International Journal of Computing and Business Research (IJCBR)
ISSN (Online) : 2229-6166
International Manuscript ID : 22296166V12I1202211
Volume 12 Issue 1 January 2022

Image Forensic and Analytics using Machine Learning

Jasim Mohammed Atiyahc
Ministry of Education.
Slah AL Deen Education, Iraq.
jassimali52@yahoo.com

Zainab hammoodi noori
Ministry of Education,
Karbala Education Directorate,
Iraq
fathayrt21@gmail.com

Amanj Shihab Ahmed
KRG-Ministry of Education
Directorate of Education in
Garmian/Kifri
amanjshahab@gmail.com

Abstract

Machine learning for multimedia forensics is novel because it can detect picture fraud in a matter of seconds. Security in the forged sector may be improved by using machine learning with Deep Learning and Convolution Neural Network (CNN) approaches. SOFM, fuzzy c-means, support vector machines (SVM), and k-nearest neighbor's (k-NN) are all unsupervised classification algorithms that may be used to detect forgeries. False photographs used to disseminate racial animosity or misinformation about certain ethnic groups or political campaigns may be detected via the use of forensic image analysis tools in both criminal and civil proceedings (e.g., defamation). Image forensics is growing more dependent on machine learning. There are, however, a number of restrictions and weaknesses (e.g., how to recognise adversarial (picture) occurrences) that have real-world ramifications with machine learning-based systems (e.g., inadmissible evidence, or wrongful conviction).

Keywords : Image Forensic, Image Processing, Machine Learning

Introduction

A variety of challenges, including those connected to the Internet of Things (IoT) and data gathered in more complicated settings, have been solved using machine learning [1].

As a case study, we examine how machine learning may be applied in the present cyber security context to acquire digital evidence. Forensics may benefit from the application of machine learning in this investigation.

It's possible that our study on the existing uses of machine learning might lead to new avenues in the development of more effective and helpful digital forensics tools.

The digital revolution, which started in the latter half of the twentieth century and continues now, has had a profound impact on our contemporary world. All of our everyday activities are now dependent on digital technology, including smartphones and the Internet. Even though we didn't know it at the time, we now use these tools in both our personal and professional lives to great effect. As seen by this, our most vital data will likely be housed digitally in the future. Now that we're in the age of information technology, it's imperative that existing regulations be revised and enforced. As a result, traditional criminal activities, particularly those involving money and trade, are being transformed by the rapid advancement of technology. Computer systems and digital gadgets are becoming more vital in all investigations, and this trend is only expected to continue[2].

Machine learning may play an important role by automating time-consuming activities in this new medium of electronic evidence, where digital forensics is concerned.

The phrase "digital forensics" is used to describe the study of digital evidence. Keeping up with the rapid rate of technological change is a constant challenge in the area of data recovery from computers and other digital storage devices. The "Big Digital Forensic Evidence" is today's most daunting task for law enforcement authorities. Analyzing ever-increasing amounts of data is becoming an increasingly difficult task. Artifact detection is

a problem that has to be addressed... This section seeks to offer an overview of machine learning in digital forensics so that more research may be done.

In the 1970s, the first studies in the field of cyber security were conducted. When I initially started in this industry, building theoretical models was more essential than really putting those notions into practice. The work of De Denning has had a considerable influence on computer security machine learning.

One of the more recent ideas in machine learning-based computer security is the use of system audit data to detect breaches and penetrations in real time. It was also a sign of how the person was interacting with the subject matter they were modelling. Several computer security fields have employed machine learning applications like the one described in this paper that clusters a network-level behavioural virus using malicious HTTP traffic for structural similarities After 25 years of international efforts (and billions of dollars in expenditures), we can now say with confidence that the internet is safe[3][4].

A wide range of algorithms for machine learning have been implemented. Digital forensics may be tackled in a number of ways with the help of technologies that are designed to be intelligent. For the purpose of more easily identifying confiscated equipment, Grillo et al. proposed a technique for tracking down machine users. Internet users' habits, abilities, interests, and activities were taken into account in the process. Internet surfers, chatters, office workers, and hackers were all identified using user profiling. This strategy was used by authorities in order to give priority to the examination of confiscated devices. Investigative teams may complete their work in half the time they had budgeted by merely consulting relevant hard discs. Automated forensic inquiry is increasingly relying on metadata. Inconsistencies and suspicions about file artefacts may be discovered by thoroughly examining a vast number of data, such as the file name (extended file name),

extension (extended file name), path, size, access and modification times, hash codes, and the status flag. Using the metadata linkages shown in 2013, Raghavan and Raghavan were able to establish the origin of downloaded files. According to the work of Asaf Varol and his colleagues published in 2020, computer learning may be used to uncover criminal behaviour and intentions. It may also be used to predict future criminal behaviour. BSAF was suggested by S. Baskar and others to examine digital forensic evidence on an IoT platform using a Blockchain-assisted Shared Audit Framework (BSAF). Data scavenging attacks on virtualized resources were tracked down to their root cause and source in this analysis. Research by Francisca and her colleagues looked at how machine learning may be used to find and classify objects. Alhoseni et al. developed an IoT-enabled Optimal Deep Learning-based Convolutional Neuronal Network (ODL–CNN) for the identification of suspects. The hyperparameters of the DL-CNN model were optimised using the improved elephant herd optimization (IEHO) method [5][6].

