International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2021; 2(1): 06-19
Received: 09-11-2020
Accepted: 16-12-2020

Francisca O Oladipo
Department of Computer Science, Federal University Lokoja, Kogi, Nigeria

Emeka E Ogbuju
Department of Computer Science, Federal University Lokoja, Kogi, Nigeria

Alayesanmi F Samson
Department of Computer Science, Federal University Lokoja, Kogi, Nigeria

Abraham E Musa
Department of Computer Science, Federal University Lokoja, Kogi, Nigeria

Corresponding Author:
Francisca O Oladipo
Department of Computer Science, Federal University Lokoja, Kogi, Nigeria

The state of the art in machine learning: Based digital forensics

Francisca O Oladipo, Emeka E Ogbuju, Alayesanmi F Samson and Abraham E Musa

DOI: https://doi.org/10.33545/27076571.2021.v2.i1a.21

Abstract
Digital forensics of visual-based evidence from video surveillance systems and forensic photographs holds object detection as a key aspect of the process. Recognizing an instance of object classes over a wide range of image data using computational techniques is one of the areas that has gained continuous attention over the years due to their numerous practical applications. Several algorithms and techniques have been specified for object detection and recognition with Machine Learning gaining more prominence and ensuring the remarkable performance of object detection and recognition systems. This study presents a comprehensive review of the frameworks and applications of Machine Learning in object detection and classification with particular applications to Digital Forensics. The analysis covers a wide range of publications between 2007 and 2019 available in different indexed and non-indexed databases and the candidate papers were selected using certain exclusion criteria proposed in the Kitchenham's methodology. The study in a bid to streamline future researches categorized digital forensic researches into six knowledge areas and identified the convolutional neural network as a state-of-the-art algorithm for machine learning-based digital forensics.

Keywords: Systematic review, object detection, image recognition, crime scene reconstruction, convolutional neural networks

1. Introduction

Machine learning is a concept with multi-disciplinary application to various fields that involves data one way or the other. It is a branch of artificial intelligence that involves developing a model to learn from existing data and utilize identified patterns in these data to make decisions with limited human intervention. Machine learning may also be described as utilizing a computer algorithm to develop models (computer programs) to learn from experience with existing data relating to a task and performance measure. The performance of a model at a given task increases with experience. A Machine learning model maximizes its utility performance by learning from data during its training stage. The concept has various application areas which include recommendation systems (Paulo *et al.*, 2015) [108], stock market forecasting (Shen *et al.*, 2013) [128], speech recognition (Banumathi & Chandra, 2017) [19], fraud detection systems (Anuj & Prabin, 2012) [13], object recognition (Helen, 2009) [69], automatic text classification (Joel *et al.*, 2014) [82] and so on. Machine learning models are not necessarily machines or systems but can be integrated on intelligent software/hardware for decision making based on patterns identified from the available data used to train on a particular problem domain.

The development of machine learning is made possible through the application of standard machine learning algorithms. These learning algorithms are classified into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning involves training a computer algorithm with labelled data into specific classes in a particular task, thereby applying learned knowledge from labelled data on real-world data. Unsupervised learning in contrast to supervised learning does not involve a training dataset with labelled data. In unsupervised learning, the algorithm utilizes hidden patterns on the data to categorize data based on identified patterns. Semi-supervised learning is an approach where the computer algorithm learns from datasets with labelled and unlabelled categories of data. Algorithms in this category can gain experience from incomplete data (Paulo *et al.*, 2015) [108]. Reinforcement learning is an essential approach that involves the model giving specific insights in a particular context to maximize its performance.

The availability of these machine learning algorithms and adequate computational power to develop strong machine learning models with reasonable accuracies (Prerak & Rughani, 2017) has made extensive research in digital forensics currently feasible. Certain machine learning algorithms especially Neural Networks has played major roles in detecting and classifying objects (Voronin *et al*., 2016; Makov & Creutzburg, 2016; Prerak & Rughani, 2017) [144, 144], conversational Chabot with sentence prediction (Prerak & Rughani, 2017) and so on. Convolutional neural networks (CNN), a variant of the Neural Networks, have been around since the early '90s but just became popular with many applications due to the recent availability of computational power to train CNN models. Deep layered CNN trained on large datasets have demonstrated tremendous accuracies for classification, detection, or identification of images, objects, or faces respectively, and therefore has found various application areas including in Digital forensics. The terms "Digital forensics" and "Computer forensics" are most times used interchangeably to define forensic analysis related to computing but both concepts are quite different in terms of scope. While computer forensics involves a forensic analysis of system-related crimes like social engineering attacks on a computer system, digital forensics is much more extensive as it is the forensic analysis of any digital evidence; not just computers it may be a data from a digital fax machine, an audio file, image file, video file and so on (Derek *et al*., 2008) [54]. The basic concept of digital forensics is providing a computer output from a computational model or software to serve as a second reviewer for forensic experts during forensic data analysis due to the problem of uncertainty during forensic analysis most times. Research had recognized that the concept of machine learning can help provide a better output based on automated and intelligent analysis of pieces of evidence at crime scenes (Prerak & Rughani, 2017).

Crime scene reconstruction (CSR), an aspect of forensics that is becoming largely dominant, plays a major role to determine the actual sequence of a criminal case after it has happened. It is regarded as a forensic analysis task that tries to eliminate irrelevant details about a crime scene while analyzing the relevant patterns identified. These relevant patterns may be position or location of physical shreds of evidence, bloodstain patterns, track-trail patterns, injury or wound pattern, glass fracture patterns, scattered furniture positions, and so on. CSR predicts the actual course of a crime scene and it helps to influence decisions of the jury by avoiding exonerations of the innocent in certain court cases (Samir & Nabanita, 2015) [124]. Although CSR is an important aspect of a criminal investigation, it lacks a standardized approach as evidence patterns vary from one crime scene to another thereby the need for a machine learning model to reduce the discomfort and technicality of forensic investigations at a crime scene. Apart from the CSR, all other prominent aspects of computer forensics are leading research application areas of machine learning algorithms as stated in Davide *et al.*, (2011) [52]. Digital devices have become a part of our daily life even in professional field; however, it may be a disadvantage sometimes, being a tool for certain crimes like cybercrimes, computer intrusion, and credit card fraud and so on, necessitating the need for computer security to prevent these crimes and computer forensics to investigate these kinds of criminal cases.

Research on digital forensics became prominent in the 1980s when the pioneer software tools for forensic analysis were developed in response to the high demand for technological impact during criminal investigations (Whitcomb, 2002) [149]. During this period particularly in 1984, various law enforcement agencies like the FBI began to develop programs to examine pieces of evidence digitally in forensic laboratories. Following this development for effective investigation, a team known as the Computer Analysis and Response Team (CART) was set up by the FBI and this idea was also incorporated by other law enforcement agencies (Noblett *et al.*, 2000) [105]. These early computer forensic researches fostered great influence on law enforcement agencies as about 48 percent of this law enforcement agencies began to have computer forensic laboratories to carry out analysis on pieces of evidence and not surprising at all, about 68 percent of pieces of evidence were analyzed in these laboratories (Whitcomb, 2002) [149]. Progressively, in the last two decades, several advancements in the technological sector and development of highly advanced computer systems have immensely influenced various fields including digital forensics.

The development of these highly advanced specialized software tools for digital forensics has improved the crime event reconstruction process during onsite forensic investigations. This improved reconstruction process provides crime event clues from shreds of evidence found at a physical crime scene, but a major drawback in this field is that law enforcement agencies still lack specific standards that govern the examination of these varieties of evidence (Pollit, 2010) [112]. However, it vital to emphasize that digital forensics is a research area actively advancing even with various applications of machine learning algorithms to the field. Machine Learning algorithms and Artificial Intelligence research began many years ago, but the research area remained quite dormant until recent times when computing capabilities required to develop highly sophisticated machine learning models started becoming available (Prerak & Rughani, 2017). Advances are actively made on the research area though this depends largely on the availability of data. While Grajeda *et al.*, (2017) [65] had reviewed 715 cybersecurity and cyber forensics research articles from 2010-2015 to find the availability of 351 datasets for forensic researches, we present in Table 1 the descriptions of the datasets used by some of the works discussed in this study.

Table 1: Some available research datasets for digital forensics

Dataset	Description	Source
Pascal Visual Object Challenge (VOC) 2007 [57].	9,963 images, containing 24,640 annotated objects with over 20 classes of different objects	Everingham *et al.* (2007) [57]. Used in Christian *et al.* (no date)
VOC 2012	20 classes. The train/val data has 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations.	Everingham *et al.*, (2010) [58]. Used in Joseph *et al.* (2017) [84].
Image Net 2012 [4]	1000 categories and 1.2 million images	Olga *et al.* (2015) [106]. Used in Joseph *et al.* (2017) [84].

