

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Advancing Automation in Digital Forensic Investigations Using Machine Learning Forensics

Salman Iqbal and Soltan Abed Alharbi

Abstract

In the last few years, most of the data such as books, videos, pictures, medical and even the genetic information of humans are moving toward digital formats. Laptops, tablets, smartphones and wearable devices are the major source of this digital data transformation and are becoming the core part of our daily life. As a result of this transformation, we are becoming the soft target of various types of cybercrimes. Digital forensic investigation provides the way to recover lost or purposefully deleted or hidden files from a suspect's device. However, current manpower and government resources are not enough to investigate the cybercrimes. Unfortunately, existing digital investigation procedures and practices require huge interaction with humans; as a result it slows down the process with the pace digital crimes are committed. Machine learning (ML) is the branch of science that has grown from the field of AI. This advanced technology uses the explicit programming to depict the human-like behaviour. Machine learning combined with automation in digital investigation process at different stages of investigation has significant potential to aid digital investigators. This chapter aims at providing the research in machine learning-based digital forensic investigation, identifies the gaps, addresses the challenges and open issues in this field.

Keywords: digital forensic investigation, machine learning, evidence extraction, cybercrimes, automated data extraction

1. Introduction

Worldwide usage of mobile smart devices has increased dramatically over the past two decades and is becoming the part of our daily life. The term smart device ranges from variety of devices that includes mobile phones, smartphones, tablets, GPS and so on. The popularity of these smart devices is increased significantly due to their processing power, huge storage capabilities and less cost. Consequently, they can hold the enormous amount of commercial and private user's data. These devices are the essential part of our daily life because they contain private and essential information of users. However, these devices are also vulnerable to attackers and are often becoming the major part of criminal's activities, IP theft, intrusions, security threats, accidents reconstructions and many more. The number of digital crimes equally increases as the new technologies, i.e. digital devices and

internet, increases. As a result, we are becoming the soft target for various types of cybercrimes and digital attacks.

The Digital Forensic Research Workshop (DFRWS) has defined digital forensics (DF) as “The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations”.

Today, DF demands are increasingly important. DF investigation procedures help to capture important information from the compromised device. Nowadays, businesses deeply depend on the digital devices and on the Internet. Capturing the indispensable evidences from these devices is equally important. Digital evidence should be gathered from the system to support or deny some reasoning an investigator may have about the incident.

It is important to know that how to recover digital evidences which can be interested for investigators. However, current human power and other available resources are not enough to fully investigate the digital crimes on digital devices. Further, existing digital investigation procedures and practices require huge interaction with humans; as a result it slows down the process with the pace digital crimes are committed.

In this chapter, we have thoroughly discussed the current advancement of machine learning forensics (MLF) in digital forensic investigation (DFI). We present the latest surveys in this field and give critique comparisons of these approaches.

1.1 Historical perspective of digital forensic investigations

Digital forensic or computer forensic is first presented by 1970 [1]. In the first investigation, the financial fraud is proven from the suspect’s computer. The first prosecuted computer crime was reported in 1996. The computer crime is defined as when the computer is the major effect for offense and facilitates the tool to