Machine learning has been employed to build this new method of identifying picture forgeries because of its speed and accuracy. Machine learning outperforms prior Deep Learning and Convolution Neural Network approaches in terms of generating a unique location. SOFM and fuzzy c-means, as well as support vector machines (SVMs) and k-nearest neighbours (k-NNs) may all be used to detect forgeries. Due to the current situation, it is very necessary for us to prioritise detection speed. Pre-processing of feature extraction and reduction utilising "DWT" and "PCA" with support vector machine (SVM) data was generated using a machine learning technique to provide speedy results under various test situations. Forgeries of all kinds, as well as post-processing processes and geometric change, are all defined in this research to show how to discover and find them in a sea of forgeries.

Data and algorithms are used in machine learning, a field of artificial intelligence (AI) and computer science, to replicate human learning and improvement through time.

A lengthy history of machine learning research has been done at IBM. Arthur Samuel is commonly considered to have invented the term "machine learning" while doing research on the game of checkers. Robert Nealey, a self-declared checkers master, lost to an IBM 7094 computer in 1962. This accomplishment may appear little in light of current technology, yet it represents a turning point for artificial intelligence. Storage and computational power advances over the next two decades will enable new technologies like Netflix's recommendation engine and self-driving cars [7].

Machine learning is becoming more important in the field of data science. Data mining allows algorithms to be trained to categorise or predict using statistical methodologies. For example, greater decision-making enables organisations and applications to achieve better growth metrics. As big data continues to grow, businesses will need the help of data scientists to identify the most essential business questions and the data needed to answer them.

Because of advances in computer technology, machine learning is no longer the same as it was in the past. If computers could learn without being instructed, pattern recognition was the first step. A key advantage in machine learning is that models may self-correct when fresh data is introduced. Building on previous study, they may rely on their conclusions and results again and again. A growing number of people are beginning to believe in it [8][9].

Even while many machine learning techniques have been known for a long time, this new capacity to automatically perform complicated mathematical computations to massive datasets is very new.

Here are a few widely publicized examples of machine learning applications you may be familiar with:

- The heavily hyped, self-driving Google car. The essence of machine learning.
- Online recommendation offers such as those from Amazon and Netflix. Machine learning applications for everyday life.
- Knowing what customers are saying about you on Twitter. Machine learning combined with linguistic rule creation.
- Fraud detection. One of the more obvious, important uses in our world today.

While Bayesian analysis has grown more popular than ever before, machine learning has also seen a surge in popularity. An rise in data volume and a drop in computation costs are two examples.

A large-scale evaluation of increasingly complex data and the generation of faster, more accurate solutions is now possible via the rapid and autonomous construction of models. It's everything here. A firm may also benefit from more accurate models by seeing profitable opportunities or avoiding potential pitfalls, both of which can contribute to increased profitability.

Deep Learning vs. Neural Networks vs. Machine Learning

For the sake of clarity, we need to distinguish between deep learning and machine learning. There are many subfields of artificial intelligence that include deep learning,

machine learning, and neural networks. Neural networks and deep learning, on the other hand, are branches of machine learning.

How the algorithms learn is the main difference between deep-learning and machine-learning systems. It is now possible to extract characteristics from far bigger datasets without the need of humans because to advances in deep learning. Deep learning, according to MIT lecturer Lex Fridman (00:30), is "scalable machine learning" (link resides outside IBM). This kind of machine learning relies more heavily on human input than "deep" machine learning. Human specialists, on the other hand, frequently want more structured data in order to discern the differences in data sources [10].

Deep machine learning may be taught without the need of labelled datasets, which is known as supervised learning in the context of "deep" machine learning. Using text or photos as an example, it can automatically identify the characteristics that set distinct data types apart. Due to the fact that machine learning requires human participation in the data processing process, scaling it is more complex. AI and deep learning have had a significant impact on recent breakthroughs in computer vision, natural language processing, and speech recognition, among other things.

In neural networks with a single node layer, an input layer, one or more hidden layers, and an output layer are all common. There has been assigned a weight and a threshold to each artificial neuron or node. Any node whose output value exceeds a predetermined threshold is turned on across the network. The next tier of the network will not be able to access any data if this rule is not followed. "Deep" in deep learning refers to the number of layers in a neural network, not the total number of layers. For example, a deep neural network contains inputs and outputs in addition to the neural network's layers. A simple neural network has no more than two or three layers.

