# Analyzing Cyber Trends in Online Financial Frauds using digital Forensics Techniques

**Simran Koul, Yash Raj, Simriti Koul**

***Abstract*****:** ***Online financial frauds are one of the leading issues in the fields of digital forensics and cyber-security today. Various online firms have been employing several methodologies for the prevention of finance-related malpractices. This domain of criminal activity is becoming increasingly common in the present cyberspace. In this paper, we will try to implement an online financial fraud investigation using the digital forensics tool: Autopsy. A few existing cyber-security techniques for the investigation of such crimes, namely the Formal Concept Analysis and Confirmatory Factor Analysis; have been analyzed and reviewed. These techniques are primarily based on mathematical cyber-security concepts. Henceforth, it has been tried to find out whether the investigation of similar crimes can be done satisfactorily using the readily-accessible digital forensics tool: Autopsy. Also, it has been explored whether the aforementioned cyber-security techniques can be embedded along with the digital forensics tool to achieve the best results, through training a bot to automatically perform accurate investigations of such crimes. Thus, it has been tried to automate the process of online financial fraud investigation.***

***Keywords*****:** ***Autopsy analysis, Cyber Security techniques, Cyber Crimes, Digital Forensics, Digital Investigation, EnCase analysis, Report Generation, Software Tools, Timeline Analysis.***

## I. INTRODUCTION

A few existing cyber-security techniques for the investigation of such crimes, such as the Formal Concept Analysis, Cross Drive Analysis, and Confirmatory Factor Analysis; have been analyzed and reviewed. These techniques are primarily based on mathematical cyber-security concepts. The main real-time issues that affect the current generation is missing from the studied literary papers and our paper will be novel and up to date with the latest cases and methodologies. Henceforth, it will be tried to find out whether the investigation of similar crimes can be done satisfactorily using the two readily-accessible digital forensics tools: Autopsy and Encase. The major features, as well as limitations of each of the techniques, will be highlighted. Also, it will be explored whether the aforementioned cyber-security techniques can be embedded within the digital forensics concepts to achieve the best results.

The terminologies are:

- Online Frauds and Financial Frauds:

Online frauds refer to the usage of Internet services or other open-source software requiring Internet access to frame users or to otherwise take advantage of them. Finance-related flaws are becoming quite commonplace today. The most common types of online financial frauds include:

Phishing: Here, the fraudsters acquire users' sensitive data such as passwords and credit card credentials through email messages, fraud websites, and phone calls.

Card Skimming: This crime involves the illegal extraction of the user's sensitive financial details on the magnetic stripe from ATMs, debit, and credit cards. This is usually done by the installation of malware on the card reader used by the victim.

SMiShing: It involves the extracting of a user's bank account details through the exchange of text messages over the cell phone.

Vishing: It is the theft of sensitive data using information interchange over the phone; either by messaging, instant messaging, or phone calls, etc.

SIM Swap fraud: Here, the attackers replace the victim's SIM card with a false one; they can somehow install spyware on the original SIM card – which enables them to access all of the user's phone data, including those related to finances.

Identity theft: In this crime, the criminal, using some basic stolen credentials of the victim, such as Date of Birth, phone number, addresses, and credit card numbers, builds up a fake identity posing as the original victim. Internet-based financial crimes steal millions of dollars each year from the victims and continue to terrorize the Internet through various approaches.

- Digital Forensics:

Digital forensics is a subsidiary of the main discipline "Forensic Science". While forensic science includes all studies, techniques, and findings related to various types of crimes in all domains; the field of digital forensics is solely concerned with the investigations involved in the sphere of cyber-crime. It is used for the recovery and analysis of the various computational devices suspected to be involved in the crime; or found at the crime scene.

- Cyber Security:

Cyber-security is a discipline that is primarily aimed at the overall protection of computer systems within an organization. This includes the security of the system software, hardware as well as data stored in the system database. Thus, it protects computational assets from online and cyber-attacks. There are various techniques employed for the fulfillment of the purpose of safety of the cyberspace, such as computer access control, insertion of safety codes, compulsory authentication, encryption, firewall, etc.

