INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Firearm model identification based on fired bullet cartridges

Sérgio Alexandre Pascoal Valentim

Master's in Computer Engineering

Supervisor:
PhD Tomás Gomes Silva Serpa Brandão, Assistant Professor
Iscte - Instituto Universitário de Lisboa

Co-supervisor:
PhD João Carlos Amaro Ferreira, Assistant Professor with habilitation
Iscte - Instituto Universitário de Lisboa

May, 2022

TECHNOLOGY
AND ARCHITECTURE

Department of Information Science and Technology

Firearm model identification based on fired bullet cartridges

Sérgio Alexandre Pascoal Valentim

Master's in Computer Engineering

Supervisor:
PhD Tomás Gomes Silva Serpa Brandão, Assistant Professor
Iscte - Instituto Universitário de Lisboa

Co-supervisor:
PhD João Carlos Amaro Ferreira, Assistant Professor with habilitation
Iscte - Instituto Universitário de Lisboa

May, 2022

Acknowledgments

I wish to show my appreciation to my parents and siblings for all the support, help and encouragement that was provided throughout my academic journey. Without their support it would not be possible to be where I am at today.

To my girlfriend Mariana, I want to thank all the love and presence at all times. Her companionship and joy in moments of stress made it easier to complete this work. Her help and reassurance were fundamental during every step of this process.

I would also like to thank all the assistance and mentorship provided by my supervisors, Dr. João Ferreira and Dr. Tomás Brandão. The knowledge and the time allocated were essential to guide me through the development of the project and writing of this dissertation.

Lastly, I wish to extend my appreciation to the INOV researchers and engineers Ricardo Ribeiro and Stefan Nae, who contributed with their expertise in Machine Learning and the processes that allowed this dissertation to become a reality.

Sérgio Valentim

Resumo

A identificação do modelo de arma que disparou um cartucho é informação forense que pode ser uma prova crucial num crime. Este processo é tradicionalmente realizado por peritos em balística que comparam múltiplos cartuchos ao microscópio, o que pode ser demorado e requer múltiplos recursos humanos. Como tal, esta dissertação apresenta o desenvolvimento de uma técnica de identificação balística baseada em redes siamesas. Esta abordagem visa auxiliar na classificação, ao fornecer uma lista dos modelos de armas mais prováveis de terem provocado o disparo de um cartucho, poupando tempo e recursos humanos.

Para o desenvolvimento deste instrumento, a Polícia Judiciária Portuguesa forneceu um conjunto de imagens para a criação de um modelo de aprendizagem automática que efetue esta identificação. Uma vez que esta coleção de dados ainda estava em construção e não havia sido testada, as técnicas propostas nesta dissertação foram também treinadas com outro conjunto de dados, o NIST Ballistics Toolmark Research Database, com o objetivo de estabelecer um desempenho de referência.

Para a otimização da rede, técnicas de pré-processamento de dados, assim como de transferência de conhecimento são também analisadas.

No conjunto de dados da Polícia Judiciária, o modelo de classificação proposto atingiu valores de precisão de 57% em classificação top-1 e de 81% para classificação top-2. Embora estes resultados pareçam promissores, esta técnica atingiu uma precisão de 100% em classificação top-1 com a base de dados da NIST Ballistics Toolmark Research Database, sugerindo que podiam existir melhorias a ser realizadas no conjunto de dados da Polícia Judiciária Portuguesa.

Palavras-chave: Redes siamesas, Processamento de dados, Classificação balística

Abstract

Identifying the gun model that fired a given cartridge is an example of forensic information that can be crucial evidence in a crime. This process has traditionally been carried out by ballistics experts who visually compare multiple cartridges under the microscope, which can be very time consuming and requires multiple human resources. As such, this dissertation presents the development of a ballistics identification method based on siamese neural networks. This approach aims to aid classification by delivering a list of the most likely weapon models to have triggered the firing of a cartridge, saving time and human resources.

For the development of such instrument, the Portuguese Criminal Police has provided a dataset for training a machine learning model that performs this identification. Since this dataset was still under construction and had not been tested, the techniques proposed in this dissertation were also trained on another dataset, the NIST Ballistics Toolmark Research Database, with the purpose of establishing a benchmark performance.

For the optimization of the network, data pre-processing techniques as well as transfer learning are also analysed through a development pipeline.

Using the Portuguese Criminal Police's dataset, the proposed classification model based on siamese neural networks reached accuracy values of 57% and 81%, for top-1 and top-2 gun model identification. While these results seem promising, this technique reached an accuracy of 100% on top-1 classification with the NIST Ballistics Toolmark Research Database, suggesting that there were still improvements that could be performed on the Portuguese Criminal Police's dataset.

Keywords: Siamese Neural Networks, Data processing, Ballistics classification

Contributions

This research was funded by the Foundation for Science and Technology (FCT) through ISTAR-IUL's project UIDB/04466/2020 and UIDP/04466/2020.

Part of the results of this dissertation were presented at the *ISDA 2021* conference and published in the event's proceedings:
S. Valentim, T. Fonseca, J. Ferreira, T. Brandão, R. Ribeiro, and S. Nae, "Gun model classification based on fired cartridge case head images with siamese networks," in Intelligent Systems Design and Applications, A. Abraham, N. Gandhi, T. Hanne, T.P. Hong, T. Nogueira Rios, and W. Ding, Eds., Cham: Springer International Publishing, 2022, pp. 1281–1291, isbn: 978-3-030-96308-8.

The research carried out along this dissertation was also presented at the *Portugal Smart Cities Summit 2021* event.

Contents

List of Figures

List of Tables

Glossary

CMC - Congruent Matching Cells

CNN - Convolutional Neural Network

DA - Data Augmentation

DSRM - Design Science Research Methodology

NBIC - National Ballistics Imaging Comparison

NBTRD - NIST Ballistics Toolmark Research Database

NIST - National Institute of Standards and Technology

PCP - Portuguese Criminal Police

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analysis

ROI - Region of interest

SNN - Siamese Neural Network

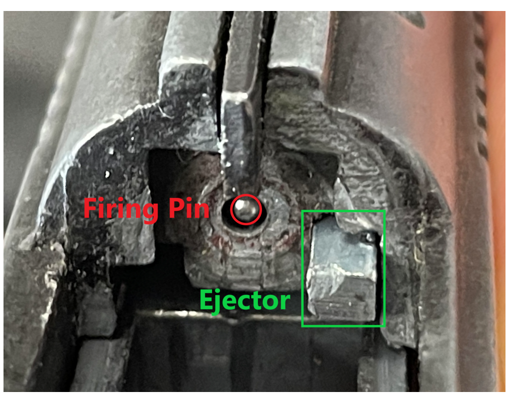
TL - Transfer Learning

CHAPTER 1

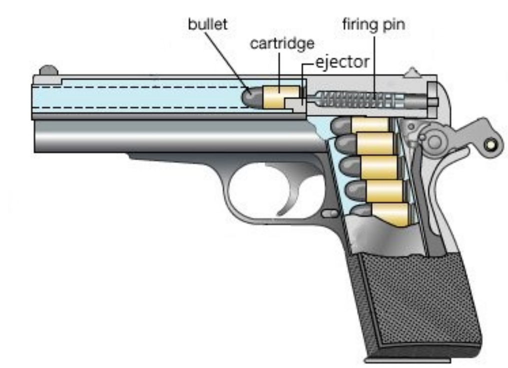
Introduction

The identification of the firearm model that fired a bullet is critical forensic information that is traditionally performed by skilled examiners using microscopes and visual inspection. At the Portuguese Criminal Police's (PCP) laboratories, this work is currently performed using the FIRETYDE database of the German Federal Criminal Police and a set of internal files with pictures of fired bullet cartridges.

When a hard surface comes into collision with a softer surface plastic deformation occurs [1]. This deformation, produced in a casing when firing a projectile, is unique to each weapon in the context of ballistics. This means that the marks imprinted by weapons on the surfaces of a bullet or casing allow for the identification of the model of the weapon that fired it [2]. Figure 1a depicts two common bullet cartridge markings: the firing pin, which strikes the cartridge and triggers the firing of the bullet, and the ejector, which causes the ejection of the bullet cartridge. Figure 1b depicts where the parts referenced in Figure 1a are located in the weapon.



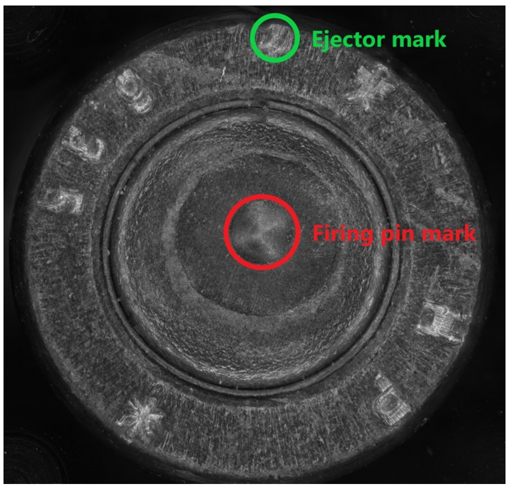
(a) Parts that may leave marks on the cartridge



(b) Pistol parts diagram [3]

Figure 1. Most relevant gun parts in the current context

The process of identifying the firearm model based on the cartridge case head is carried out by ballistics experts, who use microscopy to compare the marks found on the cartridge case head under investigation with marks found on reference cartridge case heads from multiple different gun models. Figure 2a portrays marks that are usually found in fired cartridges while Figure 2b displays a cartridge case head image captured with a microscope.



(a) Gun marks left on the cartridge after being fired

(b) Microscopy analysis

Figure 2. Fired cartridges analysis

With this method, examiners aim to find the gun model that produces the most similar marks to the ones found in the cartridge under analysis. It is a time-consuming process for the examiner, as it requires the handling of a variety of specialized equipment and the completion of numerous steps in order to collect and analyze the samples appropriately. Aside from that, the professionals must be well-trained on how to evaluate and compare specimens, as well as know what to look for when doing so.

Multiple strategies based on machine learning, image processing, and region of interest (ROI) extaction were proposed as ways of identifying the gun model that fired a given cartridge. For the development of any Machine Learning algorithm, a significant amount of data is needed to be able to train an accurate and generalizable model. For this purpose, the PCP has scanned multiple cartridge case heads as two-dimensional images and three-dimensional point clouds. With a dataset such as the one being presented, it should be possible to build a trainable model that could be integrated in a tool for assisting the examiner's work, by providing them with a list of the most likely firearm models for the gun that fired the cartridge under investigation.

1.1. Motivation

Spent firearm cartridges hold important information regarding the firearm that was behind the firing of the corresponding bullet. When a criminal shooting incident takes place, ballistic evidence in the form of spent bullet cartridges are collected when possible [4]. If that evidence cannot be retrieved from the scene, it can be photographed for further investigation.

According to Kara in [5], when cartridges are collected for investigation, firearm experts are frequently asked two questions:

- What firearm model was behind the shooting of the cartridge?
- What was the specific firearm used?

The answers to these questions make it possible to determine if there is a relation between two or more incidents that involve shootings.

To be able to identify the weapon model that was behind a shooting, ballistics research laboratories produce comparison marks on bullet cartridges by firing them in a controlled environment. The crime scene evidence is then compared to the reference control material as well as cartridges from other crime scenes [6]. The process of comparing reference cartridges from multiple weapon models to the cartridges found in a crime scene is what allows an examiner to draw a conclusion regarding the weapon model that was associated to the shooting. This process relies on the ability of ballistics specialists in being able to visually inspect and find relevant marks in the cartridges. It is also required that experts compare different sample cartridge cases under the microscope with the purpose of finding matching marks and, consequently, the matching weapon or weapon model.

According to the PCP, the manual identification of the firearm model that was associated to a shooting incident is only successful 16% of the time. This low success rate, associated with the fact that this is a very laborious task that requires experts to allocate multiple hours of their time, is what motivates the need for an automatic ballistic identification system adequate to the most common weapons used in Portuguese crimes. It is also important to mention that such system does not aim to replace the examiners' work but rather help them reach conclusions faster and with higher success rate.

1.2. Objectives

The main purpose of this research is to develop and train a semi-automatic firearm model identification system, using the cartridge case head image dataset provided by the PCP. This classification system could be subsequently integrated on a tool that would help Portuguese ballistics experts carry out firearm models identification, by providing them with a list of the most likely firearm models to have fired a bullet cartridge.

Since the dataset provided by the PCP is experimental and under development at the time of this dissertation's work, another dataset, the NIST Ballistics Toolmark Research Database (NBTRD), was used as benchmark for the developed technique. With this dataset, we aim to demonstrate the performance of the proposed technique on an established ballistic image repository.

The outcome of this work should be regarded as a proof of concept for a technique that aims to automatically provide examiners with a list of the most likely gun models to have fired a given cartridge given its image, consequently reducing the time spent on such task.

1.3. Research Questions

The identification of the firearm model that was behind the shooting of a given cartridge is a challenging process that yields an important piece of information for criminal investigation purposes. In the context of the development of a system that does this task automatically, it is important to answer the questions:

- Can an automatic ballistics classification system help improve the success rate achieved by the PCP's ballistic experts?
- Can the use of pre-processing on the provided data positively impact the automatic gun model identification?

1.4. Methodology

The development of the proposed work was carried out following an adapted version of the methodology proposed by Peffers et al. in 2007 [7] designated "Design Science Research Methodology (DSRM)". This methodology aims to provide guidance and a mental model for the presentation of the outcome of digital science research artifacts [7].

With this methodology, it was is possible to systematically identify and create a solution for the problem identified in this dissertation using a digital artifact. For this process, a number of steps proposed by Peffers et al. in [7] were followed:

- Problem identification and motivation;
- Defining the objectives for a solution;
- Design and development;
- Demonstration;
- Conclusions.

In Figure 3 it is possible to visualize the model proposed by Peffers adapted to the project's context.