Researchers at the University of California, Berkeley believe that machine learning algorithms' learning systems may be categorised into three basic types (link is external to IBM). Machine Learning Algorithms are often used to predict or categorise data. Your method will offer an estimate of a pattern in the data based on some input data, either labelled or unlabeled.

Error functions are used to evaluate a model's predictions. In order to assess the model's precision, a comparison with an error function might be used.

An Optimisation Method Using Modeling Weights are adjusted to eliminate discrepancies between the known example and the model forecast if the model can better match the data points in the training set. When the system achieves a certain level of accuracy, the weight updates will continue.

Classification and prediction algorithms may be trained using labelled datasets in supervised learning, also known as supervised machine learning. It is necessary to adjust the weights when new data is introduced to the model so that the model is properly adapted. We do this step in the cross-validation procedure to make sure our models aren't misfits. A broad variety of real-world issues may be addressed with guided learning, such as establishing a distinct spam category. Regression, linear regression, logistic regression, random forest, and support vector machine are all examples of supervised learning techniques [11].

Unsupervised machine learning uses machine learning methods to analyse and cluster data while dealing with unlabeled datasets. To find patterns or clusters in huge amounts of data, automated approaches use machine learning. For activities like as exploratory

data analysis, cross-selling techniques, client segmentation, and the detection of pictures and patterns, it is a great tool. Singular value decomposition and principle component analysis (PCA) are two more dimensionality reduction approaches that may be used to minimise the number of features in a model (SVD). Neural networks and k-means clustering are two methods for unsupervised learning.

Students get the best of both worlds with semi-supervised learning, which sits somewhere in the middle of the supervised and unsupervised learning spectrums. A tagged dataset is used to train on a larger, unlabeled dataset during the training phase. Algorithms are manually designed when supervised learning can't be employed because of the absence of labelled data (or funds).

Robotics and Reinforcement Learning

Reinforcement machine learning takes a different approach to algorithm training than traditional supervised learning does. As the model learns, it relies heavily on trial and error to get new knowledge. It's always best to build on prior triumphs when facing a new challenge.

A good illustration of this is IBM Watson®, which triumphed in the 2011 Jeopardy! Competition. It was found that using reward-learning algorithms, it was possible to predict whether or not a player should try an answer, which square on the board to choose, and how much money to gamble [12][13].

In the real world, use cases of machine learning. The following are just a few examples of how machine learning is being used today: Human speech may be translated into written form by using natural language processing (NLP), which is called "speech recognition" or "automated speech recognition" (ASR) (or voice-to-text). A mobile device

having a speech-recognition feature is likely to have been used by someone who has used Siri.

Chatbots are gradually taking the role of human customer service representatives. Since websites and social media platforms answer FAQs (commonly asked questions), the way we see consumer communication on these channels is changing. A growing number of businesses, including e-commerce sites, chat platforms like Slack and Messenger, and even certain professions, are using virtual and voice assistants.

Data may be extracted from photos, videos, and other visual inputs by computers and systems. It has the ability to make suggestions, which sets it apart from other types of image recognition. Digital imaging systems, such as radiography and self-driving automobiles, may benefit from the computer vision capabilities of convolutional neural networks.

In the future, cross-selling efforts may benefit from AI algorithms that start with prior data on customer consumption habits.. This is used by online businesses to propose relevant add-ons to customers during the checkout process.

AI-driven high-frequency trading algorithms make hundreds or even millions of deals a day without the need for human participation in order to maximise stock portfolios.

Artificial intelligence has its share of challenges.
Our lives have been improved by advances in machine learning. Because of the extensive usage of machine learning in the workplace, ethical issues have emerged.

Technological Concept of Singularity

The idea of artificial intelligence (AI) exceeding human intellect has received a great deal of media attention, although many academics aren't convinced that this will happen anytime soon. "Superintelligence" is, according to Nick Bostrum, "any mind that considerably outperforms the finest human brains in nearly every field," including scientific ingenuity, general knowledge, and social abilities. As we examine the usage of autonomous systems like self-driving cars, strong AI and superintelligence provide some fascinating challenges. If an accident involving an autonomous vehicle occurs, who is responsible and liable? So, what's more important: making semi-autonomous vehicles safer for drivers or pushing towards complete autonomy? Cutting-edge AI is igniting these and other ethical questions, but the verdict is yet out [14][15].

