 1950s. The cards were organized into a matrix by manufacturer and type of bullet, chamber marks, extractor mark shape, ejector mark, firing pin impression, and breech face marks. In some instances, Kennington was able to narrow the possibility of manufacturers down to one with the caveat there could exist a model of pistol which was not in the database.\(^\text{38}\)

Practically, the police detective who is armed with the hypothesized manufacturer at the time of interrogating a suspect is more likely to get an investigative lead which could then lead to other evidence such as the actual firearm itself to be used for test firing as well as additional incriminating evidence such as the buried body which might be needed to secure a guilty verdict if a case were to go to trial. However, the format of using cards made updating the database inconvenient as new firearms entered the market and there exists a need for an electronic version which still captures the essence of the original method but in a more quantitative format for easier searching and objectivity.

For court purposes, as opposed to investigative, a 1993 article by Cassidy highlighted the need for extra caution in acknowledging the limitation of one’s conclusions. Specifically, a court challenge had led to an earlier caliber conclusion to be incorrect since the 5L twisted crime scene bullet had class characteristics consistent with both .303 British caliber Enfield rifle as well as the .30-06 Springfield caliber Enfield rifle.\(^\text{39}\)

False identifications or false positive matches must be minimized for court and this concern was again expressed by Komar and Scala in 1993 with respect to tool marks generated by Chinese bolt cutters.\(^\text{40}\) Mistaking class characteristics for individual characteristics resulted in false identifications and once again, knowledge of the tool and manufacturing processes are key for drawing the line between class and individual characteristics.

In 1996, Kennington published an article to draw attention to a similar body of work regarding the .380 ACP caliber. Two brands of firearms discussed were the popular Lorcin and the Bryco, both of which had breech face marks described qualitatively as “billowed arcs,” an exaggerated version of the more common “arches” breech face marks. Kennington presented a case study of recognizing a cartridge case as having been ejected in a firearm which used a firing pin as the ejector which led to considering five possible makes of the gun.\(^\text{41}\)
There is a preference amongst practitioners for the features used for type determination to be independent of the substrate or ammunition. In 1997, for bolt cutters, Cassidy emphasized the importance of using a heat-treated aluminum alloy for test cuts since a softer lead substrate would lead to false eliminations when compared to the evidence marks on fence wire and a padlock shackle.42

Examiners looking at the same cartridge cases may see new marks which were previously not seen by other examiners as being useful for classification. In 1997, Silverwater and Shoshani published an article on 5.56 x 45 mm caliber cartridge cases with marks produced by the barrel extension of the M-16 rifle. These barrel extension lug marks were found to have both class as well as individual characteristics useful for comparative purposes.43

Kennnnington published another article in 1999 to attempt to again draw greater attention to the importance for examiners to use class characteristics in a way to discriminate firearms from one another, noting the importance of differentiating marks on a cartridge case prior to and after discharge.44

1.9 Class Characteristic Use Since the 2000’s

Greater attention to the problem of there being class characteristic marks not unique to the class of firearm but to the class of cartridge case manufacturer was addressed by Tam in 2001 through the use of a visual database kept over a period of three years.45 Tam considered the possibility of both false identifications as well as false eliminations if these ammunition manufacturer non-firing marks were mistaken for individualizing characteristics.45

Type determination was once again a focus in a 2002 study of three brands of 12 gauge shotguns (Breda, Remington, and Beretta) by Shoshani and colleagues of the Israeli National Police laboratory with respect to ejector cut-out marks on fired shotshell cartridges.46 These marks were found to have both class characteristics as well as individualizing characteristics.46 Marks produced by various ejectors in Smith & Wesson (S&W) firearms was also studied by Thompson and Wyant in 2002 by visiting the manufacturer’s workshop.47 The use of a simple eye loupe to examine 28 OEM ejectors revealed a combination of class and more narrow sub-class family characteristics.

Perhaps the last major mention of earlier type determination efforts via class characteristics in the literature was a book review of Kennington’s previously discussed works on 9 mm and .380 ACP by Dutton, of the Tasmania Police, in 2003.48 According to Dutton, Kennington’s decade old work contained not only a great deal of technical information but was still relevant to the current time. Specifically, Dutton considered the potential workflow of how shooting crimes should be investigated and how laboratories encountering 9 mm Luger and .380 ACP on a regular basis ought to implement the method to expedite investigations.48

Class characteristics for type determination became less of a concern in the literature than the need to differentiate between subclass characteristics and individualizing characteristics and the
knowledge of the manufacturing process required to do so. Coffman, in 2003, studied ten
Remington 870 shotgun breech bolts from a single production run to look for subclass carryover
and whether or not this subclass carryover could be traced back to the manufacturing process
of this batch. Specifically, Coffman studied the bolt faces for similarities to the tool surfaces
used in a particular manufacturing process (Computer Numerical Controlled).

The persistence of class and individual characteristics left on cartridge cases as a firearm is
used over time was studied by Vinci and colleagues in 2005. Vinci and colleagues fired 2500
cartridges through a .45 caliber pistol and collected one cartridge case after every 100 fired. It
was found the mechanical wear on the breechblock, ejector, extractor, and firing pin was not
significant enough to prevent any of the 25 collected cartridge cases from being identified to the
firearm despite there being some differences between the cartridge cases. Bailey and
colleagues came to the same conclusion in 2006 for 1000 rounds of .177 (4.5 mm) caliber air
pellets fired through an air rifle, where every 25th pellet was sampled for rifling comparison.
Once again, although rifling wear was responsible for differences in the pellets, identifications
could still be made to the rifle due to the persistence of both class and individual
characteristics.

The use of class characteristics to eliminate must always consider the possibility of there being
circumstances which have not been accounted for, such as the swapping of barrels. Although
uncommon enough to the point where examiners use the terms firearms and barrels
interchangeably with respect to identification purposes, Milam, in 2007, drew attention to the
custom aftermarket stainless steel barrels manufactured by Bar-sto Precision Machine for Glock
pistols which shared the general rifling class characteristics of Smith & Wesson pistols.

One of the most recent advancements in the use of class characteristics to determine the
potential type or brand of firearm was done by Bati of the Istanbul Crime Laboratory of Police in
2007. Bati used a quantitative method to measure the width of the firing pin impression (FPI) in
a 7.62 x 39 mm cartridge case in order to link to the brand of AK 47 which fired it and argued for
the reproducibility of the method in differentiating Chinese and Polish manufactured classes.
Of additional significance was the development of a modified FP acquisition protocol
(specifically focused on the smaller firing pin imprint as opposed to the entire crater of the firing
pin impression) for the computer-based Integrated Ballistics Identification System (IBIS®)
system which improved the performance of the correlation.

For the AK-47, the use of individualizing characteristics to identify a particular firearm is difficult
due to a lack of imperfections in the firearm’s parts and a lack of transfer due to variable
hardnesses of cartridge cases made of steel as opposed to softer brass, as well as varying
primer hardness. Thus, known matches are often compared and classified as inconclusive and
it is difficult to determine the number of guns used at a shooting scene. For a massive scene
with many cartridge cases, numerous individual entries are made into the open case file. In
Turkey, the popularity of the AK-47 has resulted in this firearm accounting for about 70% of the
open case file of crime scene cartridge cases.
Bati hypothesized the firing pin impression would allow the determination of firing pin width and he segmented the firing pin impression into three distinct regions. The first region was where the firing pin tip contacted the primer, the middle region was what corresponded to the firing pin diameter, and the third region which contains the first two regions plus the curved area around the firing pin impression bordering on the undamaged portion of the primer. Unfortunately, the IBIS® FP acquisition protocol (FPAP) measures the third region, which is largely dependent upon variations in primer hardness, powder and pressure of the cartridge, as well as strength of the AK-47’s firing pin spring and other mechanical factors. The second region can be manually selected by the operator in IBIS® or even measured manually and stored into a database.

When examining two test fired cartridge cases from each of five Polish AK-47s and five Chinese AK-47s using the IBIS® FPAP, the measurements were highly variable to the point where no class separation was seen. Using the modified FPAP on the third test fired cartridge case resulted in class separation where the diameter of the firing pin impression was constant within each class, at a mean of 1816 vs. 1920 microns respectively.

For a single cartridge case, correlation scores were higher between two images both using the modified FPAP than the IBIS® FPAP. Again while two images using the modified method both returned the same firing pin diameter, the same operator using the IBIS® FPAP for the same cartridge case returned a difference of 11 microns between images. Correlation scores were also negatively affected by small differences in lighting.

Five cartridge cases were test fired from a single AK-47 and each entered as different scene cartridges into IBIS® using the IBIS® FPAP and another five using the modified FPAP. One of the five IBIS® FPAP cartridge cases was selected as the query to be searched against over 30,000 cartridge cases of the same 7.62 x 39 mm caliber where nearly all were imaged using the IBIS® FPAP. As a result, only one of the other four test fired IBIS® FPAP cartridge cases even made it to the ranking list at position 138 and its image even had the same measured firing pin diameter, a fairly uncommon occurrence. Meanwhile, four out of the five imaged using the modified FPAP at least showed up in the top 50.

In contrast, first rank recognition was seen for the modified FPAP cartridge case image query and only one false positive was observed within the top five ranked images. Finally, four cartridge cases from two actual crime scenes identified as having been fired by the same AK-47 via firing pin impression were used to compare the IBIS® FPAP with the modified and the modified protocol once again performed better in a query where the other three cartridge case images ranked at positions 5, 78, and 126.

