SoK: Exploring the State of the Art and the Future Potential of Artificial Intelligence in Digital Forensic Investigation

Xiaoyu Du
University College Dublin
Ireland
xiaoyu.du@ucdconnect.ie

Chris Hargreaves
University of Oxford
United Kingdom
christopher.hargreaves@cs.ox.ac.uk

John Sheppard
Waterford Institute of Technology
Ireland
jsheppard@wit.ie

Felix Anda
University College Dublin
Ireland
felix.anda@ucdconnect.ie

Asanka Sayakkara
University College Dublin
Ireland
asanka.sayakkara@ucdconnect.ie

Nhien-An Le-Khac
University College Dublin
Ireland
an.lekhac@ucd.ie

Mark Scanlon
University College Dublin
Ireland
mark.scanlon@ucd.ie

ABSTRACT

Multi-year digital forensic backlogs have become commonplace in law enforcement agencies throughout the globe. Digital forensic investigators are overloaded with the volume of cases requiring their expertise compounded by the volume of data to be processed. Artificial intelligence is often seen as the solution to many big data problems. This paper summarises existing artificial intelligence based tools and approaches in digital forensics. Automated evidence processing leveraging artificial intelligence based techniques shows great promise in expediting the digital forensic analysis process while increasing case processing capacities. For each application of artificial intelligence highlighted, a number of current challenges and future potential impact is discussed.

KEYWORDS

Digital Forensics, Machine Learning, Deep Learning

ACM Reference Format:
Xiaoyu Du, Chris Hargreaves, John Sheppard, Felix Anda, Asanka Sayakkara, Nhien-An Le-Khac, and Mark Scanlon. 2020. SoK: Exploring the State of the Art and the Future Potential of Artificial Intelligence in Digital Forensic Investigation. In *The 15th International ARES Conference on Availability, Reliability and Security, June 25–28, 2030, Dublin, Ireland.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Digital Forensic Science involves the recovery of evidence from digital devices, and is sometimes defined in terms of process models that capture the stages of the investigation [26]. For the purposes of this paper, the process is divided up into stages that assist in the discussion of where AI techniques have been applied to a digital investigation. These are: acquisition, examination, analysis, and presentation, similar to the Interpol guidelines [29]. This overall digital investigation process can be applied to a variety of data sources: traditional computers, mobile and other embedded devices (such as UAVs, smart home devices and other IoT devices). It can also apply to network forensics, cloud forensics, and live forensics.

This paper is structured into three main parts. Section 2 provides a brief introduction to artificial intelligence (AI) techniques, including the most useful references for a digital forensics researcher to become familiar with the area. Section 3 provides a series of sections, each describing a sub-area of the digital forensics field where AI techniques have been already applied. Each of these subsections has a consistent structure which starts with an introduction to the subtopic, a overview of the current AI applications in that area, and finishes with current challenges and future directions. Finally, Section 4 provides a general discussion of challenges and future directions for AI applications in digital forensics.

The contribution of this paper is therefore a comprehensive systematisation of AI research in digital forensics that can be used by digital forensic researchers and practitioners to identify the latest applications of AI in particular sub-areas, but also an important resource for AI researchers looking for real-world application areas for their new techniques and the challenges that are unique to applying AI in the specific field of digital forensics.

2 BACKGROUND ON ARTIFICIAL INTELLIGENCE

AI, or machine intelligence, is the discipline studying intelligent agents, i.e., an agent that reacts to its environment to achieve an optimal path to its goal. In Computer Science, AI can be split into two primary fields; Machine Learning (ML) and Deep Learning (DL). The success of AI is data-driven in so far as no explicit code controls the precise output. The datasets used for training the models are

ARES '20, August 25–28, 2020, Dublin, Ireland

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

critical, and data pre-processing is a key step in ML. An overview of the datasets available for training AI models in Digital Forensics is provided by Grajeda et al. [41].

2.1 Machine Learning

ML has been widely applied to digital forensic investigation for data discovery [23, 115], device triage [72, 73], network forensics [81], etc. Flach [33] outlined the ML ingredients as: **tasks**, the problems that can be solved; **models**, the output of ML; and **features**, the workhorses of ML. There are three steps for ML applications: 1) task definition; 2) feature construction; 3) evaluation and optimisation.

An ML task is an abstract representation of the problem. For a prediction problem, it can be defined to be either a classification/clustering or regression problem, depending on the type of target labels. Take age estimation as an example. If age is considered categorical, it can be defined as a classification task; while it could be a regression task if the age is numeric.

Feature construction is crucial for the success of ML application [33]. There are different kinds of features: categorical, ordinal and quantitative. For text analysis, the raw data is a sequence of symbols cannot be fed directly to algorithms, *bag-of-word* representation is applied. For image data, patch or contiguous patches can be extracted. During the experiment, features are transformed and selected to reducing over-fitting, improve performance or reduce training time. The *No-Free-Lunch* theorem implies that there is no ultimate feature learner; it is variable depending on the data distribution and learning algorithm [97].

Models are the output of ML [33]. Model evaluation enables its refinement, and the process is iterated until the performance is sufficient. A confusion matrix is able to show the accuracy of a classification task, where the classification performance of each class can be found. The F1 score is an average accuracy of each class, which shows the average performance of the model. Precision and recall are usually used in the evaluation matrix.

2.2 Overview of Deep Learning

The key differentiator of DL from ML is that the features are not designed by human engineers. Instead, they are learned from data using a general-purpose learning procedure [61]. ML tasks require input that is computationally convenient to process. However, it is often difficult to engineer features of real-world data such as images, video, and sensor data. Representation (feature) learning techniques employed by artificial neural networks (ANNs) allows a system to automatically discover the representations needed for feature detection or classification from raw data [61].

A DL model can be described in two stages; optimisation and inference. The optimisation process, known as training, is used to update the weights connecting the layers of neurons defined in the model. The process of weight update is achieved by a back-propagation algorithm [61]. Before training a DL model, a loss objective is defined to measure the difference/error between the predicted outputs and the targets. The model updates its weights with the objective of minimising the loss function through many times of iterations. To make it closer to the objective, the mathematics under the hood are gradient descent algorithms for minimising the loss [89]. After completing the optimisation, then the model is applied for inference, namely, making predictions on data that are unseen during training. One key performance metric e is the generalisation ability. That says if the model generalises well, it performs on the unseen (test) data as well as the training data.

DL is often applied for natural language processing (NLP) and computer vision (CV), but more specific applications include content filtering [116], e-commerce recommendations [98], and search result relevancy scoring [121]. Other applications that are discussed later include camera sensor model identification, image forgery detection, facial detection and recognition, text clustering, etc. Digital forensic specific applications include malware classification, network intrusion detection, file fragment typing, watermarking, steganalysis, pattern recognition, timeline analysis, etc.

3 APPLICATIONS OF AI IN DF

3.1 Data Discovery and Recovery

One of the early stages of a digital investigation is making the digital evidence obtained available in a human readable form [55] (extraction). This can include extracting information from known file systems and file types, but also recovering deleted data.

3.1.1 *State of the Art of AI in Data Discovery.*

Files deleted within a file system may be recoverable deterministicly if some metadata remains. However, in some cases this metadata is absent and the file content resides in the unallocated parts of a volume. File carving is the process of recovering such files without the metadata. However, it is also possible that such files may be fragmented over the disk and partially overwritten. Garfinkel [38] reported on fragmentation statistics collected from over 350 disks containing FAT, NTFS and UFS file systems. While fragmentation on a typical disk is low, the fragmentation rate of forensically important files such as email, JPEG and Word documents is relatively high.

As the search space for fragments belonging to a particular file is so large, distinguishing the file type of a fragment can shorten the search time. One approach proposed for file fragment classification used NLP [32]. In this research, a supervised learning approach is taken based on the use of support vector machines (SVM) combined with the *bag-of-words* model. File fragments are represented as "bags of bytes" with feature vectors consisting of unigram and bigram counts as well as other statistical measurements (including entropy). Chen et al. [23] proposed a novel scheme based on fragment-to-grayscale image conversion and DL to extract hidden features and therefore improve the accuracy of classification. This CNN model was trained and tested on the public *GovDocs* dataset. The average classification accuracy achieved was 70.9%. Vulinovic et al. [115]. Vulinović et al. [115] applied a CNN model using 1D convolution on the original byte block. Both feedforward neural networks (FFNN) and CNNs are tested. FFNNs achieved better results using selected bigrams as input the highest macro-average F1 score being 0.8138.