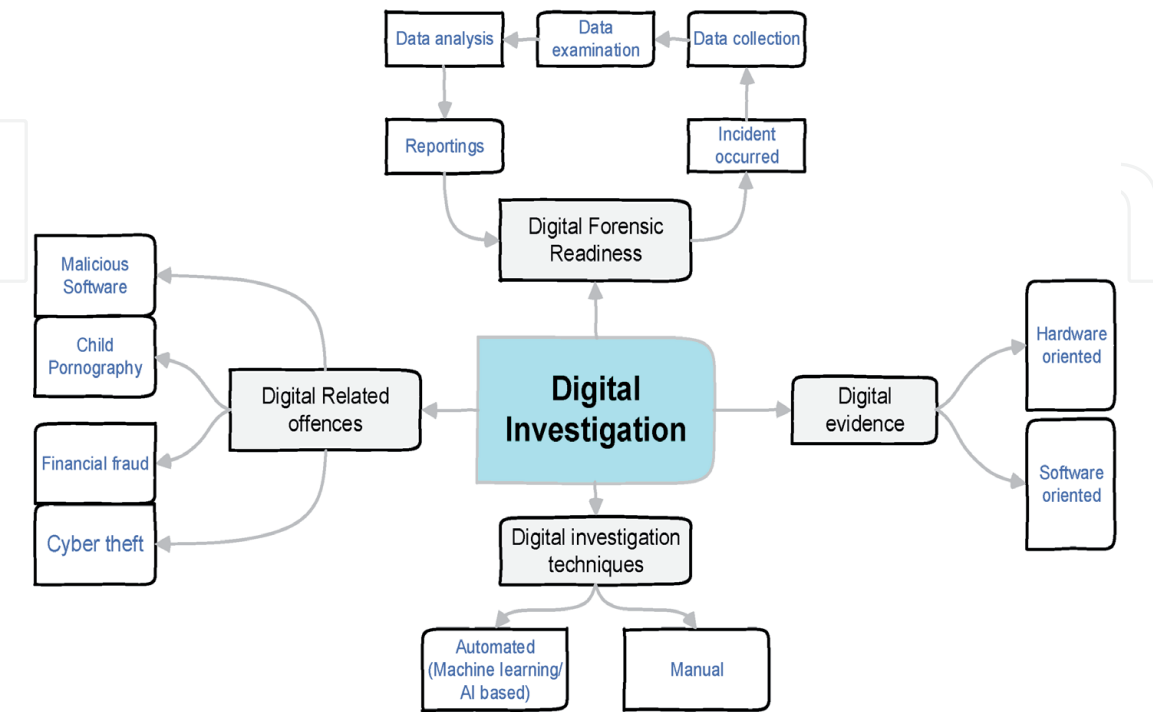


Figure 1.
Taxonomy of digital investigations.

commission a crime [2]. The first prosecuted computer crime was reported in Texas, USA, in 1996 [3] and resulted in a 5-year sentence. In 1990, computer-based digital crimes started to grow with the increasing popularity of the computers and the Internet. The computer forensic is developed as the independent field in the late 1990s and in the early 2000s. The CSI surveys report that almost 46% among the respondents were affected by some kind of computer crimes [4]. The 2010 Gallup surveys reports that 11% of the American adult become victim of computer- or Internet-related crimes in their homes. This ratio is 6–8% more than the last 7 years. A survey conducted by “Australian Company Crime Survey” [5], estimated that A\$ 2,000,000 financial fraud and information breaches occurs in 2006. Company Crime Survey, its estimated A\$ 2,000,000 financial fraud and information breaches in lost revenue. The term digital forensic is used nowadays with the advent of new digital devices with increasing number of frequency of use for investigation purposes (**Figure 1**).

2. Artificial intelligence (AI), machine learning (ML) and deep learning

It's important to examine how actually AI, ML and deep learning (DL) methods can help in solving the problems of DF and how these methods differentiate with each other's.

a. Artificial intelligence

AI is the science of making things smart or the capability of the machines (e.g. visual recognition, NLP, etc.) to perform human tasks. The important point is that AI is not machine learning or smart things. AI can be viewed as the things that can carry the human tasks and make these tasks easy. The AI technology is increasing day by day, and its enormous use also significantly increases the number of malicious activities.

Artificial intelligence programs are called intelligent agent. Intelligent agents are used to interact with the environment. The agent uses the technique to identify the environments through its sensors, and then it can take the action to affect the state through its sensors.

The important aspects in the AI technologies are how the sensors are used to collect the data and how it maps to the actuators; this is how the functions within the agents can perform these consequences. The ultimate goal of the AI is to develop the machine that acts just like humans. This task can be accomplished by only using the learning algorithms to which it is aimed to try to make a sketch of the human brain learnings. AI technologies give very good advantages and have a bright future ahead. However, these technologies are also unavoidably used for execution of some serious crimes that can be dangerous for people.

b. Machine learning

ML is one of the approaches of AI that uses a system that can be learned by itself from experience. It is not used for only AI purposes such as copying human behaviour but also needs to reduce the human efforts and time spent to perform the difficult and even the simple tasks. ML can be viewed as a system that can learn from experience and examples rather than from programming. Thus, if the system learns constantly and makes a decision based on the data rather than programming, then it's called ML. ML is developed as a new technology to provide new functionalities for computers and is used for industry and science. There are many autonomous solutions based on ML for medical science, robotics, engineering and so on.