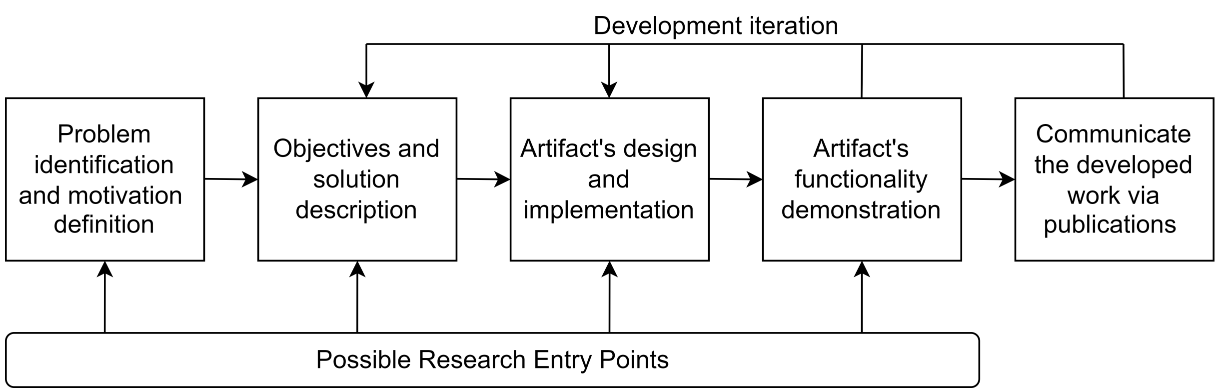


Figure 3. Adapted DSRM process model [7]

The original methodology proposes a development process that aims to iteratively improve the developed artifact by having it evaluated by experts, which would be the PCP examiners, in the last step of the iterative process and then modifying it according to their feedback. Since the end result of the carried out research was only a proof of concept for the used technique and not a finished product, the development iteration process was adapted to end with the prototype's demonstration and validation regarding its functionality.

This model shows that any given research can start off in either one of four possible starting points. The first step in this process has already been accomplished by identifying and describing the project's problem and motivation at the beginning of this chapter.

In the second step, the methodology states that objectives for the digital artifact should be established, which have been defined in section 1.2.

It is in the next step that the development of a solution starts. In this phase, the proposed solution starts to take shape and is incrementally improved over various iterations. In this case, the solution was developed using a predefined pipeline structure using both the PCP and NTBRD datasets.

After the development of a solution, it has to be put into practice. In step number four, the artifact built in the previous step was tested and its outcome was demonstrated by predicting test set images.

The final step in this process is to document and publish the research carried out so the scientific community can consider it in future scientific investigations.

1.5. Dissertation Structure

After defining the objective, methodology and motivation the dissertation for this work, it is organized as follows:

In **Chapter 2** a literature review is conducted regarding the state-of-the-art in ballistics identification tools and techniques. This literature review uses the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) methodology for systematic reviews and meta-analyses [8].

Chapter 3 describes the proposed solution. It provides descriptions for the used datasets, the classification approach, the image processing techniques as well as the used pipeline for the project's concretization.

In **Chapter 4** the performance of the developed technique is demonstrated for both datasets. A final evaluation is also carried out.

Chapter 5 is the final chapter where a conclusion regarding the research is presented. A discussion concerning future work that can be carried out to further fulfil the realized project is introduced as well.

CHAPTER 2

Related Work

Facing the research questions and objectives presented in sections 1.2 and 1.3 the presented related work aims to find the best approach to be able to accurately predict what gun model was behind the shooting of a cartridge. Additionally, reference datasets used in this context are also in the scope of this search to provide reference metrics for the developed technique.

Having this in consideration this related work shows what techniques have been employed in the context of ballistics identification with the objective of finding what is the best way to approach the stated problem.

2.1. Search Methodology and Criteria

A systematic literature review was conducted using the PRISMA [8] flow Methodology in order to find the answer to the question "What is the state of the art regarding automatic identification of firearms based on spent cartridges images?".

The papers to be analysed had to fulfil the following criteria:

- Source of paper: Conference Paper, Conference Review, Article, Review;
- Year: From 2010 to 2021;
- Language: English.

2.2. Research Query

In order to retrieve the publications related to this work, a query was formulated in order to get all the results regarding firearm model identification, the use of cartridge case images and the use of machine learning as well as image processing. As such, the search query used was: ("Weapon" OR "Firearm" OR "Gun") AND ("Cartridges" OR "Casings" OR "Shells" OR "Firing Pin" OR "Breech Face" OR "Tool Mark" OR "Ballistic") AND ("Classification" OR "Siamese Neural Network" OR "Neural Networks" OR "Convolutional

Neural Network" OR "Machine Learning" OR "Identification" OR "Artificial Intelligence" OR "Dataset" OR "Deep Learning" OR "Image Processing" OR "Image enhance" OR "Feature extraction").

The search for this literature review was conducted within the Scopus repository using the mentioned query. Google Scholar was also used for searches that were considered relevant, outside the scope of this query.

2.3. Selection of Studies

The selection of studies was conducted considering some important aspects of this work: Machine learning applications in the automatic identification of firearms based on images of spent bullet cartridges, other automatic techniques used for ballistics automatic identification and image processing techniques used to process cartridge images. In Figure 4 the results of this search are detailed in a number of documents included and excluded for each step.

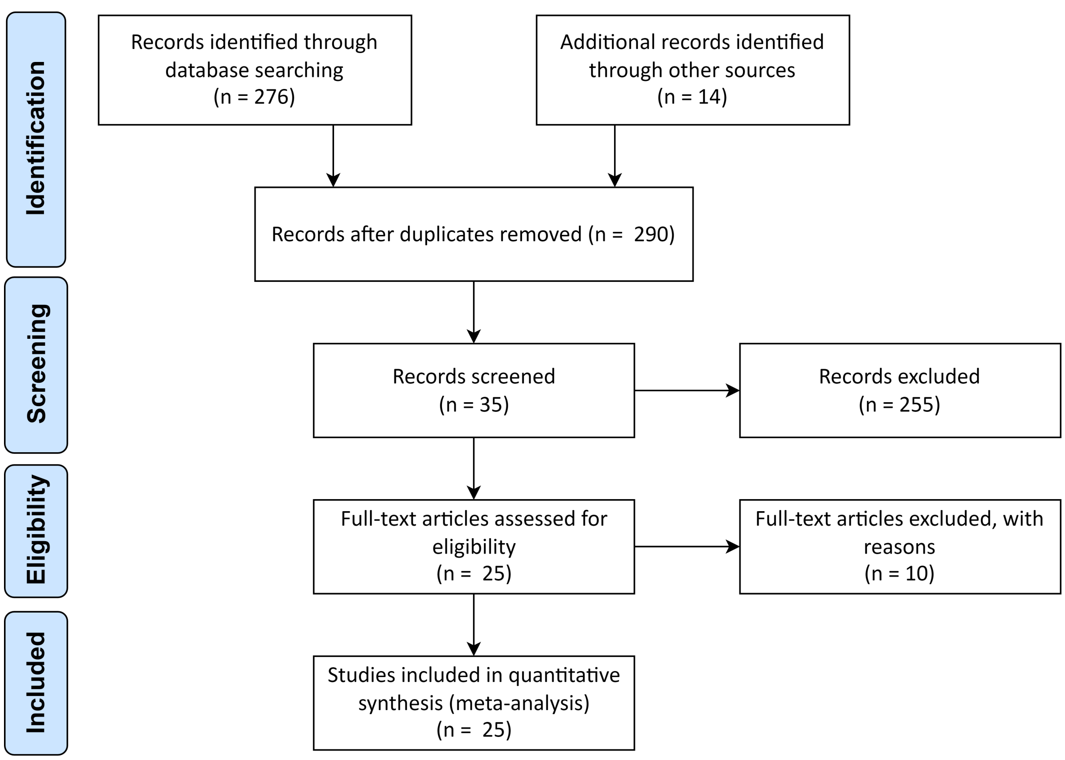


Figure 4. PRISMA workflow diagram (adapted) [8]

The first criteria to consider for document selection were its title and abstract. An additional full document analysis was carried out when that information alone was not conclusive. After selecting the studies for analysis, it is important to understand what topics studies were mostly focused on. As such, the VOSviewer [9] tool was used to make a map and analyse the bibliometric similarities within the analysed literature. Figure 5 shows a graphical representation of the bibliometrical map.

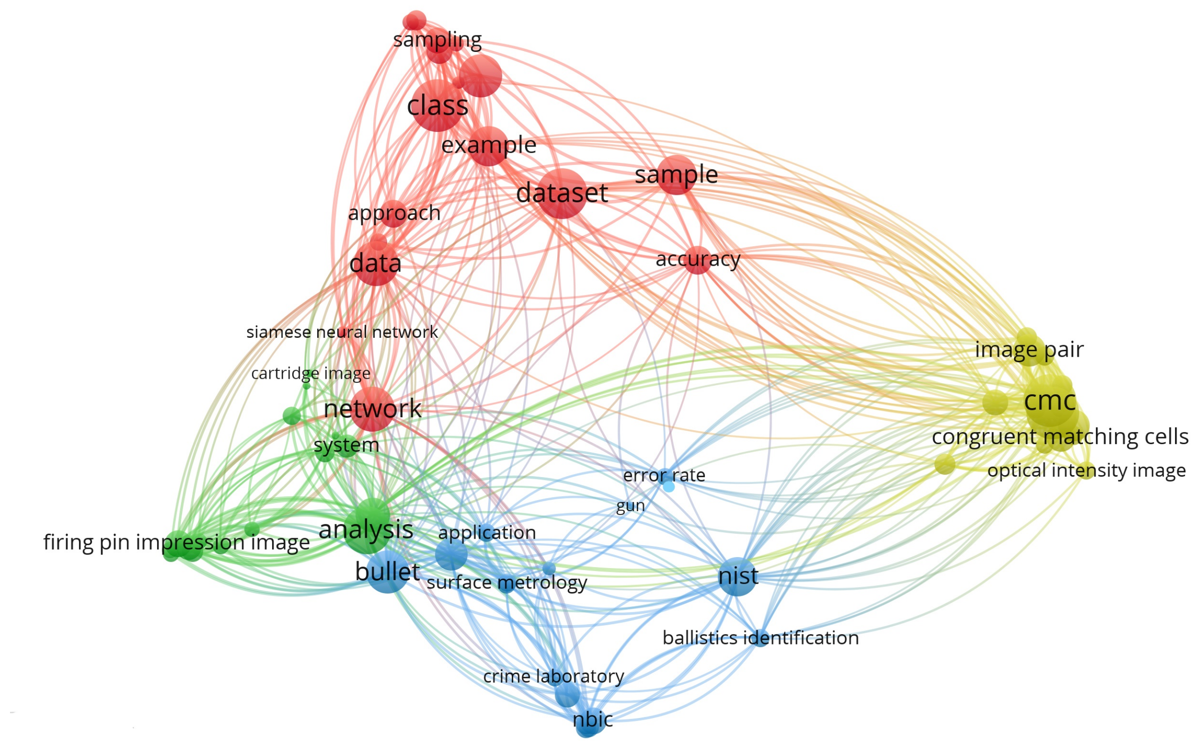


Figure 5. Graphical bibliometric representation [9]

By looking at the relative size of the nodes and cluster colors on the representation it is possible to understand what topics are mentioned in the literature and their relevance. With this in mind, we can see that the main topics are:

- Congruent matching cells method;
- The ballistics dataset;
- The content type of the used images;
- Laboratories influence on ballistics identification;

- Neural networks.

One of the topics that we were not able to identify in the collected literature were the tools that are already available in the market and partially accomplish what we're trying to achieve, in other ways. These tools are already in use in laboratories and help analyse and automatically identify firearms via ballistics imaging. Considering this, we found it is important to discuss these methods, although no evidence was found during the surveying process.

2.4. Literature Review

Congruent Matching Cells Method

In the reviewed literature, it was found that one of the most frequent methods used to automatically find what gun fired a bullet is the congruent matching cells (CMC) method. In [4], Tong et al. explains that the CMC technique was developed by the National Institute of Standards and Technology (NIST) [10]. This method uses the cartridge's 3D topography and aims to correlate small correlation cell pairs, instead of the whole cartridges. In the approach developed in [4], a reference breechface impression is divided into 7x7 cells, each cell being a correlation area, and each of these areas is compared to similarly sized areas. Figure 6 presents the described technique being applied to two breechface images, with the color representing topography values.

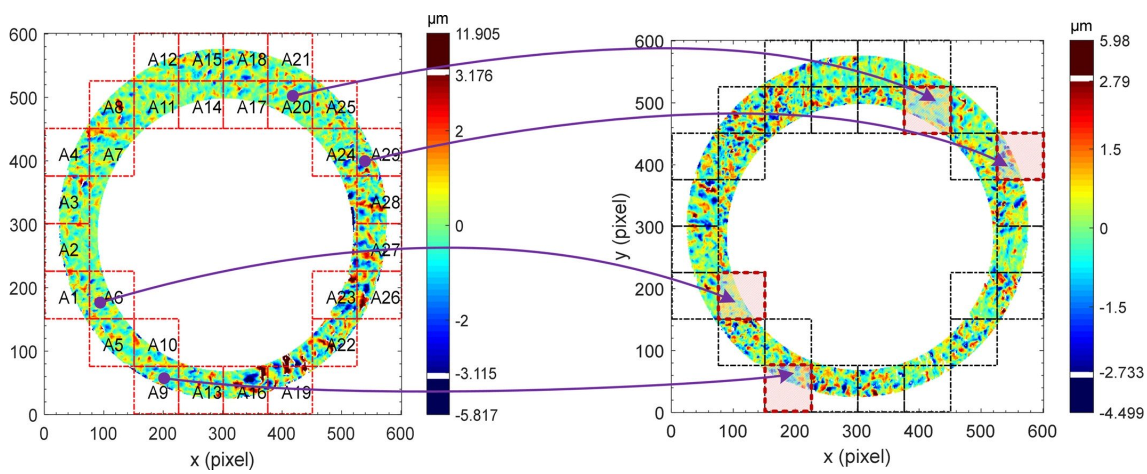


Figure 6. Breechface correlation cells [11]

For the correlation cells comparison, the cell of the known class image stays the same while the image under analysis is fully rotated, changing the correlation cell at each angle shift and making sure that the right correlation cell is not be missed during the correlation process. Each pair of correlation areas is then classified as valid (matching cells) or invalid (not matching cells) according to four parameters values. In [12], the same author improves the used technique by applying correlations at a common angle and making use of the correlation cells pairs in both directions to improve the identification capability. Figure 7 shows the improvement of the new technique correlation over the original technique.