Many people fear that their careers will be at risk as a result of advances in artificial intelligence. It's possible that this is overstating things. Certain occupational roles' market demand changes with the advent of a disruptive new technology. For environmental reasons, an increasing number of automakers are refocusing their efforts on the development of electric cars (EVs). Even while the source of energy isn't going away, the change to an electronic economy is. When seen in this manner, artificial intelligence will broaden the kinds of jobs for which humans will be needed. As data increases and evolves, there will be a need for people to assist in the management of these systems. Although the need for employees is expected to alter in the future, even in customer service jobs, workers need tools to deal with more complicated scenarios. AI's impact on the labour market will need assistance for humans in adjusting to these new areas of demand. [16]

As a result, data privacy, data protection, and data security have made significant advances in recent years. EU and EEA people now have greater control over their personal data according to the GDPR (General Data Protection Regulation), which was

implemented in 2016. The Consumer Privacy Act of California, for example, is one of many state initiatives aimed at safeguarding the privacy of American consumers (CCPA). As a result of the new regulation, businesses will have to reevaluate how they handle and manage customer data (PII). As a consequence, companies are more concerned about securing their networks and avoiding monitoring, hacking, and other forms of cyberattack.

Anti-AI sentiments have led to a number of instances of bias and discrimination in intelligent systems. In order to avoid prejudice and discrimination, how can we ensure that the training data is not biassed? Reuters (link outside IBM) points out some of the unintended implications of adding AI into hiring procedures, no matter how well-intentioned such organisations' automation ambitions may be.

Due to the unbalanced pool of candidates, Amazon had to give up on the initiative. AI recruiting strategies have brought to light new challenges, such as determining what data may be utilised to evaluate a job application[2] [17][18].

Bias and prejudice are not limited to face-recognition technologies and social media algorithms.

Artificial intelligence (AI) is attracting the attention of a growing number of businesses. "IBM strongly opposes and will not condone any technology, including that offered by other vendors, that is used for mass surveillance," racial profiling and violations of basic human rights and freedoms," stated CEO Arvind Krishna. IBM has discontinued its general-purpose facial recognition and analysis products.

Forensics and Machine Learning

One day in November of 1984, a passerby in Milwaukee found the mutilated body of a lady who'd been brutally murdered. Based on their results, forensic dentists believe that someone with a missing front tooth was responsible for the eight bite marks on the victim's body. Police discovered Robert Lee Stinson, a nearby resident, matching this description after a neighbourhood check. He had been formally charged with murder and presented in court.

"Teeth similar to Stinson's" were the cause of the bite marks during his trial, according to one expert witness. "Overwhelming evidence," a second expert witness said. The bite marks on Stinson's torso led to him being sentenced to life in prison. After 23 years in prison, DNA evidence from the victim's clothing established his innocence and overturned his conviction.

Any reader of Steven Mark Chaney or Keith Harward's descriptions of bite-mark evidence will not be surprised by this difficulty. Bite-mark analysis has serious flaws, as shown by these erroneous conclusions. There are others in this evidence, though.

Pattern-evidence domains, according to the National Academy of Sciences (NAS), suffer from subjectivity and lack scientific validity, as was noted in a research published in 2009.

Bite marks are a kind of "pattern evidence," which also includes "image evidence" such as handwriting, gun and tool marks, fingerprints, and shoe prints. Scientific validity was judged to have been lacking in the pattern-evidence sectors, according to a 2009 National Academy of Sciences (NAS) investigation.

It is necessary to determine whether the victim's attacker was responsible for the bite marks on the victim. The bloody footprint may have been left by the suspect's shoe at the

crime site. Regardless of whether the response is yes or no, it does not necessarily mean that the suspect committed the crime, since there may be other (innocent) reasons for their connection to the site of the crime. Forensic investigators begin by determining whether the evidence originates from the subject of their investigation [4][19].

There may be some resistance to this stage even if there isn't any pattern. In this instance, a latent print was located at the crime scene, and a reference print was taken from the suspect's shoe, as illustrated in Figures. What we want to know is whether or not we can be certain that the latent print came from the suspect's shoe.

Figure 1 : Shoe Print for Forensic Analysis

There are many factors at play when comparing photographs like the one shown in Figure, making it difficult to draw conclusions. There is no generative model for reducing the dimensionality of imagegraphs while establishing an approved testing method. Since the picture size, rotation, and translation all vary over time, a pixel-by-pixel comparison is wrong. Another thing to consider is if the picture data can be reduced to to a few comparable measures [5][20].

Investigators rely heavily on forensic examiners' expertise and knowledge when evaluating and deciphering the vast majority of patterns found in crime scene evidence. The two samples are compared to see whether there are enough similarities between them to establish that they come from the same region.

However, even the pattern evidence judgements of forensic investigators with decades of expertise are open to interpretation. There isn't much consensus on what defines "similar enough" in the pattern disciplines. Error rates may be difficult to measure, both for individual examiners and as a whole. In other words, just because an examiner hasn't been brought to account in court doesn't imply they haven't messed up somewhere else. Various examiners or even the same examiner at different times in time examining the same material might lead to inconsistent judgments. In 2009, the National Academy of Sciences issued a proposal that the scientific and statistical foundations of any field of research should be strengthened promptly and often. In the decade that followed, machine learning technology was able to deal with the problem of subjective pattern evidence judgement.