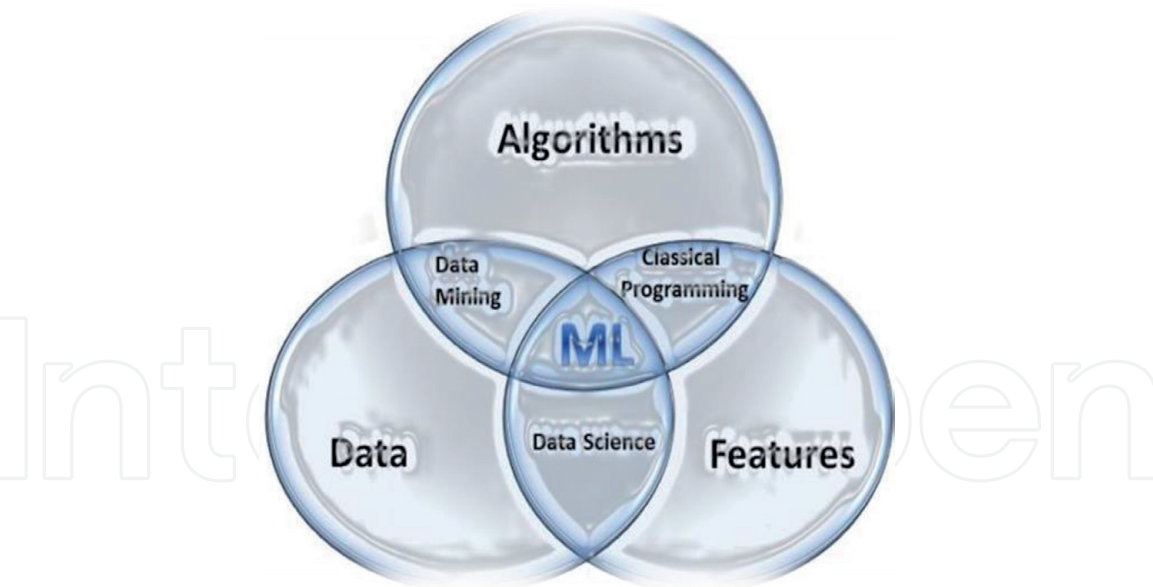


Figure 2.
Machine learning essentials.



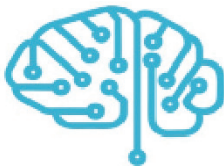
Artificial intelligence	Machine learning	Deep learning
		
Ability of a machine to imitate intelligent human behaviour	Application of AI that allows a system to automatically learn and improve from experience	Application of ML that uses complex algorithms and deep neural to train a model
Originated around the 1950s	Originated around the 1960s	Originated around the 1970s
Represents simulated intelligence in machines	Getting machines to make without being programmed	Process of using artificial neural networks to solve complex problems
Subsets of data science	Subset of AI and data science	Subset of ML, AI and data science
Building machines that are capable of thinking like humans	Make machines that can learn through previous experience to solve problems	To build neural networks that automatically discovers patterns for feature detection

Table 1.
Difference between artificial intelligence, machine learning and deep learning.

c. Deep learning

Deep learning combines the set of techniques used to implement ML methods to recognize patterns of patterns such as image recognition. First of all the system is used to identify the object edges, structure of the object, object type and then the object itself (Figure 2) (Table 1).

3. Approaches to machine learning forensics

Usually two main approaches are used to define the ML forensics, that is, inductive reasoning and deductive reasoning:

a. Inductive learning

Inductive reasoning is obtained from the general knowledge of specific information. The obtained knowledge is new and not truth preserving. That means the knowledge obtained can be invalidated from new information. There is no well-founded theory. In this area there are a large number of goals such as it is important to discover general concepts from a limited set of examples. The examples are called experience. The basis of this is to search for similar characteristics among examples. The methods used in these are based on the inductive learning.

b. Deductive learning

Deductive reasoning obtains the knowledge from well-established methods called logic. Deductive reasoning obtains from the knowledge by using well-established methods. The knowledge is not new. But it is implicit in the initial knowledge. New knowledge cannot invalidate the existing knowledge obtained and its basis on the mathematical logic.