Another problem faced during file carving is to determine the ownership of carved information when the storage media is used by more than one user. An automated solution to the multi-user carved data ascription was proposed by Garfinkel et al. [39]. The features used by the automated ascription system are 1) file system metadata (MAC timestamp, file owner), 2) file placement (i.e., sector, fragment) information , 3) embedded file metadata (JPEG camera model, Word file save time, etc.). The data used to verify this system is disk images from the Real Data Corpus [37], a collection of more than 2,000 disk images made from hard drives that were purchased on the secondary market. The result shows accuracy of classification is from 65.66% to 99.83%. In the end, this approach achieved a low accuracy (0%) considering no discernible difference between the activity patterns of each user.

3.1.2 Current Challenges and Future Directions. The current literature shows automation in digital forensic investigation employing statistical measurement for data representation and ML algorithms for classification. ML techniques have the potential to acquire useful information for investigations more efficiently – leveraging the accumulation of experience learned from the previous digital evidence analysis. Adversarial attacks are one of the challenges of AI model development. It has been suggested that the existence of adversarial attacks may be an inherent weakness of DL models [69]. The adversary can manipulate the input resulting in incorrect output. Adversarial attacks could also be used as a counter forensics technique. As a result, any pre-trained model could loose its effectiveness during an investigation. To this end, anti-counter-forensics for adversarial attacks remains an open question.

3.2 Device Triage

With the proliferation of digital evidence, the data volumes encountered in investigations is a significant challenge faced by Law Enforcement Agencies (LEAs). Digital evidence triage was proposed for the timely identification, analysis, and interpretation of digital evidence, with a process model proposed in Rogers et al. [88]. Currently, the prioritisation of device acquisition and processing at a crime scene is determined by the investigative officer. As more AI based techniques are developed, on-scene preliminary inspections could quickly focus the analysis towards the devices most likely to contain case-progressing information first.

3.2.1 State of the Art of AI in Device Triage. With the increasing significance of mobile device forensics, Marturana et al. [72] proposed an approach for device prioritisation leveraging data mining and ML theory. This work presents the result of a study concerning mobile phone classification in a real child abuse investigation case. The features used consisted of the phone model, phone contacts, calls made, text messages sent/received/read, number of video/audio/photo files, URL, email, memos. The experimentation tested the performance on the feature value represented as numeric (a number) and category (the number is low, medium or high).

In some subsequent work, Marturana and Tacconi [73] expanded the triage approach to detect the device's relative importance using features from: *1) the timeline of events, 2) the crime's specific features, and 3) the suspect's private sphere (habits, skills and interests).* The experimentation in this work was conducted on a copyright infringement and a CSEM exchange case. The dataset applied consisted of 23 cell phones for the CSEM case with 13 digital media files and 45 copyright infringement-related features. A result of 99% correctly classified samples on both cases was achieved.

3.2.2 Current Challenges and Future Directions. The lack of a sufficiently large, shared dataset is a challenge for developing AI triage models. As the triage task consists of a quick, simple examination and analysis to help investigators to reduce the noise and identify relevant information quickly, the development of a emulated, realistic dataset is a substantial task.

Future digital investigation may heavily rely on efficient device triage. The report of serious digital forensics backlogs [94] indicates comprehensive examination of all digital devices is almost impossible. Increasing the accuracy of triage would result in less resources wasted processing non-pertinent data. In addition, and common for all ML approaches, the training dataset determines the performance of the model. The higher the volume and quality of data used to train the model, the better the model will perform.

Investigations involving multiple devices is common, if not the norm. Multiple device analysis and triage can be integrated. For example, when determining the importance of the devices, the actions between them can be considered, e.g., file sharing, device connections, information exchanging, etc.).

3.3 Network Traffic Analysis

The voluminous nature of data associated with Network Traffic Analysis (NTA) makes it an excellent candidate for the application of AI techniques to help filter redundant information and automate the detection of crimes or other forms of misconduct.

3.3.1 State of the Art of AI in Network Traffic Analysis. Network investigations often form a part of a bigger investigation involving incident response, cloud, IoT, mobile devices, wearable technologies and fraudulent monetary activities. These investigations tend to involve multiple devices or technologies which have been communicating with each other. A wealth of literature is available to investigators for the use of Intrusion Detection techniques to network data offline in batch mode after the fact. Surveys on the use of AI in IDS can be found in [81], [92], [14] and [2].

Feature selection techniques impact heavily on the models produced by AI techniques. The most up to date datasets for intrusion detection include the CICIDS 2017, CICIDS 2018 [99] and CICDDoS2019 [100] whose features are constructed using CICFlowMeter-V3 [58]. Principal Component Analysis (PCA) has been applied to these datasets. AI techniques used to model the CICIDS datasets include SVMs and DL [3, 110, 113]. Auto-Encoders and PCA were used for dimensionality reduction in [1]. The reduced datasets were evaluated using classifiers such as Random Forest (RF), Bayesian Network (BN), Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). AI techniques have successfully been employed for Botnet Detection using ML on DNS requests [12, 101], while [4] used traffic reduction with Reinforcement Learning (RL).

Elrawy et al. [28] surveyed the challenges of security in IoT and presented a comprehensive review of current anomaly-based IoT IDSs. Deng et al. [24] proposed a lightweight ML NIDS for IoT environments using a combination of fuzzy c-means clustering (FCM) and PCA, while Amouri et al. [6] presented a NIDS with low computational and resource requirements using decision trees. A data mining approach for an IoT NIDS using PCA and suppressed fuzzy clustering (SFC) techniques proved to be well suited to high dimensional spaces producing high levels of accuracy [65]. Pour et al. [83] applied PCA for feature selection, with clustering techniques, to IoT data to infer exploited IoT devices, and IoT coordinated probing campaigns.

Indicators of Compromise (IOC) are evidence artefacts that are indicative of a system or network being attacked. Useful sources of this data can be network packets or network logs. In [85], network data was collected and features extracted before applying clustering techniques to extract IOC rules for malware detection.

Android network traffic was modelled for malware detection in [78]. RF, K-Nearest Neighbour (KNN), decision Tree (DT), Random Tree (RT) and Regression were all applied to the CICAndMal2017 dataset which was generated and made available by the authors. An updated version of the dataset was evaluated by Taheri et al. [106] using Random Forest classification. This also utilised API call data for classification. A DT approach for the detection of cryptocurrency miners is presented in [112].

3.3.2 Current Challenges and Future Directions. Network traffic analysis is becoming increasingly hierarchical – providing better potential for correlation of user data over multiple networks or devices. Modern networks and devices allow for the broader profiling of individual suspect users and their actions. Correlation of incidents in new and emerging environments should also be inter-device dependent. This is of particular importance in areas such as in the event of an modern automobile crash. Network traffic analysis will also see growth in inter-user correlation, e.g., the correlation of mobile phone communication through applications over data networks to identify who a suspect is in contact with most frequently or most recently relative to a certain time period.

One of the biggest challenges network analysis faces is the huge increase in volumes of data from new and emerging devices that needs to be gathered, stored and modelled. Classification and prediction models require accurate and up to date datasets with the correct features. Datasets traditionally used in this area have suffered from problems such as a lack of relevant or real-world data, bias and disproportionate classes. In the area of IDS, these datasets by their nature are always behind the curve in terms of up to date attacks. This creates issues in the creation and evaluation of accurate models. GDPR legislation also raises issues around user privacy for inter-event correlation. Novel protocols associated with emerging devices can result in previously undocumented network traffic patterns. This can affect the performance of existing pre-trained AI models, which obviously may not have taken this new activity into account during training. Encryption poses a challenge to network traffic analysis but does not hinder it completely. Even with encrypted networks, AI techniques can still be used to model statistical information of a network.