In 2008, another more extensive persistence study was done by Gouwe et al. for class and individual characteristics with respect to breech face and firing pin marks with 10,000 cartridge cases fired from a single .40 S&W caliber Glock 22 pistol. Every 10th cartridge case was sampled and it was found even the 10,000th cartridge case could be identified to the first cartridge case, cementing the validity and utility of the discipline in spite of tool surface wear.

The validity and utility of the discipline in recent times with advancements in manufacturing and
science has been viewed as the ability to individualize a firearm from other firearms belonging to the same class.

In 2009, Sarybey and Hannam examined 10 consecutively manufactured firearms of the same model and make for six different manufacturers. Each of these sixty firearms was used to fire 10 cartridge cases for a total of 600 cartridge cases. The cartridge cases were compared via their firing pin, breech face, and ejector marks and while agreement of class characteristics within each set of ten consecutive serial number firearms was observed as expected, each firearm did produce individual characteristics which would allow for identification.

One should note while identification is certainly an important research topic, additional attention should be given to determining simply the class or manufacturer of firearm. A detailed class characteristics database in electronic format to facilitate easier information sharing amongst examiners is long overdue as GRC databases are often insufficient for no-gun cases. Oberg published an article in 2010 on a case with a 9 mm caliber right twist bullet with seven lands and grooves and a cartridge case with hemispherical firing pin impression, two extractor marks, and one ejector mark. The initial search of a GRC database returned zero hits and it was additional inquiry using the online AFTE forum which suggested the firearm may have been made by SCCY. Oberg then had to reach out to the manufacturer SCCY in order to determine the type of firearm as being a model CPX-1, the only model SCCY makes, with class characteristic 4 o’clock extractor and 7-8 o’clock ejector.

In 2010, Monturo noted investigative value in understanding the class and individual characteristics produced by drills which are used to modify firearms, destroy serial numbers, and other general machining processes such as construction of pipe bombs. The recognition of a mark on a substrate surface as having likely been produced by a drill is an important investigative lead. For example, investigators would search a suspect’s home for not only a drill but also for trace amounts of shavings which could be later determined as having potentially originated from the same substrate material.

Huertas published a warning in 2010 to examiners to be aware of mislabeled class characteristic caliber head stamps as having the potential to be incorrect and the need to use a caliper to measure the diameter of cartridge cases. In his crime scene, a cartridge case with a .40 S&W caliber head stamp recovered from a crime scene was actually a 9 mm.

One of the most recent highlights in the literature on using class characteristics to differentiate types of firearms was in 2012 by Kosachevsky and Bokobza, of the Israel National Police, and focused on differentiating the Tavor TAR-21 rifle from the older M-16 rifle. Kosachevsky and Bokobza considered firing pin impression, ejection port, ejector, extractor, chamber, lug, cutout, and neck marks and found the the neck marks on 5.56 x 45 mm cartridge cases fired by the TAR-21 as being the key class characteristic to differentiate the two rifles.

In 2013, Maxwell examined the effects of humidity on tool marks made on wood surfaces commonly encountered in burglaries and concluded marks made by the same tool may be unable to be identified due to moisture altering mark shape over a two week period.
In 2013, Holliday and Rankin, of the Center of Forensic Investigation in the UK, recognized how firearms investigations involving air rifles and pellets are greatly slowed down since the manufacturer and model of ammunition can not even be determined due to the nature of projectile deformation upon impact. Thus, Holliday and Rankin created a reference catalog of images depicting the deformation patterns for various known brands and models of pellets and was able to validate the method using blind classification of ground truth knowns. The class characteristic breech face recess and firing pin aperture shear for three 9 mm Walther P99 pistols was examined by the University of Science, Malaysia and was found to have excellent persistence, reproducibility, in addition to being suitable for identification via individualizing characteristics by Yong and colleagues in 2014.

From an investigative standpoint, class characteristics may allow for the recognition of cartridge cases as having been reloaded and in 2014, Bruce found the primers of 9 mm cartridges recycled from the Australian Federal Police had unique manufacturing marks as well as marks indicative of reloading. Also in 2014, Bokobza and colleagues studied class characteristics of cartridge cases fired by homemade submachine guns and the ability for IBIS® to differentiate these from conventionally manufactured firearms such as the Uzi® and Carl Gustav® submachine guns.

In 2016, Siso and colleagues, also of the Israeli National Police, re-examined the class characteristic differences between the Tavor and M-16 rifles with respect to extractor marks. Siso and colleagues discussed the increased prevalence of the use of these rifles in crime, how only cartridge cases are recovered in many instances, and the investigative need to determine the type used, number of different rifles involved, and any connections to past open case files.

While roughly 95% of the time examiners can easily distinguish between the M-16 and Tavor using primarily neck marks as previously discussed by Kosachevsky and Bokobza in 2012, around 5% of the time the cartridge cases are so badly deformed only the protected extractor marks remain intact.

Siso and colleagues had recently received a single cartridge case from two different criminal cases where it was difficult to determine whether the firearm used was an M-16 or a Tavor. Upon consulting the previous open case file of cartridge cases known to have been from an M-16 or a Tavor, Siso and colleagues noted while both had “banana” shaped extractor marks, the ends of the “banana” are much more curved or rounded for M-16 extractor marks. For the Tavor, the edge of the extractor mark at each end of the “banana” is a straight line traveling to the rim of the cartridge case and has a more pointed appearance.

Upon consideration of these data, perhaps this could best be quantified by measuring the angle formed between two lines: one line tangential to the end of the extractor mark and the other line tangential to the rim of the cartridge case. Using this approach, the angles for the Tavor extractor marks are much smaller and would cluster around roughly 30 degrees while the M-16 marks would have an angle of roughly 60 degrees.
Regardless, Siso and colleagues proceeded to validate their subjective qualitative method.\textsuperscript{66} Five cartridge cases were test fired in a total of 20 rifles (10 M-16 and 10 Tavor) for a total of 100 cartridge cases and 20 plastic bags corresponding to each rifle. Seven examiners who were unaware of the different class characteristic extractor mark differences were selected to classify unknowns.

The base, neck, and shoulders of all cartridge cases were then intentionally damaged to simulate case work and to force examiners to rely solely upon the extractor marks. One cartridge case was randomly selected from each of the 20 bags and each of the seven examiners was asked if they could determine the type and if yes, M-16 or Tavor. In other words, 140 determinations were requested.

As expected, the examiners only gave 17 determinations and considered the remaining 123 to be inconclusive. Interestingly, for the 17 determinations, while an accuracy of 50% would have corresponded to a coin toss, only 5 were actually correct for an accuracy of roughly 29%. After responses were collected, Siso and colleagues gave them a tutorial using extractor shape and returned the same cartridge cases to each examiner.\textsuperscript{66} Now, the examiners gave 138 determinations and considered only 2 to be inconclusive. Of the 138 determinations, 133 were correct for an accuracy of roughly 96%. While individual examiners were able to introduce error, Siso and colleagues concluded when other marks which may be present such as neck marks in addition to extractor cut out, barrel extension lug, ejector and ejector cut-out, firing pin impression, and breech face marks, correct type determination is quite promising.\textsuperscript{66}

In 2017, Warren and Pitts placed various Glock and Smith & Wesson Sigma 9 mm cartridge case images in a 4X4 grid in a Microsoft PowerPoint® slide in order to quantify the location of an ejector mark in a meaningful manner with respect to distance from the center of the cartridge.\textsuperscript{67} Potential limitations of their methods include the imprecision relative to a software such as ImageJ® (a free software distributed by the National Institute of Health), the fact ejector marks are often not present and the center of mass of irregularly shaped ejector marks is more difficult to determine as compared to a firing pin impression which is always present and fairly symmetric, and finally, the relative rarity of the two class problem presented as opposed to the common multi-class problem which does not involve a Glock or Sigma type firearm with well known rectangular firing pin apertures. It is interesting to note how Warren and Sheets published another article in the same year which explored the possibility of creating an algorithm to classify ejector mark shapes.\textsuperscript{68} Again, the same limitations apply related to the infrequency of encountering ejector marks and the limited utility of the two class problem as a one-to-one comparison of ejector mark shapes may be more desirable than using an algorithm which may not be able to handle a multi-class problem as well as the human eye. An algorithm could be useful for considering other features which can be used to rank firearms by likelihood of being the correct one which discharged a particular or given set of questioned cartridge cases and the determination would ultimately be made by an examiner via comparison to ground truth known images filed away in a database.
The algorithm or classifier considered for this thesis study is Quadratic Discriminant Analysis (QDA), otherwise known as the square of the Mahalanobis distance, and is defined as a classification metric based on statistical distance to a certain class (i). A certain class is defined by the mean \( (m_i) \) of the class, as well as how the measured features vary and co-vary with respect to each other (variance-covariance matrix) for the particular class. The variance-covariance matrix of the class, or simply the covariance matrix \( (S_i) \) for short, can also be expressed as an inverse \( (S_i^{-1}) \).

QDA builds upon the univariate statistical distance to a class mean which is simply the distance or difference between a questioned sample \( x \) and class mean \( m_i \), normalized or standardized (since units are canceled out) by dividing by the standard deviation of the class. A small difference between a questioned sample and the class mean coupled to a large class standard deviation or class spread results in a short distance between the questioned sample and the class. Thus, the equation for the square of multivariate Mahalanobis (QDA) distance \( (d_i) \) is, as defined by Schott:

\[
d_i = (x - m_i)^T S_i^{-1} (x - m_i)
\]

The order of the terms in Eq. (1) and the use of the transpose (T) are used to allow conformable vector-matrix multiplication consistent with the rules governing the linear algebra required to handle multiple measured variables. The end result, like a univariate statistical distance to the mean, is a single unitless number, otherwise thought of as a 1X1 scalar value.