Public dataset	4,500 images of more than 30,000 license plate characters	Rayson *et al.* (2017)
MS-COCO	3, 000 images and 80 object categories with multiple objects per image.	Surajit *et al.* (2017) [133]
ImageNet Room objects	1, 345 images with 12 object categories that can be found in a bedroom	Surajit *et al.* (2017) [133]
Karina	16 videos of 3 minutes in 7 different rooms containing 40 categories	Surajit *et al.* (2017) [133]
UCI Glass Identification	214 labeled glass instances with 10 features	Jose and Antonio (2016) [11]

There have been traditional tools and techniques that solve digital forensic problems. These traditional digital forensic tools can be multipurpose in operations, that is, they can perform detailed functions such as memory forensic analysis, hard drive forensic analysis, forensic image exploration, forensic imaging, and mobile forensics. Some of the open-source traditional digital forensic tools are presented in Table 2.

Table 2: Some open-source traditional digital forensic tools

S. No	Tools	Usage
1.	Autopsy	Analyze hard drives and smartphones
2	Encrypted Disk Detector	Check encrypted physical drives
3	Wireshark	Network capture and analyzer
4	Magnet RAM Capture	Analyze artifacts in memory
5	Network Miner	Network forensic analyzer
6	NMAP	Networks and security auditing
7	RAM Capturer	Dump data from a computer's volatile memory
8	FAW	Acquire web pages for forensic investigation
9	USB Write Blocker	Use the Windows registry to write-block USB devices.
10	Crowd Response	Gather system information for incident response and security engagements.
11	NFI Defraser	Detect full and partial multimedia files in the data streams.
12	Dumpzilla	Extract all interesting information from some browsers
13	Sleuth Kit	Investigate and analyze volume and file systems to find evidence
14	CAINE	To analyze, investigate and create an actionable report
15	Volatility	Incident response and malware analysis
16	The Coroner’s Toolkit	Aid analysis of computer disasters and data recovery.
17	Bulk Extractor	To extract useful information for solving cyber crimes
18	Xplico	To extract applications data from internet traffic
19	USB Historian	Gives a list of all USB drives that were plugged into the machine.
20	SIFT	To carry out a detailed forensic analysis or incident response study
21	Oxygen Forensic Suite	To collect evidence from a mobile phone.

Adapted from Infosec (2020) [78] and GFI (2018) [62]

However, the usages of these traditional tools are most effective when applied for a single forensic investigation case. On the other hand, most of the available forensic tools cannot handle heterogeneous big data which almost all current investigative cases deal with. Again, in instances where multiple tools are used for a single investigative case, it has been found that there has been an inability to cross-correlate the findings thereby often leading to inefficiencies in processing and identifying evidence (Mohammed *et al.*, 2016). This gap necessitates the engagement in this study to systematically extract the state of the art in applying machine learning algorithms in developing solutions for digital forensics. In this study, we present a systematic review of the literature concerning research in machine learning-based digital forensics to detect and recognize object classes over a wide range of image data using findings from publications over 12 years. This paper will advance the notion of conducting research by literature and making substantial contributions to knowledge through establishing the state of the art in any research field. Building research on existing knowledge and benchmarking with previous research is a very essential ingredient in all research activities.

2. Methodology

A systematic review involves analyzing, evaluating, and interpreting available research and contributions relevant to a particular topic of discussion (Kitchenham, 2004) [85]. Though sometimes not so obvious and not widely accepted as significant contributions to knowledge by some in the core research communities, systematic review and evaluation have the potential to contribute significantly to knowledge in a chosen domain. Available scientific literature in the given domain is usually used in the review. There are several approaches and methods of conducting a literature review depending on the reasons for undertaking the review. Table 3 summarizes some popular approaches (Snyder, 2019) [68].

Table 3: Approaches to literature reviews (Snyder, 2019) [68]

Approach	Systematic	Semi-systematic	Integrative
Typical purpose	Synthesize and compare evidence	Overview research area and track development over time	Critique and synthesize
Research questions	Specific	Broad	Narrow or broad
Search strategy	Systematic	May or may not be systematic	Usually not systematic
Sample characteristics	Quantitative articles	Research articles	Research articles, books, and other published texts
Analysis and evaluation	Quantitative	Qualitative/quantitative	Qualitative

Examples of contribution	Evidence of effect Inform policy and practice	State of knowledge, Themes in literature, Historical overview, Research agenda, Theoretical model	Taxonomy or classification Theoretical model or framework

We adopted the systematic search strategy in this study. The search was performed using two keywords: “machine learning” and “digital forensics” from 2007 – 2019. The search engines used were the Association for Computing Machinery’s (ACM) Digital Library and Google Scholar. While the former returned 82 articles, the later returned 3170 articles from which we choose the top 18 articles sorted by relevance and indexed at Scopus, Science Direct, and EBSCOhost. We had a total of 100 articles from both search engines including books, periodicals, proceedings, theses, and technical reports. The Selection and exclusion criteria proposed in Carvalho and Wangenheim (2019) [37] were adopted in choosing the literature. For each candidate, we find the problem identification and definition, whether the work had established the state-of-the-art in the research direction; and what gaps and deficiencies were observed in previous approaches. We identify the methodologies, tools, techniques, and materials adopted, and the justification of the choice. Finally, we examine the obtained results and how (if) they significantly differ from the expectations. We aim to extract the state of the art in different areas of machine learning-based digital forensics which we grouped into six (6) knowledge areas as follows: Object Detection & Video/Audio Forensics; Image Recognition & Classification; Printer, Camera & Mobile Device Forensics; Computer, Network & Web Forensics; Crime Scene Reconstruction; and Reviews).

3. Result and Discussions

The result of this study is a presentation of a clear research direction or advances in machine learning-based digital forensics. We observed that while we grouped the reviewed works into six (6) knowledge areas (see Table 4), some works intersect into another group and there may be no clear boundaries in the grouping. For instance, an algorithm used in object detection may be useful in image classification. Likewise, an algorithm in object detection may be applied in a crime scene reconstruction work. The main note is that all the investigations in the works applied one machine learning algorithm or another in solving the forensic problem in the current big data domain.

Table 4: Research advances in machine learning-based digital forensics

Object Detection, Video/Audio Forensics	**Image Recognition & Classification**	**Printer, Camera, & Mobile Device Forensics**	**Computer, Network & Web Forensics**	**Crime Scene Reconstruction**	**Reviews**
Platzer *et al.* (2014) [111]	Zwanger *et al.* (2013) [153]	Barmpatsalou *et al.* (2018) [20]	Tang & Fidge (2010) [34]	Levett *et al.* (2010) [86]	Luciano *et al.* (2018) [91]
Shi *et al.* (2007) [129]	Bouchaud *et al.* (2018) [27]	Choi *et al.* (2010) [42]	Cuzzocrea & Pirrò (2016) [50]	Mohammed *et al.* (2018) [99]	Awad *et al.* (2018) [16]
Al-athamneh *et al.* (2016) [3]	Brennan *et al.* (2012) [28]	Wang & Rountev (2017) [147]	Ariu *et al.* (2011) [14]	Aron *et al.* (2016) [15]	Matthew *et al.* (2015) [94]
Chen *et al.* (2017) [39]	Boroumand & Fridrich (2016) [26]	McLaughlin *et al.* (2017) [95]	Chou *et al.* (2012) [44]	Saikia *et al.* (2017) [122]	Anderson *et al.* (2011) [8]
Calderara *et al.* (2009) [34]	Loia *et al.* (2009) [90]	Carter (2013) [36]	Petrik *et al.* (2018) [110]	Liu *et al.* (2018) [132]	
Taranta *et al.* (2016) [135]	Tariq *et al.* (2018) [136]	Bondi *et al.* (2017) [25]	Sayakkara *et al.* (2018) [125]	Yang *et al.* (2016) [151]	
Thurau *et al.* (2010) [139]	Choudhary & Nain (2016) [43]	Sameer (2017) [123]	Popov *et al.* (2018) [113]	Surajit *et al.* (2017) [133]	
Guang *et al.* (2015) [66]	Nguyen *et al.* (2018) [104]	Cozzolino & Verdoliva (2020) [49]	Wang *et al.* (2018) [146]	Jose and Antonio (2016) [11]	
Chunhui *et al.* (2014) [46]	Liu *et al.* (2011)	Amel (2016) [6]	Michalas *et al.* (2017) [97]		
Wang & Zhang (2016) [145]	Bogen *et al.* (2013) [43]	Muhammad *et al.* (2017) [101]	Dhanalakshmi & Chellappan (2010) [55]		
Huo & Zhu (2019) [75]	Gloe & Böhme (2010) [63]		Maiya *et al.* (2013) [92]		
Bulbule *et al.* (2019) [33]	Collomosse *et al.* (2018) [47]		Therdphapiyanak & Piromsopa (2013) [137]		
Bakas & Naskar (2018) [17]	Michiel & Claudia (2014) [98]		Steinebach *et al.* (2018) [132]		
Hong-Wei (2015) [73]	Steven *et al.* (2015) [131]		Merkle (2008) [96]		
Christian *et al.* (no date) Joseph *et al.* (2017) [45, 84].	Ren *et al.* (2017)		Adarsh & Ajeena (2014) [1]		
Rayson *et al.* (2017).	Chen (2018) [40]		Natércia *et al.* (2018) [102]		
Felix *et al.* (2018) [159]	Qian *et al.* (2016) [117]		Carlos *et al.* (2017) [35]		
Hoo-Chang *et al.* (2016) [74]	Anda *et al.* (2018) [7]		Shujun *et al.* (2018) [130]		
	Moreira & Fechine		Alhawi *et al.* (2018) [5]		