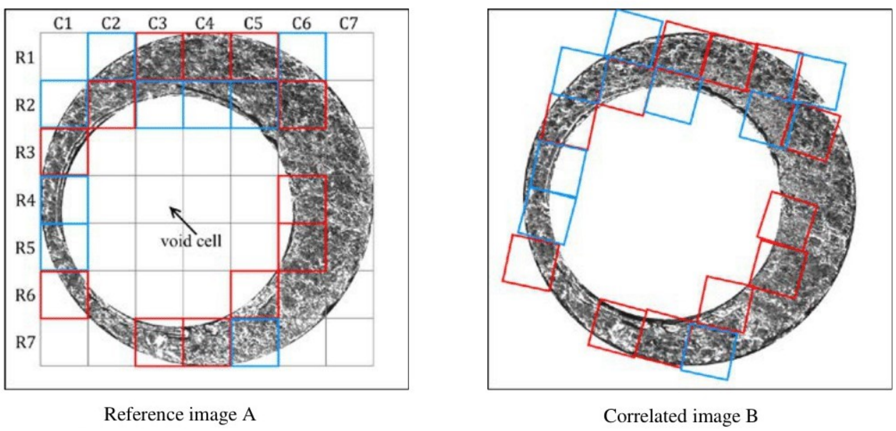


Figure 7. Improved CMC method [12]

Figure 7 shows in red the cells that represent the original technique, while the blue cells represent the additional cells that the improved CMC method could identify.

Besides the congruent matching cells identification problem, it is also essential to be able to classify the image pairs as matches (same firearm) or non-matches (different firearm). For this purpose, a numerical threshold, C, was suggested for cartridge case matching. This parameter as a CMC classification criterion was initially proposed by Song in [13] with a value of 6 and its value was kept for the research carried out in [12]. This criterion indicates that image pairs with 6 or more individual matching cells are classified as cartridge matches, while pairs with a count inferior to 6 are classified as non-matches.

Ballistic Imaging Quality Assurance

The CMC method was developed by NIST, who had also made efforts to establish a Traceability and Quality control system for ballistics applications and crime laboratories, with the National Ballistics Imaging Comparison (NBIC). This project was carried out in combination with the Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF) as demonstrated by Song et al. in [14]. The project consisted in the scanning of 2D and 3D measurements and the correlation analysis of the NIST Standard Reference Material (SRM), acquired by multiple crime laboratories based in the United States of America. Figure 8 shows one bullet example of the standard reference material used for evaluation over different laboratories:

Figure 8. SRM 2460 bullet [12]

This was done in order to establish quality assurance across the laboratories. Vorburger et al. describe in [15] that with the second NBIC project, the process of acquisition and correlation of the SRM is done again and its results collected and analysed by NIST to define control parameters and limits, with the objective of assuring the compliance with the ISO 17025 Standard.

The NIST has also developed an open-access Ballistics Toolmark Research Database (NBTRD) [16] where scanned 3D and 2D data from multiple cartridge cases and toolmark surfaces is available for researchers to be able to conduct their experiments and methods and compare results. This database eliminates the need for researchers to create their own datasets, which is an expensive and time-consuming process, and allows new algorithms and methods to be objectively evaluated.

Image Processing

Regarding Image processing techniques, Gerules et al. [17] state the importance of the use of image preprocessing techniques for correction of acquisition defects or image enhancement for use on other algorithms. Some techniques used for this purpose are described, those include: Noise reduction techniques which smooth the image, such as the Gaussian Kernel [18]–[20]. Other authors have also used the median filter for noise generated by some types of sensors [21]. The author also states that the images are processed and their background is removed to extract the relevant parts of the image. For this purpose, automatic edge detection methods that rely on sharp changes in intensity within the image such as the Sobel operator and the Canny Edge were mentioned [22]–[24].

In [25], Huang et al. develops a binarization algorithm for edge detection and compares it to previously proposed algorithms: Otsu [26], Chow and Kaneko [27] and Yanowitz and Bruckstein [28]. The results of the proposed algorithm are shown in Figure 9:

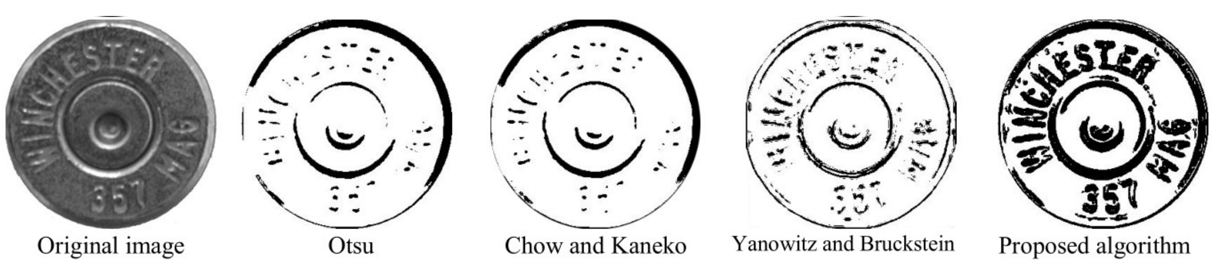


Figure 9. Comparison of the algorithm proposed in [25]

The proposed algorithm, which is shown in Figure 9, outperforms the algorithms it was tested against. The same author, states in [29] that to extract relevant features from cartridges images, it is necessary to first process them. It is also declared that, to do so, it is required to binarize the images. In their analysis, three edge detection operators are compared and it is found that the combination of the Sobel operator and the Canny operator yield the optimal results regarding edge detection.

In a study conducted by Kara [5], the author uses the Turkish BALİSTİKA 2010 system to compare similarities and differences between cartridges using firing pin impressions, capsule traces, and the combination of these areas. The results of the comparison between the different parts of the cartridges suggest that the firing pin impression is the most effective of the three for ballistics classification.

Machine Learning Classification

Regarding the use of Machine Learning in the identification of firearms it was found that in 2011, Kamaruddin et al. [30], tried using firing pin based geometric moments proposed by Ghani et al. [2] in 2010 to train a back propagation neural network with the "trainlm" algorithm and a 6-7-5 architecture, achieving 96% accuracy on firearm classification. One year after that development, Leng et al. [29] proposed a novel method for feature extraction called the "circle moment invariants". They then used the outputs of this extractor as features to train a 3 layer backpropagation Neural Network, obtaining a 98% accuracy for firearm identification.

Recently, Giudice et al. [31] suggested the use of breech face only images, generated from 3D point clouds, as input for a siamese neural network. This network showed positive results for a Top-N probability based metric.

Available Ballistic Identification Systems

In the book "Handbook of firearms and ballistics: examining and interpreting forensic evidence" [32], Heard mentions a list of ballistic identification systems that are available in the market, namely:

- ARSENAL by Papillon Systems of Russia;
- EVOFINDER by SCANBII Technology;
- FIREBALL from Australia;
- CIBLE, a French system;
- TAIS, a Russian system;
- BALLISTIKA from Turkey.

Heard also states that these systems cannot replace the examiners that do the ballistics comparisons. These systems generate a list of the most likely candidates as possibles matches. With these results, experts still have to manually compare and analyse the possible matching cartridges to the cartridges being analysed. This being said, similarly to the system that is being proposed, the ultimate decision on whether there is a match still has to come from the examiners [32].

Summary

The previously analysed techniques show that there are many ways one can approach a ballistic identification problem. There is software in the market that can be used for ballistic analysis and identification, as well as other techniques such as the Congruent matching cells technique or machine learning algorithms, which have also shown potential. In this literature review, it was possible to observe that only one siamese neural network technique has been applied in this field in recent years, using 3D point clouds. This shows that there might be unknown potential in this approach, considering the use of images instead of three-dimensional data.

2.5. BALCAT Project

The work presented in this dissertation was developed within the scope of the BALCAT project, commissioned by the Portuguese Criminal Police. The BALCAT project consists in the creation of a ballistics classification tool that is being developed by INOV - Instituto de Engenharia de Sistemas e Computadores Inovação, with who we are collaborating with by developing the presented cartridge images' classification algorithm.

The BALCAT project can be divided into three main phases:

(1) Image acquisition and labelling phase;
(2) Ballistics classification technique development;
(3) Deployment into production.

In the image acquisition phase, an application was developed by INOV for the acquired scans to be uploaded, segmented and the relevant parts of the images annotated. Although no part was taken in the development of this interface, it was necessary to do some of the annotations since there were not enough annotations to proceed with the presented work.

The second phase is where the work being described in this dissertation comes in. For this stage, an automatic ballistics identification method was developed with the capacity of returning a list with the most likely firearm models to have fired a given cartridge.

At last, it is necessary to deploy this network into production at the PCP's headquarters. For this task, it is intended that the developed technique is integrated into an application that can be operated by the end user, developed by INOV. Besides having the pre-trained model to classify the images, it is also planned that this tool has the ability to further train this model as more data is uploaded into the system.

Regardless of the accuracy of the developed tool, it should be viewed as a technique for optimizing the examiners' search by providing indications on which cartridges to target, and not as a substitute for the examiner's work.

It is also important to note that the work presented in this dissertation is only one of the parts that make up the BALCAT project. This project involves other tasks that were outside the scope of this dissertation, such as the gathering of data and its annotation as well as the creation of the tool that can be used by the examiners. With this in mind, this dissertation presents the development and test process behind the method used to classify the images that are uploaded into the final system.

CHAPTER 3

Design and Development

This chapter presents the development process of a siamese neural network architecture with the ability to classify cartridge case images. From the literature review presented in the previous chapter, the most promising machine learning based approach for firearm model identification uses a siamese neural network whose inputs are 3D point clouds [31]. Since this technique achieved good results when compared with other state-of-the-art algorithms, it should be worth to apply a similar concept to the two dimensional imagery domain and to evaluate its outcome.

On the other hand, due to limitations on the amount of available images, the use of SNNs is potentially a good fit for the application since these networks usually require a smaller number of images for training when compared with traditional classification approaches [33]. Thus, the PCP has digitized a portion of its archives with the purpose of building a tool that is able to help them with the identification process.

Since the PCP's dataset is experimental and has not been tested before in other applications its effectiveness cannot be taken for granted. In the current state of the art it was also found that the NIST had setup and published an open source dataset, the NBTRD [16]. Furthermore, it was used by several other scientific studies, making it a relevant dataset to use as reference. The works [1] [31] [34] [35] [36] [37] are examples of studies that have used this dataset to carry out their ballistics related scientific investigation. With this in mind, this dataset will also be described and used as a benchmark for the proposed technique's performance.

3.1. Image Acquisition and ROI extraction

PCP's Dataset

Regarding the collection of the images gathered by the PCP, the ToolScan imaging system

[38] was used to acquire 2D and 3D data of several fired cartridges in a single scan. The 2D scanning of the cartridges results in a matrix of casings as illustrated in Figure 10.

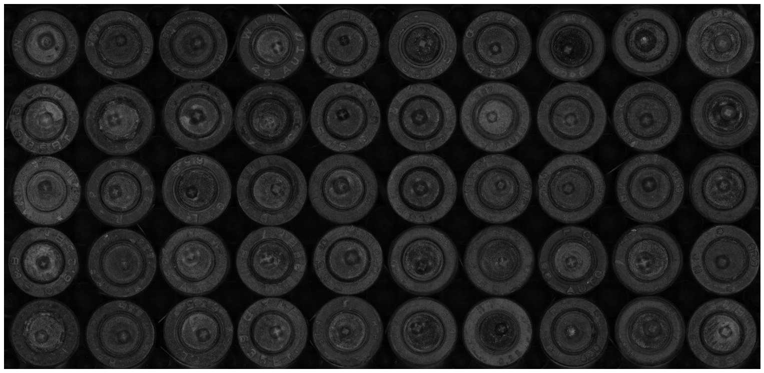


Figure 10. PCP’s scan output image

These images then needed to be segmented into multiple singular cartridge case head images for it to be useful in a neural network training approach. To address this issue, a threshold was applied to the images to binarize them, and then the Hough Circles algorithm was used to determine the positions of the cartridges. Figure 11 shows an example with the cartridges’ suggested final positions.

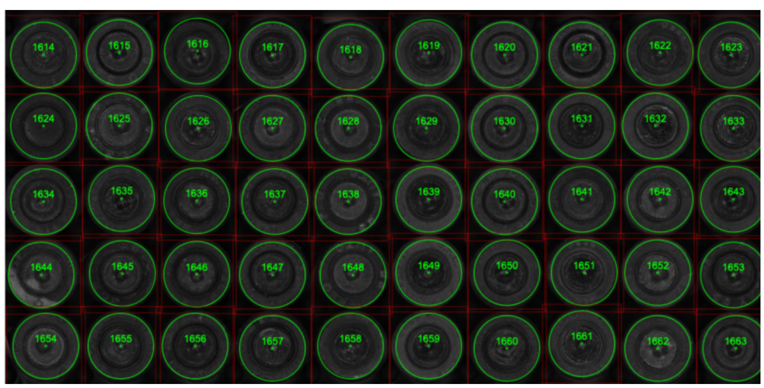


Figure 11. Proposed cartridges positions

For the segmentation task, a 10% size increase was applied to the circle’s boundaries and the corresponding bounding box is cropped according to each circle’s size and position. The identification of the cartridges’ position (using binarization and the Hough Circles method) and their segmentation were both done with OpenCV[39]. In addition to this acquisition, PCP examiners also annotated the majority of the images’ cartridge case outline, breechface impression, ejector mark and firing pin impression.

All the mentioned work regarding the scanned images' segmentation was carried out by researchers at INOV who have shared and allowed the publication of the developed dataset.

NBTRD

The database developed by NIST also includes 2D and 3D data from cartridges. According to Zheng et al. in [16], the 3D data was collected using a disc-scanning confocal microscope. For firing pin impressions, a 20X objective was employed and 10X objective was used for the breech face.

The 2D data was collected using a stereo microscope, which allowed the rendering of 2D images from different points of view. A 4X objective was used to collect firing pin images while a 2X objective was employed for the cartridge case breechface impressions. Besides NIST, other entities have also uploaded data but for the scope of this project only NIST collected images were used. Figure 12 shows example cartridge images from the NBTRD.