The final \( d_i \) from Equation 1 is a chi-square distributed variable with degrees of freedom (df) equal to the number of measured observable features. As the number of features increases, so does df and the resulting chi-square distribution curve adapts by shifting towards the right. The corresponding probability density function (pdf) curve for a given df is integrated from the chi-square value \( (d_i) \) on forward to result in smaller distances corresponding to larger probabilities. The resulting area under the curve is the probability of having a QDA distance equal to or greater than this \( d_i \).

With respect to the influence of sample size on QDA, if an individual does not have enough samples, samples tend to cluster in an overly-elliptical manner where the smaller axis or eigen value is shorter and leads to misclassifications since this makes it harder for a sample to get classified into a class. Mathematically, smaller eigen values drive up or increase the distance metric by reducing the spread of the correct class which is equivalent to how dividing by a smaller denominator results in a larger end value. One way to get around the problem of not having enough samples to adequately train an algorithm is to use a single combined or pooled covariance matrix, again and again, for all classes. In other words, \( S_i \) in Equation 1 is always substituted with a constant \( S_{pooled} \). This is known as linear discriminant analysis (LDA). To
calculate $S_{pooled}$ for five classes, one multiplies all of the class covariance matrices ($S_i$) by the number of samples in each respective class, than divides by five less than the total number of samples. However, LDA is not without its own set of limitations, as noted by Brereton\textsuperscript{70}. If we assume variance is proportional to information, information is lost when we average out the $S_i$'s. Thus, LDA is only appropriate if the class data clusters similarly.

LDA, along with regularized discriminant analysis (RDA), a hybrid approach between LDA and QDA, reduces the interpretability of the data since the algorithm’s output is no longer directly tied to a single class’s mean and covariance. Similarly, information is lost from an examiner’s interpretability perspective when using methods such as Canonical Discriminant Analysis to reduce the five, tangible and visually observable features down to two abstract feature axes which are mathematical manipulations or combinations of the five. QDA stands out as an algorithm which has the ability to actually be easily explainable to the layperson since it is simple and transparent.
2. Materials and methods

2.1. Materials
The 9 mm Luger cartridge cases stored as entries within the 2D Heritage IBIS® database at West Virginia University’s Department of Forensic & Investigative Science had unique identifying prefixes and numerical suffixes corresponding to various test fires of different modern 9 mm firearms (where serial numbers were known). These entries or primer images with marked IBIS® determined firing pin center and breech face area or primer centers were exported as screenshots into Microsoft Paint® and saved as .bmp files, uploaded to organized folders within Google Drive®. These images could subsequently be downloaded and opened using ImageJ® for measurements of x-y pixel coordinates while Google Drive Sheets® was used to record measurements as spreadsheets which could then be imported into R² and RStudio³.

2.2. Basis for measurements using ImageJ®
IBIS® determines centers via a proprietary process which considers the full impression or circular crater created by the firing pin and the impressed circular areas which made contact against the breech-face. Likewise, via a similar process, IBIS® also determines the diameter of the full firing pin impression and breech face areas and this, in conjunction to the known diameter of the primer, served as a ground truth known in order to serve as a control to validate subsequent measurements made using the human eye and ImageJ®. For example, it was visually determined 468 pixels corresponds to the known diameter of a small pistol primer, 4.45 mm, as illustrated in Table 1. Thus, 1 pixel equals approximately 0.0095 mm, or approximately 3.7 x 10⁻⁴ inches, equivalent to approximately 2700 pixels per inch (ppi). This resolution is perfectly good enough for discriminating distances as user variability or misplacement by a pixel is trivial. Thus, if IBIS® claims a cartridge case has a full firing pin impression and breech face diameter of approximately 0.056 and 0.153 inches, respectively, this can be compared to the user-determined 0.060 and 0.150 inches based on the previously discussed scale or conversion factor utilizing the primer diameter, as seen in Table 2. In this example, assuming IBIS® gives only true measurements, the relative errors are -1.96% and 7.14% for the breech face areas and firing pin impressions, respectively.

Table 1: Basis for measurements illustrated for a given IBIS® image

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top of Primer</td>
<td>428</td>
<td>162</td>
</tr>
<tr>
<td>Center of Primer</td>
<td>428</td>
<td>396</td>
</tr>
<tr>
<td>Primer Radius</td>
<td>234</td>
<td></td>
</tr>
<tr>
<td>Primer Diameter</td>
<td>468</td>
<td>pixels</td>
</tr>
<tr>
<td>Known Diameter</td>
<td>4.45</td>
<td>mm</td>
</tr>
</tbody>
</table>

1 pixel = ~0.0095 mm
Table 2: Basis for measurements illustrated for a given IBIS® breechface area

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center of Breechface Area</td>
<td>428</td>
<td>396</td>
</tr>
<tr>
<td>Bottom of Breechface Area</td>
<td>428</td>
<td>597</td>
</tr>
<tr>
<td>Radius</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>Diameter</td>
<td>402 pixels</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.82 mm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.15 in</td>
<td></td>
</tr>
<tr>
<td>Known Diameter</td>
<td>0.153 in</td>
<td></td>
</tr>
<tr>
<td>Relative Error</td>
<td>-1.96 %</td>
<td></td>
</tr>
</tbody>
</table>

2.3. Tests to determine adequate representation of variability present

The next goal was to find whether or not a working sample size of 10 cartridge cases was enough to characterize the variability present within a single firearm. Since the most variable feature was the distance between the center of the primer and the center of the firing pin impression, this was the feature which served as the basis of subsequent tests. Two Ruger SR9s, serial numbers X96719 and X69363, as well as two Springfield XD9s, serial numbers X17802 and X17841, had 99 or 100 test fires previously entered into IBIS®. The first test was to create 91 sets of ten test fires i.e. 1 to 10, 2 to 11,...,99 to 100 and to then compare each set of distances to the complete set of 99 or 100 distances.

Assuming distances were normally distributed and using a t-test at the 95% confidence level, the worst performing individual firearm was one of the Ruger SR9s where 13 sets out of 91 sets (of 10 test fires) were found to have a significant difference with the complete set. Interestingly, creating sets of 15 test fires resulted in an even worse performance of 18 sets being incorrectly found to have a significant difference with the complete set of 100. The other SR9 only had 5 and the two XD9s only had 6 and 2 out of 91 sets of 10 test fires found to have a significant difference with the complete 100 test fires.

To obtain a better understanding of the data, a simulation where 10 cartridge cases were randomly selected, using the replicate() and sample() functions within R, to be compared to the complete set was repeated 1000 times. The success rates for not finding any statistical difference at the 95% confidence level (small p-values less than 0.05), assuming normal distributions and using a t-test of groups where variances are not assumed to be the same, ranged from 92 to 97%. However, since this simulation involved the possibility of there being outliers which would skew the data when replaced back into the complete set in such a way it could result in a much higher proportion of not finding any statistical differences between sampled and complete data sets, another simulation to sample without replacement was performed. Thus, to avoid “double counting,” the simulation was repeated using the sample() function to sample all 99 or 100 distances, effectively randomizing the order and allowing comparisons to be made between the first 10 random samples and the remaining 89 or 90 samples. The success rates for not finding any statistical difference ranged from 90 to 96%, as
seen in Table 3. Thus, a sample size of 10 appears to be justified and this led to the building of a model based upon 10 test fires for three guns per class of firearm. It is noted the t-test using a p-value (or alpha cut-off $\alpha = 0.05$) by definition means $\Pr(\text{Reject } H_0 | H_0 \text{ True}) = 0.05$ and thus, the $\Pr(\text{Fail to Reject } H_0 | H_0 \text{ True}) = 0.95$. Thus, it is no surprise the values in Table 3 are around 95%. Ultimately, the justification for sample size lies in the performance of the overall model or method of classification. Namely, a method which will rely upon five observable, measured features to comprise a variance-covariance matrix used to inform an algorithm about a given class would generally require a sample size of at least 25 samples per class. This would be in order for the algorithm to perform properly, if the rule of five times the number of observable features is followed.  

Thus, if three firearms per class are used with 10 test fires per firearm, there would be 30 samples to characterize each class (or 29 even if performing hold-one out cross validation) which still satisfies the five times observable features rule of thumb. It is noted there is no universally accepted rule for the number of samples relative to the number of observed features as the end performance of the algorithm is what ultimately is of interest. The selection of a sample size of 10 test fires per firearm corresponds to a practical working sample size for examiners considering single stack magazines for a compact 9 mm handgun commonly and conveniently hold 10 cartridges. Likewise, it is noted several states, such as California, impose a magazine-capacity restriction of 10 cartridges which arguably increases the chance of an individual choosing to purchase a firearm with this exact capacity.