	(2018) [100]				
	Sharma et al. (2016) [126]		Irvin and Panagiotis (2017) [79]		
	Brown *et al.* (2005) [31]				
	Yiyu *et al.* (2017) [152]				

Again, a closer look at the research advances shows that there is less concentration on works that focus on the Printer, Camera & Mobile Device Forensics as well as on the Computer, Network & Web Forensics. However, most of the advances show a high concentration on Object Detection & Video/Audio Forensics as well as Image Recognition & Classification which find most of their applications in the Crime Scene Reconstruction domain. This section discusses some of the works with the most advances (object detection and classification) to examine the algorithms applied, the dataset used and the key findings from each of the works. Other works in the fewer groups were also discussed to achieve the same aim.

3.1 Object detection/Classification

A comprehensive description and development of a novel method for object detection in images using Deep Neural Networks were provided by Christian *et al.* (no date). In the research, a machine learning model was developed to successfully classify images and localize various object positions detected in the image. The research comprehensively explains how deep neural networks outperformed other classification techniques and it presented neural networks as a more powerful and robust algorithm suitable for classification problems because of its deep architecture. The validity of the approach used in the model developed is analyzed by using it on VOC as the test dataset. The experiment conducted with the model on this dataset utilizes boundary boxes to detect a significant object in these test images and its conclusive accuracy is compared with three related approaches which include: sliding window version of a DNN classifier (Alex *et al.*, 2012), a 3 layer compositional model by Long *et al.* (2010) and the DPM by Pedro *et al.* (2010) [109] and Girshick *et al.* (2013) to evaluate the achieved results of the model.

Hong-Wei (2015) [73] developed a model for crowd analysis for digital forensics over video surveillance during the 2015 Emotion Recognition in the Wild contest. Due to the large number of resources required to train the neural networks from scratch on a pre-trained deep CNN using a small dataset of a static image from movies, the Transfer Learning approach was deployed. The pre-trained model was originally trained on a large dataset of 1.2 million images thereby contributing to the accuracy of the model. The overall accuracy of the model after being trained and tested on the new dataset for emotion recognition on static images is 48.5%, which is quite low and inefficient. A proposed solution to this accuracy drawback is the expansion of the labeled dataset to a reasonable extent while retraining the model.

An object detection model on digital images, based on a novel CNN was developed by Joseph *et al.* (2017) [84]. The architecture of the model involves a single CNN to classify boundary boxes of objects in a resized input image of 448 x 448, unlike other object detection framework that uses a sliding window approach which the CNN is applied on several spaced locations on the entire image. The model was trained on several datasets including VOC 2007 & 2012 and ImageNet 2012 [4] thereby making the model generalized on several categories of objects. However, this model still has a drawback in that it still struggles with a small and clustered image like a flock of birds in the sky.

Similar research based on the YOLO-CNN was conducted by Rayson *et al.* (2017). The researchers developed a real-time end to end automatic license plate recognition system in support of the state-of-the-art YOLO-CNN algorithm for object detection. The final layer of the model was trained and fine-tuned using a large public dataset. The dataset was prepared with three different digital cameras used to take photographs of the license plates at different angles, thereby preparing a large dataset with a variety of images in terms of image quality which contributes to the level of accuracy achieved by the model. The experimentation was done using the Darknet Framework on GPU due to the computational power required to train YOLO on a large dataset. The accuracy achieved with the developed model is 78.33%. The research model was not created from scratch for object detection, they retrained CR-NET, YOLOV2, and Fast YOLO which are pre-trained state-of-the-art models for object detection (Transfer Learning) and compare their results.

A novel approach for object detection and localization in digital images for forensic crime scene investigation was developed by Surajit *et al.* (2017) [133]. Crime scene reconstruction starts with the identification of shreds of evidence at a crime scene, therefore the research presents a Faster Region-based Convolutional Neural Network (R-CNN) to optimize existing deep learning algorithms for object detection. The R-CNN model was pre-trained with the MS-COCO dataset. It was tested on two different test sets; the ImageNet Room objects and Karina dataset. The accuracy of the model turns out to be excellent on the ImageNet dataset but low on the Karina dataset due to poor image quality of the videos.

An evaluation and comparison of models design for age estimation from facial features extracted from digital images and videos was conducted by Felix *et al.* (2018) [59]. The researchers developed a model to support digital forensic expert with an automated investigation of images of child abuse and child pornography. To solve the prominent problem of the non-availability of datasets, they went further to develop a dataset generator that creates a structured dataset from images contained in various semi-structured datasets. The comparative analysis compares the accuracy of the following machine learning models (online and offline) based on their mean absolute error: Amazon Recognition (Artificial Neural Network), Deep Expectation, DEX (CNN), Kairos (Support Vector Machine, SVM), and Microsoft Azure Cognitive Service (Deep Learning).

Jose and Antonio (2016) [11] proposed research on various categories of glasses that could be found shattered or broken at a crime scene as a primary source of evidence if a proper and accurate identification of these glass types is achieved. A dataset from glass identification of the USA Forensic Science Service which is available on the University of California, Irvine (UCI) repository was used. They performed a comparative analysis of four machine learning classification algorithms (Decision Tree, Naïve Bayes, Artificial Neural Network, and K-Nearest Neighbours).

An image mining system to detect illicit images with criminal behaviors for digital forensics was developed by Brown *et al.* (2005) [31]. Using SVM and a Bayesian classifier algorithm, the researchers developed a model to filter relevant features of images based on grammar queries from the user. The model developed was trained to classify appropriate and inappropriate images based on clad and nude content respectively, using a training set of 214 images in three different color spaces to train the SVM model. The performance result of the model is quite remarkable considering the time of the research with 92% true positive results and 79% true negative detection rate, however, incorrect classification can be corrected by fine-tuning the grammar-based query and feedback feature provided by the system for communication with the user.

A different dimension to digital image forensics was introduced by Muhammad *et al.* (2017) [101] whereby the validation of the authenticity of images is conducted through the recognition of the particular camera used to take the image. A justification for this approach is the affordances of a large variety of mobile phones today with highly sophisticated cameras and it becomes highly expensive in terms of time and computational power to carry out digital forensic analysis on a single PC. They carried out the source camera identification methodology presented on a 6000 image dataset taken by six (6) mobile phone cameras using Hadoop for feature extraction and Manhout Random Forest Classifier for image classification, thereby achieving a very efficient forensic analysis process in terms of speed and an accuracy of 85% to 95% across various mappers.

An application of CNN for art painting identification to detect copyrighted images used without content provider's permission for commercial purposes was created by Yiyu *et al.* (2017) [152]. The research is an instance of *Scale-Invariant Feature Transform* (SIFT); a state-of-art handcrafted image descriptor. It deployed an artwork dataset of 100 main art images distorted in various forms to expand the dataset to 30000 distorted images. The resultant model was trained on 25000 distorted images and tested on 5000. The model produced a CNN with a 2% test error rate.

An extensive analysis and evaluation are carried out on different CNN architectures on a computer aid detection problem for thoracoabdominal lymph node (LN) and interstitial lung disease (ILD) detection by Hoo-Chang *et al.* (2016). The Alex Net CNN (seven-layered), Cifar-CNN (three-layered) and Google Net CNN (eleven layered) are the CNN evaluated in the research. Two major challenges using CNN to classify medical images include the unavailability of an extensive dataset for training and testing. However, they utilize publicly available thoracoabdominal lymph node and interstitial lung disease image datasets labeled by radiologists, and secondly, CNNs are trained on natural images, unlike medical images which are 2D or 2.5D images. Pre-trained CNN can still classify medical images effectively by using the approach of transfer learning which involves fine-tuning the higher layers of these CNN. The conclusive accuracy of various CNN evaluated in the research creates a state-of-the-art machine learning model that can be used to develop high-performance CAD systems in further research works.