Pattern-recognition and Machine Learning

To categorise data, "pattern recognition" refers to a scientific technique that automatically detects and uses patterns in data using computer algorithms. Algorithms for pattern

recognition, sometimes referred to as "learning algorithms" in statistics, may be used to make inferences about the future as well as predictions.

It's possible to categorise pattern recognition algorithms using the supervised learning and unsupervised learning techniques. In order to use supervised learning, a large number of units with known labels, known as a training set, is expected to be available If you prefer the term response rather than label or feature in statistical language, you might use the terms predictor or independent variable. f(X) is an approximated function that describes the relationship between feature X and label Y. This function is then used to train the algorithm. Algorithms must give an estimate of f in order to anticipate the labels of previously unknown items once they have been taught.

Unsupervised learning differs from supervised learning in that it does not reliant on pre-labeled data. It's possible that a big number of units have a certain set of measurements or qualities that are missing, but the associated response or label is not. In order to properly classify an unlabeled item, a computer approach is used to choose these features.

Learning algorithms have been developed in the field of weapon identification to answer queries about the provenance of a firearm.

In the case of two bullet samples, one from a crime scene and one from a suspect's handgun, comparing the two is standard practice. A final determination of whether or not samples can be discriminated rests with the examiner.

Bullet or cartridge casing markings are used to determine this information. There are a plethora of ways to make your mark on just about everything. There are several reasons

for this, but rifling, the spiral grooves that let bullets to spin as they depart, might be one (lands being the areas between each groove). Small imperfections in the finish of a rifle's barrel may also result in striations on the bullets themselves. The base of the cartridge case is marked by the breech face when a shot is fired. The breech face markings may be seen on the cartridge casing. For the most part, gun identification is predicated on the small sub-classes and individual striations seen on ammunition and the firing pin depression in cartridge casings [6][21].

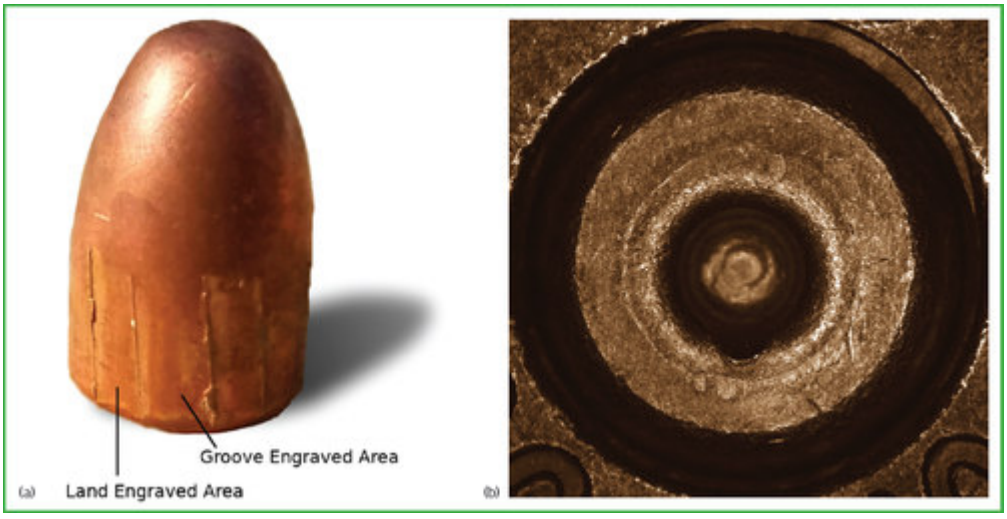

Figure 2 : Rifle Marks Analysis

Science and art are mixed together in the modern practice of firearms inspection. As a consequence of the fact that weapons examiners do not incorporate ideas like false-positive or false-negative rates in their assessments since they do not attach an estimate of uncertainty to any of their categorical distinct results [7][22].

Imaging technologies have been used to examine the surface topology of the etched areas on bullets and cartridge case bases over the last 20 years in order to objectively compare two samples. Strictly speaking, De Kinder and Bonfanti3 employed a laser

profilometer in order to measure the distance between two samples and determine the correlation between them. Since then, a slew of new methods have emerged for comparing the similarity of two objects using two- and three-dimensional images of bullets and cartridge cases. In 2017, Hare et al. established a similarity score for comparing bullets using supervised learning methods [8][23].

Arcs And Pinnacles

A (x, y) grid is used to measure the land engraved areas of bullets in micrometer-level increments. The resolution of pictures may be determined by the microscope. A total of 1.5625 nm by 1.5625 nm scans are included in Hare and coworkers. Each point on the grid has its (x, y, z) coordinates recorded, and the data collected here spans a surface area of 2.2 mm 0.6 mm.