3.1 Supervised, unsupervised and reinforcement ML

Supervised and unsupervised are the most commonly used techniques in ML algorithms.

a. Supervised and unsupervised

On the other hand, the reinforcement learning is complex and difficult to implement. Supervised learning is the most common type of ML paradigm. This type is easy to understand and implement. The data in this type is in the form of examples with labels. The data can be called as training data. The learning algorithms can be feed to these example-label pairs one by one. This allows the algorithms to predict the label for each example. Further, it provides the feedback whether this gives the right answer or not. In this type the model is first trained by using lots of training data (input and targets). This process is really fast and accurate. With the passage of time, the algorithms are able to learn in order to approximate the concrete nature of the relationship between examples and their labels. The trained supervised learning can see the totally new and never seen before data and predict the good label for it. Supervised learning is the most widely used and easiest to implement. Supervised learning is the most popular technique used for machine learning.

The unsupervised learning does not have a well-structured format. There are no targets for the training data. Therefore, the system does not know where to go. The system needs to understand itself from the given data. The unsupervised learning is the opposite of supervised learning. There are no labels in it. The algorithms are fed up with a lot of data, and the tool is given to understand the properties of the data. In this way, the task of the system is to learn to group, cluster and/or organize the data in the similar way as the human can organize the data. The unsupervised learning is much more interesting in a way that the overwhelming majority of data in this world is unlabelled. This type can make benefit of industries in a way that we have terabytes of unlabelled data, and organizing this data can be beneficial for the industry and potential profits for making it organized without minimal or no human effort (**Table 2**).

	Supervised learning	Unsupervised learning
Definition	Data set labeled with predefined classes	Data set labeled without predefined classes
Method	Data classification	Data clustering
Example	Support vector machine Decision tree	K-means clustering, ant clustering algorithms
Known attack detection	High	Low
Unknown attack detection	Low	High
Unsupervised learning is not easy and is not used as widely as supervised.		

Table 2.
Supervised vs. unsupervised learning.

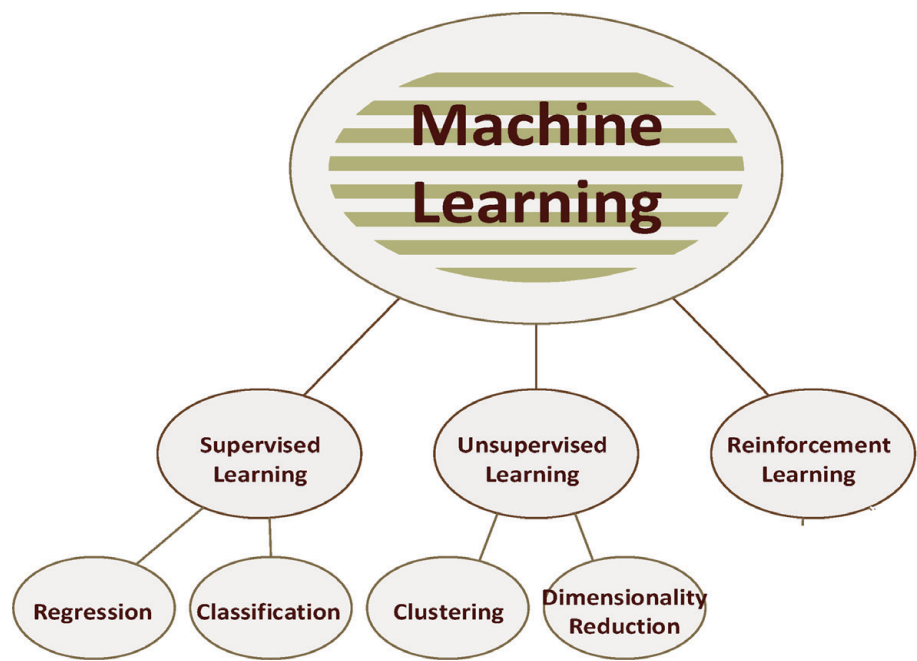


Figure 3.
Machine learning types.

b. Reinforcement learning

The reinforcement learning is totally different from both supervised and unsupervised ML. The relationship among supervised and unsupervised can be related with each other with the presence and absence of labels. However, the reinforcement learning learns from the mistakes. When deploying the reinforcement learning algorithms in any type of environment, it will make a lot of mistakes at the beginning. The signals to the algorithms are provided that can associate the good behaviour with positive signals and bad behaviour with negative label. The algorithms can reinforce algorithms to prefer good behaviour and bad behaviors. With the passage of time, the algorithm can learn to make fewer mistakes as it was initially (Figure 3).