Table 3: Results of 1000 set simulation involving sampling without replacement

<table>
<thead>
<tr>
<th></th>
<th>Ruger SR9</th>
<th>Springfield XD9</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Significant Difference</td>
<td>X96719</td>
<td>X17802</td>
</tr>
<tr>
<td></td>
<td>X69363</td>
<td>X17841</td>
</tr>
<tr>
<td></td>
<td>96.20%</td>
<td>89.60%</td>
</tr>
<tr>
<td></td>
<td>94.80%</td>
<td>93.50%</td>
</tr>
</tbody>
</table>

2.4. Construction of a model system of classification measurements

In addition to the distance between the center of the firing pin impression and the center of the primer (radius $r$), retained as a result of using ImageJ® to acquire x-y pixel coordinates is information regarding the spatial orientation of the firing pin impression in the form of an angle or theta ($\theta$) with respect to the x-axis (from the firearm’s perspective). This was determined by taking the inverse tangent of the y-distance over the x-distance for each firing pin impression, after accounting for the default tradition for IBIS® entries as having a 12 o’clock (firearm’s perspective) drag oriented towards 3 o’clock (image orientation). The drag mark is potentially useful information when it is indeed present to indicate the 12 o’clock position of the cartridge case (from the firearm’s perspective) when it was discharged in the firearm since certain types of firearms are known to always or never leave a drag mark. The presence or absence of a drag mark was recorded for each test fire as a categorical variable which could be used to later segment the data, or to inform the algorithm as a point estimate of a probability of leaving a drag. When considering various differences between cartridge cases without drag marks, a
mental decision tree was created involving the measurement of the firing pin imprint as well as the aperture, as seen in Figure 1. To clarify, the imprint corresponds to the tip of the firing pin impression and the flatter this tip is, the more clearly defined the imprint circle would be. From a categorical perspective, the type of aperture as well as the texture of the breech face are also features needed to mentally screen out certain classes of firearms (i.e. circular apertures would never be compared to rectangular apertures).

Thus, the quantitative inputs for any type of algorithm are the distance from the center of the firing pin impression, the angle with respect to the x-axis from the perspective of the firearm, the width of the firing pin imprint, as well as the width of the aperture. It should be noted the imprint and aperture are not only visually distinct from the rest of the firing pin crater, but correspond to changes in grayscale values. This has the potential to be more accurately quantified in future studies utilizing a 3D microscope to scan the surface of the primer to detect these changes in depth in the z-direction. The width of the smaller firing pin imprint in the center, as opposed to the full impression crater measured by the default IBIS® FPAP, seemed to be more constant within one class of gun and more different between classes of guns.

The grayscale values of the IBIS® images do seem to support the visual observation of there being a smaller lighter gray circle (corresponding to larger grayscale values) in the middle of the crater. This may be an actual feature of the firing pin (likely to be the case of the LC9 in Fig. 1) or the illumination setup where light is perfectly reflected coaxially and specular reflectance (as the result of this setup) is what is primarily responsible for this circular ring.

Fig. 1. Sample binary decision tree using categorical and continuous variables.
The categorical variables which would be used to segment the model are:

- Presence or absence of a drag mark (where absence would essentially indicate the angle is invalid in the absence of ejector and extractor information)
- The type of breech-face (linear or not apparent),
- Shape and type of firing pin aperture (circular, circular with shear marks, circular with ring, rectangular, light-bulb, etc.).

More subjective features such as the shape of the ejector mark are not being considered for use in a model or algorithm since they can still be used at the end of an algorithm result in the form of a user one-to-one comparison of images in the database, if present and available. Doing so avoids the subjectivity of assigning an ejector mark to a particular class or shape which could hypothetically result in inflated performance when using hold one-out cross validation, since the user is not blind to the true class.

In order to test the performance of a basic model limited to five classes (models), the previous measurements were taken for 10 test fires for three firearms for the Ruger SR9 (# XX9723, X69363, X96719), Springfield XD9 (# X60916, X17802, X17841), Hi-Point C9 (# X80728, X55429, X55457), Taurus 24-7 G2 (# X55720, X45398, X45399), as well as the SCCY CPX (# XX7569, XX7570, XX7571). For the XD9 and SR9, one of the firearms, XX9723 and X60916, respectively, were from actual test fires and all of the other test fires were comprised of IBIS® entries. This was done in attempt to challenge the model with mixed ammunition as it was hypothesized the features selected, in particular where the firing pin strikes, will not be affected by potential differences in cartridges such as primer hardness which may have resulted from different manufacturers (Winchester vs. Federal American Eagle).
3. Results and discussion

3.1. Data analysis
It is proposed the data from the above mentioned five models of firearms is first minimally segmented according to categorical variables in order to compensate for the obvious fact the model is based on a limited data set. Obviously, the SR9 is the only model considered which lacks a linear breech face and nearly always leaves a dragmark while the XD9 has aperture shear marks in addition to nearly always leaving a dragmark. However, the goal is not to have perfect model determination but to see the underlying weakness or limitations of a model.

A algorithm such as Mahalanobis Distance/Quadratic Discriminant Analysis (QDA) should be used to consider only the quantitative measurements of distance, angle, imprint width, and aperture width. However, what still does need to be decided is perhaps a particular chi-square distance metric to use as a cut-off for inclusion into a model or class (i.e. the $d_i$ where 95% or 80% of distance metrics are included in a class). One can vary the cut-off depending on how conservative one wishes to be for inclusion. One potential drawback of QDA distance metrics is the resulting chi-square probability values are still somewhat non-intuitive to understand. One could attempt to force the probabilities to sum up to 1.0 by dividing each Chi-square probability by the sum of all five Chi-square probabilities, thus enabling the concept of a more understandable Likelihood Ratio-type approach. Another potential drawback with the particular form of QDA discussed in Equation 1 is a lack of consideration of prior probability information. Prior probabilities become important if the probabilities of encountering a class prior to observing the questioned sample are unequal. In other words, one particular class may be much more prevalent in the relevant population. One distance metric which allows us to consider prior probabilities is the Gaussian Maximum Likelihood (GML). In addition to means and covariance matrices, GML also needs prior probabilities of encountering each class and how one decides on the priors is also subject to debate. In this model, priors would have no influence on ranking if they were determined by dividing the number of samples per class (30) by the total number of samples (150). In future studies, priors could be determined by assuming difficulty in gathering samples of a particular class is related to the lower likelihood of encountering the class in the relevant population.

However, the ultimate problem with GML is it, as an estimate of the posterior probability, requires dividing by the determinant of $S_i$. This is an impossible calculation if the determinant is zero, as is when there is zero variance observed for even one feature, as was the case with presence or absence of drag.

Of course, the absence of drag marks could be a reason to segment the data to only allow an algorithm to consider the individual cartridge case members of a class which did not leave a drag mark and for the angle to not be considered. However, ultimately, to see the limitations of an algorithm, no segmenting was performed and drag was inputted as the fifth observed variable which was binary “0” for no drag and “1” for with drag. Hold-one out cross validation...
was used since highly subjective features such as the shape of an ejector mark were not considered.

3.2. Hold-one out cross validation

Hold-one out cross validation ensured the data set used to inform the algorithm about a class was independent of the questioned sample. Each test fire was withheld as a questioned sample and the other 29 samples of its class, along with the rest of the 120 test fires, were used to train the algorithm (QDA) as to what the mean and covariance structure of a class appears to be. The results with respect to correct first rank recognition (i.e. correct class having the lowest Mahalanobis distance) are summarized in Table 4. Overall correct first rank class recognition was \((29+26+23+27+23)/150\), or about 85%. As expected, the algorithm struggled the most with the Hi-Point C9 and SCCY CPX. The C9 was never observed to leave a drag mark (thus all angle inputs were essentially subject to the random orientation chosen by the operator when it was imaged into IBIS® and only served as noise to confuse the QDA classifier. The CPX is likewise notorious for exhibiting a large degree of visual variability since it, like the 24/7, only sometimes leaves a drag mark. Thus, when no drag is present, the angle inputs served as an additional challenge to the algorithm. Table 4 shows how ten cartridge cases were misclassified as a CPX and seven CPX cartridge cases were misclassified as another firearm. Four cartridge cases were misclassified as a C9 and seven C9 cartridge cases were misclassified as another firearm. Six cartridge cases were misclassified as an XD9 while four XD9 cartridge cases were misclassified as another firearm.

Table 4: Results of hold-one out cross validation with respect to first-rank recognition

<table>
<thead>
<tr>
<th>Predicted Class</th>
<th>Actual Class</th>
<th>SR9</th>
<th>XD9</th>
<th>C9</th>
<th>24/7</th>
<th>CPX</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR9</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>XD9</td>
<td>0</td>
<td>26</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>0</td>
<td>1</td>
<td>23</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>24/7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>CPX</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

% Correct 85.33%

Table 4 only provides information about first-rank recognition, but none about for example, how often the correct class might be in the second rank which is still useful if the first hit can quickly by visually eliminated by the examiner (i.e. a questioned cartridge case with no drag is not likely to be a XD9). Table 5 provides information about this top-two-rank recognition. However, more generally, the questions remain 1) can a threshold for inclusion be proposed and 2) how often does a randomly selected known match rank higher than a random known non match?
Table 5: Results of hold-one out cross validation with respect to top-two-rank recognition

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>SR9</th>
<th>XD9</th>
<th>C9</th>
<th>24/7</th>
<th>CPX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Class</td>
<td>SR9</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>XD9</td>
<td>0</td>
<td>28</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C9</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>24/7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>CPX</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

% Correct 94.67%

3.3. Generation of Known Match and Known Non Match Density Plots

To address the first question, based on the above data set which generated known match distances and known non match distances, the 150 known match (KM) distances and the 600 known non match (KNM) distances were extracted to generate the density plot seen below as Figure 2, where R’s smoothing of an empirical histogram resulted in illogical left-hand tails with distances less than 0. Since the Mahalanobis Distance (MD) is squared in the algorithm, there can be no negative distance values. Based on where the lines representing the KM and KNM intersect, it appears a Mahalanobis distance about 15 corresponds to a likelihood ratio of 1 where \( \text{Pr}(\text{MD}=15|\text{KM}) = \text{Pr}(\text{MD}=15|\text{KNM}) \) and could be proposed as a cut-off which excludes very few of the known matches. It should be noted, this is not a true likelihood ratio since actual functions would be needed to account for points near MD= 0. Logically, the LR should only increase and be at a maximum where \( \text{Pr}(\text{MD}=0|\text{KM}) >> \text{Pr}(\text{MD}=0|\text{KNM}) \). One way to deal with R’s illogical smoothing is to change the default bin or band-width (the size of the x-axis interval used to count empirical observations within the interval to estimate a frequency) from 1.061 and 17.81 for KMs and KNMs to 0.75 and 0.15, respectively. As expected, the smaller band-widths reduced the bias towards smaller values causing the illogical negative left-hand tails but reduced the smoothness and increased the variability of the curves (Fig. 3). Conceptually, at a Mahalanobis Distance of 0, an LR of 0/0 is not possible and would occur if the means of the KM and KNM are the same and equal to the observed questioned cartridge case, suggesting equal odds. This is logical for a two-class problem, however if KNM is multi-class and refers to all of the other classes, it is difficult to conceive of how the means of the KM and KNM could coincidentally be the same.