The 3D graphic shows a Smith & Wesson rifle bullet with a land engraving on it. The red striations are the most useful. Measurement and comparison with the signatures of other bullets will be used to ascertain the "signature" of this bullet.

Figure 3 : Pattern Evaluation

Show a scanned picture of a bullet's land indented region in PowerPoint's figure viewer. The striations in the red portion are very telling.

Using a value of y where the striations seem to be stable, Hare and colleagues' technique looks at the average height of striations on several successive cross-sections of an incised region. Even if the grooves are clearly visible on both panels, automated identification may be difficult [9].

When a group of points is matched up to a curve, the term "Loess" is used to describe the process of fitting basic models to a small portion of the data. The loess function's residuals represent the real striations that go into creating a signature in the first place. Keep in mind that the residuals have a width of just 6 micrometres! Depending on the ammunition and rifle, the striations may be more or less visible.

Quantitative comparisons may be made after the signatures of two projectiles have been retrieved Cross-correlation and variations in peak and valley height or depth are some examples of quantitative data that may be used to compare two bullet signatures. Look closely, and you'll see that this weapon fired two separate rounds. In contrast, the signals' similarity suggests that they came from the same source.

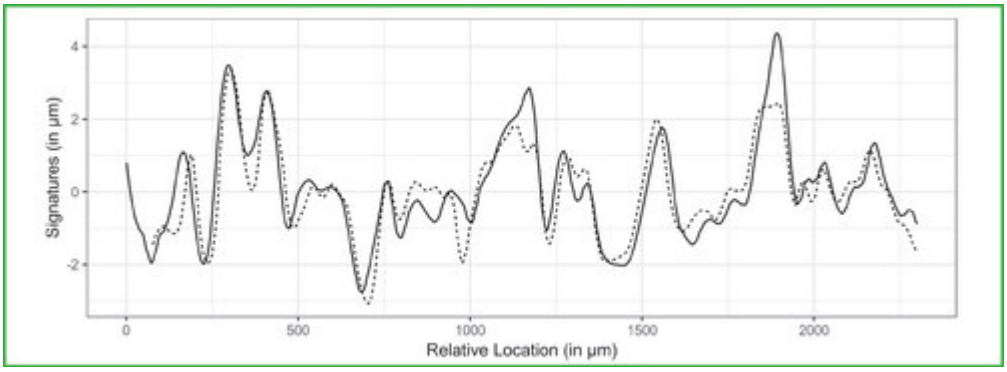

Figure 4 : Signature Overlapping and Analysis

For example, as seen in Figure, overlapping traces of two bullets fired from the same gun

Here is an experiment that was done to determine whether there was any way to tell if two rounds were shot from the same firearm by Hare et al.

Hamby and colleagues employed 10 sequentially rifled 9mm Ruger barrels to fire shots, and all possible combinations of photographs of land-etched areas were used to construct the photos used in the research study. 5 Certain land engraved regions on bullets fired from the same barrel (known matches) were discovered to match land engraved regions on bullets fired from other barrels (unknown matches) (known non-matches). For each pair of photographs, Hare et al. analysed the distribution of each characteristic's value between pairs with known matching and non-matching land engraving regions. The first graph contrasts pairs of known matching lands (light blue) with known non-matching lands (dark blue) to demonstrate examples of empirical distributions of the number of sequentially matching striae [12].

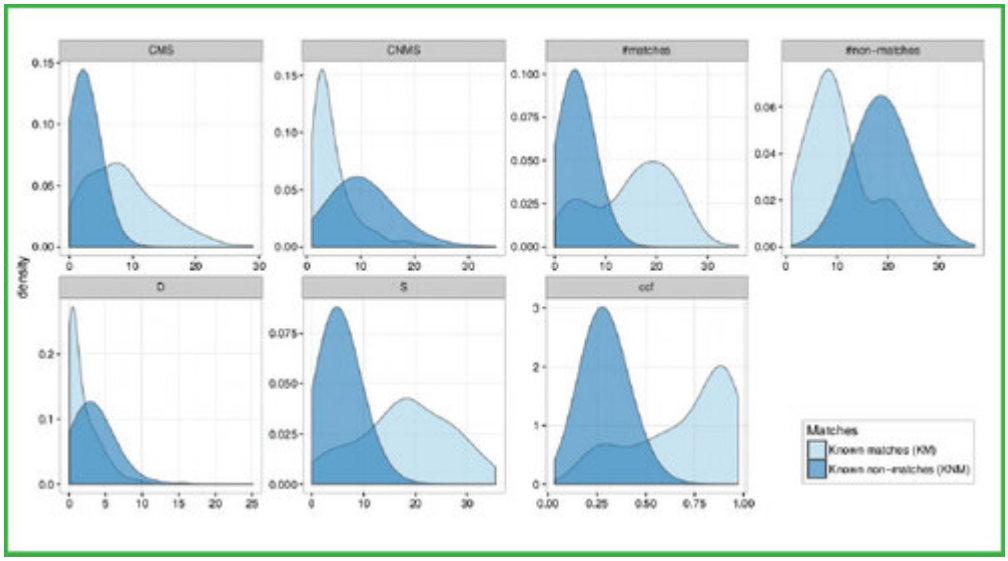

Figure 5 : Evaluation of Land Marks

Features that are known to match and that aren't known to match in pairs of land engraving sites (dark blue). Additional metrics include: striae (CMS and CNMS), peak and valley numbers, average depth difference between the two signatures (D), total area between the two signatures (S), and cross-correlation, in addition to the previously mentioned metrics.