**Revised Manuscript Received on June 30, 2020.**
***** Correspondence Author

**Simran Koul***, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Email: simran.koul@yahoo.com

**Yash Raj**, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Email: yashraj.vit01@gmail.com

**Simriti Koul**, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Email: simriti.koul@yahoo.com

*Retrieval Number: I7185079920/2020©BEIESP*
*DOI: 10.35940/ijitee.I7185.079920*

*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

- Formal Concept Analysis:

Formal Concept Analysis is a mathematical cyber-security technique. It works by first structuring the input dataset into a concept lattice, and then the division of these formed lattices into a binary lattice. This binary lattice can be then used to verify which data is fake; and which one is presumably fraud. This technique has already been applied to the field of online financial fraud investigations, with satisfactorily good performances.

- Cross Drive Analysis:

In the Confirmatory Factor Analysis, the primarily quantitative-type data analyzed by the investigators are then checked with the previously-existing similar data records; to "confirm" the consistency of the result interpretation. In case the deviations are very small, this technique is extremely useful to predict the forensic results. This technique has also already been used in the field of online financial crime investigation.

- Autopsy:

Autopsy is a fairly popular digital forensics tool. It serves as a platform as well as a graphical interface for digital-crime investigation purposes. It is a much simpler part of the Sleuth Kit forensics software. It is primarily used by corporate organizations, law enforcement agencies, and to some extent, by the military forces – for online crime inspection. It can be used to extract the past events that occurred on a particular computer system. It provides features such as the creation of disk images to prevent evidence loss, analysis of user activity, analysis of the discovered data on the system, retrieval of the deleted data items, etc. It is supported by most operating system platforms; such as Windows, Ubuntu, Linux, Unix, etc. It can also be run on the cell phone platform of Android; using the specialized "Autopsy: Mobile Forensics" toolkit.

- EnCase:

Encase is the most popular and top-performing software tool used in the field of computer forensics. It is the leading tool used by professional investigator teams, law agencies, corporate CSIRT teams, military intelligence officers, etc.

## II. LITERATURE SURVEY

Matthew Kul and Nick Waler (2017) in the publication "Cyber-security and Fraud Management Convergence" talked about the growing importance of cyber-security techniques and tools, concerning the rapid spurt of growth of cyber frauds, especially financial frauds. They also discussed the various challenges involved the domain of cyber-security, such as legal permissions, unresolved technical issues, etc. [1]

Benjamin E. Onodi, et al (2015) in the paper "The Impact of Forensic Investigative Methods on Corporate Fraud Deterrence in Banks in Nigeria", explored Garfinkel's technique of Cross Drive Analysis for investigation of financial cyber-crimes. For the implementation of this experimentation, data comprising of credit card numbers, email addresses, and other kinds of confidential information was accumulated from various victimized hard drives and other sources; and their correlation with the perpetrator's communication messages, geographical coordinates, etc. was found out. This gave a significantly clear idea about the actual perpetrator. [2]

Nisarg Trivedi and Dhruv Patel (2015) in the paper "Digital Evidence Handling Using Autopsy", discussed the various features provided by the forensic tool: Autopsy. They analyzed the software's efficiency using various test cases. They have also described the functioning of the software for the cases they investigated using it. They concluded that Autopsy was fairly well-performing when it came to conduction of digital investigations; with limited number of issues. [3]

Peter Prudon (2015) in the journal article "Confirmatory Factor Analysis as a Tool in Research Using Questionnaires: A Critique" provided a detailed explanation and criticism related to the usage of the Confirmatory Factor Analysis technique in investigation procedures. He talked about how the methodology of calculating deviation between the predicted results and the previously-known results (relating to similar research cases) could be used to determine the accuracy of a particular prediction. Such a prediction can be used in investigative sciences. Also, the extent of accuracy of results was discussed; which was satisfactory. [4]

Tommie W. Singleton (2006), in the publication "Digital Evidence in a Fraud Investigation" talked about the significance of cyber evidence in various criminal activities, including financial malpractices. He concluded that digital forensics is an increasingly important area in the investigation of several crimes using forensic sciences. According to him, digital evidence should not be neglected as it can give new directions to any legal proceedings. Also, the digital investigation should not be limited to just the victim and accused's computational devices; but also, be extended to other peripherals. [5]