3.4 Forensics on Encrypted Data

One of the most significant issues facing digital forensics investigators around the globe is encrypted data. The prevalence of cryptographically protected devices and data poses an inevitable threat to digital forensic investigation. If the device under investigation uses disk encryption, the forensic disk image becomes unusable [64]. Currently, the law of many countries demands that the owner of the device has to surrender their passwords/keys to LEAs under warrant. However, unavailability/non-compliance often brings the investigation of the encrypted device to a halt [114]. Due to the large bit length used in modern cryptographic algorithms a successful brute-force attack is computationally infeasible.

Side-channel attacks on encryption algorithms have been proven as effective key attack vectors [102]. An electromagnetic side-channel analysis (EM-SCA) attack is performed by observing the EM emissions over time of a device under test (DUT) while it is performing data encryption/decryption. A single such observation is called an *EM trace* containing the three signal characteristics; *Amplitude*, *Phase*, and *Frequency*. Once a sufficient number of EM traces are collected, they are fed into an EM-SCA algorithm, e.g., differential electromagnetic analysis (DEMA) or correlation electromagnetic analysis (CEMA), to extract the underlying cryptographic key [53, 54]. These algorithms require the EM traces to be precisely aligned in the time-domain in order to succeed. Due to the nature of EM trace extraction, minor misalignments often force the attacker to extract more EM traces – this alignment issue can be greatly improved by ML [93]. Furthermore, these algorithms can take a considerable time to complete, making them difficult to be used in live device investigation scenarios in digital forensics [60, 107, 124].

3.4.1 State of the Art of AI on Handling Encrypted Data. There are two potential avenues for EM-SCA that can be assisted by AI techniques; gaining useful insights without accessing the encrypted content, and performing cryptographic key retrieval attacks. Towards the first goal, various AI approaches have been applied using power and electromagnetic side-channel observational data [93]. Knowing whether a target device is running the expected software/firmware can be useful to the investigator, i.e., a malicious user may have modified the firmware. DL algorithms such as multilayer perceptron (MLP) and long short-term memory (LSTM) have been used to detect anomalies in IoT devices through power consumption side-channels [117]. Furthermore, various insights such as the identification of the specific hardware device or software application, and the behaviour of the software are shown to be identifiable with DL methods [15, 16, 57, 62, 79, 103].

Ronald Rivest, one of the co-founders of the RSA algorithm, discussed the inter-relationship between cryptography and ML three decades ago [86]. Cryptanalysis attempts to retrieve cryptographic keys by analysing a large amount amount of information, i.e., plaintexts and ciphertexts, connected by an unknown key. There exists an interesting similarity between this and ML that eventually paved way to cryptographic key retrieval attacks leveraging ML and DL.

Template attacks are a common key retrieval approach whereby an attacker has a testing device similar to the target device [21]. A template can be built for the test device and subsequently used to attack the target device. It has been shown that SVMs are applicable in similar circumstances and provide comparable performance in key retrieval attacks [45]. Experimental studies show that ML and DL methods can succeed even when cryptographic implementations use side-channel mitigation techniques to counter attacks [70]. Furthermore, DL architectures, e.g., CNNs, are increasingly being applied to this problem [11].

3.4.2 Current Challenges and Future Directions. It is reasonable to expect that a large percentage of computing devices encountered in digital forensic investigations in the future will be encrypted. Therefore, cryptography is turning into a critically important challenge in digital forensics. With the increasing popularity of software defined radio (SDR) hardware, acquisition of EM traces becomes easier and more affordable [10, 68]. Meanwhile, with the rapid increase of computational resources, ML and DL methods that are capable of performing key retrieval attacks can be expected to be more and more sophisticated. This can lead to a significant reduction in the time required for key retrieval.

Various side-channel mitigation techniques exist to defend against attacks to cryptographic implementations e.g., randomisation of operations, masking variables with random values, accessing critical variables indirectly via pointers, and hardware shielding [52, 91, 122]. Furthermore, secure cores that are dedicated for cryptographic operations are increasingly present in modern computer processor chips. Operations performed inside such cores lower the side-channel information leakage, forcing attackers to use more sensitive measuring equipment and sophisticated pre-processing of EM traces in order to perform key retrieval attacks [35, 74].

Software implementations of cryptographic algorithms tend to evolve over time due to updates carrying bug fixes and improvements. Such changes to software tend to impact the corresponding EM emission patterns. Therefore, ML and DL models that are trained to recognise patterns or retrieve keys can get affected by these changes. The ability of ML and DL models to generalise the minor changes of EM traces needs to be explored further. Meanwhile, tools and frameworks are needed to facilitate the application of ML/DL techniques for law-enforcement.

3.5 Timeline/Event Reconstruction

Event reconstruction in digital forensics has been defined in terms of finite state machines by Gladyshev and Patel [40]. However, it less formally refers to a process that can "convert the state of the [digital] objects into the events that caused the state" [17]. This can include simply being able to determine that some event occurred, or more precisely that an event occurred at a certain time. This second, more detailed event reconstruction would be achieved by looking at the timestamps recoverable from digital forensic artefacts. Sources of timestamps would include times from the file system, e.g., file modified, accessed, created, entry modified, etc., but can also include timestamps from inside more complex file formats, e.g., Windows Registry, SQLite databases, event logs, etc.

The state of the art in terms of timestamp extraction is Plaso (log2timeline), which has many plugins and parsers[1]. However, the challenge is that the analysis of a system, even with minimal user activity, would generate millions of these timestamps. Attempts have been made to perform automated analysis of this high volume of timestamps and infer a usable activity history from this data. One approach by Hargreaves and Patterson [44] involved manually coding the pattern of low-level timestamps associated with a 'higher level' event, e.g., a user opening a file on a Windows system produces a series of 'low level' artefacts including entries in the Windows registry, link file changes, jump lists, and others. This can be manually encoded and pattern matched. However, this can be a time-consuming process to identify these changes, code and test them. It is also potentially error prone as subtle differences in behaviour of operating system versions could produce incorrect inferences. There are also representation problems for events, something that was examined by Chabot et al. [19], with a correlation of events also discussed in [20].

3.5.1 State of the Art of AI in Event Reconstruction. Despite the potential of ML approaches in this area, there are relatively few papers on ML applied to pattern matching in timeline data. Muhammad Naeem Khan and Young [76] and Khan [50] discuss a neural network-based approach for event reconstruction using file system times and describe that neural networks are appropriate for dealing with the large volumes of data because of their parallelism and generalisation capabilities. They tested both feedforward and recurrent neural networks. Turnbull and Randhawa [109] developed ParFor, which as a result of the explainability problems of other ML techniques, use Symbolic AI based on an ontological representation of forensic artefacts and implemented inferences such as computer on/off. However, Studiawan et al. [104] used DL techniques to highlight events of interest in a timeline based on positive or negative sentiment in the text-based representation of events (specifically operating system logs), e.g., 'failed password' or 'authentication failure'.

3.5.2 Current Challenges and Future Directions. There are a number of challenges in this area. Performing event reconstruction using timestamps inherently makes the assumption that the timestamps are correct. There are many reasons why this may not be the case, e.g., clock drift, manual changing of the system clock, overwritten timestamps as part of normal system processes, or anti-forensic techniques. There is some work in mitigating some of these, e.g., Marrington et al. [71] developed a rule-based approach to detecting timestamp inconsistencies, but there may be merit in testing a ML based approach to this problem too. It may also be possible to use many of the approaches developed in a network forensics and traffic analysis context and apply to artefact timeline analysis. Correct inference of user activity is also a challenge; the low-level events generated for one version of an operating system are not necessarily the same in other versions and evaluation of the correctness of the inferences is critical Jeyaraman and Atallah [48]. Finally, labelled and verified datasets for forensic timelines are very difficult to obtain and time consuming to generate at scale [25].

There are many areas to explore in the application of AI to event correlation. Aside from identification of individual events, it could also be possible to have higher level 'anomaly detection' applied to a system. This is a very difficult problem given the multipurpose nature of typical computer systems and that the difference between legal and illegal activity may be very subtle. Nevertheless, in terms of applications, timelines have great potential to allow easier analysis across multiple applications, e.g., chat messages in multiple clients, or across devices [43]. Finally, a timeline-based view of activity is just one view of a data set, but it can provide a useful entry point into a dataset, allowing the view to be then 'pivoted' to file system views, similar content and back to timelines.