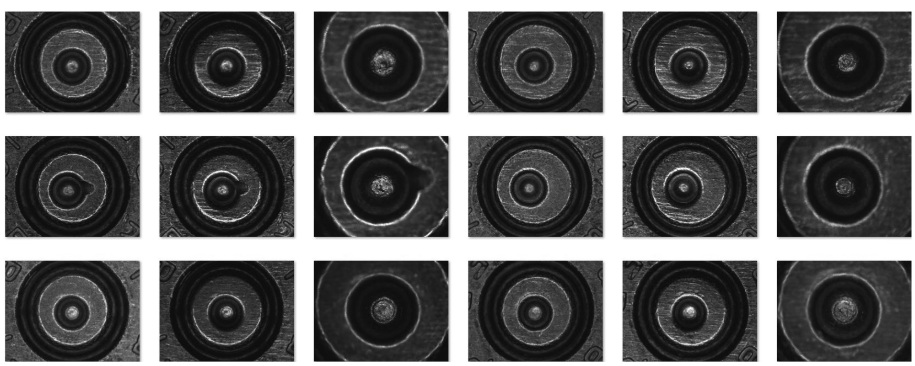
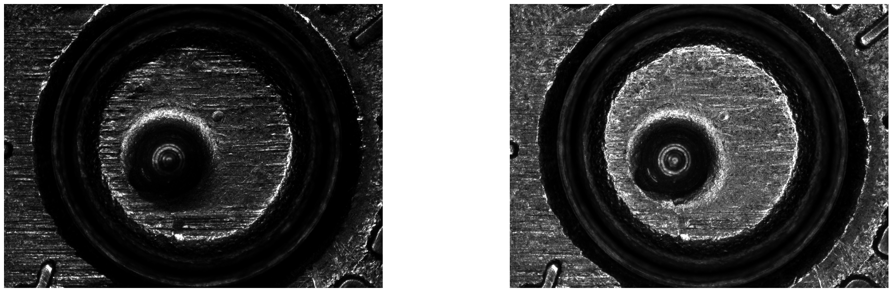


Figure 12. NBTRD example images

Since there are very few firing pin images for each gun model, only breechface images were considered for this part of the project. Every breechface image was acquired under two different light conditions, which enhance the image in distinct ways. Figure 13 shows the difference between image acquisition under a ring light and a 6 o'clock light. From this image it can also be observed that, while the firing pin impression is not in focus, it is possible to observe its impression outline.



(a) Ring Light image

(b) 6 o'clock light image

Figure 13. NBTRD cartridge images captured under different light conditions

3.2. Dataset Characteristics

PCP's Dataset

After the images were segmented and labelled the dataset ended up width a total of 1295 images distributed throughout five different classes (five different gun models). Table 1 shows the number of cartridges per class and how many unique weapons were used to fire those rounds.

Table 1. PCP's base dataset characteristics

Gun model	Number of images	Unique firearms
GT28	746	235
P6	150	46
315 Auto	150	38
950B	149	42
Baby	99	30

It is possible to understand that the dataset is significantly unbalanced, especially due to the large amount of "GT28" cartridge samples in contrast to the other classes. Furthermore, both the number of images and the firearm variability within each class is limited, taking into consideration that reference machine learning datasets such as the MNIST [40] or ImageNet [41] often contain classes with thousands of images.

Additionally, the dataset must be divided so that the Training Set, Validation Set, and Test Set do not share any unique guns. Not doing this could cause the network to learn characteristics of the weapons and not generalize the learning to the weapon model.

Another factor that was considered is the number of available annotations. Considering the low amount of images, it was important to focus the network's training on the most relevant parts, which has been suggested to be the firing pin impression in a study conducted by Kara in [5]. Table 2 shows the number of annotations for each class.

Table 2. PCP's dataset number of annotations

Gun model	Breechface	Firing pin	Ejector
GT28	741	740	631
P6	80	79	4
315 Auto	146	148	88
950B	70	70	0
Baby	40	40	0

From Table 2 it is possible to observe that, since the dataset is not yet in a finished version, not all of the images have annotations for the ejector, breechface and firing pin. These are important components to annotate due to the fact that not every part of the cartridge image holds relevant information to its classification. Due to the low amount of samples, the training should be more focused on the important parts of the image. Figure 14 demonstrates the different possible annotations in the dataset collected by the PCP.

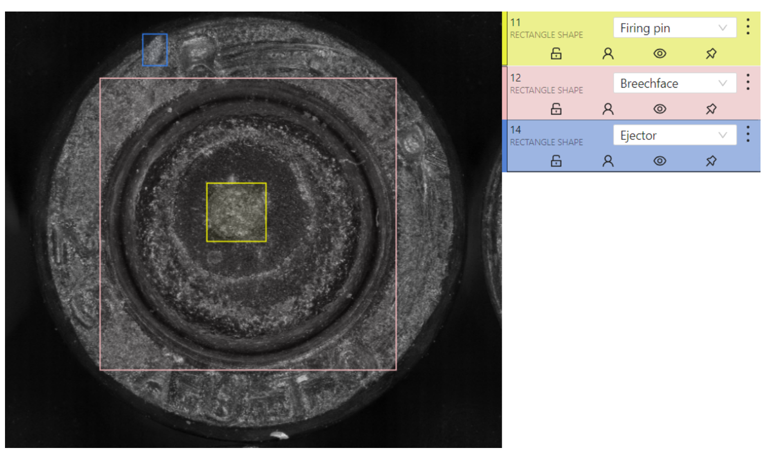


Figure 14. Annotated cartridge image

NBTRD

The NBTRD, mostly consisting of data collected by NIST, is an open source dataset containing 3D data as well as 2D data. For the purpose of this dissertation, the number of cartridges collected from that dataset were arranged with the objective of being smaller than or equal to the PCP's dataset classes' frequency. The collected dataset number of images per class is described in Table 3.

Table 3. NBTRD characteristics

Gun model	Number of images	Unique firearms
Ruger P95DC	100	10
Ruger P9PR15	80	10
SW 10-10	72	12
Hi-Point C9	60	10
SW 40VE Sigma	60	10

Table 3 shows that similarly to the PCP's dataset, this collection of images is also unbalanced and even more limited in both quantity and variability. This way, the obtained results can confirm whether the dataset size has a significant impact on the network's performance. Contrary to the homologous dataset, the images from the NBTRD do not have any annotations associated. Similarly to what was done for the PCP's dataset, it must be divided so that the Training Set, Validation Set, and Test Set do not share any unique guns.

3.3. Proposed Classification System

Unlike conventional neural networks, a Siamese Neural Network (SNN) does not directly predict classes for a given input sample, instead, it takes two inputs and outputs the probability of both belonging to the same class, according to Koch in [42]. Since the current problem deals with images, the followed approach uses two identical parallel Convolutional Neural Networks (CNN) to process the input images and output the match probability.

The network training aims to reduce the distance between the outputs of the convolutional networks, for images of the same class, while increasing it for different class images. This sort of network is advantageous since it operates by computing a probability for pairs of images rather than individual pictures, allowing the same data to be utilized several times by pairing the dataset in different ways. Figure 15 demonstrates the general architecture of an SNN based on CNNs that was used in this problem, adapted from [42].

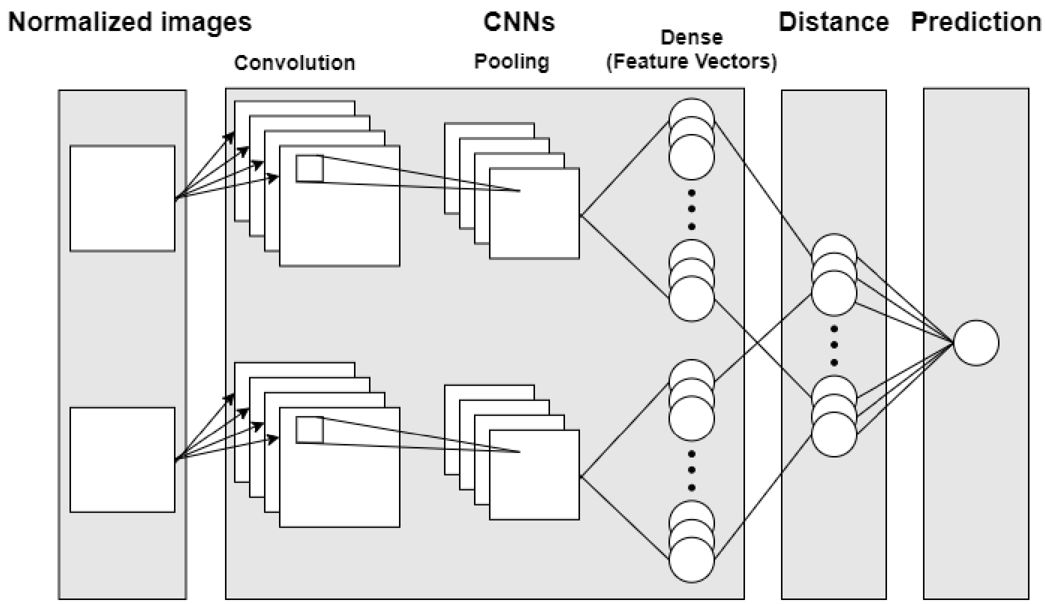


Figure 15. Proposed SNN architecture in [42] (adapted)

Since SNNs do not directly classify the images' classes, a way to do predictions needs to be applied. With this in mind, for any technique used for this purpose, it is necessary to first select a set of reference images from each gun model for comparison. After this selection two methods were tested: Distance classification and Probability classification, which are going to be described.

Distance Classification

Assuming that any two images of the same class should have a low distance between them on a well performing SNN (measured at the output of the CNN part of the SNN), the cartridge under analysis can be classified by measuring the distance between the CNN feature vector of the mentioned image to the feature vector of every reference cartridge (computing the distance). This way, it is then possible to classify it as the class of the images with the average lowest distance or output a list ordered by relative distance to each class. Figure 16 demonstrates how this process was carried out.

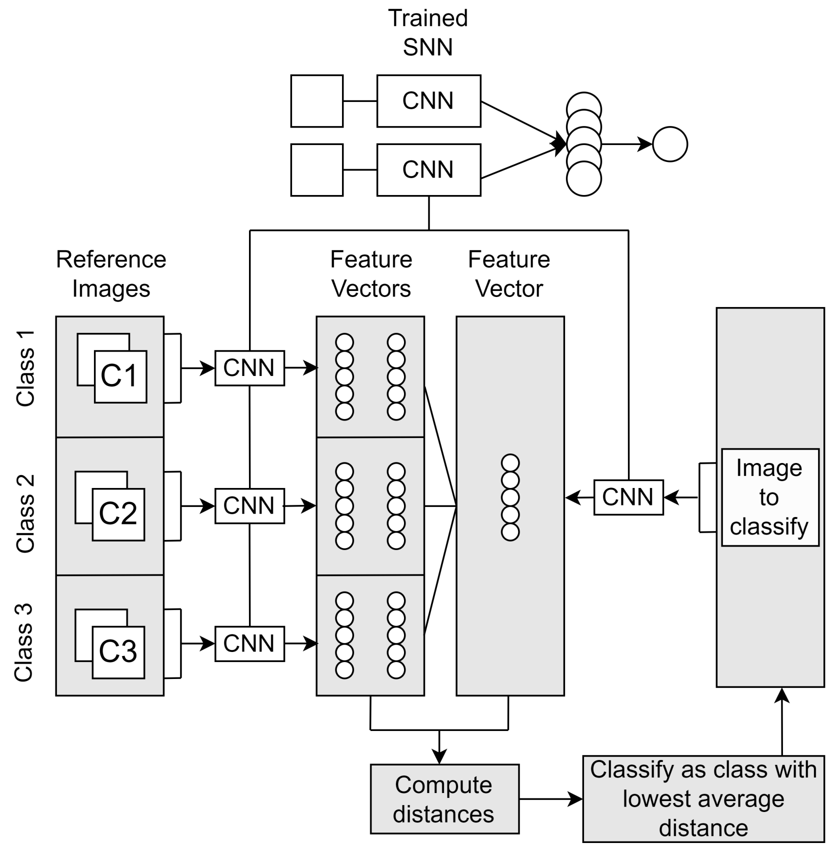


Figure 16. Proposed distance classification method

Probability Classification

Similarly to the distance classification, the probability classification compares the output of an image to the output of multiple reference images. As previously mentioned, an SNN outputs the probability of a match. With this in mind, the image to classify is paired with every reference image and it is possible to form a list with the average matching probability per class. The most likely class to be a match will be on the top of this list and therefore it will be considered as the predicted class. Figure 17 demonstrates how the probability classification process was implemented.

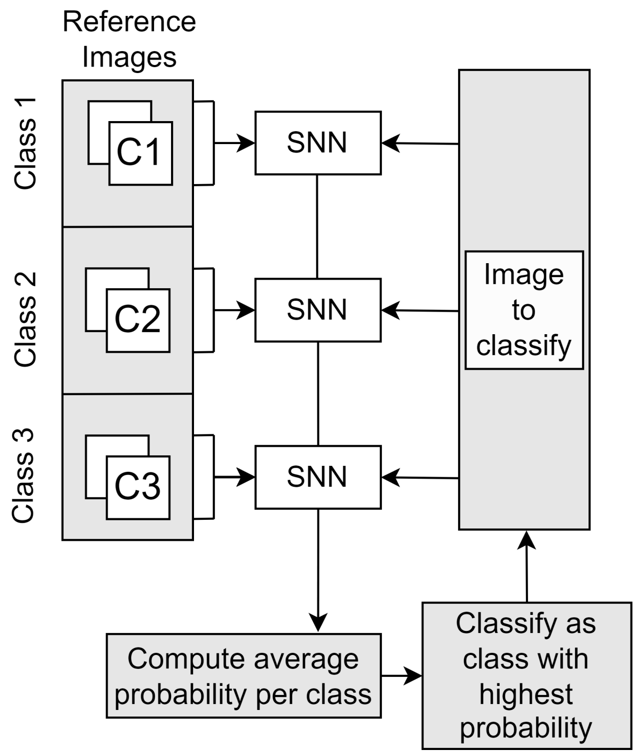


Figure 17. Proposed probability classification method

3.4. Proposed Pipeline

In order to methodologically develop the most adequate SNN for the problem at hand, a procedure pipeline was defined for this work. Since an SNN is composed of two equal CNNs, it is suggested that this pipeline is used to develop a well performing CNN and, by removing its classification layer, use it as a feature extractor for the structure of the target siamese neural network, which was then retrained.

Figure 18 depicts the proposed pipeline. This pipeline shows the followed process used to find a CNN with a good performance, within the tested parameters. For each phase, the best approach is chosen and used in the following step.

