BOOK CHAPTER | *"The Deeper You Go The More You Learn"*

Deep Learning (DL) Oriented Forensic Analysis

Herbert Cyril Dodoo
Digital Forensics & Cyber Security Graduate Programme
Department Of Information Systems & Innovations
Ghana Institute of Management & Public Administration
Greenhill, Accra, Ghana
E-mail: herbert8686@yahoo.com
Phone: +233243216312

ABSTRACT

Cyber-attacks are now more prevalent than ever before in all aspects of our daily lives. As a result of this circumstance, both individuals and organizations are fighting cybercrime on a regular basis. Furthermore, today's hackers have advanced a step further and are capable of employing complex cyber-attack strategies, exacerbating the problem. Some of these approaches are minute and undetectable, and they frequently masquerade as genuine requests and directives. To combat this threat, cyber security professionals, as well as digital forensic investigators, are constantly compelled to filter through massive and complicated pools of data, also known as Big Data, in order to uncover Potential Digital Forensic Evidence. that can be used as evidence in court. Potential Digital Evidence can then be used to assist investigators in reaching certain conclusions and/or judgments. The fact that Big Data frequently comes from various sources and has diverse file formats makes cyber forensics even more difficult for investigators. When it comes to the processing of vast amounts of complicated data for forensic purposes, forensic investigators typically have less time and budget to fulfil the rising demands. This paper will be studying how to incorporate Deep Learning cognitive computing approaches into Cyber Forensics

Keywords: Deep Learning, Forensic Analysis, Artificial Intelligence, Online Safety, Evidence

BOOK Chapter | Research Nexus in IT, Law, Cyber Security & Forensics. Open Access. Distributed Free

Citation: Herbert Cyril Dodoo (2022): Deep Learning (DL) Oriented Forensic Analysis
Book Chapter Series on Research Nexus in IT, Law, Cyber Security & Forensics. Pp 320-328
www.isteams.net/ITlawbookchapter2022. dx.doi.org/10.22624/AIMS/CRP-BK3-P51

1. INTRODUCTION

The Digital Forensic Research Workshop (DFRWS) has defined digital forensics as "the application of scientifically derived and proven methods to 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 criminal events, or assisting in the prevention of unauthorized actions that disrupt planned operations."

Deep learning is a machine learning and Artificial Intelligence (AI) technique that mimics how humans acquire knowledge. Data science, which also encompasses statistics and predictive modelling, contains deep learning. Deep learning is highly useful for data scientists who are responsible for gathering, analyzing, and interpreting massive amounts of data; it speeds up and simplifies the process. Deep learning can be regarded as a means to automate predictive analytics at its most basic level. Deep learning algorithms are piled in a hierarchy of increasing complexity and abstraction, whereas typical Machine Learning algorithms are linear.

1.1 Machine & Deep Learning Concept

It's critical to consider how Machine Learning and deep learning approaches might assist in solving digital forensics difficulties, as well as how these methods differ from one another.

Machine Learning

Machine Learning is an Artificial Intelligence strategy that employs a system that can learn on its own through experience.
It is not just utilized for Artificial Intelligence reasons like mimicking human behaviour but also to reduce human effort and time spent on complex and even basic jobs. Machine Learning is a system that learns through examples and experience rather than from programming. Machine Learning is defined as a system that continuously learns and makes decisions based on data rather than programming. Machine Learning is a new technology that is utilized in industry and science to give new functionality to computers. Many autonomous solutions for medical research, robotics, engineering, and other fields are based on Machine Learning.

Deep Learning

Deep learning is a collection of approaches used to create Machine Learning methods for pattern identification, such as picture recognition. First, the system is used to identify the object's edges, structure, and type, followed by the object itself.

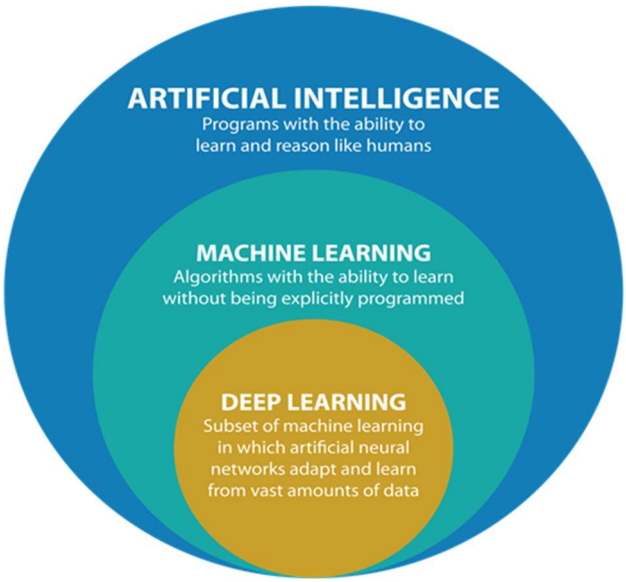

Fig 1: Relationship between Artificial Intelligence Machine Learning and Deep Learning

Source: Stack Exchange

Machine Learning's Importance in Digital Forensic Investigations Machine Learning Forensic is based on Artificial Intelligence and is used to process large amounts of data, analyze it for criminal activity and risk, and segment it to find criminal activity and behaviour. Intelligence systems without an intelligent component cannot perform actual learning capabilities or be true intelligence systems. Digital Forensic Investigation through Machine Learning is the most recent trend in leveraging Artificial Intelligence as a key security solution capability.