3.6 Multimedia Forensics

Multimedia forensics is a branch of digital forensics that studies content such as audio, video and images that have been obtained as part of a digital forensics investigation and can include not just computers and mobile devices, but also CCTV analysis. There are a number of aspects to explore in this topic. The first is the problem of volume. Typical devices will contain thousands of media files and identifying those that are relevant can be a challenge as they

[1] https://plaso.readthedocs.io/en/latest/

cannot simply be keyword searched. The second area is analysis to determine the media's provenience, which could provide a link to a suspect. The third area is forgery detection as digital images can be easily tempered with.

Object detection also has a role to play. A social media crowd sourcing approach by Europol has been used to trace objects to combat child abuse. The organisation explains that even the most innocent clues on photographs can aid investigations. Their aim is that once the origin of an object is identified, the LEA of the country involved will be informed to further investigate the lead and speed up the identification of both the offender and the victim.

3.6.1 State of the Art of AI on Computer Vision. In terms of identifying relevant images from a large set, the search for objects of interest in digital images is arduous due to the large volume of seized devices. The need for automated object detection, specially in low quality images is required and has also triggered the need to develop effective image mining systems for digital forensics purposes [13].

In terms of general approaches, object identification can be tackled with CNNs. In 2016, Grega et al. [42] presented the automated detection and recognition of dangerous situations such as events where firearms and knives are present in CCTV footage. The algorithm proposed for knife detection is based on visual descriptors and ML. The algorithm for firearm detection is limited to a pistol and is based on a PCA approach. In 2017, a crowd-sourced and CV based approach to fight sex trafficking was proposed; hotel identification with a search-by-image based on features extracted from neural networks was implemented [105]. Later in 2019, Xiao et al. [119] proposed a DL-based object detection and tracking algorithm to identify potential suspects from footage. Their approach for low quality video/image analysis is based on contrast limited adaptive histogram equalisation that improve CCTV quality and is used for Digital Forensic Investigations. Similarly, Jasmine and Annadurai [47] proposed a real-time video quality enhancing method using adaptive histogram equalisation. Also, regarding the use of illicit substances, Yang and Luo [120] proposed the tracking of drug dealing and abuse on the Instagram social network by using multimodal analysis including methods such as multi-task learning and decision-level fusion; this approach enabled the ability to identify drug-related posts and the examination of behaviour patters of drug-related user accounts.

Specifically related to CSEM investigations, many practitioners are unfamiliar with AI, but demand automated nudity, age and skin tone detectors [90]. This is unsurprising as it has been reported that some law enforcement personnel have suffered ill effects due to the continuous exposure of CSEM [118], and it has been proven to affect some groups by causing secondary traumatic stress disorder [67, 90]. To lessen the exposure to CSEM, multiple approaches have been considered. Skin detection algorithms could potentially sift unnecessary images and flag inappropriate content. In 2005, Ap-Apid [8] developed a skin colour distribution model based on RGB. The aforementioned nudity detection algorithm had a 95% recall with a 5% false positive rate. Later in 2016, Deep CNNs were used by Nian et al. [80]. The latter demonstrates the advantage of using AI over hand-engineered visual features that are hard to analyse and select. The notable trend of CNNs has been flooding research topics in the past years.

Another relevant CV area is age estimation. In 2020, Anda et al. [7], proposed the segregation of the age component from a CSEM investigative model. This approach tackles specifically the facial age estimation problem for underage subjects, which can be further consolidated with a nudity component to create a CSEM ensemble. This approach has achieved a mean absolute error (MAE) rate of 2.73 years. In order to tackle unbalanced dataset and bias, a balanced dataset generator was used [31]. Age estimation is a challenging task for both humans and computers. The range of factors that influence age prediction are considerable. Environment, habits, diets, use of anti-ageing products, smoking, drinking, drug abuse, skin tone, gender, etc. are only some identifiable parameters that can change the course of the appreciation of age. Nevertheless, in certain age groups (newborn and children), the influence of these factors has less of an impact. Age prediction may also have other applications for digital forensics including suspect and victim identification. Missing children cases could benefit from Generative Adversarial Networks that are able to estimate images of victims creating aged versions from an input image.

3.6.2 State of the Art of AI in Forgery Detection. Finally, as mentioned above, detecting forgeries in images is also a challenge. A digital image has been accepted as a "proof of occurrence" of an event [51, 96] and so it is important to demonstrate that it is authentic. Farid [30], classified tools to detect image forgery into five categories:

(1) Pixel-based techniques that detect statistical anomalies introduced at the pixel level.
(2) Format-based techniques that leverage the statistical correlations introduced by a specific lossy compression scheme.
(3) Camera-based techniques that exploit artefacts introduced by the camera lens, sensor, or on-chip post-processing.
(4) Physically based techniques that explicitly model and detect anomalies in the three-dimensional interaction between physical objects, light, and the camera.
(5) Geometric-based techniques that make measurements of objects in the world and their positions relative to the camera.

The previous techniques are mainly statistically, geometrically and physically-based scientific methods, and are solely for the validation of the integrity of images. Nevertheless DL based techniques have also been used to detect image manipulation [9, 22, 84, 123]. The first three studies employ CNNs and the final study implements a Stacked Auto-encoder (SAE) approach.

3.6.3 Current Challenges and Future Directions. Automated mechanisms to detect CSEM have been used in the past with skin tone detection algorithms or hash comparisons. Nevertheless, either the performance has not been adequate or the approach used has been trivial. The rise of CNNs has enabled impressive and promising results. There are still a myriad of application to explore that could improve the performance of algorithms to detect CSEM. Images with low resolution and visibly challenging to the human eye could be tackled with the application of specific models trained on low quality data. Objects that have been found on CSEM with low quality can benefit with the creation of ensembles for different type of items matching certain quality standards.

However, the need for shared, well curated datasets in the research community is clear. Data pollution present in datasets that are already being shared in the community may present a risk to further research. Keeping big data under control may become a challenge and could be subject to data protection acts that would hinder certain types of longitudinal research. Unavailability of information due to ethical concerns and lack of transparency can impede the creation of reliable models. Non-robust models could also be subject to adversarial attacks that could bypass certain systems such as nudity detectors, and age limit systems.

Nevertheless, automation in multimedia forensics could help alleviate the digital forensic backlog by optimising analysis and prioritise artefacts in an intelligent manner. As previously highlighted, the usage of CNNs in digital forensics shows great promise. Proposed models should emphasise expertise while focusing on solving smaller problems rather than generalising to attempt to solve several problems. Ensemble models are considered to be more stable and, most importantly, predict better than single classifiers [63]; therefore, an ensemble of expert models would improve the performance while decreasing errors.

3.7 Fingerprinting

Device fingerprinting is a growing area of digital forensics. It ranges from server-side browser fingerprinting [27] (based on the unique set of browser and extension metadata/configuration sent to a web server), camera sensor identification [66] (based on subtle imperfections of camera sensors), to malware behavioural analysis and classification [59] (based on program execution patterns).

3.7.1 State of the Art of AI in Fingerprinting. The task of fingerprinting lends itself well to AI classification techniques. For example, malware classification has been a popular application area, with significant existing work in this area [36]; both static [59, 77] and dynamic analysis [46, 56]. With respect to the aforementioned media provenience issue, Tsai et al. [108] were able to obtain highly accurate predictions with SVM on similar photographed scenes generated both by traditional and mobile-phone cameras. Also, CNNs have shown promising results on image recognition, video analysis and NLP.

Similarly to how scratches on a bullet facilitate the identification of the weapon that shot it, subtle imperfections in digital camera sensors leave their imprint on the resultant digital photos and videos. This allows the subsequent association of this content with a specific camera sensor [66]. This approach can be used to identify both the specific make/model of the source camera, e.g., an iPhone 11, and potentially the specific camera, e.g., *this* iPhone 11. Identifying the camera model with which a video has been taken can provide valuable insight in an investigation. Freire-Obregon et al. [34] implemented a Source Camera Identification (SCI) method that is able to infer the noise pattern of mobile camera sensors/fingerprints. Their CNN approach has achieved over a 90% of accuracy in determining not only the brand of the phone but also identifying if the front or rear camera was used. CNNs are capable of performing image manipulation detection as well as camera model identification.