3.2 Machine learning forensics for law enforcement, compliance and intelligence

Standardization is still a big challenge for DFI. The DI experts perform DI on the basis of their experience, the company’s policies and basis on their previous

experience. This is due to the lack of any universal standard for digital evidence collection. The law enforcement is continuously changing in this information technology age. The traditional crimes such as financial and commerce are also gaining the benefits of technology advancements and continuously upgrading with the latest development in the technology.

These days, law enforcement techniques are also changing.

DFI is a very common practice in law enforcements and commerce industry. The way in which the use of information technology is increasing by the government sectors, public and corporate agencies, has also increases the victimology of cyber-attacks through the internet.

4. Literature review

The work of [6] is one of the earliest efforts to make an application for expert systems for digital forensic to automate the analysis process. The expert system is used with decision tree in order to detect network anomalies automatically. The expert system is used to analyze the log files.

The Open Computer Forensic Architecture (OCFA) [7] is a well-organized forensic platform of automating the digital forensic tasks. This toll provides the scalability, modularity and openness in digital forensic process. This framework consists of different modules, and each module works independently on a specific file type in order for content extraction of the file for digital evidence. It creates the searchable index of text and metadata. It is a pluggable module that recursively processes the evidence according to the dispatching entity which decides which module needs to be invoked by seeing information in evidence. However, the OCFA follows the pre-extracted data and is not designed to search and recover files. The examination is done by an IT expert on the extracted data to generate indices for text and metadata.

Another effort is made by [8] of automating the disk forensic process. They name their tool “fiwalk” which is used to automate the processing of the data in order to assist the user for the development of the program which automatically processes disk images. This tool also integrates the command line tool of [9]. However, this toll only works for file system data only without any integration of AI techniques. Expert examiner tasks become easy by using this tool.

The research work of Hoelz et al. develops the MultiAgent Digital Investigation toolKit (MADIK) toolkit [9]. The tool provides the multiagent systems which helps the experts in computer forensic examinations. The authors apply the AI-based methods to the problem of digital forensics applications by assigning the tasks to each agent. Every agent has specialized in different tasks such as hashing, keyword search, Windows registry agent and so on. However this tool is not focused on building the new knowledge during investigations. It is used to learn from the previous investigations for any future investigation purposes. Moreover, this work cannot be used for nonexpert users.

The chapter [10] presents the machine learning-based digital triage model for selective pre-examination and statistical classification of digital data. This data can be deployed both on the crime scene and on digital forensic labs. The work is able to provide the quick actionable intelligence on the crime scene in time-critical systems, reduce the burden on forensic labs and protect suspect privacy when a huge amount of data is needed to be analyzed. As advantages the framework provides the minimum manual work and also produces measurable and reproducible error rate.

Existing methods for digital evidence extraction are not coherent to provide the readiness of process support with standardized integrated implementation system which provides guidance and technical knowledge to nonexpert investigators.