Conceptually, the LR is better expressed as the ratio of areas underneath a chi-square curve from given distances on forward. For an extreme illustrative example, suppose for a single questioned evidence cartridge case, the algorithm returns a distance of 0 to the SR9 and 15 to the XD9, C9, 24/7, and CPX. The corresponding chi-square probability for the SR9 is 1.0, the result of integrating a chi-square curve from 0 on forward, or the entire plot equal to 1, the cumulative density up to this point. The corresponding chi-square probability for the XD9 is 0.01, the result of integrating a chi-square curve with five degrees of freedom from 15 on forward, or 1
minus 0.99, the cumulative density up to this point. Likewise, each of the other three classes has a chi-square probability of 0.01.

Thus, the numerator of the LR, or \( \Pr(\text{Evidence: } d_{SR9} = 0 \cap d_{XD9} = 15 \cap d_{C9} = 15 \cap d_{24/7} = 15 \cap d_{CPX} = 15 | \text{SR}9) \), is \( \frac{1.0}{1.0 + 4 \times 0.01} = \frac{1.0}{1.04} = 0.962 \). The denominator of the LR, or \( \Pr(\text{Evidence} | \text{SR}9^c) \), is \( \frac{0.04}{1.04} = 0.038 \). \( \text{SR}9^c \) refers to the complement of \( \text{SR}9 \), or all other models outside of the \( \text{SR}9 \). The LR for the \( \text{SR}9 \) to \( \text{SR}9^c \) is \( \frac{0.962}{0.038} = 25 \), or 1.0 divided by the sum of the chi-square probabilities for the other four classes, which is still equal to \( \frac{1.0}{0.04} = 25 \). One might quickly see this approach of forcing all probabilities to sum up to 1 only prevents the denominator from ever being greater than 1. It is not a true LR since if we had 200 other classes to consider, the LR for the \( \text{SR}9 \) to not \( \text{SR}9 \) would by default be 1.0 divided by 200(0.01), or \( \frac{1.0}{2.0} = 0.5 \). An LR of 0.5 masks the significance of the numerator suggesting a perfect match to the \( \text{SR}9 \) class. However, say, if only the \( \text{SR}9 \) and \( \text{XD}9 \) are of interest. Than the LR would be \( \frac{1.0}{0.01} \), or 100. Overall, the less an examiner segments the data and the more classes which need to be simultaneously considered, the more classes a questioned cartridge case sits on the edge of, the more this pseudo LR is “washed out.” This is even if the questioned cartridge case is simultaneously sitting perfectly at the mean or center of the “top hit” class. This suggests future research could explore the possibility of examining the ratio of the “top hit’s” actual distance to each of the other distances, using the F probability distribution to decide whether the differences are significant, and using the frequency of significance as a way of quantifying certainty in the “top hit” as an approach which is more desirable than this pseudo-LR discussed.
Fig. 2. Distribution Plot of KM and KNM Mahalanobis Distances. The five KM distance outliers (greater than 20) originated from SR9 X69363 (25.2), XD9 X60916 (55.8), CPX XX7569 (80.9), and CPX XX7571 (22.4, 124.9).
As expected, the plot of Known Match Mahalanobis Distances resembles the regular central chi-square plot for five degrees of freedom with variable x, as shown in Fig. 4. In both plots the peak with height or probability of approximately 0.15 corresponds to an x-value of approximately 3, which means the most frequently expected x-value is about 3. In other words, the most common Mahalanobis distance or mode for the known matches is about 3.
Since the distribution for Known Non Matches does not follow a distribution with a known non-centrality parameter (ncp), one method of conceptually representing or visualizing a likelihood ratio (where low distance is in favor of a known match and vice versa) is to convert the plots for the Known and Known Non Match distances in Figure 2 into one minus the cumulative density and cumulative density, respectively. This is depicted in Figure 5, and now the LR is 1 when the distance is 11 where LRs to the left of this intersection are logically much greater than 1, while LRs to the right are much less than 1. To determine the ncp, it would be most appropriate for one to assume a questioned cartridge case belonged to a class (central Chi-square, ncp = 0) and calculate the four ncp’s with respect to the four means and

![Chi-Square Density Graph](image)

Fig. 4. Distribution Plot of Chi-square distributed variable x with five degrees of freedom.
covariances of the other four classes. Using these ncp’s to generate the pdfs and upper tail probabilities would theoretically result in large LRs when the assumed class it belonged to is correct. A cutoff for inclusion would then be determined based on these LRs, rather than simply Mahalanobis distances. This was not performed as the complexity of such an approach was found to be beyond the scope of this thesis research.

9mm Firearm Class Distance Distributions

Fig. 5. Cumulative Densities for Known and Known Non Matching Distances.

3.4. Generation of Receiver-Operating-Characteristic Curve
To answer the second question of how often does a random known match have a lower distance than a random known non match (and to draw additional conclusions about potentially
determining a threshold for inclusion), a receiver operating characteristic (ROC) curve needed to be plotted and the area underneath the curve (AUC) determined.

In order to simplify the process behind what a ROC curve (traditionally used for a two-class problem) actually means for a more complicated multi-class problem, the plot was generated by simply ranking all the distances from small to large (from high chance of being included to low chance). The simplicity for what a ROC curve represents as opposed to attempting to derive the non-centrality parameter for the KNM chi-square distribution was found to be useful for the practitioner since it has broader and more general applications. The ability for an individual to make intuitive sense of where an AUC comes from is why the AUC and ROC curve approach is proposed for quantifying uncertainty for this specific application but plausibly for any other pattern-based forensic problem where a continuous set of scores can be generated. Namely, even if subjective scores of similarity are generated by human practitioners for comparisons in the same arbitrary manner as points are summed in academia to form letter grades (i.e. >90 = A, <60 = F), this might allow the average American-educated juror/lay person to have a better understanding of what KM vs. KNM and “difficult” vs. “easy” comparisons mean in relation to conclusions of identification (“A”), elimination (“F”), and inconclusive (“D through B”) where there might be more “A” and “F” grades than the other grades, perhaps forming a bimodal distribution corresponding to the overlap between ground truth KMs and KNMs.

Going down a list of small to large distances, the rule for plotting was simply: if the distance was a KM, move up 1/150 along the y-axis and if the distance was a KNM, move over to the right 1/600. Perfect classification where all KMs are a lower distance than the KNMs would result in moving up 150 times towards (0,1) followed by moving over 600 times to reach (1,0). This can even be done in a typical spreadsheet, already familiar to practitioners. The y-axis corresponds to the True Positive Rate (TPR) while the x-axis corresponds to the False Positive Rate (FPR). The square formed would have an AUC of 1, corresponding to the probability of a randomly selected KM having a lower distance (or higher ranking) than a randomly selected KNM equal to 100%. This is otherwise known as perfect stochastic dominance. This process was performed for the KM and KNM distances and the plot which resulted can be seen in Figure 6. Also plotted is a dashed diagonal line from the origin to (1,1) corresponding to an AUC of 0.5 and the expected chance of true and false positives being equal to each-other, a useless decision process.
Fig. 6. ROC curve for performance of Mahalanobis Distance where the dashed diagonal orange line represents a useless decision making rule.

The utility of the plot in Fig. 6 is if an individual says he is prepared to accept a False Negative Rate of 5%, it corresponds to a y-axis True Positive Rate of 95% which corresponds to the data point (0.197, 0.953). This data point is the first to cross the 95% True Positive Rate desired and is the 261th distance in the list of distances ranked from small to large. This 261th distance had a Mahalanobis distance of 15.472. Thus, using a threshold rule of “include if it’s below or equal to 15.472” corresponds to a True Positive Rate of over 95%. It is noted the selection of 95% is a subjective selection but resonates as a “solid A” conception ingrained from academia.

Since we know the distribution of KM Mahalanobis Distances follows a centrally distributed chi-square function derived from the square of a Normally distributed variable centered at a mean of zero since the expected value of a variable is equal to the mean of a class ($E(x-m_i) = E(x) - m_i = 0$) we can integrate the chi-square function from 0 up to 15.472 to determine the percentage of KM Mahalanobis Distances which are less than or equal to this cut-off threshold.