In a similar vein, fingerprinting techniques have also been used for authorship attribution. This authorship attribution ranges from open source intelligence/social media attribution [87], source code attribution [49] and malware attribution [5]. Authorship attribution relies on identifying unique programming or language traits of the individual behind the keyboard.

3.7.2 Future Directions. Device and user behavioural fingerprinting can greatly aid in anomaly detection. For networked devices, this can result in more accurate host-based and network-based intrusion detection. Modelling the usage/behavioural fingerprint of each user on a system can similarly be used as an indicator of account compromise. In online video streaming scenarios, camera sensor detection combined with device fingerprinting can be used to identify the source of the stream.

4 CONCLUSION

This paper has shown how a range of AI techniques are currently used across different areas of digital forensics. It has also highlighted common challenges including availability of data sets in some areas, specific difficulties in explaining the results when certain techniques are used, and even challenges in releasing models where potentially restricted training data could be inferred from the models [82]. However, despite these challenges, there is enormous potential for future work. As discussed above, this is both in terms of improving the performance of some of the current techniques, but also that there are some approaches that have not yet been tested in individual areas. These gaps should now be more easily identified as a result of this systematisation of knowledge and help accelerate developments in this field.

4.1 Future Directions

The previous sections have shown that there is significant existing work in the application of AI to specific areas of digital forensics. This section discusses general challenges, and potential opportunities including unexplored areas where existing and emerging AI techniques have not yet been applied.

In terms of general challenges, improving the accuracy of techniques is an obvious focus area. Specific to digital forensics, training models and measuring the accuracy is a challenge because of the lack of large, clean, labelled datasets in some areas or existing datasets not being publicly available. While there are extensive datasets that can be used to train computer vision based approaches, the availability of sensitive datasets, such as CSEM datasets, are understandably and necessarily restricted. Datasets for whole hard disk approaches, e.g., needed for timeline analysis, do not exist in a useful manner; where user activity is clearly documented and labelled allowing DL techniques to identify relevant features. Even if such disk images were produced, having sufficient background 'noise' is also difficult, meaning that techniques developed in a research setting are unlikely to work when exposed to data from a real investigation. Automated digital 'story' generation is needed to address these issues.

While Explainable AI is a general computer science problem, for digital forensics this is paramount for the court admissibility and understanding of evidence. However, it should be highlighted that there is some subtlety to this. For example, an AI process reporting that a system containing criminal activity needs to be able to produce a very clear explanation of why that is the case. However, a human-in-the-loop approach that is designed to highlight to an investigator data that is likely to be relevant does not necessarily have the same explainability requirement and is something called for by law enforcement [90]. There is still a danger here in that bias can be a problem in investigations in general [75], but a system that is promoting 'relevant evidence' has the potential to bias an investigator. A related problem is validation – increasingly necessary for methods used in a digital forensic context. This means that a technique should be applied to known data and produce an expected result. Subsequently, once validated, that technique can be used. New versions of software, or in the case of AI models, new models mean the techniques should be re-validated. The edge case is in a scenario where an AI model is updating live, for example learning from on-going case processing to expedite evidence discovery in future cases. In this case, the result the technique may produce may change on a daily basis, posing a significant validation challenge.

Finally, it should also be considered whether sharing models is appropriate in some contexts. AI trained models in the context of GDPR and summary of attacks such as 'model inversion' and 'membership inference' is discussed by Veale et al. [111]. This is therefore worth considering when developing digital forensic solutions using AI and potentially sensitive training data.

Despite these challenges, there are many opportunities to enhance AI applications and to apply AI to additional areas of digital forensics. These include inference of behaviour from data obtained from novel sources including smart homes, IoT sensors, vehicle forensics, and combinations thereof. Indeed AI techniques could potentially assist any time there is a need to correlate data from multiple sources, either from multiple suspects, devices or cases. Non-AI based efforts such as standard form of representations, e.g., CASE [18] will be critical for such efforts.

There will also be significant opportunities in the future for the investigation of AI based systems themselves. Determining the cause of a decision made by a self-driving car, a smart building, or a SCADA system, will be a new area for digital forensics, although the concept is discussed by Schneider and Breitinger [95]. The investigation of these systems will require significant effort on behalf of the investigator in terms of understanding the models, their training data, and the state of the model's inputs when the decision was made. This will also require a reasonable level of explainable AI. Of course, the investigation of digital forensic AI systems themselves will be far from exempt from this scrutiny.

REFERENCES

[1] Razan Abdulhammed, Hassan Musafer, Ali Alessa, Miad Faezipour, and Abdelshakour Abuzneid. 2019. Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection. 8 (March 2019).

[2] Mohiuddin Ahmed, Abdun. Naser, and Jiankun Hu. 2016. A Survey of Network Anomaly Detection Techniques. *J. Netw. Comput. Appl.* 60, C (Jan. 2016), 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

[3] Dogukan Aksu, Serpil Ustebay, Muhammed Ali Aydin, and Tülin Atmaca. 2018. *Intrusion Detection with Comparative Analysis of Supervised Learning Techniques and Fisher Score Feature Selection Algorithm.* 141–149. https://doi.org/10.1007/978-3-030-00840-6_16

[4] Mohammad Alauthman, Nauman Aslam, Mouhammd Al-kasassbeh, Suleman Khan, Ahmad Al-Qerem, and Kim-Kwang [Raymond Choo]. 2020. An efficient reinforcement learning-based Botnet detection approach. *Journal of Network and Computer Applications* 150 (2020), 102479. https://doi.org/10.1016/j.jnca.2019.102479

[5] Saed Alrabaee, Paria Shirani, Mourad Debbabi, and Lingyu Wang. 2016. On the feasibility of malware authorship attribution. In *International Symposium on Foundations and Practice of Security*. Springer, 256–272.

[6] Amar Amouri, Vishwa Alaparthy, and Salvatore Dominic Morgera. 2018. Cross layer-based intrusion detection based on network behavior for IoT. In *2018 IEEE 19th Wireless and Microwave Technology Conference (WAMICON)*. 1–4. https://doi.org/10.1109/WAMICON.2018.8363921

[7] Felix Anda, Nhien-An Le-Khac, and Mark Scanlon. 2020. DeepUAge: Improving Underage Age Estimation Accuracy to Aid CSEM Investigation. *Forensic Science International: Digital Investigation* 32 (04 2020), 300921. https://doi.org/10.1016/j.fsidi.2020.300921

[8] Rigan Ap-Apid. 2005. An algorithm for nudity detection. In *5th Philippine Computing Science Congress*. 201–205.

[9] Belhassen Bayar and Matthew C Stamm. 2016. A deep learning approach to universal image manipulation detection using a new convolutional layer. In *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*. ACM, 5–10.

[10] Andrei Cristian Bechet, Robert Helbet, Iulian Bouleanu, Annamaria Sarbu, Simona Miclaus, and Paul Bechet. 2019. Low Cost Solution Based on Software Defined Radio for the RF Exposure Assessment: A Performance Analysis. In *2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*. IEEE, 1–4.

[11] Ryad Benadjila, Emmanuel Prouff, Rémi Strullu, Eleonora Cagli, and Cécile Dumas. 2018. Study of deep learning techniques for side-channel analysis and introduction to ASCAD database. *ANSSI, France & CEA, LETI, MINATEC Campus, France. Online verfügbar unter https://eprint. iacr. org/2018/053. pdf, zuletzt geprüft am* 22 (2018), 2018.

[12] Anuradha D. Biradar and B. Padmavathi. 2020. BotHook: A Supervised Machine Learning Approach for Botnet Detection Using DNS Query Data. In *ICCCE 2019*, Amit Kumar and Stefan Mozar (Eds.). Springer Singapore, Singapore, 261–269.

[13] Ross Brown, Binh Pham, and Olivier Vel. 2005. Design of a Digital Forensics Image Mining System. *Lecture Notes in Computer Science.* https://doi.org/10.1007/11553939_57

[14] Anna L. Buczak and Erhan Guven. 2016. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys and Tutorials* 18, 2 (2016), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

[15] Robert Callan, Farnaz Behrang, Alenka Zajic, Milos Prvulovic, and Alessandro Orso. 2016. Zero-overhead profiling via em emanations. In *Proceedings of the 25th International Symposium on Software Testing and Analysis*. ACM, 401–412.