Paper title	Problems addressed	Methods used	Proposed solution	Implementation	Open problems
Building an intelligent assistant for digital forensics [11]	Supports investigations conducted by non-IT expert and expert investigators	Series of experiments comparing it with a human investigator as well as against standard benchmark disk images	Proposed AUDIT, an automated disk investigation toolkit	Systematically examine the disk in its totality based on its physical and logical structures	Seizure of an entire hard disk drive is a complex task
A Machine Learning-based Triage methodology for automated categorization of digital media [12]	Defines a list of crime-related features	Populates an input matrix and processes it with different machine learning mining schemes to come up with a device classification	Crime features extracted from available devices and forensic copies	Classified digital media using Bayes networks or support vector machines	Extract data in its raw form without the nature of the information
A machine learning-based approach to digital triage [13]	Identifies test objects allegedly used for exchanging child pornography material	Mobile handset classification on the basis of the 5MF technique	Multiclass categorization for classifying objects on the basis of owner's usage profile	Data corpus with binary categorization	Most files of forensic interest are not fragmented
Artificial intelligence applied to computer forensics [9]	Assists the computer forensic expert on its examinations	Set of rules and a knowledge base	MultiAgent Digital Investigation toolKit based on the experience of the expert	Six specialized intelligent agents implemented: HashSetAgent FilePathAgent FileSignatureAgent TimelineAgent WindowsRegistryAgent KeywordAgent	The method is very heavyweight to be practical
Automating disk forensic processing with SleuthKit, Xml and Python [8]	Automation to perform disk forensic	XML methods used to describe partitions and files on a hard drive or disk image	Creating special-purpose forensic tools	SleuthKit, XML and the Python programming language	Capturing every aspect of a live system is not feasible

Paper title	Problems addressed	Methods used	Proposed solution	Implementation	Open problems
Automated analysis for digital forensic science: Semantic integrity checking [6]	Automates data collection	Expert system with a decision tree	Predetermined invariant relationships between redundant digital objects to detect semantic incongruities	Collection of C programs and Perl scripts	
Android forensics: Automated data collection and reporting from a mobile device [14]	Broadcasts receiver, content observer and alarm	Forensic collection, local SQLite storage, HTTP transfer and clear local SQLite DB	Collects, stores and transfers forensically valuable Android data to a remote Web server without root privileges	DroidWatch is an automated system prototype composed of an Android application and an enterprise server	Architecture models of Android applications are complex and diverse in nature
An automated approach for digital forensic analysis of heterogeneous big data [15]	Understanding the relationships between artifacts	Metadata to solve the data volume problem, semantic web ontologies to solve the heterogeneous data sources	Automated identification and correlation	Artifacts to reduce the burden placed upon the investigator	Not given any particular implementation details
Data mining methods applied to a digital forensics task for supervised machine learning [16]	Glass identification in the context of multi-class supervised learning	Decision trees, Bayes classifiers, based on rules, artificial neural networks and based on nearest neighbors	Empirical overview of the performance with classifiers from different machine learning approaches	Uses two metrics like accuracy and Cohen's kappa for training and test stages	Abstraction errors can occur when representations of the system are not accurate
Data mining methods applied to a digital forensics task for supervised machine learning	Multi-class classification	Supervised machine learning techniques	Decision trees, Bayes classifiers, based on rules, artificial neural networks and based on nearest neighbour techniques	Nondeterministic algorithms	The algorithms implemented are complex in nature and system needs careful understanding of the extracted data
Android forensics: Automated data collection and reporting from a mobile device [14]	Enterprise monitoring system for Android smartphones	Comprehensive guide of data sets available for collection without elevated privileges	First open-source Android enterprise monitoring prototype	Continuously collect many data sets of interest to incident responders, security auditors, proactive security monitors and forensic investigators	Increasing interoperability among Android devices

Paper title	Problems addressed	Methods used	Proposed solution	Implementation	Open problems
Automated forensic analysis of mobile applications on Android devices [17]	Inter-component string propagation, string operations (e.g. append) and API invocation	Inter-component static analysis on Android APKs	Identifies how the information is stored by parsing SQL commands	Fordroid: builds control flow and data dependency graphs	Inter-component string propagation
Automated inference of past action instances in digital investigations [18]	Detects multiple instances of a user action	Signature-based methods	Integrating time into event reconstruction	Detected using signature-based methods during a postmortem digital forensic analysis	Aligning time stamps from different systems and analyzing complex events with incomplete time information
An automated timeline reconstruction approach for digital forensic investigations [19]	Extracts low-level events to a SQLite backing store	Pattern matching to automatically reconstruct high-level, human understandable events.	Automatically analyzed for patterns	Automatically reconstruct high-level events (e.g. connection of a USB stick) from this set of low-level events	Do not cover all aspects of forensic analysis between events
Automated event and social network extraction from digital evidence sources with ontological mapping [20]	High-level analysis based on low-level digital artifacts	Automatically derived “events” from the base forensic artifacts	Information fusion and homogenization techniques are used to reconstruct social networks	Standardized knowledge representations techniques and automated rule-based systems to encapsulate expert knowledge for forensic data	Validation of extracted ontologies and correctness of the data is a big issue

Table 3.
Comparison of literature surveys.