Doing so using five degrees of freedom corresponding to our five observed features results in capturing over 99% of KM Mahalanobis Distances. In other words, for this data set, 99% of KM Mahalanobis Distances are less than or equal to 15.472. Thus choosing 15.472 as a cut-off for
inclusion means our Likelihood Ratio of our True Positive Rate divided by our False Positive rate is 0.953/0.197 or 4.8.

3.5. Calculation of AUC
To calculate the AUC for our ROC curve, one can simply calculate the Mann-Whitney U statistic as noted by Brown and Davis\textsuperscript{74}, and divide by the product of the number of KM and KNM distances, or 150*600 = 90,000. To calculate the U-statistic, we first rank the distances jointly from small to large distances or high to low chance of inclusion, just as they have already been ordered to generate the ROC curve.

The U-statistic is calculated based on the group with the smaller sum of joint ranks, in this case, the KM distances, since they are generally smaller than and ranked higher than the KNM distances. Computation of the U-statistic can be seen in Eq. 2 below where \( n^+ \) is the number of KM distances (150), \( n^- \) is the number of KNM distances (600), and \( W^+ \) is the sum of joint ranks for the KM distances. The relationship between U and the AUC, as first noted by Bamber\textsuperscript{72}, is described Eq. 3.

\[
U = n^+ \cdot n^- + \frac{n^+(n^+ + 1)}{2} - W^+ \\
AUC = \frac{U}{n^+ n^-} \tag{3}
\]

Upon re-arrangement and substitution:

\[
U = [n^+ \cdot n^-] - [W^+ - \frac{n^+(n^+ + 1)}{2}] \tag{4}
\]

\[
AUC = 1 - \frac{W^+ - \frac{n^+(n^+ + 1)}{2}}{n^- n^+} \tag{5}
\]

The AUC can directly be computed in three simple steps, possible even using a typical spreadsheet already familiar to practitioners, as opposed to an extra package in R which may have a steep learning curve. First, calculate the difference between the sum of the KM ranks in the complete list of 750 distances and the sum of 1 to the number of KM distances or \( 1+2+\ldots+150 = \frac{[150(150+1)]}{2} = 11,325 \). Thus, if all 150 KM distances were ranked in the top 150 overall positions, the AUC = 1.0 - (11,325-11,325)/90,000 = 1.0 - 0/90,000 = 1.0.

In this data set, the sum of the KM ranks was 15,533. Thus the difference is 15,533 - 11,325 = 4208. Second, divide this difference by the product of the number of KM and KNM distances, or 150*600 = 90,000. Third, subtract this value from 1.0. Thus, the AUC is 1.0 - (4,208/90,000) = 0.953. With respect to the AUC, the stochastic dominance or the probability of a randomly selected KM distance ranking higher than a KNM distance is 95.3%. It should be noted, had one
not subtracted from 1.0, one would have the area above the curve up to a boundary of \( y = 1.0 \). Mathematically, one could have directly calculated the AUC without having to subtract from 1.0 if he had instead re-ranked from large to small distances or low to high chance of inclusion. Thus, if all 150 KM distances were ranked in the bottom 150 overall positions, the sum of these new ranks would be 601+602+...+750 = 101,325. Therefore, the AUC = \((101,325-11,325)/90,000 = 90,000/90,000 = 1.0\).

Using the approach of re-ranking from low to high chance of inclusion, the new sum of ranks is 97,736 and the AUC is similarly equal to \((97,736-11,325)/90,000 = 0.960\). Thus, the AUC can be estimated as being between 0.95 to 0.96.

3.6. Potential use in case work via blind classification
A set of 17 physical cartridge cases meant to simulate 17 crime scene cartridge cases were analyzed in the context of the above-mentioned MD algorithm, where 4 cartridge cases were from two models which were outside of the data set considered above. These were two cartridge cases from a Sig Sauer P250 and two cartridge cases from a Ruger P95. Since these were physical cartridge cases as opposed to IBIS images, each of the crime scene cartridge cases was photographed. Each resulting cartridge case image was analyzed independently of the other 16, as would be the situation if each crime scene had only one cartridge case recovered. The resulting distances are shown in Table 6.

**Table 6**: Mahalanobis distances for blind case work samples (first rank distances in bold)

<table>
<thead>
<tr>
<th></th>
<th>SR9</th>
<th>XD9</th>
<th>C9</th>
<th>24/7</th>
<th>CPX</th>
<th>Ground Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc_01</td>
<td>145.30568</td>
<td>12.532923</td>
<td><strong>0.4471987</strong></td>
<td>144.01843</td>
<td>12.719062</td>
<td>P250</td>
</tr>
<tr>
<td>cc_02</td>
<td>153.495567</td>
<td>13.719992</td>
<td>15.5507706</td>
<td>250.52758</td>
<td><strong>10.244268</strong></td>
<td>CPX</td>
</tr>
<tr>
<td>cc_03</td>
<td>74.923033</td>
<td>58.631548</td>
<td>226.1202951</td>
<td><strong>11.06164</strong></td>
<td>17.196175</td>
<td>24/7</td>
</tr>
<tr>
<td>cc_04</td>
<td>36.496729</td>
<td>9.742935</td>
<td>25.5262666</td>
<td>121.17224</td>
<td>5.055131</td>
<td>CPX</td>
</tr>
<tr>
<td>cc_05</td>
<td>4.523519</td>
<td>22.642758</td>
<td>114.075941</td>
<td>68.36839</td>
<td>16.222403</td>
<td>SR9</td>
</tr>
<tr>
<td>cc_06</td>
<td>148.466549</td>
<td>16.058284</td>
<td>11.7964973</td>
<td>153.72635</td>
<td>8.635851</td>
<td>C9</td>
</tr>
<tr>
<td>cc_07</td>
<td>137.668787</td>
<td>3.739622</td>
<td>4.7790693</td>
<td>121.18954</td>
<td>5.877818</td>
<td>XD9</td>
</tr>
<tr>
<td>cc_08</td>
<td>9.668035</td>
<td>16.900142</td>
<td>83.9643691</td>
<td>105.29042</td>
<td>20.777581</td>
<td>SR9</td>
</tr>
<tr>
<td>cc_09</td>
<td>66.093776</td>
<td>43.89901</td>
<td>45.664017</td>
<td>246.13178</td>
<td><strong>34.940546</strong></td>
<td>24/7</td>
</tr>
<tr>
<td>cc_10</td>
<td>149.061651</td>
<td>8.149305</td>
<td>5.853292</td>
<td>107.99733</td>
<td>7.471381</td>
<td>C9</td>
</tr>
<tr>
<td>cc_11</td>
<td>95.592056</td>
<td>7.454202</td>
<td>5.3298289</td>
<td>127.87183</td>
<td><strong>2.200853</strong></td>
<td>CPX</td>
</tr>
<tr>
<td>cc_12</td>
<td>100.99614</td>
<td>12.436083</td>
<td>27.4758771</td>
<td>85.41833</td>
<td>5.825301</td>
<td>P95</td>
</tr>
<tr>
<td>cc_13</td>
<td>8.713552</td>
<td>16.657991</td>
<td>91.6456282</td>
<td>28.00692</td>
<td><strong>7.315204</strong></td>
<td>SR9</td>
</tr>
<tr>
<td>cc_14</td>
<td>185.976777</td>
<td>5.391066</td>
<td>3.7807222</td>
<td>155.97575</td>
<td>11.344062</td>
<td>P250</td>
</tr>
<tr>
<td>cc_15</td>
<td>275.207506</td>
<td>56.359356</td>
<td>157.5169239</td>
<td>672.35033</td>
<td>65.678753</td>
<td>24/7</td>
</tr>
<tr>
<td>cc_16</td>
<td>191.43563</td>
<td>7.342959</td>
<td>9.0331769</td>
<td>161.26036</td>
<td>8.531957</td>
<td>XD9</td>
</tr>
<tr>
<td>cc_17</td>
<td>33.904701</td>
<td>8.985801</td>
<td>26.3028344</td>
<td>200.84749</td>
<td>22.138755</td>
<td>P95</td>
</tr>
</tbody>
</table>

As seen in Table 6, the Sig Sauer P250 (a recoil action operated firearm which fails to leave an expected drag mark) was misclassified as a C9. The Ruger P95, which also failed to leave a
drag mark, was misclassified as a CPX and an XD9. This was expected since while the P95 had a non-apparent breech face and the CPX and XD9 have parallel linear breechfaces, the algorithm was not informed of this.

Cartridge case 2 is a good example of the algorithm not only successfully classifying it to the CPX class at the first rank but also three other classes would not be listed since they all have distances greater than the proposed 15.472 cut-off at the 99th percentile of the central Chi-square distribution which known matches follow. The same can be said for cartridge case 3 and 5 being classified as a 24/7 and SR9 respectively, where all four of the other classes are not included.

In contrast, cartridge case 4 is correct first rank wise, however an investigator would also need to consider the XD9 since the distance is less than the 15.472 cut-off. Cartridge case 6 was incorrectly classified as a CPX by first rank, however the correct class (C9) would also be included on a list to investigators since it’s distance is less than the cut-off while the other three classes would fail to be put on the list.

Cartridge case 7 was correctly classified as an XD9 via first rank, however both the C9 and the CPX would be put on the list as well. This is fine since while cartridge case 7 had a drag mark, it is possible for the C9 to leave a drag mark from ejection and this frequency or proportion is a subject of future research requiring a much larger sample size of test fires. Cartridge case 8 was correctly found to be an SR9 and the algorithm would not include the other four classes which fail to meet the threshold.