[16] Robert Locke Callan. 2016. *Analyzing software using unintentional electromagnetic emanations from computing devices.* Ph.D. Dissertation. Georgia Institute of Technology.

[17] Brian Carrier and Eugene H Spafford. 2004. An event-based digital forensic investigation framework. In *Digital forensic research workshop*. 11–13.

[18] Eoghan Casey, Sean Barnum, Ryan Griffith, Jonathan Snyder, Harm van Beek, and Alex Nelson. 2018. The evolution of expressing and exchanging cyber-investigation information in a standardized form. In *Handling and Exchanging Electronic Evidence Across Europe*. Springer, 43–58.

[19] Yoan Chabot, Aurélie Bertaux, Christophe Nicolle, and M-Tahar Kechadi. 2014. A complete formalized knowledge representation model for advanced digital forensics timeline analysis. *Digital Investigation* 11 (2014), S95–S105.

[20] Yoan Chabot, Aurélie Bertaux, Christophe Nicolle, and Tahar Kechadi. 2015. An ontology-based approach for the reconstruction and analysis of digital incidents timelines. *Digital Investigation* 15 (2015), 83–100.

[21] Suresh Chari, Josyula R Rao, and Pankaj Rohatgi. 2002. Template attacks. In *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 13–28.

[22] Jiansheng Chen, Xiangui Kang, Ye Liu, and Z Jane Wang. 2015. Median filtering forensics based on convolutional neural networks. *IEEE Signal Processing Letters* 22, 11 (2015), 1849–1853.

[23] Qian Chen, Qing Liao, Zoe L Jiang, Junbin Fang, Siuming Yiu, Guikai Xi, Rong Li, Zhengzhong Yi, Xuan Wang, Lucas CK Hui, et al. 2018. File fragment classification using grayscale image conversion and deep learning in digital forensics. In *2018 IEEE Security and Privacy Workshops (SPW)*. IEEE, 140–147.

[24] Lianbing Deng, Daming Li, Xiang Yao, David Cox, and Haixiang Wang. 2018. Mobile network intrusion detection for IoT system based on transfer learning algorithm. *Cluster Computing* (31 Jan 2018). https://doi.org/10.1007/s10586-018-1847-2

[25] Xiaoyu Du, Christopher Hargreaves, John Sheppard, and Mark Scanlon. 2020. TraceGen: User Activity Emulation for Digital Forensic Test Image Generation. *Forensic Science International: Digital Investigation* (09 2020). Proceedings of DFRWS APAC 2020.

[26] Xiaoyu Du, Nhien-An Le-Khac, and Mark Scanlon. 2017. Evaluation of Digital Forensic Process Models with Respect to Digital Forensics as a Service. In *Proceedings of the 16th European Conference on Cyber Warfare and Security*

(ECCWS 2017). ACPI, Dublin, Ireland, 573–581.

[27] Peter Eckersley. 2010. How unique is your web browser?. In *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 1–18.

[28] Mohamed Faisal Elrawy, Ali Ismail Awad, and Hesham F. A. Hamed. 2018. Intrusion detection systems for IoT-based smart environments: a survey. *Journal of Cloud Computing* 7, 1 (04 Dec 2018), 21. https://doi.org/10.1186/s13677-018-0123-6

[29] EURPOL. 2019. *Global Guidelines for Digital Forensic Laboratories*. https://www.interpol.int/content/download/13501/file/INTERPOL_DFL_GlobalGuidelinesDigitalForensicsLaboratory.pdf

[30] Hany Farid. 2009. Image forgery detection. *IEEE Signal processing magazine* 26, 2 (2009), 16–25.

[31] Nhien-An Le-Khac Felix Anda, David Lillis and Mark Scanlon. 2018. Evaluating Automated Facial Age Estimation Techniques for Digital Forensics. In *2018 IEEE Security and Privacy Workshops (SPW)*. 129–139.

[32] Simran Fitzgerald, George Mathews, Colin Morris, and Oles Zhulyn. 2012. Using NLP techniques for file fragment classification. *Digital Investigation* 9 (2012), S44–S49.

[33] Peter Flach. 2012. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press.

[34] David Freire-Obregon, Fabio Narducci, Silvio Barra, and Modesto Castrillon-Santana. 2018. Deep learning for source camera identification on mobile devices. *Pattern Recognition Letters* (2018). https://doi.org/10.1016/j.patrec.2018.01.005

[35] Takeshi Fujino, Takaya Kubota, and Mitsuru Shiozaki. 2017. Tamper-resistant cryptographic hardware. *IEICE Electronics Express* 14, 2 (2017), 20162004–20162004.

[36] Ekta Gandotra, Divya Bansal, and Sanjeev Sofat. 2014. Malware analysis and classification: A survey. *Journal of Information Security* 2014 (2014).

[37] Simson Garfinkel, Paul Farrell, Vassil Roussev, and George Dinolt. 2009. Bringing science to digital forensics with standardized forensic corpora. *digital investigation* 6 (2009), S2–S11.

[38] Simson L Garfinkel. 2007. Carving contiguous and fragmented files with fast object validation. *digital investigation* 4 (2007), 2–12.

[39] Simson L Garfinkel, Aleatha Parker-Wood, Daniel Huynh, and James Migletz. 2010. An automated solution to the multiuser carved data ascription problem. *IEEE Transactions on Information Forensics and Security* 5, 4 (2010), 868–882.

[40] Pavel Gladyshev and Ahmed Patel. 2004. Finite state machine approach to digital event reconstruction. *Digital Investigation* 1, 2 (2004), 130–149.

[41] Cinthya Grajeda, Frank Breitinger, and Ibrahim Baggili. 2017. Availability of datasets for digital forensics–and what is missing. *Digital Investigation* 22 (2017), S94–S105.

[42] Michał Grega, Andrzej Matiolański, Piotr Guzik, and Mikołaj Leszczuk. 2016. Automated detection of firearms and knives in a CCTV image. *Sensors* 16, 1 (2016), 47.

[43] Christopher Hargreaves and Angus Marshall. 2019. SyncTriage: Using synchronisation artefacts to optimise acquisition order. *Digital Investigation* 28 (2019), S134–S140.

[44] Christopher Hargreaves and Jonathan Patterson. 2012. An automated timeline reconstruction approach for digital forensic investigations. *Digital Investigation* 9 (2012), S69–S79.

[45] Gabriel Hospodar, Benedikt Gierlichs, Elke De Mulder, Ingrid Verbauwhede, and Joos Vandewalle. 2011. Machine learning in side-channel analysis: a first study. *Journal of Cryptographic Engineering* 1, 4 (2011), 293.

[46] Wenyi Huang and Jack W Stokes. 2016. MtNet: a multi-task neural network for dynamic malware classification. In *International conference on detection of intrusions and malware, and vulnerability assessment*. Springer, 399–418.

[47] J. Jasmine and S. Annadurai. 2019. Real time video image enhancement approach using particle swarm optimisation technique with adaptive cumulative distribution function based histogram equalisation. *Measurement* 145 (2019), 833 – 840. https://doi.org/10.1016/j.measurement.2018.12.105

[48] Sundararaman Jeyaraman and Mikhail J Atallah. 2006. An empirical study of automatic event reconstruction systems. *digital investigation* 3 (2006), 108–115.

[49] Vaibhavi Kalgutkar, Ratinder Kaur, Hugo Gonzalez, Natalia Stakhanova, and Alina Matyukhina. 2019. Code authorship attribution: Methods and challenges. *ACM Computing Surveys (CSUR)* 52, 1 (2019), 1–36.

[50] Muhammad Naeem Ahmed Khan. 2012. Performance analysis of Bayesian networks and neural networks in classification of file system activities. *Computers & Security* 31, 4 (2012), 391–401.

[51] Mehdi Kharrazi, Husrev T Sencar, and Nasir Memon. 2004. Blind source camera identification. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, Vol. 1. IEEE, 709–712.