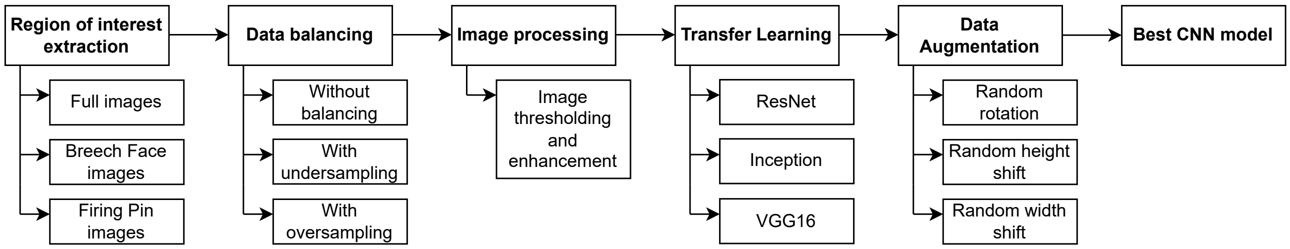


Figure 18. Proposed pipeline for the PCP's Dataset

Region of interest extraction

The cartridge case head images needed some consideration regarding the portion of the image that was used for training and classification. For this reason, the first stage of the pipeline evaluates three possibilities: breech face image region, firing pin image region and full image. The missing firing pin impressions and breech face impressions were annotated in order to increase the data available for this assessment. The ejector mark had to be discarded for this project due to a lack of information and consistency while annotating it.

Data Balancing

One of the problems present in both datasets is the imbalance between classes frequency. This presents an issue since most classification algorithms tend to have a high bias toward the majority class [43], which results in a tendency to classify more images as the most frequent class and, therefore, decreasing performance. Considering this issue, the developed pipeline presents two data balancing techniques that were tested and compared to a network trained on unbalanced data.

One of the tested techniques is **Over-Sampling**. This technique, presented by Japkowicz in [44], consists in randomly duplicating images from the minority class (possibly applying preprocessing function to differentiate the image) until it reaches the same amount as the majority class. This causes the overall dataset to grow, at the expense of having a considerable amount of duplicates.

The other technique that was tested is **Under-Sampling**, which is also mentioned in [44]. This technique consists in randomly eliminating samples from the majority classes until the amount of examples equals the minority class. This approach ensures that no duplicate data is present at the expense of a considerable loss of data in the predominant classes.

After the two methods were tested, the one that produced the best performing CNN was used to test the following techniques.

Image Processing

The majority of machine learning problems include data preparation techniques before feeding it to networks. Images, for example, can be preprocessed using computer vision techniques that extract features, remove noise, highlight regions of interest, and facilitate the generalization process of the models to get superior performances.

Due to the lack of images in the current problem, an image processing technique was used with the objective of highlighting the regions of interest and removing parts of the images that do not contain relevant information. This way, the training can be focused on the most relevant parts of the image.

On the other hand, CNNs are developed and trained to learn optimized image filters that, when applied to the images, should highlight key features in the images, as the network advances into deeper layers [45]. Having this in consideration, if a CNN is already trained to find the best filter parameters for an image, using image processing may not be useful.

Transfer Learning

The ideal scenario for a regular machine learning application is when there is a large number of labelled examples with equal distribution. However, for a dataset to have the mentioned characteristics is sometimes very costly, time-consuming, or even impossible to get [46]. Typically, gathering adequate data is difficult and it is often not equally distributed, which is the case for the datasets used in the context of this work.

In real use case scenarios datasets are often not perfect, either due to imbalance or shortage of datapoints. Transfer Learning (TL), which focuses on knowledge transfer across domains, is a potential machine learning paradigm for addressing this challenge, as it usually uses reference neural networks to learn from standard large and diversified datasets. This already obtained knowledge is then transferred to other problems by training new layers on top of the network with the least ideal dataset, keeping the pre-trained weights, a technique called fine tuning.

It is important to note that transferred knowledge does not always have a positive impact on new domains, especially if there is little in common between areas [46].

In this application, three deep learning models were tested: ResNet [47], Inception [48] and VGG16 [49]. The models were chosen based on the fact that they have been widely used for general classification applications and their performance is frequently used as a benchmark for other architectures [50]. Additionally, although these designs all have different architectures, they all share the same basic working principle, being made up of convolutional, pooling and other types of layers while classifying at the image level [51].

Data Augmentation

When adopting convolutional neural networks, one of the most common preprocessing applications is data augmentation (DA). DA is a technique that consists of artificially expanding and diversifying a dataset. It generally improves model performance, contributes to the overall data's heterogeneity and aims to provide improved generalization on the trained model [52]. This operation works by taking the training images and applying different image transformations before every training epoch. These transformations can include flipping, scaling, rotations, width and height shifts and other operations like brightness.

Some examples of data augmentation techniques such as rotation, height and with shift applied on the NBTRD can be observed in Figure 19.

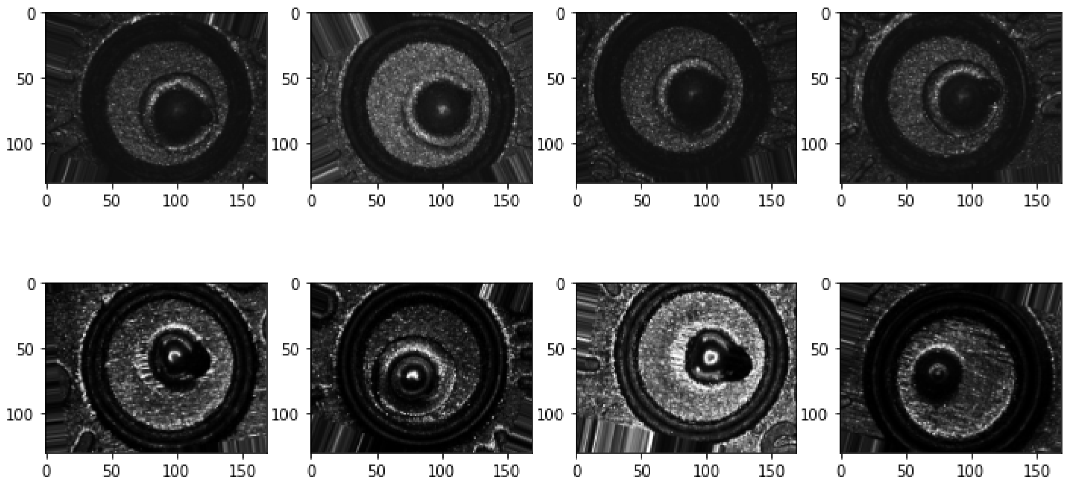


Figure 19. Rotation, height and width shift data augmentation on the NBTRD

This procedure is usually beneficial for studies that have smaller datasets and do not have the ability to generalize well, such as this one. Data augmentation also allows their models to be trained on the same data multiple times, delaying the time it takes to reach a state of overfitting and generalizing better.

Despite data augmentation approaches that enrich a dataset with label-preserving changes, hundreds of datapoints are generally still required for the successful training of a deep neural network, depending on the complexity of the subject under investigation [52] (i.e. number of classes, image complexity, etc.).

CHAPTER 4

Experimental Results

In this chapter, the steps that were carried out to follow the pipeline proposed in section 3.4 are presented, as well as the parameters and techniques used to obtain the results. Over the course of this portion of the dissertation, some determinations are stated regarding what approach should be followed in each step of the pipeline. A section highlighting final remarks concerning the obtained results is also presented at the end of the chapter.

For the development of every CNN that is presented in this chapter, the python library “Keras Tuner” [53] was adopted. This library allows for the automated tuning of the network’s hyper parameters within a set of predetermined bounds. After the ideal hyperparameters have been found, the best performing model is returned (already trained).

For the developed networks’ hyperparameter tuning, the parameters presented in Table 4 were used for both datasets.

Table 4. Parameters used for hyperparameter tuning

Nº Layers	Layer	Filters/Neurons	Filter Size/dropout	Activation
1-5	Convolutional	4-64	2-4	relu
1-5[a]	MaxPooling	-	1-3	relu
1	Flatten	-	-	-
1-3	Dense	2-256	-	sigmoid or relu
0-3[b]	Dropout	-	0-0.3	-
1	Dense (classification)	5	-	softmax

[a] Interspersed with convolutional layers
[b] Interspersed with dense layers

Throughout the experiments an image size of 150x150 px was used. This value was found to achieve a good balance between the hardware memory limitations, the network’s training speed and the visible detail on the images.

After the development of the CNNs, the SNNs were constructed and the two evaluation methods, the distance method and the probability method, were compared. In the end, the results for both datasets were evaluated and a discussion regarding the performance of the datasets is presented.

4.1. Region of interest extraction

For the first section of the work, only the PCP's dataset was considered, since the NBTRD does not have annotations to enable the segmentation of the images. In this first part, a custom network was built and trained for each of the possible regions. Since not every image was annotated with the breech face and the firing pin impressions, there was still some work in this regard.

For the breech face images, since the coordinates for the center of the cartridges were known (by using the Hough Circles method mentioned in Section 3.1) and because the breech face is centered relative to the cartridge, it was possible to get this annotation for all the images, by applying a fixed radius at the center of the cartridges.

For the firing pin region extraction it was not as simple, due to the fact that the firing pin is not at a fixed point. Therefore, the firing pin impressions that were annotated were automatically segmented, while the others were segmented by hand.

With the techniques mentioned above, it was possible to have firing pin impression images and breech face images for all of the labelled data.

An example for each one of these images can be seen in Figure 20.

(a) Full image | (b) Breech face image | (c) Firing pin image

Figure 20. PCP's example images used in the first phase of the pipeline

Although the data for this step is unbalanced, the test and validation data do need to be balanced, otherwise, the results would not be reliable. This happens because, if there is a majority class in the test or validation data and the network is biased towards the majority class, the accuracy of the network can be inaccurately high. Due to the fact that there was also a need to do these splits in a way that unique firearms are not shared between them, it was not possible to keep the validation and test sets perfectly balanced. With this in consideration, the training, test and validation data frequency can be seen in Table 5.

Table 5. Number of images per set for each class

Number of images for:	315 Auto	950B	Baby	GT28	P6	Total
Training	108	105	58	700	100	1071
Validation	28	30	27	29	32	146
Testing	14	14	14	17	18	77

After running the hyperparameter tool on the splits presented in Table 6 the accuracies and losses achieved for each region were the following:

Table 6. Accuracy and loss results for region of interest extraction on the PCP's dataset

Image type	Validation loss	Test loss	Validation accuracy	Test accuracy
Full images	1.80	1.82	40%	37%
Breech face	3.11	3.31	44%	47%
Firing pin	1.34	1.73	65%	53%

Table 6 shows that every result showed a relatively high loss, especially for breech face images. Overall a slight increase in accuracy can be noticed from the full images to the breech face, although the losses are also higher. The best results were achieved for the firing pin images, where a significant increase in validation accuracy (less noticeable in the test accuracy) and a slight decrease in loss can be observed. These results confirm the conclusions drawn by Kara in [5], where the firing pin images showed the most promising results for firearm identification.

From this experiment's results, the following tests with the PCP's dataset were conducted considering the firing pin images only.

4.2. Data Balancing

In this section, two data balancing techniques were tested and compared to the datasets without any data balancing. This allows to understand if there is any advantage in using balanced datasets over the unbalanced datasets. Regarding data balancing approaches, an under sampling technique was applied to the dataset. For this method, each class contains about the same amount of images as the class that contains the lower amount of samples. An oversampling technique was also applied. For this approach, each classes' images were randomly duplicated until their frequency is similar to the majority's class frequency. Both techniques were presented by Japkowicz in [44].

In this section, the networks were trained and tuned using the keras tool with the parameteres depicted in Table 4.

It is also important to mention that the training set, validation set and test set do not share different samples associated to the same firearm. In other words, different samples coming from the same gun are assigned to only one of these sets.

PCP

The training, validation and test set splitting setups for the PCP's dataset are presented in Table 7.

Table 7. Number of images per set for each class considering data balancing on PCP's dataset

Type	Number of images for:	315 Auto	950B	Baby	GT28	P6	Total
No balancing	Training	108	105	58	700	100	1071
	Validation	28	30	27	29	32	146
	Testing	14	14	14	17	18	77
Undersampling	Training	74	74	74	74	74	370
	Validation	13	17	16	15	15	76
	Testing	10	10	9	11	9	49
Oversampling	Training	700	700	700	700	700	3500
	Validation	28	30	27	29	32	146
	Testing	14	14	14	17	18	77

Table 8 shows that the undersampling technique achieved the best overall results. It reached a lower loss value and an higher test accuracy, performing better than the unbalanced and oversampled data, which has a very high loss.

Table 8. Accuracy and loss results for PCP's dataset using data balancing techniques

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
No balancing	1.34	1.73	65%	53%
Undersampling	1.26	1.27	63%	57%
Oversampling	2.8	4.67	62%	47%

These results made sense as an unbalanced dataset could cause the model to have a tendency towards classifying images as the majority class, lowering its performance. Alternatively, oversampling produces an even lower performance. These results might be due to the fact that by oversampling, the network is essentially training on the same images multiple times, not introducing any variability. In this case, where the oversampling is high, multiplying the images around 7 times for some classes, the results show that this approach is not beneficial. Considering these results, the next experiments with PCP's dataset were carried out considering the undersampled data.

NBTRD

For the NBTRD images to keep their aspect ratio of 4/3, but hold about the same amount of pixels as the PCP's dataset, the size was set to 170x130 px.

The dataset spliting strategies presented in Table 9 were done so they resemble the same procedure applied to the PCP's dataset splits.

With the NBTRD it was found that the validation and test accuracy was the same throughout all the experiments. The only differentiating factor was the value of the loss function, which was very high for the undersampled test set and the lowest in the unbalanced set. Taking this into account, the next tests with the NBTRD made use of the unbalanced dataset, which showed the best performance.

Table 10 shows the performance of the network for each of the tested sampling techniques on the NBTRD dataset.