Francesco *et al.* (2015) [61] presented a method to prevent counterfeit images from being detected by state-of-the-art forgery detectors by modifying certain micro-pattern in these images. The research developed a strategy for counter-forensics which overrules the operation of techniques used for forgery detection which uses the statistical distribution of micro-patterns in images which are optimized through high-level filtering and summarized in some image descriptor used for the final classification. We review the statistical algorithm proposed for counter forensics of images or videos which they describe as Greedy Sampling Algorithm, they analyzed its efficiency when it has limited knowledge or perfect knowledge of the feature used by a forgery detection algorithm for classification (genuine or not). They propose the success of the research in the paper by analyzing the experimental result of the counter-forensics algorithm on 100 images; thereby the output images are indistinguishable by the forgery detector. However, the result is remarkable when the algorithm presented has complete knowledge of the feature extraction technique used by the detector compared to the limited knowledge scenario which consumes more CPU time.

3.2 Fraud detection

A Naive Bayesian algorithm-based data mining model to detect fraudulent transactions was defined by Bhowmik (2009) [23]. The researcher reviewed various machine learning classification techniques for fraud detection and presented a probabilistic supervised learning classification technique which considers each instance in the dataset independent with certain attribute thereby classifying an instance with an unknown class with the highest probability given its attribute or features as the most appropriate for the problem domain. The model was applied over a dataset for detecting fraud in automobile insurance, which consists of a training set with 20 instances (3 fraud and 17 legal) with 6 features for each instance, thereby assigning a new instance to a specific class (fraud or legal) with the highest probability. The performance of the model was validated with a confusion matrix and visualized using the Relative Operating Characteristic curve which compares the performance of various classifiers reviewed in the research.

A neural networks-based machine learning model for forensic activities on various cases of credit card fraud was developed by (Divya *et al.* 2014) [56]. The model was applied over an unlabelled dataset containing a summary of 20000 active credit cardholders over six months using the Neuroph IDE, a neural network framework implemented in Java. In addition to credit card fraud forensics, the research aimed to demonstrate that neural network models are more robust and highly optimized than Naive Bayes, Hidden Markov Model, or other classification machine learning algorithms due to their multiple layers. Although the classification result is quite impressive, the model can still be improved by expanding the dataset used and improving the neural network by adding more layers.

3.3 Computer and network forensic

Machine learning techniques were also deployed by Irvin and Panagiotis (2017) [79] in identifying and predicting protocols carried through a DNS channel to aid network forensic analysis thereby reducing the time required for identification, analysis, and reconstruction of a network-related digital crime. The research seeks to improve on the approach used for network forensics due to the lack of an existing universally standardized approach to identify various network protocols through DNS tunnels; and developed a novel approach to identify four protocols (HTTP, HTTPS, FTP, and POP3), an extension of previous work for two protocols (HTTP and FTP) by (Homem *et al.*, 2016) [71], using K-Nearest Neighbours, Decision Trees, Support Vector Machine and Neural Networks, thereby comparing accuracies of these machine learning algorithms on the available. However, a collection of a real-world summary of DNS tunneling into a single dataset for training and testing these machine learning algorithms is an almost impossible task, thereby they create a dataset to address this problem. The results of the comparative analysis give the Multi-layered Neural Network with the highest accuracy (95%) while K-Nearest Neighbours taking K as 5 is the least accurate (90%).

Ikuesan *et al.* (2017) [77] integrated user attribution which involves identifying a human user based on specific thinking style and behavior on a digital medium into digital forensics. Reoccurring patterns of 43 users over network traffic were collected, analyzed, and classified into a specific thinking style using various machine learning algorithms. The decision tree was considered most accurate for user attrition with the lowest error rate, thereby developing a graphical model with Unified Modelling Language to describe its forensic application. However, we consider the research not very elaborate as various other parameters could be used to identify a user on a digital medium like personality trait amongst others.

3.4 Mobile and text forensics

Homem (2016) [71] presented an architecture for automation of digital forensic in mobile and cloud environments to ensure the soundness of digital evidence in the judicial system and reduce human intervention in the forensic investigation of this evidence. The solution presented in this research is a technological solution to automate a large amount of the entire forensic analysis process. The research includes a review of various tools used in the digital investigation to improve the efficiency of the process; however, these tools still require human expertise. They explore four research ideologies for digital forensic automation which all sums up to the entire architecture presented in the paper. The Life Evidence Information Aggregator (LEIA) architecture presented in this research is considered a hypervisor-based and a peer-to-peer distributed system with a cloud-based backend for digital evidence acquisition, however, they develop a prototype for experimentation.

Priyanka and Prashant (2014) [115] developed a data mining technique for digital forensic investigations based on the Generalised Sequential Pattern algorithm for digital forensics; the algorithm is an application of sequence mining. A text dataset for forensic investigation is used with the algorithm introduced in the research; however, they optimize its operation by adding a statistical test analysis and the Self-Organizing Kohonen maps (SOM) classification technique. SOM is a neural network model that is used to map high dimensional input data to a lower-dimensional space thereby giving a more unsupervised learning sequence from textual data used in the paper; it is used mostly for clustering and visualization of high dimensional data. The accuracy of the data mining methodology presented in this paper for digital forensics is tested on the data contained in a USB drive and it is accuracy is very impressive with 98.3%, which is higher than a method used to compare its accuracy in the research. Thereby, the research conclusively presents a standardized approach for digital forensics using data mining techniques.

3.5 Tools and Frameworks

A multi-agent-based artificial intelligence system named Multi-Agent Digital Investigation toolkit (MADIK) was introduced to forensics by Hoelz *et al.* (2009). The research was aimed at reducing the time required to investigate and correlate a large number of files on a computer drive that can serve as evidence in criminal cases, thereby giving a computer forensic examiner a precise direction for its analysis. The MADIK system has six intelligent agents which perform specific forensic analysis to assist forensic experts. This system is used on real forensic data and gives a commendable result; however, the system is not considered perfect as more intelligent agents can be included to improve its performance and reasoning process.

A generic framework for divergent Deep Learning (DL) cognitive computing techniques into Cyber Forensics (CF) hereafter referred to as the DLCF Framework was proposed by (Kariea *et al.*, 2019) [103]. The authors believed that Deep Learning has the potential to help in the fight against cybercrime. The research developed a generic DL framework that can be integrated into Cyber Forensics (CF) to realize effectiveness during a forensic investigation. The authors recommended more research towards improving their prototype DLCF framework.

A machine learning-based approach to cyber forensics in an Internet of Things (IoT) environment was proposed by (Chhabra *et al.*, 2018). The authors first identified size constraints as the major limitation to forensic analysis in IoT systems and proposed an approach that generally takes the size benchmarking into full consideration through the use of a Google paradigm. Using open-source tools that support scalability and parallel processing, and dataset from the Center for Applied Internet Data Analysis (CAIDA), the proposed forensic framework was validated and a result with 99% sensitivity was obtained from the performance metrics of the model.

Research by Costantini *et al.* (2019) [48] explored the possible application of artificial intelligence and computational logic to digital forensic based on the automation of evidence analysis through the Answer Set Programming (ASP) approach. The research demonstrated how significant complex investigations that are difficult to solve by human experts and investigators, are expressed as optimization problems belonging in many cases to the $\mathbb{P}$ or $\mathbb{NP}$ complexity classes. Sample ASP programs were deployed to define and formalize complex problems to demonstrate the formulation of tangible investigative hypotheses.

4. Summary of review findings

Digital forensic involves various tedious processes due to a large number of digital devices and data available during criminal investigations and various research studies suggest ways to mitigate this problem. From the different research examined in this study, the CNN model has played a major role in forensics research and is still the model of choice in real-time object detection. Having essentially explored various works that applied machine learning algorithms and techniques to solve problems in digital forensic, we found that works on image classification, object detection, crime evidence analysis for event reconstruction are prominent in the field. As an active research area, the criminal investigation requires analysis of a large amount of data mostly visual (images and videos) for physical crimes, which can be very technical and error-prone when done manually, hence the application of the state of the art CNN algorithm. CNN and its various deep learning classifier variants have become a gold standard for image classification, thanks to the large dataset of images explored in this study. Table 5 summarizes the algorithms and data applied in each of the works validating the fact that CNN is the algorithm of choice across the works.