The lack of the automated intelligent systems for digital evidence extractions is another big issue. Further, digital evidence are difficult to handle and cannot be easily understandable even for experts. Extracting digital evidence from different storage media may require several layers of transformations (**Table 3**).

5. The significance of machine learning in digital forensic investigations

MLF is originating from AI to perform the huge amount of data, analyse the data to discover any criminal actions and risk and to segment the data to find criminal activity and behaviour. The intelligence systems which do not have any intelligent part cannot perform true learning capabilities and be a true one. DFI through ML is the latest trend to seize the potential of AI as leading security solutions capabilities.

ML behavioral analytics is the core part of modeling, profiling and prediction in medical, manufacturing, advertising and business intelligence and is recently used in law enforcement mechanism. In order to discover the criminal behaviour, MLF uses the wireless or wired networks via web or cloud computing. Thus MLF aims are to provide the new knowledge and skills and provide organized knowledge structure in order to produce progressive improvements in its own performance.

Originating from AI, ML algorithms can be used to analyze the huge amount of data to identify the risk, segment the data and detect criminal behaviour. ML algorithms enable the investigators to interrogate the vast scattered data sets which are placed in social and wired networks and web or cloud computing. In essence, ML algorithms contain the pattern recognition software that are used to analyse huge amount of data which are used to predict some behaviour. ML algorithms seek to learn from historical perspectives which are then used to predict future behaviour. MLF gains the capability to recognize the patterns of criminal activities through ML algorithms, in order to learn from the historical data about when and where the crime will take place. The malicious activities from extracted data set can be from burglaries, money laundering or intrusion attacks. This task can be achieved by formalizing and analyzing the servers, suspect's devices, wireless devices, the Internet and other kinds of data for visualization, link association, segmentation and predicting criminal activities. Nowadays, the industry is facing more advance cyber threats that cannot be tracked though traditional security measures. Attackers have designed more sophisticated ways to attacks on the system and become complicated over time. System administrator would not be able to detect these attacks each time. On the other hand, human expertise and competences have some limits, and this leads to the fact that industry is lacking in poor speed of incident occurrence, longer delay in detection and prevention of cyber threats and takes more advanced expertise to remove these cyber threats. Therefore, developing more advance machine learning models may help to prevent and protect form these cyber threats. Nowadays, there are many automated software available that can help the human to perform complicated and scientific tasks. In the next step, these automated tools need to be more advanced and should have the capability of AI and ML techniques.

6. Discussion and future prospects

From literature survey, it has been observed that there are many challenges which can be faced by the forensic experts when performing the test.

First of all there is an ultra-exponential growth in the data due to the inexpensive storage devices such as hard drives, CD, USB stick and so on. This makes it almost impossible for the individuals to perform the forensic in a short period of time.

Consequently, it is almost impossible for the forensic experts to perform the proper data analysis of each machine individually and also perform the cross-check on each machine's process. That limits the capability of the human works. In this line of reasoning, a huge amount of data needs to be sent to laboratory for forensic purposes with limited time and available resources. In a real-time digital forensic investigation, it is very difficult to determine in early stages which evidence is more important and relevant for investigating the crime, as an example, if we consider the cybercafé or a network of computers where several computers share the same IP address.

On the other hand, the intelligent tools are the main part of the MLF. However, these tools also show the problem for investigation in the pre-analysis phase. For that reason the lack-ness in the collection of large amount of data from distributed machines is need to be examined. Some of the existing tools are not helpful in solving the problem and even increases the time of investigation. The need is to make more intelligent methods and tools so that the automatic investigation of the suspects machines or malicious activity can be analyzed and determined in accurate time. The data can be stored and placed in any place for destructive purposes. Therefore, MLF techniques are the best sources for storing, evaluating and using this data in a productive way to anticipate and harmful activities. MLF methods can perform the meta-analysis on the meta-knowledge from different sources, and it can simplify the complex tasks into understandable and manageable data formats in a short period of time. MLF can provide the well-formed repository that can contain the well-sanitized data of digital investigation with well-known properties and results.