Table 9. Number of images per set for each class considering data balancing on the NBTRD

Type	Number of images for:	P95DC	P9PR15	10-10	C9	40VE	Total
No balancing	Training	82	62	54	42	42	282
	Validation	12	12	12	12	12	60
	Testing	6	6	6	6	6	30
Undersampling	Training	42	42	42	42	42	210
	Validation	12	12	12	12	12	60
	Testing	6	6	6	6	6	30
Oversampling	Training	82	82	82	82	82	410
	Validation	12	12	12	12	12	60
	Testing	6	6	6	6	6	30

Table 10. Accuracy and loss results for the NBTRD using data balancing

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
No balancing	0.00010	0.46	100%	93%
Undersampling	0.00054	10.51	100%	93%
Oversampling	0.034	0.56	100%	93%

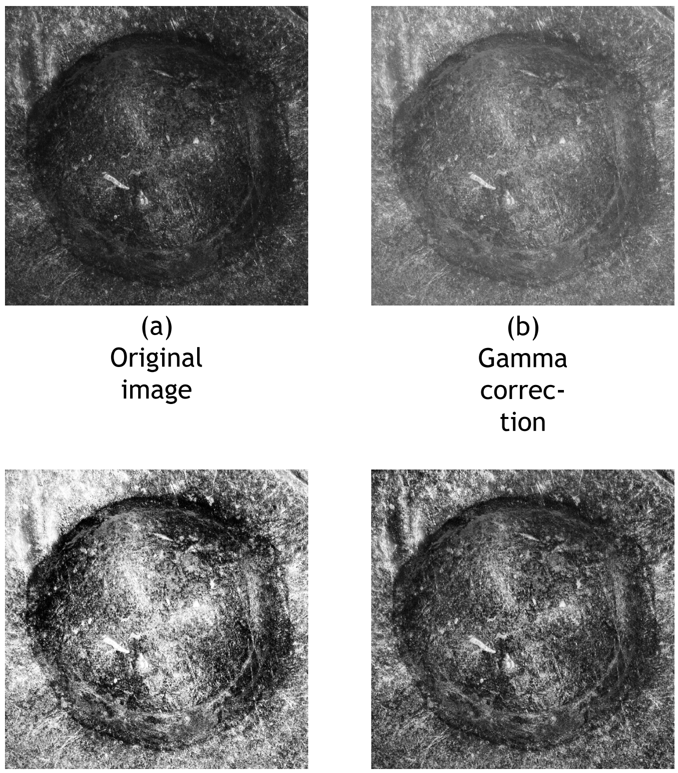
4.3. Image Processing

In this section, an image processing technique based on the work described in [54] is presented. It aims to be able to identify and enhance the most relevant areas in the images such as firing pin impressions and breechface marks while also eliminating irrelevant marks from the images. As such, the developed technique consists in:

(1) Resizing to the target size;
(2) Applying a median blur;
(3) Enhancing the contrast via an adaptive histogram equalization;
(4) Binarizing the image using Otsu's thresholding method [26];
(5) Applying image erosion followed by a dilation to reduce mask's noise;
(6) Applying the obtained mask to the contrast enhanced image.

The original method proposes gamma correction as a means of contrast enhancement. By experimenting with different approaches it was found that this method would not yield satisfactory results regarding contrast increase. Therefore, two other methods were tested: histogram equalization and adaptive histogram equalization. By analysing

Figure 21 it is clear that the visual differences between these image operations are substantial. The experiments with the gamma correction method showed that, while the image is brighter, the contrast does not significantly improve. The histogram equalization technique did show potential in terms of contrast, but it resulted in an over exposed image. The method that is believed to produce the most balanced results was the adaptive histogram equalization, which significantly improved the contrast on the image without over exposing it.



(a) Original image

(b) Gamma correction

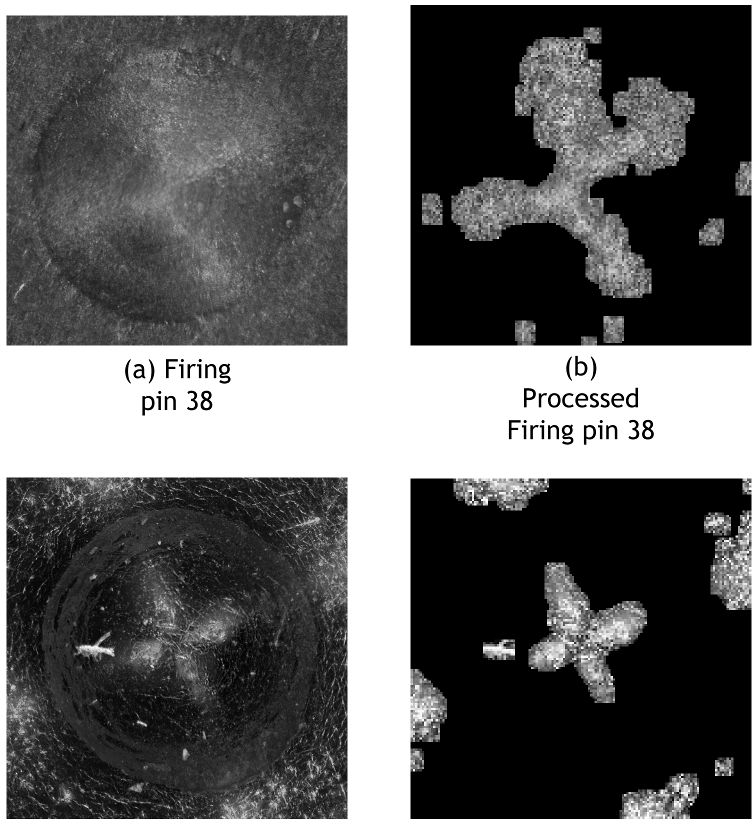
(c) Histogram equalization

(d) Adaptive histogram equalization

Figure 21. Contrast improvement techniques

PCP

The proposed technique was applied to the PCP's dataset. Figure 22 illustrates the result of the proposed image processing technique for two firing pin images.



(a) Firing pin 38

(b) Processed Firing pin 38

(c) Firing pin 42

(d) Processed Firing pin 42

Figure 22. PCP's dataset image processing

The application of this method to the images seemed to have a strange impact by highlighting a predominant cross on the center of the firing pin image, as can be observed in Figure 22.

In cooperation with the examiners at the PCP, it was found out that these light crosses, which are present in most of the images of this dataset, are not a characteristic of the cartridges themselves, but a result of the imaging acquisition process caused by the scanning instrument.

Because of this issue, it was decided that for the PCP's dataset, an approach without binarization step should also be tested, as the binarization would not produce the desired results. With this in consideration, a network with only contrast enhanced images was also trained, as shown in Figure 21d.

It was found that using the proposed image processing technique did not result in an improvement of the network's accuracy, while using contrast enhancement alone also did not cause the network's accuracy to increase. Table 11 compiles the results of the best performing network against two networks trained on an image processed dataset, one following the full method, and the other only applying the technique up to the contrast enhacement step. The results show a decrease of 9% accuracy for the full image processing procedure, while almost tripling the networks' loss. Regarding the images with enhanced contrast, it is possible to see that the test accuracy remained the same while the test loss decreased.

Table 11. Accuracy and loss results for PCP's dataset using image processing

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
No processing	1.26	1.27	63%	57%
Full technique	3.31	3.81	61%	48%
Contrast enhancement	1.22	0.92	58%	57%

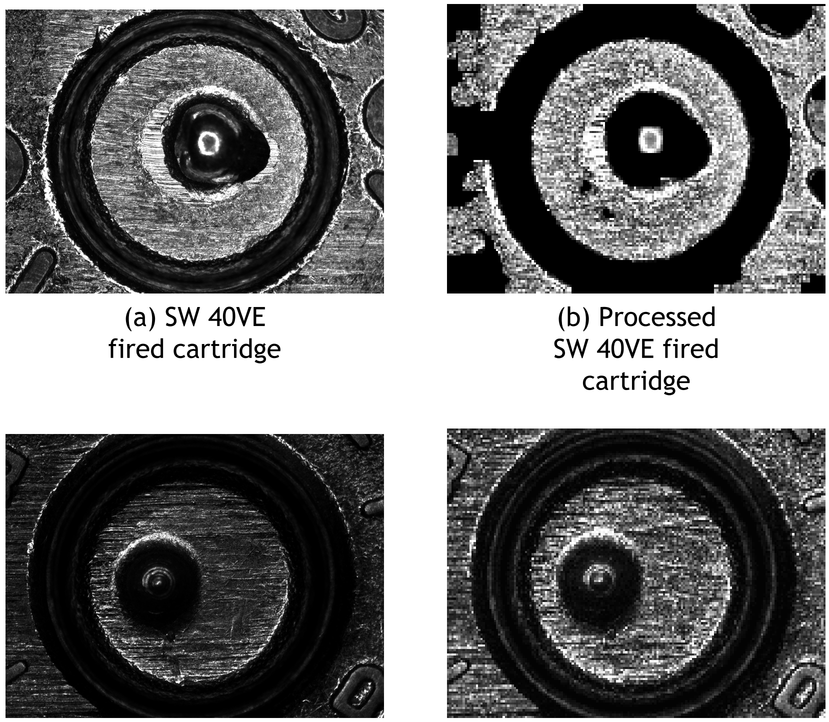
Since the performance on the test set for the contrast enhanced technique improved the trained model by reducing loss, the next tests were carried out with this preprocessing applied.

NBTRD

On the NBTRD, the same image processing methods were applied. Figure 23 shows how the proposed technique and only contrast enhancement affect the NBTRD images.

From these images, it is possible to understand that the processed data has higher contrast with less visible detail due to the image resizing. It is also possible to state that the general firing pin shape remains in the image as well as the breech face marks, while the darker areas with less detail are removed.

The outcome demonstrated in Table 12 suggests that the use of the image processing techniques slightly affected the performance of the network regarding loss, without any noticeable decrease in its accuracy.



(a) SW 40VE fired cartridge

(b) Processed SW 40VE fired cartridge

(c) Hi-Point C9

(d) Contrast enhanced Hi-Point C9

Figure 23. NBTRD Image Processing

Table 12. Accuracy and loss results for NBTRD using image processing

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
No processing	0.00010	0.46	100%	93%
Binarization	0.0068	0.78	100%	93%
Contract enhancement	0.00015	1.14	100%	93%

Although the processed images show visually promising results, the network's performance did not improve relative to the unprocessed images. Thus, the next experiments were carried out using the original unprocessed images.

4.4. Transfer Learning

For the Transfer learning section three pre trained network architectures were used: ResNet V2, Inception V3 and VGG16. These networks were used with the imagenet dataset weights. Although this dataset is not similar to the NBTRD or the PCP's dataset these weights can be useful as feature extractors, given their vast generalization power across different classes.

Since this approach is based on previously developed networks, the keras tuner parameters had to be modified. In this context, not many additional layers are required to train the network, given the fact that the used architectures already contain a vast number of layers. As such, for the tuning of the network, an additional Convolutional and Pooling layers, as well a small number of fully connected layers were considered so their outputs could be used as feature vectors. Table 13 shows the parameters used for the fine tuning of the transfer learning networks.

Table 13. Parameters used for hyperparameter tuning with transfer learning

Nº Layers	Layer	Filters/Neurons	Filter Size/dropout	Activation
1	Transfer Learning	-	-	-
1	Conv	4-32	2-5	relu
1	MaxPool	1-2	-	-
1	Flatten	-	-	-
1-3	Dense	2-256	-	sigmoid or relu
1	Dense (classification)	5	-	softmax

PCP

The use of TL in the PCP's dataset had an overall positive impact in the results. Of the three tested networks, the Inception V3 had the poorest performance, even significantly worse than the approach without transfer learning. While both the ResNet V2 and the VGG16 outperformed the network without transfer learning, it is clear that the VGG16 had a very significant impact, improving the model's test accuracy by 12% and decreasing its loss. Table 14 shows how each TL based architecture network affected the model's performance.

Table 14. Accuracy and loss results for the PCP's dataset using TL

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
Without Transfer Learning	1.22	0.92	58%	57%
ResNet V2	1.63	1.25	65%	58%
Inception V3	2.44	1.73	51%	46%
VGG16	1.06	0.83	70%	69%

NBTRD

One of the problems that had been noticed throughout the development of the network for the NBTRD was the fact that, although the validation accuracy was always 100%, the testing accuracy never reached a value above 93%. The use of TL shows a significant impact in the results regarding the network's loss, especially for the test loss. It is also possible to see that the VGG16 architecture was the only one that was able to reach 100% accuracy both on the validation and test sets, as well as reaching a very significant decrease in loss. Considering these results, for the siamese neural network training, a network using the VGG16 was used.

Table 15. Accuracy and loss results for NBTRD using TL

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
Without Transfer Learning	0.00010	0.46	100%	93%
ResNet V2	0.00063	0.50	100%	93%
Inception V3	0.0022	0.12	100%	93%
VGG16	0.0076	0.018	100%	100%

4.5. Data Augmentation

For the training of the network using data augmentation the following parameters were used:

- Random vertical 20 pixel variation;
- Random horizontal 20 pixel variation;
- Random rotation up to 25 degrees.

Data augmentation can help reduce overfitting [55] by augmenting the training data without using other information. In this case, where the data is limited, it is also important so that the network has more time to learn without overfitting on the training set.

PCP

The application of the proposed approach for data augmentation using the PCP's dataset resulted in images such as the ones seen in Figure 24.

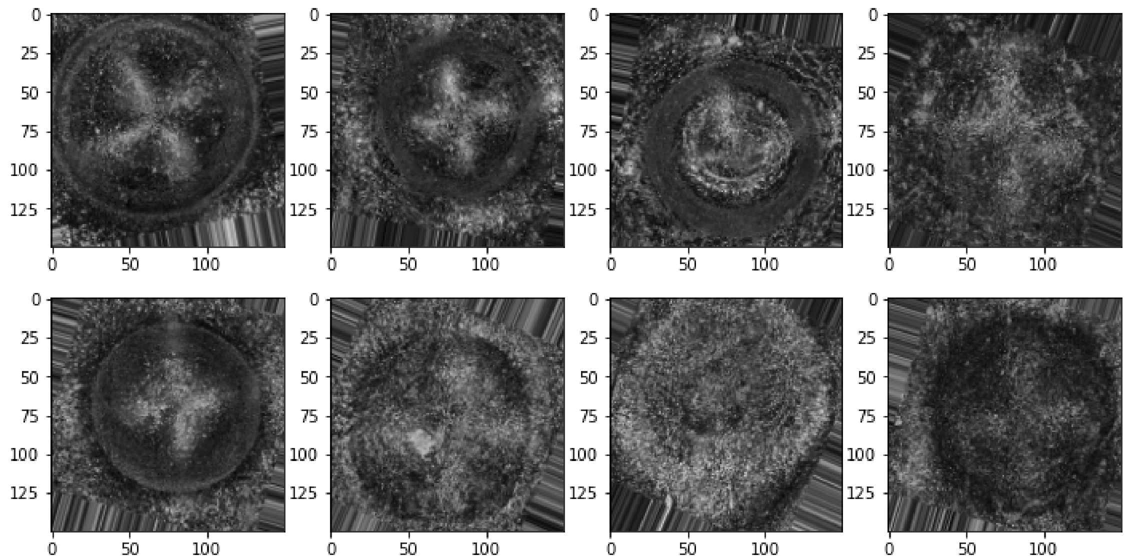


Figure 24. Data augmentation example on the PCP's Dataset

In this case, looking at the results depicted in Table 16, it is clear that using DA did not positively impact the classifier. Since the images from the dataset are very consistent in terms of positioning, it is believed that this variation might not be beneficial to the end results.