Table 5: Summary of Reviews

Study	Aim/Objective	Data	Settings /Method
Christian *et al.* (n.d.)	The development of a machine learning model for classifying images	Pascal VOC 2007[57]	Deep Neural Network
Hong-Wei (2015) [73]	The development of a model for crowd analysis	Image Net 2012 [4]	Transfer learning
Joseph *et al.* (2017) [84]	The development of an object detection model on digital images	Pascal VOC 2007 & 2012, Image Net 2012 [57, 4]	CNN
Rayson *et al.* (2017)	The development of a real-time automatic license plate number recognition system	Images of licensed plate character	YOLO-CNN
Surajit *et al.* (2017) [133]	The development of a novel approach for object detection and localization in digital images for forensic crime scene investigation	MS-COCO, ImageNet Room Objects, Karina	Faster Region-based CNN
Felix *et al.* (2018) [159]	The development of a model that supports digital forensic expert with an automated investigation of images of child abuse and child pornography	A dataset generator	ANN, CNN, SVM
Jose and Antonio (2016) [11]	The development of a model that categorizes shattered or broken glasses at a crime scene	UCI glass identification dataset	Decision Tree, Naive Bayes, KNN, and ANN
Brown *et al.* (2005) [31]	The development of an image mining system that detects illicit images with criminal behaviors	The researcher created a dataset of 214 images	SVM, Naive Bayes
Muhammad *et al.* (2017) [101]	The development of a forensic approach for source camera identification	The researcher created a dataset of 6000 images	Mahout Random Forest
Yiyu *et al.* (2017) [152]	The development of a deep learning model for art painting identification to detect copyrighted images	The Researcher created a dataset of 30000 distorted images	CNN
Hoo-Chang *et al.* (2016) [74]	Computer-aided detection for thoracoabdominal lymph node and interstitial lung disease	The researcher created a dataset from radiologists	CNN
Chhabra *et al.* (2018)	The development of a machine learning-based approach for cyber forensic approach in an IoT environment	Dataset from CAIDA	Google paradigm
Bhowmik (2009) [23]	The development of a data mining model for the detection of fraudulent transactions	Fraud detection dataset with 20 instances	Naïve Bayes
Divya *et al.* (2014) [56]	The development of a deep learning model for analyzing the cases of credit card fraud	A dataset of 20000 active credit cardholders over six months	Neural Network
Irvin & Panagiotis (2017) [79]	The improvement on the approach used for network forensics	The researcher created a dataset of DNS tunneling	KNN, Decision Trees, SVM, Neural Networks

As identified in Banumathi and Chandra (2017) [19], the various types of deep learning classifiers which include Recurrent Neural Network (RNN), Restricted Boltzmann Machines (RBM), Deep Belief Network (DBN), Deep Convex Nets (DCN), Deep Neural Networks (DNN), Deep Auto Encoder, and Deep Stacking Network (DSN) should be the go-to algorithms for any machine learning-based digital forensics investigation involving object detection, image classification, and crime scene reconstruction. Although other algorithms like SVM, Decision Trees, Random Forest, and Naïve Bayes are applied in the works on the other knowledge areas (Printer, Camera, & Mobile Device Forensics; and Computer, Network & Web Forensics); CNN still played significant roles in them.

5. Conclusion

We have established in this systematic review that machine learning, a branch of artificial intelligence gives machines or computers the ability to learn from an existing dataset and utilize this experience on new data items to make predictions and decisions based on data patterns learned during the training stage. We have highlighted various extensive researches in the areas of object recognition and classification with several approaches from trivia to complex in this study. This study has established that the state-of-the-art algorithm in machine learning-based digital forensics is the CNN algorithm and its various deep learning classifier types. All the reviewed works validate the finding and show that there are no significant differences from the expectation of CNN as the state-of-the-art algorithm for digital forensics.

6. References

1. Adarsh SVN, Ajeena BAS. A log-based strategy for fingerprinting and forensic investigation of online cybercrimes. 2014 International Conference on

Interdisciplinary Advances in Applied Computing (ICONIAAC ’14). Association for Computing Machinery, New York, NY, USA 2014;7:1-5. DOI: https://doi.org/10.1145/2660859.2660912
2. Ahmad UK, Asmuji NF, Ibrahim R, Kamaruzaman NU. Forensic classification of glass employing refractive index measurement. Malaysian Journal of Forensic Sciences 2012, 1-4.
3. Al-athamneh M, Kurugollu F, Crookes D, Farid M. Video authentication based on statistical local information. 9th International Conference on Utility and Cloud Computing (UCC ’16). Association for Computing Machinery, New York, NY, USA 2016, 388-391. DOI: https://doi.org/10.1145/2996890.3007857
4. Alex K, Ilya S, Geoff H. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 2012.
5. Alhawi OMK, Baldwin J, Dehghantanha A. Leveraging machine learning techniques for windows ransomware network traffic detection. In: Dehghantanha A, Conti M, Dargahi T. (Eds) Cyber Threat Intelligence. Advances in Information Security 2018, 70.
6. Amel TA. Forensic source camera identification by using features in machine learning approach. Université Montpellier 2016.
7. Anda F, Lillis D, Le-Khac N, Scanlon M. Evaluating automated facial age estimation techniques for digital forensics. IEEE Security and Privacy Workshops (SPW), San Francisco, CA 2018, 129-139.
8. Anderson R, Walter S, Terrance B, Siome G. Vision of the unseen: Current trends and challenges in digital image and video forensics. ACM Computational Survey 43 2011;4(26):42. DOI: https://doi.org/10.1145/1978802.1978805
9. Android. Android Studio 2018. developer.android.com/studio/index.html
10. Aniruddha T. Transfer learning for image classification and plant phenotyping. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 2016;5(11):2664-2669.
11. Antonio JTB, José CR. Data mining methods applied to a digital forensics task for supervised machine learning. Seville: Department of Languages and Computer Systems, University of Seville 2016.
12. Antonio TB, Riquelme J. Data mining methods applied to a digital forensics task for supervised learning. Seville, Spain 2014.
13. Anuj S, Prabin PK. A review of financial accounting fraud detection based on data mining techniques. International Journal of Computer Applications (0975 – 8887) 2012;39(1):37-47.
14. Ariu D, Giacinto G, Roli F. Machine learning in computer forensics (and the lessons learned from machine learning in computer security). 4th ACM workshop on Security and artificial intelligence (AISec ’11). Association for Computing Machinery, New York, NY, USA 2011;99-104.
15. Aron M, Nils T, Niloy JM. SMASH: physics-guided reconstruction of collisions from videos. ACM Trans. Graph. 35 2016;6(199):14. DOI: https://doi.org/10.1145/2980179.2982421
16. Awad AR, Beztchi S, Smith JM, Lyles B, Prowell S. Tools, techniques, and methodologies: a survey of digital forensics for SCADA systems. 4th Annual Industrial Control System Security Workshop (ICSS’18). Association for Computing Machinery, New York, NY, USA 2018, 1-8. DOI: https://doi.org/10.1145/3295453.3295454
17. Bakas J, Naskar R. A digital forensic technique for inter–frame video forgery detection based on 3D CNN. In: Ganapathy V., Jaeger T., Shyamasundar R. (eds) Information Systems Security. ICISS 2018. Lecture Notes in Computer Science. Springer, Cham 2018, 11281.
18. Banumathi A, Dr. Chandra E. Deep learning architectures, algorithms for speech recognition: An overview. International Journal of Advanced Research in Computer Science and Software Engineering 2017, 213-220.
19. Banumathi C, A, Chandra E. Deep learning architectures, algorithms for speech recognition: An overview. International Journal of Advanced Research in Computer Science and Software Engineering 2017;7(1):213-220.
20. Barmpatsalou K, Cruz T, Monteiro E, Simoes P. Current and future trends in mobile device forensics: A survey. Association for Computing Machinery, Computational Survey 51 2018;3(46):31. DOI: https://doi.org/10.1145/3177847
21. Bevel T. The evolution of crime scene reconstruction from proto-analysis to holistic analysis: A court case that assisted in this evolution. Journal of the Association for Crime Scene Reconstruction 2011, 25-29.
22. Bhatt P. Machine learning forensics: A new branch of digital forensics. International Journal of Advanced Research in Computer Science 2017, 217-222.
23. Bhowmik R. Data mining techniques in fraud detection. Journal of Digital Forensics, Security and Law 2009;3(2):35-54.
24. Bogen PL, McKenzie A, Gillen R. Redeye: A digital library for forensic document triage. In Proceedings of the 13th ACM/IEEE-CS Joint conference on Digital libraries (JCDL ’13). Association for Computing Machinery, New York, NY, USA 2013, 181-190. DOI: https://doi.org/10.1145/2467696.2467716
25. Bondi L, Baroffio D, Güera P, Bestagini EJD, Tubaro S. First steps toward camera model identification with convolutional neural networks. IEEE Signal Processing Letters 2017;24(3):259-263. DOI: 10.1109/LSP.2016.2641006
26. Boroumand M, Fridrich J. Boosting steganalysis with explicit feature maps. 4th ACM Workshop on Information Hiding and Multimedia Security (IH & MM Sec ’16). Association for Computing Machinery, New York, NY, USA 2016, 149-157. DOI: https://doi.org/10.1145/2909827.2930803
27. Bouchaud F, Grimaud G, Vantroys T. IoT Forensic: identification and classification of evidence in criminal investigations. 13th International Conference on Availability, Reliability and Security (ARES 2018). Association for Computing Machinery, New York, NY, USA, Article 2018;60:1-9. DOI: https://doi.org/10.1145/3230833.3233257
28. Brennan M, Afroz S, Greenstadt R. Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. ACM Trans. Inf. Syst.