Machine Learning Forensics, which are based on Artificial Intelligence, can be used to evaluate large amounts of data in order to identify danger, segment data, and detect illegal behaviour. Machine Learning methods allow investigators to comb through massive amounts of data distributed across social and wired networks, as well as web and cloud computing. Machine Learning algorithms, in essence, are pattern recognition software that is used to analyze large amounts of data in order to predict behaviour. Machine Learning algorithms attempt to learn from past experiences in order to predict future behaviour.

Through Machine Learning algorithms, Machine Learning Forensics gains the ability to discern patterns of criminal activity in order to learn from prior data about when and where the crime will occur. Burglaries, money laundering, and infiltration attacks are examples of hostile behaviours that can be carried out using the extracted data set. This can be accomplished by formalizing and analyzing servers, suspect devices, wireless devices, the Internet, and other types of data for visualization, link association, segmentation, and criminal activity prediction. Cyber threats are becoming increasingly sophisticated, and old security measures are no longer effective.

Attackers have devised more sophisticated methods of attacking the system, which is becoming increasingly complex over time. The system administrator would be unable to identify these attacks on a consistent basis. Human experience and competencies, on the other hand, have some limitations, resulting in a slow rate of incident occurrence, a longer delay in detecting and preventing cyber threats, and the need for more advanced expertise to eliminate these cyber dangers. As a result, developing more advanced Machine Learning models could aid in the prevention and mitigation of cyberattacks. Nowadays, there is a lot of automated software that can assist humans in performing difficult and scientific jobs. These automated tools will need to be more advanced in the next level and should have the ability to do more. Attackers have devised more sophisticated methods of attacking the system, which are becoming increasingly complex over time.

The system administrator would be unable to identify these attacks on a consistent basis. As a result, developing more advanced Machine Learning models could aid in the prevention and mitigation of cyberattacks. Nowadays, there is a lot of automated software that can assist humans in performing difficult and scientific jobs. The next stage is for these automated systems to become more complex, with the potential to use Artificial Intelligence and Machine Learning approaches.

2. RELATED LITERATURE

A number of research works have been carried out by different scholars on the use of Deep Learning in Forensic analysis, the table below shows a review of some of the published works

Table 1: Review of Related Literature.

Paper Title	Journal Name	Author	Purpose of the Study Research	Gaps/Findings	Direction for Future Works
Digital Forensic Analysis of Files Using Deep Learning	*2020 3rd International Conference on Signal Processing and Information Security (ICSPIS)*, 2020	Mohammed Al Neaimi, Hussam Al Hamadi, Chan Yeob Yeun, M. Jamal Zemerly	the existing forensic approaches to identify the file type and developed a new approach based on deep learning and overcome previous approaches' limitations.	The need to run the Deep Learning on an algorithm on supercomputers hence cannot be used in the field	our model could be deployed on a portable device for on-field accessibility Several file types could be added to our dataset pool size to have a wide range of files identified. Possibly a more accurate model, with much more file types to identify
1 **"PEDA 376K A Novel Dataset for Deep-Learning Based Porn-Detectors**	2020 International Joint Conference on Neural Networks (IJCNN)	D. C. Moreira, E. Torres Pereira and M. Alvarez	Using deep learning to determine whether an image is pornographic or non-pornographic to determine not-safe-for-work files/media	The Paper sought to provide a baseline for future experiments in the Deep learning field, there is a lot of room for future improvement	provide a baseline for future experiments using the same data
Enabling Trust in Deep Learning Models: A Digital Forensics Case Study	2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering	A. K., S. Grzonkowski and N. A. Lekhac	design and implement a domain-independent Adversary Testing Framework (ATF) to test security robustness of black-box DNN's.	Current the solutions and theory proposed has not been fuly matched against otherin in its class	extend the Adversary Testing Framework to test the robustness of Adversary Testing Framework algorithms used in other security domains mainly malware classification,

3. DEEP LEARNING AND CYBER SAFETY IN AFRICA

The African Union, does not have an institution that coordinates cyber security efforts throughout the continent. In 2014, the African Union (AU) ratified the Malabo Convention on Cyber Security and Personal Data Protection. The Malabo Convention intends to create a legislative framework for cybersecurity, data protection, and electronic transaction security. It establishes the essential principles for criminalizing and combating cybercrime and related challenges. It has been signed by 15 member states and will enter into force after 15 member states have rectified it. The Convention is silent on the adoption of Deep Learning Tools to combat cybersecurity threat being experienced on the continent. The Government of Ghana has formed the Cyber Security Authority (CSA) to regulate cybersecurity operations in the country, encourage the development of cybersecurity in the country, and provide for relevant matters under the Cybersecurity Act, 2020 (Act 1038).