Seven characteristics in Figure seem unable to tell the difference between pairings that match and those that don't. [13][24].

The random forest's categorization of Hamby bullet pairings into one of two categories had no false positives or false negatives. According to the results of the calculations, each pair that matched and each pair that didn't received different scores.

An overfitting tendency of learning algorithms may lead to much worse performance when categorising fresh sets of units and major misclassifications. One of the disadvantages of algorithms is that they might be difficult to recall. The misclassification error is underestimated even if a tiny portion of the training data is used for testing.

The random forest model was tested on thousands of pairs of bullets from crime labs around the country to see how well it predicted the identification of the Hamby bullets. No false positives or negatives were possible due to conventional (rather than polygonal) rifling, obviously visible striations, and bullets coated in polymer that flakes when it comes into contact with the barrel. However, significant testing and validation are required before these methods can be put to use in the real world due to a number of roadblocks.

In order to understand what it means to be "non-matched," we must first define the term "non-matched." It's vital to remember that the distribution of scores between unmatched pairings is directly affected by these parameters. A big collection of 3D photos of bullets whose source is known is the only way to overcome this challenge.

Second, the grooves on the discs themselves are difficult to recognise, making automation of the whole process challenging. Current methods of groove recognition rely heavily on human judgement and hence have a higher chance of inaccuracy and a longer processing time.

There will be a considerable shift in the techniques and methods utilised by weapons examiners to evaluate evidence as a result of this new technology. As a result, they are unlikely to be put into effect anytime soon.

With the pioneering work of the Center for Statistics and Applications in Forensic Evidence (CSAFE), machine learning is destined to revolutionise pattern evidence and other sorts of evidence (CSAFE).

Convolutional neural networks and multi-channel deep features are also used by Kong and colleagues to identify the model and markings of the shoe that left a print at the crime scene in addition to the bullet matching work presented below [16].

With the suitable datasets and thorough testing and validation, algorithms may help eliminate some of the subjectivity that permeates most forensic disciplines. They may also serve to quantify the degree of uncertainty in forensic findings.

Conclusion

As evidence, media records are crucial in digital forensics studies. It is critical when working with photos to be able to identify the subject and identify any forensic patterns. It is common practice to employ picture clustering algorithms in forensic investigation to categorise and compare image clustering techniques in various publications. For the benefit of digital forensics researchers, we've identified and discussed the most serious difficulties in the area, as well as potential solutions and future study avenues. SSOM, Kernel K-means and subject-based, LDA, and various more clustering methods have been evaluated. Machine learning approaches provide more efficient picture forensics and analytics, as well as more accurate image assessment.

References

[1] Omer Kaspi, Olga Girshevitz, Hanoch Senderowitz, PIXE Based Machine-Learning (PIXEL), Talanta, 10.1016/j.talanta.2021.122608, (122608), (2021).

[2] Shaodong Wang, Xiao Zhang, Yi Zheng, Beiwen Li, Hantang Qin, Qing Li, Similarity evaluation of 3D surface topography measurements, Measurement Science and Technology, 10.1088/1361-6501/ac1b41, 32, 12, (125003), (2021).

[3] Yiqun Jiang, Shaodong Wang, Hantang Qin, Beiwen Li, Qing Li, Similarity quantification of 3D surface topography measurements, Measurement, 10.1016/j.measurement.2021.110207, (110207), (2021).

[4] Hyunmin Kim, InSeok Kim, Kyounggon Kim, AIBFT: Artificial Intelligence Browser Forensic Toolkit, Forensic Science International: Digital Investigation, 10.1016/j.fsidi.2020.301091, 36, (301091), (2021).

[5] Soyoung Park, Alicia Carriquiry, The effect of image descriptors on the performance of classifiers of footwear outsole image pairs, Forensic Science International, 10.1016/j.forsciint.2021.111126, (111126), (2021).

[6] H, O. Ali., Abu, N. A., Abidin, Z. Z., & Darwish, S. M. (2022). Realistic Smile Expression Recognition Approach Using Ensemble Classifier with Enhanced Bagging. CMC-COMPUTERS MATERIALS & CONTINUA, 70(2), 2453-2469.