Secur 15 2012;3(12):22. DOI: https://doi.org/10.1145/2382448.2382450

29. Brian D, Carrier EH. Defining event reconstruction of digital crime scenes 2004.
30. Brian D, Carrier EH. Defining event reconstruction of digital crime scenes. West Lafayette 2004.
31. Brown R, Pham B, De Vel O. Design of a digital forensics image mining system. Australia 2005.
32. Bruno WH, Célia GR, Rajiv G. Artificial intelligence applied to computer forensics. Hawaii: Association for Computing Machinery 2009.
33. Bulbule SS, Sutaone MS, Vyas V. Component-based face recognition using CNN for forensic application. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India 2019;1-7. DOI: 10.1109/ICCCNT45670.2019.8944841.
34. Calderara S, Prati P, Cucchiara R. Video surveillance and multimedia forensics: an application to trajectory analysis. In Proceedings of the First ACM workshop on Multimedia in forensics (MiFor '09). Association for Computing Machinery, New York, NY, USA 2009;13-18. DOI: https://doi.org/10.1145/1631081.1631085
35. Carlos ER, Josephine L, Adam D, Ziming Z, Gail-Joon A. Mutated policies: towards proactive attribute-based defenses for access control. In Proceedings of the 2017 Workshop on Moving Target Defense (MTD '17). Association for Computing Machinery, New York, NY, USA 2017;39-49. DOI: https://doi.org/10.1145/3140549.3140553
36. Carter JM. Locating executable fragments with Concordia, a scalable, semantics-based architecture. In Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop (CSIIRW '13). Association for Computing Machinery, New York, NY, USA 2013;24:1-4. DOI: https://doi.org/10.1145/2459976.2460004
37. Carvalho LE, Von Wangenheim A. A 3D object recognition and classification: a systematic literature review. Pattern Analysis Application 2019;22:1243-1292. DOI: https://doi.org/10.1007/s10044-019-00804-4
38. Chatfield K, Simonyan K, Zisserman A. Return of the Devil in the details: delving deep into convolutional nets. British Machine Vision Conference.
39. Chen B, Luo W, Li H. Audio Steganalysis with Convolutional Neural Network. In Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security (IHMM Sec '17). Association for Computing Machinery, New York, NY, USA 2017;85-90. DOI: https://doi.org/10.1145/3082031.3083234
40. Chen Q. File fragment classification using grayscale image conversion and deep learning in digital forensics. IEEE Security and Privacy Workshops (SPW), San Francisco, CA 2018;140-147. DOI: 10.1109/SPW.2018.00029.
41. Chhabra GS, Singh VP, Singh M. Cyber forensics framework for big data analytics in IoT environment using machine learning. Multimedia Tools Appl. DOI: https://doi.org/10.1007/s11042-018-6338-1
42. Choi J, Lee H, Lee H, Suh Y. Color laser printer forensics with noise texture analysis. In Proceedings of the 12th ACM workshop on Multimedia and Security (MMSec '10). Association for Computing Machinery, New York, NY, USA 2010;19-24. DOI: https://doi.org/10.1145/1854229.1854235
43. Choudhary P, Nain N. A four-tier annotated urdu handwritten text image dataset for multidisciplinary research on urdu script. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 15 2016;4(26):23. DOI: https://doi.org/10.1145/2857053
44. Chou J, Lin S, Cheng C. On the effectiveness of using state-of-the-art machine learning techniques to launch cryptographic distinguishing attacks. In Proceedings of the 5th ACM workshop on Security and artificial intelligence (AISec '12). Association for Computing Machinery, New York, NY, USA 2012;105-110. DOI: https://doi.org/10.1145/2381896.2381912
45. Christian S, Alexander T, Dumitru E. (n.d.). Deep neural networks for object detection. Google, Inc 2012.
46. Chunhui F, Zhengquan X, Wenting Z, Yanyan X. Automatic location of frame deletion point for digital video forensics. In Proceedings of the 2nd ACM workshop on Information Hiding and Multimedia Security (IHMM Sec '14). Association for Computing Machinery, New York, NY, USA 2014;171-179. DOI: https://doi.org/10.1145/2600918.2600923
47. Collomosse J, Bui T, Brown A, Sheridan J, Green A, Bell M, *et al.* Archangel: Trusted archives of digital public documents. In Proceedings of the ACM Symposium on Document Engineering 2018 (DocEng '18). Association for Computing Machinery, New York, NY, USA 2018;31:1-4. DOI: https://doi.org/10.1145/3209280.3229120
48. Costantini S, De Gasperis G, Olivieri R. Digital forensics and investigations meet artificial intelligence. Annals of Mathematics and Artificial Intelligence 2019;86:193-229 DOI: https://doi.org/10.1007/s10472-019-09632-y
49. Cozzolino D, Verdoliva L. Noiseprint: A CNN-based camera model fingerprint. IEEE Transactions on Information Forensics and Security 2020;15:144-159, DOI: 10.1109/TIFS.2019.2916364.
50. Cuzzocrea A, Pirrò G. A semantic-web-technology-based framework for supporting knowledge-driven digital forensics. In Proceedings of the 8th International Conference on Management of Digital Eco Systems (MEDES). Association for Computing Machinery, New York, NY, USA 2016;58-66. DOI: https://doi.org/10.1145/3012071.3012099
51. Dai Q, Hoiem D. Learning to localize detected objects. Urbana-Champaign: University of Illinois 2011.
52. Davide A, Giorgio G, Fabio R. Machine learning in computer forensics (and the lessons learned from machine learning in computer security). Cagliari, Italy: University of Cagliari 2011.
53. Davis ML. Mobile crime scene applications: an evaluation of their use and future direction. Marshall University 2013.
54. Derek B, Francine F, Ewa H, Oscar B. Computer forensics - past, present and future. Journal of Information Science and Technology 2008;44-59.
55. Dhanalakshmi R, Chellappan C. An intelligent technique to detect file formats and e-mail spam. In Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India (A2CWiC '10). Association for Computing Machinery, New York, NY,