As a Government Agency under the Ministry of Communications and Digitalisation, the Cyber Security Authority

- Regulate cybersecurity activities in the country;
- Prevent, manage and respond to cybersecurity threats and cybersecurity incidents;
- Regulate owners of Critical Information Infrastructure in respect of cybersecurity activities, cybersecurity service providers and practitioners in the country;
- Promote the development of cybersecurity in the country to ensure a secured and resilient digital ecosystem;
- Establish a platform for cross-sector engagements, on matters of cybersecurity for effective co-ordination and cooperation between key public institutions and the private sector;
- Create awareness of cybersecurity matters; and
- Collaborate with international agencies to promote the cybersecurity of the country

To boost Ghana's regional and international response to cybercrime and improve cybersecurity, the country has accepted the Economic Community of West African States(ECOWAS). Regional Cybersecurity and Cybercrime Strategy & Regional Critical Infrastructure Protection Policy. The International Telecommunication Union (ITU) recognized the Ghana's cybersecurity efforts with a score of 86.69 percent in the latest Global Cybersecurity Index (GCI) - a significant mprovement above prior scores of 32.6 percent and 43.7 percent in 2017 and 2018, respectively.

The Cyber Security Authority (CSA)currently does not employ the use of Deep Learning tools it its Cyber Security safety management and detection The European Union Agency for Network and Information Security (ENISA) is the EU's cybersecurity agency. ENISA's function is strengthened under the European Union Cybersecurity Act. The agency now has a permanent mission and is tasked with assisting the EU in improving operational collaboration and crisis management. It also has increased financial and personnel resources. ENISA however acknowledge the use of Deep Learning tools to help in it cybersecurity battle to ensure a safer Cyber European Union

4. LIMITATIONS AND CHALLENGES

The fact that deep learning models learn through observation is their biggest flaw.
This implies they only have access to the data they used to train. The models will not learn in a generalizable fashion if a user has a little amount of data or if it originates from a single source that is not necessarily representative of the broader functional area.

Biases are also a significant concern for deep learning algorithms. If a model is trained on data with biases, the model's predictions will reflect those biases. Deep learning programmers have struggled with this since models learn to differentiate based on minor differences in data items. The crucial factors it determines are frequently not made plain to the programmer. This means that a facial recognition model, for example, could draw assumptions about people's attributes based on race or gender without the programmer's knowledge.

Deep learning models may potentially face difficulties due to the learning pace. If the rate is too high, the model will converge too soon, yielding an inferior result. If the rate is too low, the process may become stalled, making it much more difficult to find a solution. Deep learning models' hardware needs can also impose restrictions. To ensure enhanced efficiency and reduced time consumption, multicore high-performance graphics processing units (GPUs) and other similar processing units are necessary. However, these devices are costly and consume a lot of energy. Random-Access Memory (RAM) and a hard disk drive (HDD) or RAM-based solid-state drive(SSD). are also required

5. CONCLUSION

Deep Learning when optimized has a big potential to list the man-hours spent in the forensic analysis since the advanced neural algorithm created could process and produce results quickly and accurately in the in a shorter timeframe

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
21. M. A. Neaimi, H. A. Hamadi, C. Y. Yeun and M. J. Zemerly, "Digital Forensic Analysis of Files Using Deep Learning," *2020 3rd International Conference on Signal Processing*

and Information Security (ICSPIS), 2020, pp. 1-4, doi: 10.1109/ICSPIS51252.2020.9340141.
22. S. K. Konaray, A. Toprak, G. M. Pek, H. Akçekoce and D. Kılınç, "Detecting File Types Using Machine Learning Algorithms," *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2019, pp. 1-4, doi: 10.1109/ASYU48272.2019.8946393.
23. D. C. Moreira, E. Torres Pereira and M. Alvarez, "PEDA 376K: A Novel Dataset for Deep-Learning Based Porn-Detectors," *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206701.
24. A. K., S. Grzonkowski and N. A. Lekhac, "Enabling Trust in Deep Learning Models: A Digital Forensics Case Study," *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 2018, pp. 1250-1255, doi: 10.1109/TrustCom/BigDataSE.2018.00172.
25. 2.5 Digital Forensic Research Workshop (DFRWS) Research Road Map, Utica, NY. (2001) http://www.dfrws.org/archive.html
26. https://dfrws.github.io/dfrws2019-EU-workshops/topic-03-Forensic-Intelligence-Workshop/index.html
27. "The EU Cybersecurity Act | Shaping Europe's Digital Future." *Shaping Europe's Digital Future*, digital-strategy.ec.europa.eu, 23 May 2022, https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-act.
28. "Is Machine Learning Required for Deep Learning? - Artificial Intelligence Stack Exchange." *Artificial Intelligence Stack Exchange*, ai.stackexchange.com, 11 Oct. 2019, https://ai.stackexchange.com/questions/15859/is-machine-learning-required-for-deep-learning.
29. "CSA || Home." *CSA || Home*, www.csa.gov.gh, 23 May 2022, https://www.csa.gov.gh/.