[52] Taesung Kim, Seungkwang Lee, Dooho Choi, and Hyunsoo Yoon. 2016. Protecting secret keys in networked devices with table encoding against power analysis attacks. *Journal of High Speed Networks* 22, 4 (2016), 293–307.

[53] Paul Kocher, Joshua Jaffe, and Benjamin Jun. 1999. Differential power analysis. In *Advances in Cryptology (CRYPTO '99)*. Springer, 789–789.

[54] Paul Kocher, Joshua Jaffe, Benjamin Jun, and Pankaj Rohatgi. 2011. Introduction to differential power analysis. *Journal of Cryptographic Engineering* 1, 1 (2011), 5–27.

[55] Michael Donovan Kohn, Mariki M Eloff, and Jan HP Eloff. 2013. Integrated digital forensic process model. *Computers & Security* 38 (2013), 103–115.

[56] Bojan Kolosnjaji, Apostolis Zarras, George Webster, and Claudia Eckert. 2016. Deep learning for classification of malware system call sequences. In *Australasian Joint Conference on Artificial Intelligence*. Springer, 137–149.

[57] Gierad Laput, Chouchang Yang, Robert Xiao, Alanson Sample, and Chris Harrison. 2015. Em-sense: Touch recognition of uninstrumented, electrical and electromechanical objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 157–166.

[58] Arash Habibi Lashkari, Gerard Draper-Gil, Mohammad Saiful Islam Mamun, and Ali A. Ghorbani. 2017. Characterization of Tor Traffic Using Time Based Features. In *In the proceeding of the 3rd International Conference on Information System Security and Privacy, SCITEPRESS* (Portugal).

[59] Quan Le, Oisín Boydell, Brian Mac Namee, and Mark Scanlon. 2018. Deep learning at the shallow end: Malware classification for non-domain experts. *Digital Investigation* 26 (2018), S118–S126.

[60] Thanh-Ha Le, Jessy Clédière, Christine Serviere, and Jean-Louis Lacoume. 2007. Efficient solution for misalignment of signal in side channel analysis. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, Vol. 2. IEEE, II–257.

[61] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[62] Liran Lerman, Gianluca Bontempi, and Olivier Markowitch. 2011. Side channel attack: an approach based on machine learning. In *Proceedings of 2nd International Workshop on Constructive Side-Channel Analysis and Security Design (COSADE)*. Schindler and Huss, 29–41.

[63] Stefan Lessmann, Bart Baesens, Hsin-Vonn Seow, and Lyn C Thomas. 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research* 247, 1 (2015), 124–136.

[64] David Lillis, Brett Becker, Tadhg O'Sullivan, and Mark Scanlon. 2016. Current Challenges and Future Research Areas for Digital Forensic Investigation. In *The 11th ADFSL Conference on Digital Forensics, Security and Law (CDFSL 2016)*. ADFSL, Daytona Beach, FL, USA, 9–20.

[65] Liqun Liu, Bing Xu, Xiaoping Zhang, and Xianjun Wu. 2018. An intrusion detection method for internet of things based on suppressed fuzzy clustering. *EURASIP Journal on Wireless Communications and Networking* (2018).

[66] Jan Lukas, Jessica Fridrich, and Miroslav Goljan. 2006. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security* 1, 2 (2006), 205–214.

[67] Alison D MacEachern, Divya Jindal-Snape, and Sharon Jackson. 2011. Child abuse investigation: police officers and secondary traumatic stress. *International journal of occupational safety and ergonomics* 17, 4 (2011), 329–339.

[68] José Raúl Machado-Fernández. 2015. Software defined radio: Basic principles and applications. *Revista Facultad de Ingeniería* 24, 38 (2015), 79–96.

[69] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083* (2017).

[70] Houssem Maghrebi, Thibault Portigliatti, and Emmanuel Prouff. 2016. Breaking cryptographic implementations using deep learning techniques. In *International Conference on Security, Privacy, and Applied Cryptography Engineering*. Springer, 3–26.

[71] Andrew Marrington, Ibrahim Baggili, George Mohay, and Andrew Clark. 2011. CAT Detect (Computer Activity Timeline Detection): A tool for detecting inconsistency in computer activity timelines. *digital investigation* 8 (2011), S52–S61.

[72] Fabio Marturana, Gianluigi Me, Rosamaria Berte, and Simone Tacconi. 2011. A quantitative approach to triaging in mobile forensics. In *2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 582–588.

[73] Fabio Marturana and Simone Tacconi. 2013. A Machine Learning-based Triage methodology for automated categorization of digital media. *Digital Investigation* 10, 2 (2013), 193–204.

[74] Loïc Masure, Cécile Dumas, and Emmanuel Prouff. 2020. A comprehensive study of deep learning for side-channel analysis. *IACR Transactions on Cryptographic Hardware and Embedded Systems* (2020), 348–375.

[75] Christian A Meissner and Saul M Kassin. 2002. "He's guilty!": Investigator bias in judgments of truth and deception. *Law and human behavior* 26, 5 (2002), 469–480.

[76] Chris R. Chatwin Muhammad Naeem Khan and Rupert CD Young. 2007. A framework for post-event timeline reconstruction using neural networks. *digital investigation* 4, 3-4 (2007), 146–157.

[77] Hiran V Nath and Babu M Mehtre. 2014. Static malware analysis using machine learning methods. In *International Conference on Security in Computer Networks and Distributed Systems*. Springer, 440–450.

[78] Mohammed Murtaz Amir Naviq, Hassan Azwar, Syed Baqir Ali, and Saad Rehman. 2018. A framework for Android Malware detection and classification. In *2018 IEEE 5th International Conference on Engineering Technologies and*

Applied Sciences (ICETAS). 1–5.

[79] Alireza Nazari, Nader Sehatbakhsh, Monjur Alam, Alenka Zajic, and Milos Prvulovic. 2017. EDDIE: EM-Based Detection of Deviations in Program Execution. In *Proceedings of the 44th Annual International Symposium on Computer Architecture*. ACM, 333–346.

[80] Fudong Nian, Teng Li, Yan Wang, Mingliang Xu, and Jun Wu. 2016. Pornographic image detection utilizing deep convolutional neural networks. *Neurocomputing* 210 (2016), 283 – 293. https://doi.org/10.1016/j.neucom.2015.09.135 SI:Behavior Analysis In SN.

[81] Jiankun Hu Nour Moustafa and Jill Slay. 2019. A holistic review of Network Anomaly Detection Systems: A comprehensive survey. *J. Network and Computer Applications* 128 (2019), 33–55.

[82] Seong Joon Oh, Bernt Schiele, and Mario Fritz. 2019. Towards reverse-engineering black-box neural networks. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer, 121–144.

[83] Morteza Safaei Pour, Elias Bou-Harb, Kavita Varma, Nataliia Neshenko, Dimitris A. Pados, and Kim-Kwang Raymond Choo. 2019. Comprehending the IoT cyber threat landscape: A data dimensionality reduction technique to infer and characterize Internet-scale IoT probing campaigns. *Digital Investigation* 28 (2019), S40 – S49. https://doi.org/10.1016/j.diin.2019.01.014

[84] Yuan Rao and Jiangqun Ni. 2016. A deep learning approach to detection of splicing and copy-move forgeries in images. In *Information Forensics and Security (WIFS), 2016 IEEE International Workshop on*. IEEE, 1–6.

[85] Chiadighikaobi Ikenna Rene and Johari Abdullah. 2017. Malicious Code Intrusion Detection using Machine Learning And Indicators of Compromise. *International Journal of Computer Science and Information Security (IJCSIS)* 15, 9 (September 2017).

[86] Ronald L Rivest. 1991. Cryptography and machine learning. In *International Conference on the Theory and Application of Cryptology*. Springer, 427–439.

[87] Anderson Rocha, Walter J Scheirer, Christopher W Forstall, Thiago Cavalcante, Antonio Theophilo, Bingyu Shen, Ariadne RB Carvalho, and Efstathios Stamatatos. 2016. Authorship attribution for social media forensics. *IEEE Transactions on Information Forensics and Security* 12, 1 (2016), 5–33.