Table 16. Accuracy and loss results for NBTRD using Data Augmentation

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
Without Data Augmentation	1.06	0.83	70%	69%
Data Augmentation	1.45	0.99	63 %	51%

NBTRD

For the NBTRD the same DA parameters used for the PCP's dataset case were applied. Figure 25 shows examples of images taken from the NBTRD after the applying the DA parameters.

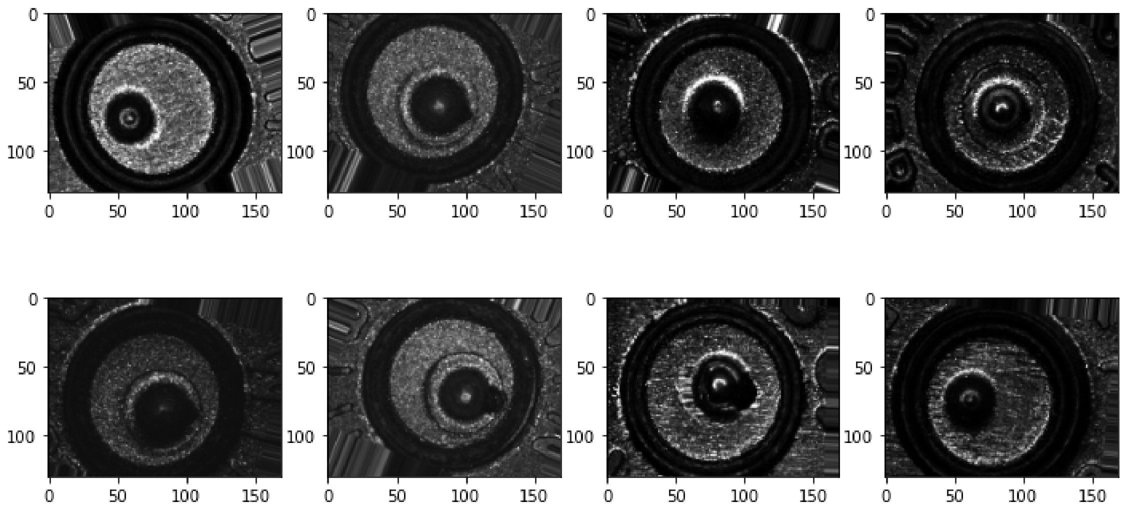


Figure 25. Data augmentation example on the NBTRD

In this case, the results presented in Table 17 show that there was no improvement in the network's performance. Similarly to the images from the PCP's dataset, the NBTRD's images are also very standard across the dataset regarding positioning and lighting, with the added factor that their rotation is the same with reference to the firing pin impression. Attending to the achieved results, it is believed that using DA on this dataset does not bring performance improvements.

Table 17. Accuracy and loss results for NBTRD using Data Augmentation

Type	Validation loss	Test loss	Validation accuracy	Test accuracy
Without Data Augmentation	0.0076	0.018	100%	100%
Data Augmentation	0.0109	0.019	100%	100%

4.6. Siamese Neural Network

For the development of an SNN for each of the datasets, the CNN that yielded the best results was used as a feature extractor. All the tested CNNs had fully connected layers outputs that can be used as feature vectors. Taking this into account, the classification layer for the networks was removed and a feature extractor based on a CNN was used in the SNN. The evaluation of the SNN's performance, which takes pairs of images as input, can be conducted in several ways:

(1) Measuring the binary accuracy of the SNN, considering that a prediction values above or equal to 0.5 means that both images belong to the same class, while predictions below 0.5 mean that the images belong to different classes;
(2) Visualizing the embeddings (output vector of the network) for the convolutional part of the network in a 2D space;
(3) Using a Top-1 and Top-2 (one of the two most likely classifications, out of five classes) accuracy metrics with a distance and probability classification, both presented in section 3.3.

For the latter assessment procedure (3), a number of reference images were needed to compare the test set against. Therefore, twenty images from each class of the training set were chosen as a reference set for the classification of other images.

PCP

The results from the experiments described along this chapter allowed to conclude that the setup leading to the best classification results for the PCP's dataset is the following:

(1) Using firing pin images;
(2) Class balancing using undersampling;
(3) Contrast enhancement;
(4) Transfer learning using the VGG-16 architecture.

After the training of the SNN with the proposed setup, which adjusted all the network's weights using the binary cross-entropy loss function (which outputs a penalization according to the classification error), images of the same class should have a similar feature vector (low distance), while images from different classes should have feature vectors with higher distances. This information allows us to comprehend that this Siamese Network is essentially clustering images from the same class. For this clustering to be visualized in a two-dimensional space, a dimension reduction was carried out from the embedding of every test set image to a two-dimensional coordinate using the Uniform Manifold Approximation and Projection technique [56] and posteriorly plotted. The resulting plot is depicted in Figure 26 and contains a point (dimensionality reduced feature vector) for each image in the test set. Furthermore, every point is plotted in the color associated to the class it belongs to.

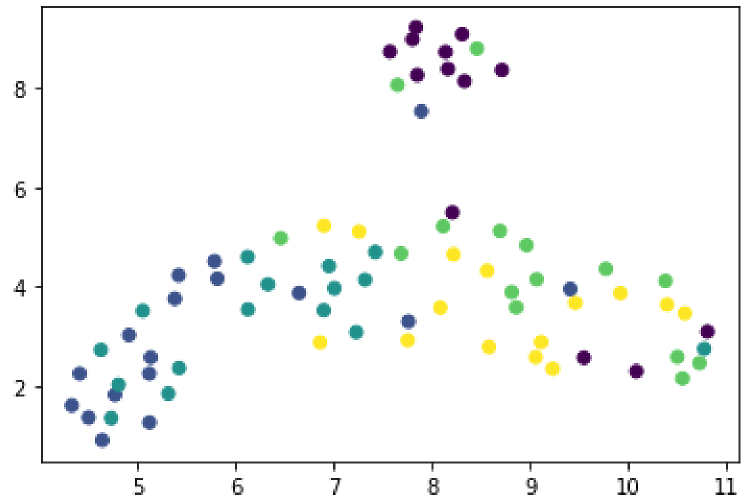


Figure 26. PCP's dataset test set images embeddings visualized as coordinates

This plot shows that the separation between classes in the PCP's dataset was not clear. Although some agglomeration is visible there is also a great amount of overlapping between different classes.

Table 18 shows the accuracy that was achieved using this SNN with different metrics. For the first row, the trained SNN is used with random pairs of test images, where predictions above or equal to 0.5 are considered the same class and predictions below 0.5 as different classes, therefore making it a binary classification metric (match or no match). The second row of Table 18 uses the convolutional layers of the trained SNN to compare feature vectors between test images and reference images, applying distance classification presented in section 3.3. This metric achieved 57% and 81% accuracy on the Top-1 and Top-2 metrics, respectively. The last row of Table 18 shows that the probability classification, which compares test images and reference images using the full SNN model, as explained in section 3.3, had the worst performance achieving 38% and 51% accuracy for Top-1 and Top-2 metrics, respectively.

Having in mind that one of the main purposes of this project is not to replace the examiners but to help them carry out the firearm identification task, these metrics also make it possible to present of a list of the most likely firearm model. Therefore, Table 18 also shows a Top-2 metric, where accuracy was measured by checking if the target gun class corresponded to one of the top two firearm models. While a technique such as this one would not be ideal, this would save examiners a great amount of time by correctly pointing them to two firearm models 80% of the time.

Table 18. Accuracy results for PCP using SNN based classifications

Type	Top-1	Top-2	Accuracy
Binary Classification	-	-	67%
Distance classification	57%	81%	-
Probability classification	38%	51%	-

In this case, the results presented for the probability and distance classifications suggest that the former performs better. While the distance classification uses the whole feature vector to compute the distance for each class, the probability classification only uses the single value output of the SNN for calculating each class's probability. This means that, by using the distance between classes, more information will be considered in the classification process. With this in mind, it is reasonable to assume that the use of more data could have lead to a greater classification performance.

NBTRD

For the construction of the Siamese Neural Network model the best performance CNN network was used without the classification layer. The network that yielded the best accuracy used the following setup:

(1) No data balancing;
(2) No image processing;
(3) Transfer learning using the VGG-16 architecture;
(4) No data augmentation;

By analysing Figure 27 it is clear that this dataset produced much clearer separation between classes. In this case, although the Siamese network did not reach 100% accuracy, the embeddings produced by the parallel convolutional neural networks resulted in a clear clustering between the different classes in the test set.

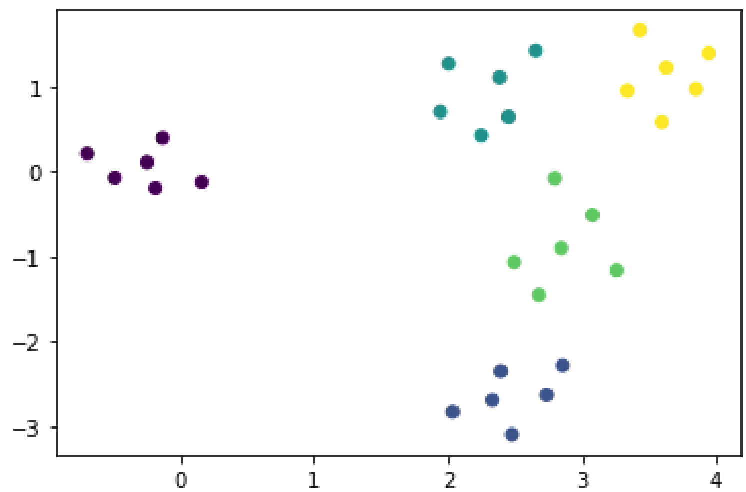


Figure 27. NBTRD test set images embeddings visualized as coordinates

The embeddings produced resulted in a distance classification that reached 100% accuracy, as shown in Table 19. Although the SNN only reached 96% accuracy, by using the probability classification technique proposed in 3.3, a 100% success classification was also achieved.

Table 19. Accuracy results for the NBTRD using SNN based classifications

Type	Top-1	Top-2	Accuracy
Binary Classification	-	-	96%
Distance classification	100%	100%	-
Probability classification	100%	100%	-

4.7. Final Prototype Evaluation

As proposed in section 1.4, one of the key steps in developing an effective digital artifact is to understand how it performs. This is what allows researchers to understand what can be further improved and modify it accordingly.

Ideally, this evaluation would be carried out by deploying a functional tool and evaluating it based on the users' feedback, in this case, the PCP's examiners. As described in section 3, the deployment of the developed machine learning model is beyond the scope of this dissertation, considering it is an instrument being developed by INOV. Given the fact that the tool under development is still in early stages and that its functionalities go beyond the scope of this dissertation, an alternative system evaluation is presented.

With this in consideration, in order to evaluate the developed prototype, the two evaluation targets that compose it are assessed.

The first evaluation target is the developed solution, the siamese neural network, and the main question is whether or not its performance would be able to help the examiners to improve their success rate while performing this task.

To address this question we can take the success rate of the examiners, which according to information disclosed by the PCP is 16%, and compare it to what was achieved by the network on the dataset provided by the PCP, which was 57% correct classification on the top match of the list and 81% on the top-2 classification. Although the accuracy achieved by the network was not very high, it is still a significant improvement that can help examiners improve their identification accuracy as well as increase their response rate. Although this shows great potential, only five classes of firearms are present in the dataset. In reality, although these classes correspond to the most common firearms models, it does not exactly match what the PCP's experts are confronted with in a real life scenario. Despite of this, a 57% accuracy still shows that this network still holds significant potential.

The other evaluation target was whether or not image pre-processing techniques would improve the performance of the classification. For this evaluation, it was possible to conclude that for the PCP's dataset, the use of image processing in the form of adaptive histogram equalization had a positive impact on the performance of the neural network along with the segmentation of the firing pin area of the images, as well as a data balancing technique via undersampling. While this is true for the PCP's dataset, the same does not happen with the NBTRD. In this case, the use of any of the tested data pre-processing techniques did not have a positive impact on the classification performance.

In this case, the results show that although the datasets were about similar in size and number of classes, the performance of this approach on the reference dataset acquired by NIST was far superior, achieving 100% correct classification both on the regular CNN and using the SNN's distance and probability metrics.

Having these results in consideration, it is clear that the developed technique was successful concerning ballistics identification, especially on the NBTRD. According to the DSRM iterative process, the next step would be to go back some steps and improve on the artifact's design and implementation.

Since the possible improvements that were identified are at the image level, we propose a second iteration of the digital artifact by re-acquiring the images with some modifications to the dataset acquisition techniques. If such dataset is developed, there is a high possibility that the results of training the same technique proposed in this dissertation with this improved dataset would result in a model with higher accuracy and robustness.

CHAPTER 5

Conclusions

In this study, an automatic ballistics identification system based on siamese neural networks was developed. For the development of such system, the PCP provided an experimental dataset with 1295 cartridge case images, which, so far, was only used in the scope of the BALCAT project and on this dissertation. Due to the experimental nature and untested characteristics of the PCP dataset, this dissertation also considered a well-established dataset for benchmarking: the NBTRD, which was developed according to NIST standards [16].

Using the pipeline shown in Figure 18 (on Chapter 3) the system was developed taking into account region of interest extraction, data balancing, image processing, transfer learning and data augmentation.

This pipeline enabled us to build a CNN model that would then be used as a feature extractor for the SNN. With this approach, the results on the NBTRD dataset show that ballistics identification is possible using Siamese Neural Networks, reaching a 100% accuracy using siamese neural networks with a distance metric on a small test set. On the other hand, the PCP's dataset showed poorer results reaching 57% top-1 classification accuracy using the SNN. Given that this system aims to be a support tool, the accuracy of the target firearm model being in the top-2 most likely classification was also evaluated, reaching an accuracy of 81%.

Considering that the first research question proposed in this dissertation was whether an automatic technique would be capable of aiding ballistics experts carry out ballistics identification, the results show that this objective was effectively accomplished. Although the results were poorer on the PCP's dataset, a 57% accuracy is still a significant improvement over the 16% success rate achieved by examiners. Thus, we believe that the use of this technique would translate into a significant positive impact on the ballistics identification, by improving accuracy and reducing the time spent doing such task.

While the results on the PCP's dataset are not ideal, they show that there is relevant information in the images that allow for their classification. This statement leads to the second research question formulated in this dissertation, which was to understand if data pre-processing could improve the results achieved by the proposed technique. This was accomplished by applying different data processing techniques such as image processing, data balancing and data augmentation.

These techniques showed some improvement on the dataset provided by PCP but no improvement on the NBTRD, which reached 100% accuracy without any data pre-processing techniques.