At the Digital Forensics Research Workshop (DRFWS)'s conference proceedings held in 2006 in the USA, Dr.Simson Garfinkel proposed the technique of Cross Drive Analysis for digital crime investigation. In this method, data could be accumulated from various suspect drives (or other sources). The accumulated data was then statistically analyzed and correlation between them was found out. This trained the model to correlate any input data with a particular pre-defined category. [6]Frano Škopljanac-Mačina, et al (2013), in the paper "Formal Concept Analysis – Overview and Applications" discussed this mathematical investigative procedure in detail. In this paper, the working of the technique and its basic principle was briefed. According to the paper, the technique generates "concept lattices" based on the input datasets, in which similar inputs are grouped into one lattice. Thus, the input data can be divided or classified under various label names, which makes it a useful tool for forensic sciences as well. [7] Waziri et al (2014) in the paper "e-Fraud Forensics Investigation Techniques with Formal Concept Analysis" discussed the application of the mathematical cyber security-based technique of using Formal Concept Analysis (FCA) for the binary classification of the input dataset into either genuine or fraud. In this model, The FCA technique was used to analyze the various data gathered from victim as well as suspects' mobile communication devices such as cell phone, tablets etc.

Then, the visualization of the relationship between the crime occurrences within different proximal geographical areas was achieved successfully. This helped to develop a pre-trained model which, when given similar crime-investigation input as well as geographical area, could classify the data as fraud or not. This would greatly help financial firm websites. [8]

Peter Prudon (2015) in the journal article "Confirmatory Factor Analysis as a Tool in Research Using Questionnaires: A Critique" provided a detailed explanation and criticism related to the usage of the Confirmatory Factor Analysis technique in investigation procedures. He talked about how the methodology of calculating deviation between the predicted results and the previously-known results (relating to similar research cases) could be used to determine the accuracy of a particular prediction. Such a prediction can be used in investigative sciences. Also, the extent of accuracy of results was discussed; which was satisfactory. [9]

Dr. Simson Garfinkel (2010) in an article named "Digital forensics research: The next 10 years" discussed about the features provided by the forensic tool: EnCase. He discussed its working, versions, features, limitations, etc. On the whole, he concluded that, as of now, EnCase is one of the topmost available forensics tools, which is heavily reliable and easily accessible for various types of cyber-crime cases. [10]

## III. PROPOSED METHODOLOGY

Most finance-related online crimes are committed by first provoking the user to somehow give out their credentials, such as credit card numbers, passwords, etc. Thus, the crime can be detected in two major steps:

- Extraction of emails and messages found to be provoking the victim to give out his credentials: This can be done on the victim's computer or phone. With the help of this step, the fraud email-id or phone number can be identified, which will ultimately help to track down the location of the criminal.
- Once the crime suspects are identified, their computer's hard drive must be scanned for the victim's credit card information as well as evidence of sending messages to the victim in the first place.

The methodologies involved in the analysis of digital crime are divided into two major steps:

i. Extraction of emails and messages found to be provoking the victim to give out his credentials:

1. First, create a new Autopsy case for the victim's computer seized in the investigation.
2. Then create the image of the victim's computer or phone hard drive.
3. Further, the image is added as the data source for the new case.
4. Perform keyword searches with the terms commonly used by fraudsters to provoke users to give out credentials.
5. Further, got to Outlook.pst >> Email and browse through the retrieved emails.
6. Go to Windows Mail and browse through the emails.
7. By now, the messages sent by the suspect to the victim are retrieved.
8. The time sent and email id is recovered. Sometimes, the cell phone number may also get recovered.
9. This email is used to track down the suspect's IP address, which is then used to track the suspect's location which is in turn used to track the suspect's location whenever he connects to the internet.
10. Finally, even if the suspect uses a VPN, the police will be able to see his activity through the suspect ISP and contact the VPN Company to disclose the suspect location.

The process carried out is illustrated in the flowchart in Fig. 1.

ii. Extraction of evidence (victim's credentials) from the suspect's computer drive:

1. First, create a new Autopsy case for the suspect's computer seized in the investigation.
2. Create an image of the suspect's hard drive.
3. Add this image as the data source for this new case.
4. Perform keyword searches for victim's stolen credentials, such as credit card numbers, passwords, etc.
5. Go to Views >> Deleted Files and browse through the emails.
6. Search for the deleted items in the unallocated disk space.
7. Go to results>> Extracted contents and browse through the web history, cookies, search history and bookmarks.
8. Any evidence found against the suspect must be carefully documented.