[88] Marcus K Rogers, James Goldman, Rick Mislan, Timothy Wedge, and Steve Debrota. 2006. Computer forensics field triage process model. *Journal of Digital Forensics, Security and Law* 1, 2 (2006), 2.

[89] Sebastian Ruder. 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* (2016).

[90] Laura Sanchez, Cinthya Grajeda, Ibrahim Baggili, and Cory Hall. 2019. A Practitioner Survey Exploring the Value of Forensic Tools, AI, Filtering, & Safer Presentation for Investigating Child Sexual Abuse Material (CSAM). *Digital Investigation* 29 (2019), S124–S142.

[91] Hendra Saputra, Narayanan Vijaykrishnan, M Kandemir, Mary Jane Irwin, R Brooks, Soontae Kim, and Wei Zhang. 2003. Masking the energy behavior of DES encryption. In *Proceedings of the conference on Design, Automation and Test in Europe-Volume 1*. IEEE Computer Society, 10084.

[92] Fadl Mutaher Ba-Alwi Saud Mohammed Othman, Nabeel T Alsohybe and Ammar Thabit Zahary. 2018. Survey on Intrusion Detection System. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)* (December 2018).

[93] Asanka Sayakkara, Nhien-An Le-Khac, and Mark Scanlon. 2019. A survey of electromagnetic side-channel attacks and discussion on their case-progressing potential for digital forensics. *Digital Investigation* (2019).

[94] Mark Scanlon. 2016. Battling the digital forensic backlog through data deduplication. In *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*. IEEE, 10–14.

[95] Johannes Schneider and Frank Breitinger. 2020. AI Forensics: Did the Artificial Intelligence System Do It? Why? (2020).

[96] Husrev T Sencar and Nasir Memon. 2009. Overview of state-of-the-art in digital image forensics. In *Algorithms, Architectures and Information Systems Security*. World Scientific, 325–347.

[97] Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.

[98] Devashish Shankar, Sujay Narumanchi, HA Ananya, Pramod Kompalli, and Krishnendu Chaudhury. 2017. Deep learning based large scale visual recommendation and search for e-commerce. *arXiv preprint arXiv:1703.02344* (2017).

[99] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani. 2018. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In *4th International Conference on Information Systems Security and Privacy (ICISSP)* (Portugal).

[100] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani. 2019. Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy. In *IEEE 53rd International Carnahan Conference on Security Technology* (India).

[101] Manmeet Singh, Maninder Singh, and Sanmeet Kaur. 2019. Detecting bot-infected machines using DNS fingerprinting. *Digital Investigation* 28 (2019), 14 – 33. https://doi.org/10.1016/j.diin.2018.12.005

[102] Raphael Spreitzer, Veelasha Moonsamy, Thomas Korak, and Stefan Mangard. 2018. Systematic classification of side-channel attacks: a case study for mobile devices. *IEEE Communications Surveys & Tutorials* 20, 1 (2018), 465–488.

[103] Barron Stone and Samuel Stone. 2016. Comparison of Radio Frequency Based Techniques for Device Discrimination and Operation Identification. In *11th International Conference on Cyber Warfare and Security: ICCWS2016*. Academic Conferences and Publishing Limited, 475.

[104] Hudan Studiawan, Ferdous Sohel, and Christian Payne. 2020. Sentiment Analysis in a Forensic Timeline with Deep Learning. *IEEE Access* (2020).

[105] Abby Stylianou, Jessica Schreier, Richard Souvenir, and Robert Pless. 2017. Traffickcam: Crowdsourced and computer vision based approaches to fighting sex trafficking. In *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 1–8.

[106] Laya Taheri, Andi Fitriah Abdul Kadir, and Arash Habibi Lashkari. 2019. Extensible Android Malware Detection and Family Classification Using Network-Flows and API-Calls. In *2019 International Carnahan Conference on Security Technology (ICCST)*. 1–8.

[107] Qizhi Tian and Sorin A Huss. 2012. On clock frequency effects in side channel attacks of symmetric block ciphers. In *2012 5th International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 1–5.

[108] Min Jen Tsai, Cheng Liang Lai, and Jung Liu. 2007. Camera/mobile phone source identification for digital forensics. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* 2 (2007), 221–224. https://doi.org/10.1109/ICASSP.2007.366212

[109] Benjamin Turnbull and Suneel Randhawa. 2015. Automated event and social network extraction from digital evidence sources with ontological mapping. *Digital Investigation* 13 (2015), 94–106.

[110] Serpil Ustebay, Zeynep Turgutand, and Muhammed Ali Aydin. 2018. Intrusion Detection System with Recursive Feature Elimination by Using Random Forest and Deep Learning Classifier. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*. 71–76. https://doi.org/10.1109/IBIGDELFT.2018.8625318

[111] Michael Veale, Reuben Binns, and Lilian Edwards. 2018. Algorithms that remember: model inversion attacks and data protection law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2133 (2018), 20180083.

[112] Vladimír Veselý and Martin Žádník. 2019. How to detect cryptocurrency miners? By traffic forensics! *Digital Investigation* 31, 31 (2019), 1–25. https://doi.org/10.1016/j.diin.2019.08.002

[113] R. Vinayakumar, Mamoun Alazab, K. P. Soman, Prabaharan Poornachandran, Ameer Al-Nemrat, and Sitalakshmi Venkatraman. 2019. Deep Learning Approach for Intelligent Intrusion Detection System. *IEEE Access* 7 (2019), 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334

[114] Eva A Vincze. 2016. Challenges in digital forensics. *Police Practice and Research* 17, 2 (2016), 183–194.

[115] Kristijan Vulinović, Lucija Ivković, Juraj Petrović, Kristian Skračić, and Predrag Pale. 2019. Neural Networks for File Fragment Classification. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 1194–1198.

[116] Xinxi Wang and Ye Wang. 2014. Improving content-based and hybrid music recommendation using deep learning. In *Proceedings of the 22nd ACM international conference on Multimedia*. 627–636.

[117] Xiao Wang, Quan Zhou, Jacob Harer, Gavin Brown, Shangran Qiu, Zhi Dou, John Wang, Alan Hinton, Carlos Aguayo Gonzalez, and Peter Chin. 2018. Deep learning-based classification and anomaly detection of side-channel signals. In *Cyber Sensing 2018*, Vol. 10630. International Society for Optics and Photonics, 1063006.

[118] Janis Wolak and Kimberly J Mitchell. 2009. Work exposure to child pornography in ICAC task forces and affiliates. *Retrieved from Crimes against Children Research Center: http://www. unh. edu/ccrc/pdf/Law% 20Enforcement% 20Work% 20Exposure% 20to% 20CP. pdf* (2009).

[119] Jianyu Xiao, Shancang Li, and Qingliang Xu. 2019. Video-based evidence analysis and extraction in digital forensic investigation. *IEEE Access* 7 (2019), 55432–55442.

[120] Xitong Yang and Jiebo Luo. 2017. Tracking Illicit Drug Dealing and Abuse on Instagram Using Multimodal Analysis. *ACM Trans. Intell. Syst. Technol.* 8, 4, Article 58 (Feb. 2017), 15 pages. https://doi.org/10.1145/3011871

[121] Dawei Yin, Yuening Hu, Jiliang Tang, Tim Daly, Mianwei Zhou, Hua Ouyang, Jianhui Chen, Changsung Kang, Hongbo Deng, Chikashi Nobata, et al. 2016. Ranking relevance in yahoo search. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 323–332.

[122] Andreas Zankl, Hermann Seuschek, Gorka Irazoqui, and Berk Gulmezoglu. 2018. Side-Channel Attacks in the Internet of Things: Threats and Challenges. In *Solutions for Cyber-Physical Systems Ubiquity*. IGI Global, 325–357.

[123] Ying Zhang, Jonathan Goh, Lei Lei Win, and Vrizlynn LL Thing. 2016. Image Region Forgery Detection: A Deep Learning Approach. *SG-CRC* 2016 (2016), 1–11.

[124] Yuanyuan Zhou and François-Xavier Standaert. 2019. Deep learning mitigates but does not annihilate the need of aligned traces and a generalized resnet model for side-channel attacks. *Journal of Cryptographic Engineering* (2019), 1–11.