This information allows us to draw some conclusions:

- The developed technique has the potential to classify weapon models based on cartridge images with high accuracy;
- There is room for improvement in the dataset provided by the PCP, which could potentially improve the results obtained.

These conclusions were drawn by applying the same technique on both datasets and seeing that the performance on the NBTRD was higher than the performance on the PCP's dataset, showing that the developed technique is effective. Having this in consideration, the reason for the lower performance on the dataset provided by the PCP could be related to various factors, such as: image quality, acquisition method, cartridges or weapons used (that may be naturally more difficult to discriminate), which are the main differentiating factors to the NBTRD dataset. As such, some improvements could possibly mitigate some of these differences by: Using microscopy to acquire the cartridge images, acquiring every image under similar light conditions and focusing the acquisition on the firing pin area. Although the provided evidence does not unequivocally prove that reducing these differences would improve the results, since these differences in performance could also be related to the firearms and cartridges present in each dataset, they present a starting point in order to understand what can be made in order to improve the classification performance on the PCP's dataset.

5.1. Limitations

The results obtained suggest that the main limitations identified in this work could be related to the quality or characteristics of the PCP's dataset. As such, there are some factors that lead to this conclusion, namely:

- Firing pin impressions show light crosses in the center, which is not a characteristic of the cartridge itself, but an artifact caused by the illumination conditions during image acquisition;
- There was a low number of ejector mark impressions, which might have contributed to improve the firearms' identification if used;
- There were 3D scans that could solve the light cross issue but due to their size more than 128GB of RAM would be needed to segment them, which is hard to achieve;
- In the projects proposal it is stated that about 20000 individual cartridges would be scanned, but the dataset only contains 1295 images.
- The image lighting conditions varied across different scans;
- The image size used on the network could not be greater than 150x150 px because of hardware limitations. Although experiences in machines with better specifications showed that higher resolutions did not have a significant impact on the results.

Besides these issues, there is also one problem regarding the NBTRD dataset. Although it was shown that the network achieved 100% accuracy on the test set it is also true that the test set was small, containing only 60 images across the 5 classes. Furthermore, every two images corresponded to the same cartridge, only with two different light perspectives. Although the network achieved peak performance on this dataset, such high accuracy might not have been possible on a larger test set.

5.2. Future Work

Regarding work that could be done in the future, the findings in this dissertation show that the main point that could be improved upon would be the dataset. For this purpose, in a work to be carried out in the future, we propose some modifications to the image acquisition techniques that could potentially result in greater accuracy and fidelity for the PCP's dataset:

- Acquire images using microscopy technique;
- Acquire each cartridge image under constant lighting conditions;
- Focus the image acquisition on the firing pin area and its surroundings;
- Use a greater variety of unique firearms;
- Scan a greater variety of firearm models;
- Increase the number of acquired images.

One asset that is believed to have great potential in this application would be the 3D scans, also acquired by the PCP. Since these scans are very large in size, most of them above 30 GB, it was only possible to segment 398 individual cartridges, as the hardware available would not allow to segment scans with sizes above 34 GB. In the future, further investigation should be carried out to understand how to approach this issue and take advantage of this data.

Besides the images used in the project, the hardware used to train the models would also benefit from being upgraded, so that the training could be more extensive with deeper models and increased image sizes.

Bibliography

[1] X. Zheng, J. Soons, T. Vorburger, J. Song, T. Renegar, and R. Thompson, "Applications of surface metrology in firearm identification," *Surface Topography: Metrology and Properties*, vol. 2, p. 014012, Jan. 8, 2014. doi: 10.1088/2051-672X/2/1/014012.

[2] N. A. Md Ghani, C.-Y. Liong, and A. A. Jemain, "Analysis of geometric moments as features for firearm identification," *Forensic Science International*, The 47th International Conference of the International Association of Forensic Toxicologists (TIAFT), vol. 198, no. 1, pp. 143–149, May 20, 2010, issn: 0379-0738. doi: 10.1016/j.forsciint.2010.02.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0379073810000630 (visited on 10/21/2021).

[3] R. C. Rattenbury, "Semiautomatic pistol," *Encyclopædia Britannica*, Feb. 2015. [Online]. Available: https://www.britannica.com/technology/semiautomatic-pistol.

[4] M. Tong, J. Song, W. Chu, and R. M. Thompson, "Fired cartridge case identification using optical images and the congruent matching cells (CMC) method," *Journal of Research of the National Institute of Standards and Technology*, vol. 119, pp. 575–582, 2014, issn: 1044-677X. doi: 10.6028/jres.119.023.

[5] I. Kara, "Investigation of ballistic evidence through an automatic image analysis and identification system," *Journal of Forensic Sciences*, vol. 61, no. 3, pp. 775–781, May 2016, issn: 1556-4029. doi: 10.1111/1556-4029.13073.

[6] N. R. Council, *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: The National Academies Press, 2009, isbn: 978-0-309-13130-8. doi: 10.17226/12589. [Online]. Available: https://www.nap.edu/catalog/12589/strengthening-forensic-science-in-the-united-states-a-path-forward.

[7] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.

[8] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *BMJ*, vol. 339, b2535, Jul. 21, 2009, Publisher: British Medical Journal Publishing Group Section: Research Methods & Reporting, issn: 1756-1833. doi: 10.1136/bmj.b2535. [Online]. Available: https://www.bmj.com/content/339/bmj.b2535 (visited on 10/15/2021).

[9] N. J. Eck and L. Waltman, "Software survey: Vosviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 2, no. 84, pp. 523–538, 2010.

[10] J. Song, W. Chu, M. Tong, and J. Soons, "3d topography measurements on correlation cells—a new approach to forensic ballistics identifications," *Measurement Science and Technology*, vol. 25, no. 6, p. 064005, 2014.

[11] H. Song and J. Song, "Virtual image standard (vis) for performance evaluation of the congruent matching cells (cmc) algorithms in firearm evidence identifications," *Journal of Forensic Sciences*, vol. 67, no. 4, pp. 1417–1430, 2022. doi: https://doi.org/10.1111/1556-4029.15026. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1556-4029.15026. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1556-4029.15026.

[12] M. Tong, J. Song, and W. Chu, "An improved algorithm of congruent matching cells (CMC) method for firearm evidence identifications," *Journal of Research of the National Institute of Standards and Technology*, vol. 120, p. 102, Apr. 1, 2015. doi: 10.6028/jres.120.008.

[13] J. Song, "Proposed nist ballistics identification system (nbis) based on 3d topography measurements on correlation cells," *AFTE Journal*, vol. 45, no. 2, pp. 184–194, 2013.

[14] J. Song, T. V. Vorburger, S. Ballou, R. M. Thompson, J. Yen, T. B. Renegar, A. Zheng, R. M. Silver, and M. Ols, "The national ballistics imaging comparison (NBIC)

project," *Forensic Science International*, vol. 216, no. 1, pp. 168–182, Mar. 10, 2012, issn: 1872-6283. doi: 10.1016/j.forsciint.2011.09.016.

[15] T. V. Vorburger, J. Yen, J. F. Song, R. M. Thompson, T. B. Renegar, A. Zheng, M. Tong, and M. Ols, "The second national ballistics imaging comparison (NBIC-2)," *Journal of Research of the National Institute of Standards and Technology*, vol. 119, pp. 644–673, 2014, issn: 1044-677X. doi: 10.6028/jres.119.028.

[16] X. Zheng, J. Soons, R. Thompson, S. Singh, and C. Constantin, "NIST Ballistics Toolmark Research Database," en, *Journal of Research of the National Institute of Standards and Technology*, vol. 125, p. 125 004, Jan. 2020, issn: 2165-7254. doi: 10.6028/jres.125.004. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/jres/125/jres.125.004.pdf (visited on 03/03/2022).

[17] G. Gerules, S. K. Bhatia, and D. E. Jackson, "A survey of image processing techniques and statistics for ballistic specimens in forensic science," *Science & Justice*, vol. 53, no. 2, pp. 236–250, Jun. 2013, issn: 13550306. doi: 10.1016/j.scijus.2012.07.002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1355030612001050 (visited on 10/17/2021).

[18] D. Baldwin, M. Morris, S. Bajic, Z. Zhou, and J. Kreiser, "Statistical tools for forensic analysis of toolmarks," Apr. 2004. doi: 10.2172/825030. [Online]. Available: https://www.osti.gov/biblio/825030.

[19] W. Chu, J. Song, T. Vorburger, R. Thompson, T. Renegar, and R. Silver, "Optimal compression and binarization of signature profiles for automated bullet identification systems," in *Scanning Microscopy 2010*, International Society for Optics and Photonics, vol. 7729, 2010, p. 77291C.

[20] W. Chu, J. Song, T. Vorburger, and S. Ballou, "Striation density for predicting the identifiability of fired bullets with automated inspection systems," *Journal of forensic sciences*, vol. 55, no. 5, pp. 1222–1226, 2010.

[21] R. Gonzalez, R. Woods, and B. Masters, "Digital image processing," *Journal of biomedical optics*, vol. 14, p. 029 901, Mar. 2009. doi: 10.1117/1.3115362.

[22] C. Brein, “Segmentation of cartridge cases based on illumination and focus series,” in *Image and Video Communications and Processing 2005*, International Society for Optics and Photonics, vol. 5685, 2005, pp. 228–238.

[23] D. Li, “Firearm identification system based on ballistics image processing,” in *2008 Congress on Image and Signal Processing*, IEEE, vol. 3, 2008, pp. 149–154.

[24] D. Li, “Image processing for the positive identification of forensic ballistics specimens,” *Information Fusion*, pp. 1494–1498, 2003.

[25] Z. Huang and J. Leng, *A novel binarization algorithm for ballistics imaging systems*. Nov. 18, 2010, vol. 3, 1287 pp., Journal Abbreviation: Proceedings - 2010 3rd International Congress on Image and Signal Processing, CISP 2010 Pages: 1291 Publication Title: Proceedings - 2010 3rd International Congress on Image and Signal Processing, CISP 2010. doi: 10.1109/CISP.2010.5647685.

[26] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979, Conference Name: IEEE Transactions on Systems, Man, and Cybernetics, issn: 2168-2909. doi: 10.1109/TSMC.1979.4310076.

[27] C. K. Chow and T. Kaneko, “Automatic boundary detection of the left ventricle from cineangiograms,” *Computers and Biomedical Research, an International Journal*, vol. 5, no. 4, pp. 388–410, Aug. 1972, issn: 0010-4809. doi: 10.1016/0010-4809(72)90070-5.

[28] S. D. Yanowitz and A. M. Bruckstein, “A new method for image segmentation,” *Computer Vision, Graphics, and Image Processing*, vol. 46, no. 1, pp. 82–95, Apr. 1, 1989, issn: 0734-189X. doi: 10.1016/S0734-189X(89)80017-9. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0734189X89800179 (visited on 10/21/2021).

[29] J. Leng and Z. Huang, “On analysis of circle moments and texture features for cartridge images recognition,” *Expert Systems with Applications*, vol. 39, no. 2, pp. 2092–2101, Feb. 1, 2012, issn: 0957-4174. doi: 10.1016/j.eswa.2011.08.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417411011146 (visited on 10/17/2021).

[30] S. B. A. Kamaruddin, N. A. M. Ghani, C.-Y. Liong, and A. A. Jemain, “Firearm recognition based on whole firing pin impression image via backpropagation neural network,” in *2011 International Conference on Pattern Analysis and Intelligence Robotics*, vol. 1, Jun. 2011, pp. 177–182. doi: 10.1109/ICPAIR.2011.5976891.

[31] O. Giudice, L. Guarnera, A. B. Paratore, G. M. Farinella, and S. Battiato, “Siamese ballistics neural network,” in *2019 IEEE International Conference on Image Processing (ICIP)*, ISSN: 2381-8549, Sep. 2019, pp. 4045–4049. doi: 10.1109/ICIP.2019.8803619.

[32] B. J. Heard, *Handbook of firearms and ballistics: examining and interpreting forensic evidence*. John Wiley & Sons, 2011, vol. 1.

[33] Swati, G. Gupta, M. Yadav, M. Sharma, and L. Vig, “Siamese networks for chromosome classification,” in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 72–81. doi: 10.1109/ICCVW.2017.17.

[34] E. Hare, H. Hofmann, and A. Carriquiry, “Automatic matching of bullet land impressions,” *The Annals of Applied Statistics*, vol. 11, no. 4, pp. 2332–2356, 2017. doi: 10.1214/17-AOAS1080. [Online]. Available: https://doi.org/10.1214/17-AOAS1080.

[35] J. Song, Z. Chen, T. V. Vorburger, and J. A. Soons, “Evaluating likelihood ratio (lr) for firearm evidence identifications in forensic science based on the congruent matching cells (cmc) method,” *Forensic Science International*, vol. 317, p. 110 502, 2020, issn: 0379-0738. doi: https://doi.org/10.1016/j.forsciint.2020.110502. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0379073820303649.

[36] J. Roth, A. Carriveau, X. Liu, and A. K. Jain, “Learning-based ballistic breech face impression image matching,” in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2015, pp. 1–8. doi: 10.1109/BTAS.2015.7358774.

[37] E. R. Hare, “Statistical methods for bullet matching,” Ph.D. dissertation, Iowa State University, 2017.

[38] *Laboratory imaging s.r.o., "toolscan"*, Accessed Oct. 23, 2021. [Online]. Available: https://www.forensic.cz/en/products/toolscan.

[39] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[40] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[42] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, Lille, vol. 2, 2015.

[43] N. Poolsawad, C. Kambhampati, and J. Cleland, "Balancing class for performance of classification with a clinical dataset," in *proceedings of the World Congress on Engineering*, vol. 1, 2014, pp. 1–6.

[44] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. of the Int'l Conf. on Artificial Intelligence*, Citeseer, vol. 56, 2000.

[45] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6. doi: 10.1109/ICEngTechnol.2017.8308186.

[46] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[47] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[50] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[51] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill, "Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs," *Scientific reports*, vol. 8, no. 1, pp. 1–13, 2018.

[52] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70–90, 2018.

[53] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, and L. Invernizzi, *Kerastuner*, https://github.com/keras-team/keras-tuner, 2019.

[54] J. Leng and Z. Huang, "On analysis of circle moments and texture features for cartridge images recognition," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2092–2101, 2012.

[55] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.

[56] A. Diaz-Papkovich, L. Anderson-Trocmé, and S. Gravel, "A review of umap in population genetics," *Journal of Human Genetics*, vol. 66, pp. 1–7, Oct. 2020. doi: 10.1038/s10038-020-00851-4.