The process carried out is illustrated in the flowchart in Fig. 2.

## IV. MERGING CYBER-SECURITY

This section deals with "Embedding of cyber-security techniques with the functioning of Autopsy".

Design of automation (bots) for online finance-based fraud investigation:

AI bots can be trained to utilize Autopsy for email and information retrieval, and then classify the emails as suspicious or not (Formal Concept Analysis). Then, the bots may also display the accuracy of their prediction, based on the source used for email extraction (Confirmatory Factor Analysis). This algorithm devised is as follows:

1. Bot retrieves suspected emails/ messages from the victim's hard drive using the Confirmatory Factor Analysis.
2. Depending upon the location from where the emails were retrieved (Outlook.pst, Windows Mail, etc.), the emails are classified as more suspicious or less suspicious (Supervised learning- Formal Concept Analysis).
3. The most suspicious emails will be used to track the IP address (and hence location) of the suspect for further enquiry).
4. On the suspect's hard drive, the bot searches for the victim's stolen credentials or any records of money transfer; by retrieving the deleted items.
5. Then a match between the victims' stolen credentials and the information retrieved from the suspect's drive is carried out.

Using the Confirmatory Factor Analysis, the bot gives the percentage accuracy with which one can say that the suspect is the real criminal.

The functioning of the proposed bot is explained as follows:

1. Initially, the bot will create a disk image file of the victim's computer hard drive; and feed it to the Autopsy software as the input source file.

2. Then, using the steps mentioned before, the bot will retrieve all the emails which contain the suspected keywords (the emails which provoked the victim to disclose his bank credentials).

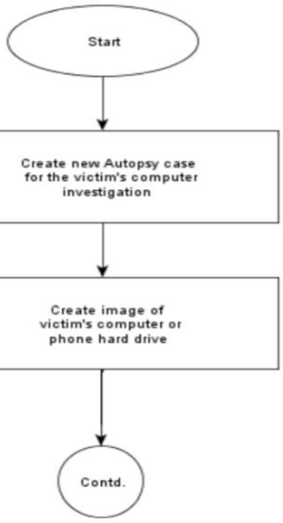
3. These emails are retrieved primarily either from the location "Outlook.pst" or "Windows mail". Based on previous similar cases, the bot has been trained (through supervised learning), to classify which emails are more suspicious and important (based on the number and type of keywords, and the location from where they are retrieved). This binary classification is a cyber-security technique called Formal Concept Analysis.

4. The most suspicious emails are then used to track the sender's IP address. Subsequently, whenever the criminal connects to the internet, his location will be disclosed. Even if he uses a VPN (Virtual Private Network), the ISP (Internet Service Provider) can retrieve which VPN is being used, and the VPN Company can disclose the suspect's location.

5. Then, after the suspect's computer drive is seized, the bot will retrieve all the deleted files, and search the entire system for information related to the victim (such as personal credentials, etc.), and the crime (such as emails relating to theft, etc.). This is done using the Autopsy software.

6. The bot then carries out a similarity test between the information retrieved from the victim as well as suspect's drives. This helps to predict the accuracy of the prediction (as to whether the suspect is the actual criminal or not). The accuracy calculation is done by the bot using the cyber-security technique called as Confirmatory Factor Analysis, in which accuracy is predicted based on past similar cases. The process that is carried out by the bot is illustrated in the flowchart in Fig. 3.

## V. FLOWCHARTS


Start
Create new Autopsy case for the victim's computer investigation
Create image of victim's computer or phone hard drive
Contd.