[7] James Rosenberger, Greg Ridgeway, Lingzhou Xue, Statisticians Engage in Gun Violence Research, Statistics and Public Policy, 10.1080/2330443X.2021.1978354, (1-12), (2021).

[8] Hassen, A., Abter, S. O., Abdulhussein, A. A., Darwish, S. M., Ibrahim, Y. M., & Sheta, W. (2021). Nature-Inspired Level Set Segmentation Model for 3D-MRI Brain Tumor Detection. CMC-COMPUTERS MATERIALS & CONTINUA, 68(1), 961-981.

[9] Xiaochen Hu, Xudong Zhang, Nicholas Lovrich, Public perceptions of police behavior during traffic stops: logistic regression and machine learning approaches compared, Journal of Computational Social Science, 10.1007/s42001-020-00079-4, 4, 1, (355-380), (2020).

[10] Oday. A., Abu, N. A., Abidin, Z. Z., & Darwish, S. M. (2021). A New Descriptor for Smile Classification Based on Cascade Classifier in Unconstrained Scenarios. Symmetry, 13(5), 805.

[11] Pattranit Pisantanaroj, Pimlapus Tanpisuth, Piyawut Sinchavanwat, Siriporn Phasuk, Phongphan Phienphanich, Parinton Jangtawee, Kittisak Yakoompai, Montri Donphoongpi, Sanong Ekgasit, Charturong Tantibundhit, Automated Firearm Classification From Bullet Markings Using Deep Learning, IEEE Access, 10.1109/ACCESS.2020.2989673, 8, (78236-78251), (2020).

[12] Oday, Ali, Hassen, "Face smile and related dimension analysis using deep learning", International Journal of Enterprise Computing and Business Systems(IJECBS), vol. 7, issue, 2, pp:1-13, 2017.

[13] Waqar Hussain, Nouman Rasool, Muhammad Yaseen, ADVIT: Using the Potentials of Deep Representations Incorporated with Grid-Based Features of

Dorsum Vein Patterns for Human Identification, Forensic Science International, 10.1016/j.forsciint.2020.110345, (110345), (2020).

[14] H, Oday A., and Nur Azman Abo. "HAAR: An Effectual Approach for Evaluation and Predictions of Face Smile Detection." International Journal of Computing and Business Research (IJCBR) 7.2 (2017): 1-8.

[15] Sophie J. Nightingale, Hany Farid, Assessing the reliability of a clothing-based forensic identification, Proceedings of the National Academy of Sciences, 10.1073/pnas.1917222117, (201917222), (2020).

[16] Abdulhussein, A. A., Kuba, H. K., & Alanssari, A. N. A. (2020, May). Computer Vision to Improve Security Surveillance through the Identification of Digital Patterns. In 2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM) (pp. 1-5). IEEE.

[17] Deisy Chaves, Eduardo Fidalgo, Enrique Alegre, Rocío Alaiz-Rodríguez, Francisco Jáñez-Martino, George Azzopardi, Assessment and Estimation of Face Detection Performance Based on Deep Learning for Forensic Applications, Sensors, 10.3390/s20164491, 20, 16, (4491), (2020).

[18] Abdulhussein, A. A., & Hassen. A. A Pragmatic Review and Analytics of Gait Recognition Techniques in Biometric Domain of Research. International Journal of Computing and Business Research (IJCB),Vol. 10 Issue 3 September - October 2020.

[19] Pablo Blanco-Medina, Eduardo Fidalgo, Enrique Alegre, Rocío Alaiz-Rodríguez, Francisco Jáñez-Martino, Alexandra Bonnici, Rectification and Super-Resolution Enhancements for Forensic Text Recognition, Sensors, 10.3390/s20205850, 20, 20, (5850), (2020).

[20] Hassen, Oday A., Nur Azman Abu, and Z. Zainal Abidin. "HUMAN IDENTIFICATION SYSTEM: A Review ".International Journal of Computing and Business Research (IJCBR), Vol. 9. Issue 3, pp. 1-26, September 2019.

[21] Xiaochen Hu, Xudong Zhang, Nicholas P. Lovrich, Forecasting Identity Theft Victims: Analyzing Characteristics and Preventive Actions through Machine Learning Approaches, Victims & Offenders, 10.1080/15564886.2020.1806161, (1-30), (2020).

[22] Alicia Carriquiry,Heike Hofmann,Xiao Hui Tai,Susan VanderPlas, Machine learning in forensic applications, 2019.

[23] F. Alkhabbas, M. Ayyad, R.-C. Mihailescu, P. Davidsson, A commitment-based approach to realize emergent configurations in the internet of things 2017 IEEE International Conference on Software Architecture Workshops (ICSAW) (2017), pp. 88-91

[24] F. Alkhabbas, R. Spalazzese, P. Davidsson, Iot-based systems of systems, Proceedings of the 2nd edition of Swedish Workshop on the Engineering of Systems of Systems (SWESOS 2016) (2016)