USA 2010;53:1-6. DOI: https://doi.org/10.1145/1858378.1858431
56. Divya M, Shailesh J, Devika J, Sreesha N. Credit card fraud detection using neural networks. International Journal of Students Research in Technology & Management 2014;2(02):84-88.
57. Everingham M, Van~Gool L, Williams CKI, Winn J, Zisserman A. Pascal Visual Object Challenge (VOC) 2007. Results. Available at http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.
58. Everingham M, Gool LV, Williams CK, Winn J, Zisserman A. The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision 2010;88(2):303-338.
59. Felix A, David L, Nhien-An LK, Mark S. Evaluating automated facial age estimation techniques for digital forensics. Dublin, Ireland: Forensics and Security Research Group, School of Computer Science University College Dublin, Ireland 2018.
60. Forenzika. Crime Scene Reconstruction. Retrieved from For enzika 2011. http://forenzika.unist.hr/Portals/6/docs/studenti/Crime%20Scene%20Reconstruction.pdf
61. Francesco M, Geraldina P, Carlo S, Luisa V. Counter-forensics in machine learning based forgery detection. Proceedings of SPIE - The International Society for Optical Engineering 2015.
62. GFI. Top 20 free digital forensic investigation tools for sys Admins 2019. Update. https://techtalk.gfi.com
63. Gloe T, Böhme R. The Dresden Image Database for benchmarking digital image forensics. In Proceedings of the ACM Symposium on Applied Computing (SAC' 10). Association for Computing Machinery, New York, NY, USA 2010, 1584-1590. DOI: https://doi.org/10.1145/1774088.1774427
64. Girshick RB, Felzenszwalb PF, McAllester D. (n.d.). Discriminatively trained deformable part models, release 5. Retrieved from 2010. http://people.cs.uchicago.edu/ rbg/latent-release5/.
65. Grajeda C, Breitinger F, Baggili I. Availability of datasets for digital forensics – And what is missing. Proceedings of the Seventeenth Annual 2017. DFRWS USA. DOI:10.1016/j.diin.2017.06.004
66. Guang H, Jonathan G, Vrizlynn LLT. Time-spread echo-based audio watermarking with optimized imperceptibility and robustness. IEEE/ACM Trans. Audio, Speech and Lang. Proc 2015;23(2):227-239. DOI: https://doi.org/10.1109/TASLP.2014.2387385
67. Guo C. The Future of Version Control" Tom Preston-Werner talks about the early days of Github. Retrieved from Medium 2015. https//:medium.com
68. Hannah Snyder. Literature review as a research methodology: an overview and guidelines. Journal of Business Research, 104, November 2019, 333-339.
69. Helen F. Machine learning applied to object recognition in robot search and rescue systems. University of Oxford 2009.
70. Homem I. Coriander: A toolset for generating realistic android digital evidence datasets. Kista, Sweden: Department of Computer and Systems Sciences, Stockholm University 2017.
71. Homem I, Papapetrou P, Dosis S. Entropy based prediction of network protocols in the forensic analysis of DNS tunnels. World Congress on Internet Security (World CIS). IEEE 2016.
72. Homem I. Towards automation in digital investigations: seeking efficiency in digital forensics in mobile and cloud environments. Sweden: Stockholm University 2016.
73. Hong-Wei N, Viet DN, Vassilios V, Stefan W. Deep learning for emotion recognition on small datasets using transfer learning. Singapore: University of Illinois at Urbana-Champaign 2015.
74. Hoo-Chang S, Holger R, Mingchen G, L, Ziyue X, Isabella N, Ronald M, S. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. Institute of Electrical and Electronic Engineers 2016.
75. Huo Y, Zhu X. High dynamic range image forensics using CNN. ArXiv preprint, arXiv:1902.10938, arxiv.org 2019.
76. IBM Foundational Data Science Methodology. IBM Whitepaper. https://tdwi.org/~/media/64511A895D86457E964174EDC5C4C7B1.PDF
77. Ikuesan RA, Shukor AR, Mazleena S, Hein S. Leveraging human thinking style for user attribution in digital forensic process. International Journal on Advanced Science Engineering Information Technology 2017;7(1):198-206.
78. Infosec. Seven (7) best computer forensics tools 2020 https://resources.infosecinstitute.com/
79. Irvin H, Panagiotis P. Harnessing predictive models for assisting network forensic investigations of DNS tunnels. Kista, Sweden: Department of Computer and Systems Sciences, Stockholm University 2017
80. JW Howarth HB. Feature-based object recognition. Palmerston north, new Zealand 2009
81. Jatto AA. Implementation of preliminary health screening software for university health centers. Lokoja: Department of Computer Science, Federal University Lokoja 2017.
82. Joel LN, Alex AF, Celso AK. Automatic text summarization using a machine learning approach. Brazil: Pontifical Catholic University of Parana (PUCPR) 2014.
83. Joseph R, Ali F. YOLO9000: better, faster, stronger 2017.
84. Joseph R, Santosh D, Ross G, Ali F. You only look once: unified, real-time object detection 2017.
85. Kitchenham B. Procedures for performing systematic reviews. Joint Technical Report TR/SE-0401 2004.
86. Levett CP, Jhumka A, Anand SS. Towards event ordering in digital forensics. In Proceedings of the 12th ACM workshop on Multimedia and security (MM & amp; Sec' 10). Association for Computing Machinery, New York, NY, USA 2010;35-42. DOI: https://doi.org/10.1145/1854229.1854238
87. Liu Y, Peng Y, Hu D, Li D, Lim K, Ling N. Image retrieval using CNN and low-level feature fusion for crime scene investigation image database. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA 2018, 1208-1214, DOI: 10.23919/APSIPA.2018.8659471

88. Liu Q, Sung AH., & Qiao, M. (2011). Neighboring joint density-based JPEG steganalysis. ACM Transactions on Intelligent Systems and Technology 2018;2(2):16. DOI: https://doi.org/10.1145/1899412.1899420
89. Long Z, Yuanhao C, Alan Y, William F. Latent hierarchical structural learning for object detection. Computer Vision and Pattern Recognition 2010.
90. Loia V, Mattiucci M, Senatore S, Veniero M. Computer crime investigation by means of fuzzy semantic maps. In Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 03 (WI-IAT '09). IEEE Computer Society, USA 2009;183-186. DOI: https://doi.org/10.1109/WI-IAT.2009.258
91. Luciano L, Baggili L, Topor M, Casey P, Breitinger F. Digital forensics in the next five years. In Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES 2018). Association for Computing Machinery, New York, NY, USA 2018;46:1-14. DOI: https://doi.org/10.1145/3230833.3232813
92. Maiya AS, Thompson JP, Loaiza-Lemos F, Rolfe RM. Exploratory analysis of highly heterogeneous document collections. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13). Association for Computing Machinery, New York, NY, USA 2013, 1375-1383. DOI: https://doi.org/10.1145/2487575.2488195
93. Malmsten CF. Evolution of version control systems. Gothenburg: University of Gothenburg 2010.
94. Matthew E, Awais R, Paul R. A Systematic survey of online data mining technology intended for law enforcement. ACM Computing Survey 2015;48(1):15, 1-54. DOI: https://doi.org/10.1145/2811403
95. McLaughlin N, Rincon JM, Kang B, Yerima S, Miller P, Sezer S, *et al.* Deep android malware detection. In Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy (CODASPY'17). Association for Computing Machinery, New York, NY, USA 2017, 301-308. DOI: https://doi.org/10.1145/3029806.3029823
96. Merkle LD. Automated network forensics. In Proceedings of the 10th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '08). Association for Computing Machinery, New York, NY, USA 2008, 1929-1932. DOI: https://doi.org/10.1145/1388969.1389001
97. Michalas A, Murray R. Mem Tri: A memory forensics triage tool using bayesian network and volatility. In Proceedings of the 2017 International Workshop on Managing Insider Security Threats (MIST '17). Association for Computing Machinery, New York, NY, USA 2017, 57-66. DOI: https://doi.org/10.1145/3139923.3139926
98. Michiel VD, Claudia H. Large-scale author verification: temporal and topical influences. In Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14). Association for Computing Machinery, New York, NY, USA 2014, 1039-1042. DOI: https://doi.org/10.1145/2600428.2609504
99. Mohammed H, Clarke N, Li F. Evidence identification in heterogeneous data using clustering. In Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES 2018). Association for Computing Machinery, New York, NY, USA 2018;35:1-8. DOI: https://doi.org/10.1145/3230833.3233271
100. Moreira DC, Fechine JM. A machine learning-based forensic discriminator of pornographic and bikini images. International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro8, 1-8, DOI: 10.1109/IJCNN.2018.8489100.
101. Muhammad F, Nor BA, Ainuddin WA, Shahaboddin S, Anthony TC. Source camera identification: a distributed computing approach using hadoop. Journal of Cloud Computing: Advances, Systems and Applications 2017, 6-18.
102. Natércia AB, Michele AB, Michele BP, Daniel HD, Mirella MM. Dealing with data from multiple web sources. In Proceedings of the 24th Brazilian Symposium on Multimedia and the Web (Web Media'18). Association for Computing Machinery, New York, NY, USA 2018, 3-6. DOI: https://doi.org/10.1145/3243082.3264609
103. Nickson Kariea M, Victor Kebande R, Venterc HS. Diverging deep learning cognitive computing techniques into cyber forensics. Forensic Science International: Synergy, 1,. Elsevier 2019, 61-67.
104. Nguyen HH, Tieu TN, Nguyen-Son H, Nozick V, Yamagishi J, Echizen I. Modular convolutional neural network for discriminating between computer-generated images and photographic images. In Proceedings of the 13th International Conference on Availability 2018.
105. Noblett M, Pollitt MM, Presley LA. Recovering and examining computer forensic evidence. Forensic Science Communications 2000.
106. Olga R, Jia Deng, Hao S, Jonathan K, Sanjeev S, Sean M, Zhiheng H, *et al.* ImageNet large scale visual recognition challenge. International Journal of Computer Vision 2015.
107. Ossama AH, Abdel-rahman M, Hui J, Li D, Gerald P, Dong Y. Convolutional neural networks for speech recognition. IEEE/ACM Transactions On Audio, Speech, And Language Processing 2014;22(10):1533-1543
108. Paulo A, Donald CD, Ivens P. The use of machine learning algorithms in recommender systems: a systematic review. Expert Systems with Applications. Expert Systems with Applications 2015, 97.
109. Pedro FF, Ross BG, David M, Deva R. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence 2010;32(9):1627-1645.
110. Petrik R, Arik B, Smith JM. Towards architecture and OS-independent malware detection via memory forensics. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18). Association for Computing Machinery, New York, NY, USA 2018, 2267-2269. DOI: https://doi.org/10.1145/3243734.3278527
111. Platzer C, Stuetz M, Lindorfer M. Skin sheriff: a machine learning solution for detecting explicit images. In Proceedings of the 2nd international workshop on Security and forensics in communication systems (SFCS' 14). Association for Computing