Cartridge case 9 is interesting as none of the distances, including the “best hit,” were low enough to meet the threshold to be put on a list, including the correct 24/7 class. It is fortunate how an “inconclusive” result at least does not mislead the investigation with the wrong class. However, the lack of inclusion of the 24/7 may be due to the imprint of the 24/7 as being not an actual feature of the firing pin, but a result of the specific lighting setup used (which was different from the IBIS® images which informed the algorithm). It is noted the full impression crater, as opposed to the imprint, measured by IBIS® actually corresponds to the width of the entire firing pin at a wider point than just the tip. The aperture mark is a best estimate made based on what one could see visually. In the absence of obvious primer flowback, as is the case with the 24/7, it is noted it may not be possible, depending on lighting, to differentiate the true diameter of the aperture from the entire firing pin impression (which is subject to variability as the result of differences in primer hardness). The same commentary applies to cartridge case 15, since although the XD9 was incorrectly returned as the “top hit,” it failed to have a small enough distance to be put on an investigative list.

Cartridge case 10 had no drag mark and was correct to the C9 by first rank, however also included on a potential list is the CPX and XD9 (at the proposed threshold), albeit at lower ranks. The 24/7 and SR9 are at distances much greater than 50. This correct C9 to C9 classification can be clearly visualized as a green dot in Figure 7, which also depicts the
distribution of all the distances less than 50 for the 17 blind cartridge cases and how they relate to a cut-off distance of 15 for inclusion as a "positive."

**Fig. 7.** Visualized distances less than 50 for the 17 blindly classified cartridge cases where non-matching distances are red dots, ground truth models are blue dots, and matching distances are green dots.

Again, more research is necessary to refine the algorithm as well as determine the frequency of the XD9 failing to leave a drag mark. The same can be said for cartridge case 11, which was correctly classified via first rank as a CPX but creates additional work for the investigator since the XD9 and C9 are also included on a list. The same can be said for cartridge case 16 (which has a prominent drag mark), where the first ranked XD9 is correct, however the CPX and C9 also would have to be put on the list, for again a total of three models of firearm. The goal is a list of five or less since this would be manageable for an investigator to work with.

Cartridge case 12 is an example of how it is important to segment the database considered by the algorithm since while it incorrectly returned the CPX and XD9 to be put on a list, one can clearly see how it not only lacks the vertical parallel breech face marks of either the CPX or the
XD9, but also has the same non-apparent texture observed in the Ruger SR9. The experienced examiner would in any case obviously not mislead the investigation and exercise his best judgement in noting how the questioned cartridge case (ground truth Ruger P95) has a Ruger SR9-like breech face but does not have the expected drag mark. Again, more research is needed into how frequently the SR9 can fail to leave a drag mark since had the algorithm classified cartridge case 12 as an SR9, it is still helpful to the investigator as the same brand. Cartridge case 13 was incorrectly classified as a CPX, however the correct SR9 distance was a close second (8.7 vs. 7.3). Again future research would consider looking at ratios of distances which might reveal how “weak” the “top hit” might be if the second rank is close.

Cartridge case 14 truly highlights the limitation of the method which is simply lack of awareness of the existence of a model which has vertical breech face marks and does not leave a drag (Sig Sauer P250). Cartridge case 17 (ground truth Ruger P95) is misclassified as a XD9, however the other four incorrect classes fail to make it on a list and thus the investigation is not misled. Once again, the experienced examiner would note how the non-apparent breech face marks observed on the questioned cartridge case is reminiscent of Ruger type firearms and never observed in XD9s (knowledge supplemented by personal database photos) and would be highly reluctant to include the XD9 on a list of one. Again, non-apparent can more specifically refer to an actual observed pattern (highly informative) which appears granular or sand blasted (in the cases observed for Ruger type firearms), or it could refer to pattern absence (not very informative) resulting from a potentially low pressure cartridge. As a general note, in spite of what an algorithm outputs, an examiner must emphasize the caveat of how there may be a model which exists outside the database informing the algorithm.

Figure 8 summarizes the results of using a cut-off of 15.4 for inclusion during the blind classification of the 17 cartridge cases, where 32 out of 85 distances were included. In this experimental design, given an inclusion, 11 are indeed a correct match, or about 34%. This posterior is consistent with Kennington’s belief in the success rate of an experienced examiner as being less than 50%. Given a known match distance, only two out of 13 failed to be included and these were the two inconclusive 24/7 cartridge cases. The true and false inclusion rates observed were about 85% and 29%, respectively. This ratio of about 3 is consistent with the intuitive understanding it is reasonably close, given the sample size, to the ratio observed from hold-one out cross validation. Figure 9 summarizes the results of using a cut-off of 11.08 (95th percentile of central Chi-Square), which resulted in case 6 also failing to be included.

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>No Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM 11</td>
<td>2</td>
</tr>
<tr>
<td>KNM 21</td>
<td>51</td>
</tr>
<tr>
<td>32</td>
<td>53</td>
</tr>
<tr>
<td>85</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 8.** Resulting distances from blind classification using 99th percentile as inclusion cut-off.
Fig. 9. Resulting distances from blind classification using 95th percentile as inclusion cut-off.

<table>
<thead>
<tr>
<th></th>
<th>Inclusion</th>
<th>No Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>KNM</td>
<td>16</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td><strong>26</strong></td>
<td><strong>59</strong></td>
</tr>
<tr>
<td></td>
<td><strong>85</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Inclusion</th>
<th>No Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>13</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>72</strong></td>
</tr>
</tbody>
</table>
4. Conclusion

4.1 Practical limitations of this thesis work
This thesis was never intended as a final end-all be-all solution to predicting models in a no-gun case. This thesis proposed what could be done and provided a model framework which can certainly be improved upon, albeit based on a limited sample set based on the materials available. The intellectual framework demonstrates how features which had been previously observed generally using the naked eye under a macro/microscope can be quantified and used to help in no-gun cases. This was the original goal which was set out to be achieved. Given the data, results, and tests, one would not expect a True Positive Rate of 95% in actual casework. However, one can be confident if one expands the number of firearm parameters considered in the system, a much better and helpful model would result. In the sense of determining what would be more helpful in a no-gun case, this was accomplished and will be elaborated upon below.

4.2 Directions for Future Research
The results of the data set suggest a MD of approximately 15 as the threshold for inclusion into a class for this set of firearms, which captures 99% of the Mahalanobis Distances in a central Chi-square distribution with five degrees of freedom. More immediate future research could examine the effect of the repeatability of manual user determined measurements in this particular thesis dataset and resulting threshold corresponding to a 95% True Positive Rate. This could be done by adding noise to the first four continuous variable measurements by multiplying each measurement by a value randomly selected from a normal distribution centered at a mean of 1 with a given standard deviation. This would model random errors where some measurements are smaller and others are larger. Hold-one out cross validation could than be replicated many times and the corresponding thresholds could be plotted as a distribution to compare to the 15 determined in this original thesis study.

Future research should look at using the method described to propose various thresholds for not only expanded data sets of firearms, but also specific sets of firearms for the purpose of segmenting a database. In this thesis research, the five models of firearms chosen were based on subjective experience of these firearms being common, fairly affordable firearms (which may be a factor in whether they ultimately end up being used in crime). The CPX was chosen due to its reputation as being highly variable with respect to the appearance of the test fires generated from different firearms. The CPX also represents the manufacturer SCCY as a whole since it is the only firearm they produce. Similarly, the C9 is the only 9 mm handgun manufactured by Hi-Point. The only other model of 9 mm firearm produced by Hi-Point is the 995TS carbine, which would also expect to produce cartridge cases of a similar appearance. The 24/7 and XD9 admittedly represent only a small portion of all the various 9 mm handgun models manufactured by Taurus and Springfield, respectively. It is noted Springfield also has an XD-S 9 mm handgun which has an entirely different aperture shape than the circular aperture observed in the XD9 and the other four models in this thesis study. The SR9 was chosen to represent Ruger, a large firearms manufacturer, which manufactures many other models of 9 mm firearms which are a
object of future study for testing whether or not the brand as a whole can be predicted based on an observed combination of features such as the sandblasted breech face pattern coupled to quantitative measurements. In this thesis study, the SR9 was the only model which did not produce vertical linear breech face marks. Future studies should emulate the goal of this thesis study in selecting models which can easily be confused with one another. Also, selection must consider how common the model is to a given jurisdiction, as potentially determined by sales data from both physical and online retailers.

Although breech face was not used in this thesis model, when breech face marks other than the most common vertical linear are clearly present and when an experienced examiner is easily able to assign a categorical state with respect to the general pattern (i.e. sandblasted/granular, arches, etc.), this should be used as a discriminating feature. Future methods should use a Bayesian network with this breech face node instantiated to a particular pattern in order to return a list of guns all sharing this pattern, and can be achieved using software packages such as Netica or Hugin as used and discussed by Taroni and colleagues for assessment of factors contributing to observed count of gunshot residue particles. Alternatively, another discrete pattern classifier which could be used is a random forest approach utilizing x and y pixel values from manual tracing of such patterns (breech face, aperture, or ejector mark) utilizing software such as tpsDig, as discussed by Warren and Sheets. This tracing might be automated as well through exploration of edge detection methods. Regardless of what pattern classifier is used as an initial “funnel" of possible guns, only data from these gun types should be forwarded and used to form the corresponding classes which a Mahalanobis distance (based on continuous measurement variables) is generated. This allows for true “apples to apples” comparisons since subsequent measurements are meaningless if, for example, breech face patterns disagree and a true elimination conclusion would be reached. The same method can be applied to type of firing pin aperture. For example, the clear presence of a “light bulb" or “teardrop" shaped aperture suggests it can actually belong to only one of three known 9 mm classes (at the time of this writing): 1) Glock 43, 2) Newer Glocks (Generation 5 and beyond), and 3) Smith & Wesson M&P9. Measurements of this aperture mark in conjunction with other measurements may allow for discrimination between these three classes.