- Machine learning forensics solutions should:
 - Have data availability to support modeling.
 - Address well-scoped problems and methodology.
 - Explain well the reasoning process.
 - Formally structure the representation of knowledge.
 - Have well-organized performance evaluation.
 - Integrate with current architecture, tools and applications.

IntechOpen

Author details

Salman Iqbal^{1*} and Soltan Abed Alharbi²

1 Department of Computer Science, COMSATS University Islamabad, Vehari, Pakistan

2 Department of Electrical and Computer Engineering, University of Jeddah, Saudi Arabia

*Address all correspondence to: simbwp@gmail.com

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Pollitt M. A history of digital forensics. In: IFIP International Conference on Digital Forensics. Springer, Berlin, Heidelberg: Springer; 2010. pp. 3-15
- [2] Raghavan S. Digital forensic research: Current state of the art. *CSI Transactions on ICT*. 2013;1(1):91-114
- [3] Dierks MP. Computer network abuse. *Harvard Journal of Law & Technology*. 1992;6:307
- [4] Richardson R, Director C. CSI computer crime and security survey. *Computer Security Institute*. 2008;1:1-30
- [5] A.C.E.R.T.A. 2006 Australian Computer Crime and Security Survey. AusCERT & Australian High Tech Crime Center (AHTCC); November 23, 2006. Available from: <http://www.auscert.org.au/render.html?it=2001>
- [6] Stallard T, Levitt K. Automated analysis for digital forensic science: Semantic integrity checking. In: 19th Annual Computer Security Applications Conference, 2003. Proceedings. IEEE; 2003
- [7] Vermaas O, Simons J, Meijer R. Open computer forensic architecture a way to process terabytes of forensic disk images. In: *Open Source Software for Digital Forensics*. Boston, MA: Springer; 2010. pp. 45-67
- [8] Garfinkel SL. Automating disk forensic processing with SleuthKit, XML and python. In: 2009 Fourth International IEEE Workshop on Systematic Approaches to Digital Forensic Engineering. IEEE; 2009
- [9] Hoelz BW, Ralha CG, Geeverghese R. Artificial intelligence applied to computer forensics. In: *Proceedings of the 2009 ACM Symposium on Applied Computing*. ACM; 2009
- [10] Fizaine J, Clarke N. A crime depended automated search and engine for digital forensics. *Advances in Communications, Computing, Networks and Security*. 2013;10:73
- [11] Karabiyik U. Building an Intelligent Assistant for Digital Forensics. 2015
- [12] Marturana F, Tacconi S. A machine learning-based triage methodology for automated categorization of digital media. *Digital Investigation*. 2013;10(2):193-204
- [13] Marturana F, Tacconi S. A machine learning-based approach to digital triage. *Methodology for Automated Categorization of Digital Media*. In *Digital Investigation*. Elsevier; 2013;10(2):193-204
- [14] Grover J. Android forensics: Automated data collection and reporting from a mobile device. *Digital Investigation*. 2013;10:S12-S20
- [15] Mohammed H, Clarke N, Li F. An Automated Approach for Digital Forensic Analysis of Heterogeneous Big Data. *The Journal of Digital Forensics, Security and Law*. 2016
- [16] Tallón-Ballesteros AJ, Riquelme JC. Data mining methods applied to a digital forensics task for supervised machine learning. In: *Computational Intelligence in Digital Forensics: Forensic Investigation and Applications*. Springer; 2014. pp. 413-428
- [17] Lin X et al. Automated forensic analysis of mobile applications on android devices. *Digital Investigation*. 2018;26:S59-S66
- [18] James JI, Gladyshev P. Automated inference of past action instances in digital investigations. *International Journal of Information Security*. 2015;14(3):249-261

[19] Hargreaves C, Patterson J. An automated timeline reconstruction approach for digital forensic investigations. *Digital Investigation*. 2012;**9**:S69-S79

[20] Turnbull B, Randhawa S. Automated event and social network extraction from digital evidence sources with ontological mapping. *Digital Investigation*. 2015;**13**:94-106