Machinery, New York, NY, USA 2014, 45-56. DOI: https://doi.org/10.1145/2598918.2598920
112. Pollit M. A history of digital forensics. International Conference on Digital Forensics 2010, 3-15.
113. Popov O, Bergman J, Valassi C. A framework for a forensically sound harvesting the dark web. In Proceedings of the Central European Cybersecurity Conference 2018 (CECC 2018). Association for Computing Machinery, New York, NY, USA 2018, 13:1-7. DOI: https://doi.org/10.1145/3277570.3277584
114. Prasad DK. Object detection in real images. Singapore 2010.
115. Priyanka K, Prashant R. An enhanced approach for digital forensics using an innovative GSP algorithm. International Journal of Computer Applications 2014;103(6):18-22.
116. Qasymphony. Agile methodology: the complete guide to understanding agile testing. Qasymphony 2018. com/blog/agile-methodology-guide-agile-testing/
117. Qian Y, Dong J, Wang W, Tan T. Learning and transferring representations for image steganalysis using convolutional neural networks. IEEE International Conference on Image Processing (ICIP), Phoenix, AZ 2016, 2752-2756, DOI: 10.1109/ICIP.2016.7532860
118. Rayson L, Evair S, Luiz A, Luiz S, Gabriel RG, William RS, *et al.* A robust real-time automatic license plate recognition based on the YOLO detector. Belo Horizonte, MG, Brazil: Department of Computer Science, Federal University of Minas Gerais
119. Recommender Systems: A Systematic Review. Waterloo, Canada: University of Waterloo.
120. Reliability and Security (ARES 2018). Association for Computing Machinery, New York, NY, USA, 1, 1–10. DOI: https://doi.org/10.1145/3230833.3230863
121. Ren, X., Guo, H., Li, S., Wang, S., Li, J. (2017). A novel image classification method with CNN-XGBoost model. In: Kraetzer C., Shi YQ., Dittmann J., Kim H. (eds) Digital Forensics and Watermarking. IWDW 2017. Lecture Notes in Computer Science, 10431. Springer, Cham
122. Saikia, S., Fidalgo, E., Alegre, E., & Fernández-Robles, L. (2017). Object detection for crime scene evidence analysis using deep learning. In: Battiato S., Gallo G., Schettini R., Stanco F. (eds) Image Analysis and Processing - ICIAP 2017. ICIAP 2017. Lecture Notes in Computer Science, 10485. Springer, Cham
123. Sameer, V.U., Naskar, R., Musthyala, N., Kokkalla, K. (2017). Deep learning based counter–forensic image classification for camera model identification. In: Kraetzer C., Shi YQ., Dittmann J., Kim H. (eds) Digital Forensics and Watermarking. IWDW 2017. Lecture Notes in Computer Science, vol 10431. Springer, Cham
124. Samir KB, Nabanita B. Interpretation of bloodstain pattern for reconstruction of crime scene. International Research Journal of Computer Science (IRJCS) 2015;2(2):18-25
125. Sayakkara A, Le-Khac N, Scanlon M. Electromagnetic side-channel attacks: potential for progressing hindered digital forensic analysis. In Companion Proceedings for the ISSTA/ECOOP 2018 Workshops (ISSTA '18). Association for Computing Machinery, New York, NY, USA 2018, 138-143. DOI: https://doi.org/10.1145/3236454.3236512
126. Sharma S, Shanmugasundaram K, Ramasamy SK. FAREC — CNN based efficient face recognition technique using Dlib. International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram 2016, 192-195, doi: 10.1109/ICACCCT.2016.7831628.
127. Shen A, Brostow G, Cipolla R. Towards automatic blood spatter analysis in crime scenes. United Kingdom: University of Cambridge 2009.
128. Shen S, Jiang H, Tongda Z. Stock market forecasting using machine learning algorithms California 2013.
129. Shi YQ, Chen C, Chen W. A natural image model approach to splicing detection. In Proceedings of the 9th workshop on Multimedia & security (MM Sec '07). Association for Computing Machinery, New York, NY, USA 2007, 51-62. DOI: https://doi.org/10.1145/1288869.1288878
130. Shujun L, Anthony TSH, Zichi W, Xinpeng Z. Lost in the digital wild: Hiding information in digital activities. In Proceedings of the 2nd International Workshop on Multimedia Privacy and Security (MPS '18). Association for Computing Machinery, New York, NY, USA 2018, 27-37. DOI: https://doi.org/10.1145/3267357.3267365
131. Steven HHD, Benjamin CMF, Mourad D. A visualizable evidence-driven approach for authorship attribution. ACM Transitions on Information System Security, 17, 3, 12, 2015, 1-30. DOI: https://doi.org/10.1145/2699910
132. Steinebach M, Ester A, Liu H. Channel steganalysis. In proceedings of the 13th international conference on availability, reliability and security (ARES 2018). Association for Computing Machinery, New York, NY, USA 2018;9:1-8. DOI: https://doi.org/10.1145/3230833.3233266
133. Surajit S, E, F, Enrique A, Laura FR. Object detection for crime scene evidence analysis using deep learning. International Conference on Image Analysis and Processing (ICIAP) LNCS 10485 2017;II:14-24.
134. Tang M, Fidge C. Reconstruction of falsified computer logs for digital forensics investigations. In Proceedings of the Eighth Australasian Conference on Information Security Australian Computer Society, Inc 2010;105(10):12-21.
135. Taranta EM, Mehran T, Maghoumi CR, Pittman La Viola JJ. A rapid prototyping approach to synthetic data generation for improved 2D gesture recognition. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16). Association for Computing Machinery, New York, NY, USA 2016, 873–885. DOI: https://doi.org/10.1145/2984511.2984525
136. Tariq S, Lee S, Kim H, Shin Y, Woo SS. Detecting both machine and human created fake face images in the wild. In Proceedings of the 2nd International Workshop on Multimedia Privacy and Security (MPS'18). Association for Computing Machinery, New York, NY, USA 2018, 81-87. DOI: https://doi.org/10.1145/3267357.3267367
137. Therdphapiyanak J, Piromsopa K. Applying hadoop for log analysis toward distributed IDS. In Proceedings of the 7th International Conference on Ubiquitous

Information Management and Communication (ICUIMC '13). Association for Computing Machinery, New York, NY, USA 2013;3:1-6. DOI: https://doi.org/10.1145/2448556.2448559
138. Tensor Flow. tensorflow. http://www.tensorflow.com 2017.
139. Thurau C, Kersting K, Bauckhage C. Yes we can: simplex volume maximization for descriptive web-scale matrix factorization. In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). Association for Computing Machinery, New York, NY, USA 2010, 1785-1788. DOI: https://doi.org/10.1145/1871437.1871729
140. Universal Class. Reconstructing a crime scene 2018. http://www.universalclass.com/articles/law/reconstructing-a-crime-scene.html
141. Ujjwal K. An intuitive explanation of convolutional neural networks. 2016. https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/
142. Vel Od, Anderson A, Corney M, Mohay G. Mining email content for author identification forensics. Brisbane, Australia 2012.
143. Venture Beat. AI Weekly: Google shifts from mobile first world to AI first world 2017. https://venturebeat.com/2017/05/18/ai-weekly-google-shifts-from-mobile-first-to-ai-first-world/
144. Voronin V, Makov S, Creutzburg R. Digital inpainting with applications to Forensic image processing. Journal of Electronic Imaging 2016.
145. Wang Q, Zhang R. Double JPEG compression forensics based on a convolutional neural network. EURASIP J. on Information Security 2016, 23. https://doi.org/10.1186/s13635-016-0047-y
146. Wang Y, Su Z, Song D. File fragment type identification with convolutional neural networks. In Proceedings of the International Conference on Machine Learning Technologies (ICMLT'18). Association for Computing Machinery, New York, NY, USA 2018, 41-47. DOI: https://doi.org/10.1145/3231884.3231889
147. Wang Y, Rountev A. Who changed you? obfuscator identification for Android. In Proceedings of the 4th International Conference on Mobile Software Engineering and Systems (Mobile Soft' 17). IEEE Press, 2017, 154-164. DOI: https://doi.org/10.1109/MOBILESoft.2017.18
148. Wei LY, Rohana M, Ram GR. An application of case-based reasoning with machine learning for forensic autopsy. Kuala Lumpur, Malaysia: Expert Systems with Applications 2014.
149. Whitcomb CM. An historical perspective of digital evidence: A forensic scientist's view. International Journal of Digital Evidence 2002.
150. Yann L, Koray K, Clement F. Convolutional networks and applications in vision. New York: New York University 2010.
151. Yang B, Li N, Lu Z, Jiang J. Event detection with convolutional neural networks for forensic investigation. In: Shi Z, Vadera S, Li G. (eds) Intelligent Information Processing VIII. IIP. IFIP Advances in Information and Communication Technology, 486. Springer, Cham 2016.
152. Yiyu H, Jongweon K. Art painting identification using convolutional neural networks. International Journal of Applied Engineering Research ISSN 0973-4562 2017;12(4):532-539.
153. Zwanger V, Felix C, Freiling FC. Kernel mode API spectroscopy for incident response and digital forensics. In Proceedings of the 2nd ACM SIGPLAN Program Protection and Reverse Engineering Workshop (PPREW '13). Association for Computing Machinery, New York, NY, USA 2013;3:1-11. DOI: https://doi.org/10.1145/2430553.2430556