With respect to data sharing in the future, the use of cloud storage systems is feasible as a supplement to storage in traditional, physical computer drives. The ease of sharing data facilitates examiner training, and measurements could be biased in the positive sense of there being more consistency amongst examiners when making blind classifications of ground truth known samples. Laboratory wide biases would also be more easily detected if conditions, such as lighting, were shared with regards to the setup which was used to make the same measurements. Potentially, uniform lighting across individuals and laboratories alike could be achieved using the same IBIS® system. The use of IBIS® images for measurement should be further studied in relation to lighting conditions affecting the the perceived firing pin imprint mark, for example, and ideally, 3D topographic data and physical examinations of the actual firing pins should also be considered. For this thesis study, the firing pins from each of the five models were eventually examined and compared to the classes of firing pins discussed by Kennington.
and considered in the context of Bati’s quantitative measurement of hemispherical AK-47 firing pins.  

Figure 10 depicts the five representative firing pins belonging to, from top to bottom, the SCCY CPX, Taurus 24/7, Hi-Point C9, Springfield XD9, and Ruger SR9. The SR9 and 24/7 firing pins were left within their corresponding striker-fired mechanisms. Figure 11 depicts a figure taken from page 12 of Kennington’s 9 mm Matrix text showcasing the various classes of firing pins from side and frontal views. Visual examination confirms the earlier belief the firing pins are not all the same. Namely, when viewed from the side, the SR9 and 24/7 do indeed have a firing pin which angles more rapidly towards a smaller tip, or what Kennington depicted as a “truncated cone.” The practical issue lies in actually being able to observe and measure the smaller tip as a firing pin imprint mark on cartridge cases, specifically with the 24/7 using a microscope with user-determined manual lighting adjustments. In contrast, the XD9 and C9 do appear to have similar hemispherical firing pins, confirming the earlier belief formed as a result of examining the XD9 and C9 test fires (and noting the visual similarity with respect to the size of the imprint). The CPX, however, has a different blunted hemi-spherical firing pin.

![Physical firing pins from a SCCY CPX, Taurus 24/7, Hi-Point C9, Springfield XD9, and Ruger SR9.](image-url)
Figure 11. Firing pin type definitions taken from page 12 of Kennington’s 9 mm Matrix.\(^{38}\)

Figure 12 depicts Figures 1 and 2 taken from Bati’s article, where it is argued the default IBIS® Firing Pin Acquisition Protocol (IFPAP) does not correspond to the actual diameter of the firing pin and rather overestimates it.\(^{54}\) On the same note, an argument is made for using a Firing Pin Modified Acquisition Protocol (FPMAP) which corresponds to the both the diameter of the firing pin as well as the diameter of the contact between the firing pin tip and primer.
It appears Bati’s modified protocol measures a region smaller than the default IBIS® (darker black grayscale values) but still larger and outside of the smaller observably lighter grey region (as measured in this thesis study as the imprint mark). Bati’s definitions seem plausible only if the firing pin was very blunted or cylindrical. It is also interesting how Bati was able to repeatedly measure exactly 1816 microns as the width of the firing pin imprint mark for one test fire per individual firearm for five firearms of the same Polish class of AK-47, which he reports as producing a hemispherical firing pin impression. This impressive level of precision and reproducibility was observed again as exactly 1920 microns for a different Chinese class of AK-47, reported as producing a different “smooth flat” firing pin impression. Future research
should clarify how this region should best be determined for various types of firing pins and photos of such firing pins should also be databased.

More blind studies are needed to evaluate the effectiveness of using standardized IBIS® entries (entering test fires of a known model and exporting the images/data to form an electronic class characteristic database) as well as using photographs taken by experienced examiners of ground truth known cartridge cases (to again form an electronic class characteristic database) to make decisions related to complicated multi-class problems. These studies must be designed to mimic the same difficulty encountered in case work and must go far beyond a simple two-class problem. Ultimately, for practical implementation to casework, the threshold for inclusion must result in a short and manageable enough list of five or less models for an investigator to consider. This would be the purpose of exploring additional statistical methods, whether they be modeling LRs or using F distributions, to see if additional ways of interpreting data might allow for a more manageable list for an investigator. This would be more plausible for crime scene cartridge cases which not only exhibit rarer features, but when assumptions can be made with respect to utilizing the GRC information (as an algorithm input) of potentially corresponding non-deformed bullets, which is also an object of future study.

Larger sample sizes are needed to estimate the frequency or proportion a firearm does or does not leave a drag mark. In the absence of this knowledge however, the algorithm should give more attention to when a drag is observed and assign less or little meaning to when drag is not observed. Namely, when a prominent drag is observed, not only is it orienting the cartridge case from the gun’s perspective and giving meaning to the angle measurement, but it also tells one the donor model must have a tilting barrel. Even if the C9 (no tilting barrel) left a drag from firing pin ejection, it would not be as visually prominent and an examiner would quickly note how the drag is unusually perpendicular to the direction of the linear breech face marks.

With respect to weighting the drag mark, the Mahalanobis distance is weighted with respect to different models having different means corresponding to the proportion of samples in the data set with a drag mark for a given model. However within a model, the five features are weighted evenly in the sense a cartridge case with a drag can still be assigned to a class where drag is never observed, if the other features are near the mean of the class. Overall, the Mahalanobis distance is an unweighted calculation which is true to the tangible measurement features observed. Using a correlation matrix (dividing covariances out by standard deviations to get rid of individual measurement units) instead of a covariance matrix would prevent any one feature from having too much influence due to units, however this was not performed since standard deviations of 0 were observed for models always or never leaving drag. Thus, absent of additional mathematical weighting techniques such as assigning drag presence to an arbitrarily determined value such as “1000” instead of “1” to inflate the distances to classes with all or mostly “0” drag, the most appropriate way to assign weight to the presence of a drag on a questioned cartridge case would be to segment out from consideration the classes which do not leave a drag as well as the individual test fires which lacked a drag for guns which can produce a drag, such as the CPX and 24/7. Thus, with another dataset comprised of only test fires with drag, only these ground truth model known test fires where the angle of orientation can be
reliably obtained would inform the algorithm about the nature of a class’s measurements and
distance to the questioned case, going back to more of an “apples to apples comparison.”

4.3 Final Conclusions
The take home message is this work has investigative value even if there may be a great
number of individual firearms in existence belonging to a given model determined (as the result
of an examiner and a future database). This is because the total number of all individual 9 mm
firearms is even greater. The analogous narrowing down or funneling process from a red sedan
to a red Honda Civic is of great value despite how many red Civics there may be. Determining
the relevant population in a criminal case is crucial for assessing the competing prosecutorial
and defense hypotheses. For individualization of a cartridge case to a specific firearm, the
defense hypothesis is not “Some other firearm of the same model discharged the cartridge,” but
rather “some other firearm.” The other firearm could be of various different class models and
work should be done to determine what other models comprise of the relevant population in
consideration. For DNA, the probability of the random match is computed not only using class
characteristic allele frequencies from a Caucasian population database (if the defendant is
Caucasian), but also utilizing the Hispanic and African American databases as well and often
reporting the statistic most favorable to the defense. This is to not make any undue assumptions
about the true ethnicity of the actual, unknown offender. If the field of firearms is to be held by
the courts to have the same advantage as DNA in being able to report a quantifiable likelihood
ratio in favor of individualization, a system must be determined for deciding what classes or
models objectively constitute the relevant population for a set of crime scene cartridge cases.
The National Institute of Science and Technology’s most recent progress towards a DNA-like
likelihood ratio approach to forensic firearms identification77 concedes in order to actually
address the traditional defense hypothesis, “practical case work would require...a database with
accurate counts of firearms manufactured by different methods with different class
characteristics...” and “a statistical procedure to combine data sets from different types of
firearms from different manufacturers – one could not generalize the results seen here to other
types of manufactured firearms.”

Overall, Mr. Kennington had the right idea of using cartridge case information to supplement
GRC, however a fully functional electronic database will require tremendous amounts of work in
order to have enough samples to characterize the quantitative variability of certain class
characteristics for many firearms, which can be altered at any point in time by manufacturing
changes. The potential of 3D optical topography to automate the measurements of certain
features is an additional object of further study, along with research into other algorithms
besides Mahalanobis distance which are also highly interpretable. The overarching theme is
there is a research need for a larger sample set of 9 mm Luger cartridge case data to be
generated, continuously updated, and evaluated based on performance with respect to blind
classification of ground truth knowns. Progress made in this area of research would directly
benefit the expediency of investigating and solving gun-related crimes, as well as advancing the
scientific reputation of the field of firearm and toolmark examination.
References

1. IBIS® is a brand manufactured by Ultra Electronics Forensic Technology WAI, Inc., of Montreal, Canada.


64. Bruce, I., Previous Cycling/Firing Marks, Reloading Marks and Primer Marks on Winchester 9x19mm Remanufactured Ammunition. *AFTE Journal 2014*, 46 (2), 152 - 156.


73. CA Penal Code Sections 12276, 12276.1, and 12276.5
75. Bamber, D., The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* 1975, 12, 387-415.