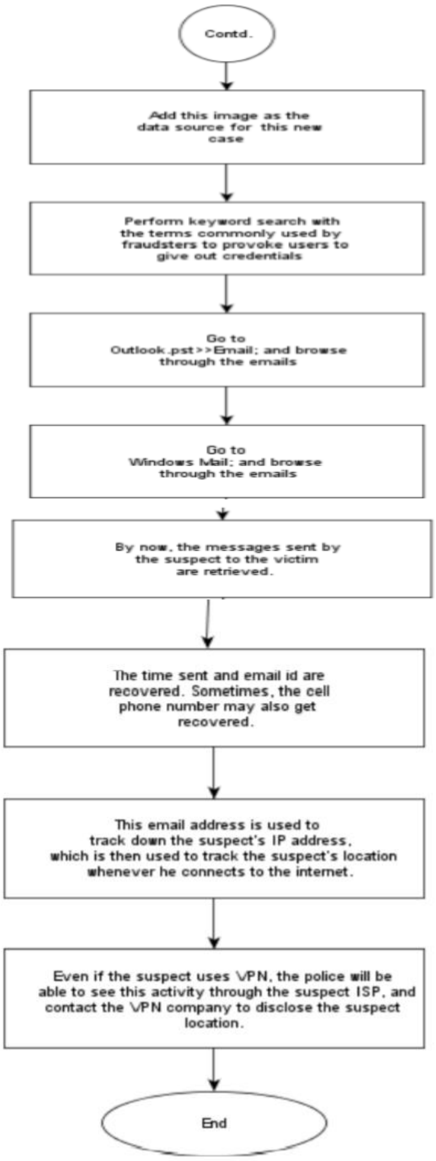
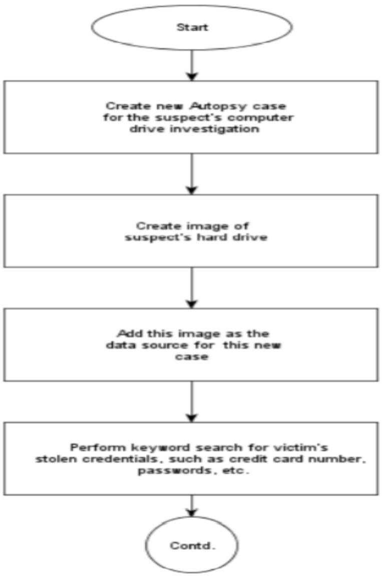

Contd.
Add this image as the data source for this new case
Perform keyword search with the terms commonly used by fraudsters to provoke users to give out credentials
Go to Outlook.pst>>Email; and browse through the emails
Go to Windows Mail; and browse through the emails
By now, the messages sent by the suspect to the victim are retrieved.
The time sent and email id are recovered. Sometimes, the cell phone number may also get recovered.
This email address is used to track down the suspect's IP address, which is then used to track the suspect's location whenever he connects to the internet.
Even if the suspect uses VPN, the police will be able to see this activity through the suspect ISP, and contact the VPN company to disclose the suspect location.
End

**Fig. 1. Extraction of emails and messages found to be provoking the victim to give out his credentials.**


Start
Create new Autopsy case for the suspect's computer drive investigation
Create image of suspect's hard drive
Add this image as the data source for this new case
Perform keyword search for victim's stolen credentials, such as credit card number, passwords, etc.
Contd.

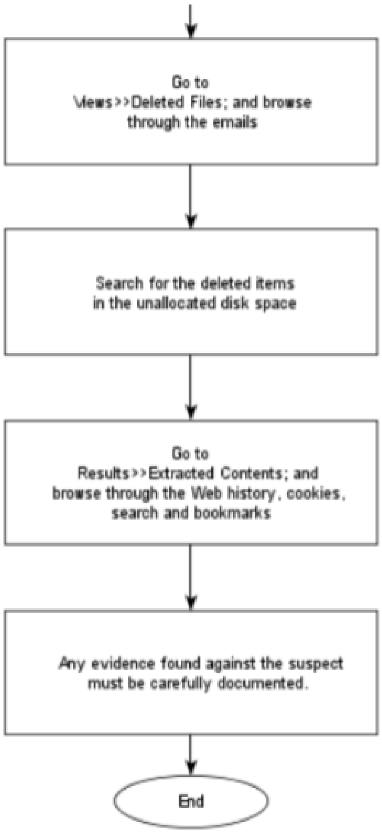

Go to
Views>>Deleted Files; and browse
through the emails
Search for the deleted items
in the unallocated disk space
Go to
Results>>Extracted Contents; and
browse through the Web history, cookies,
search and bookmarks
Any evidence found against the suspect
must be carefully documented.
End

**Fig. 2.** **Extraction of evidence (victim's credentials) from the suspect's computer drive.**

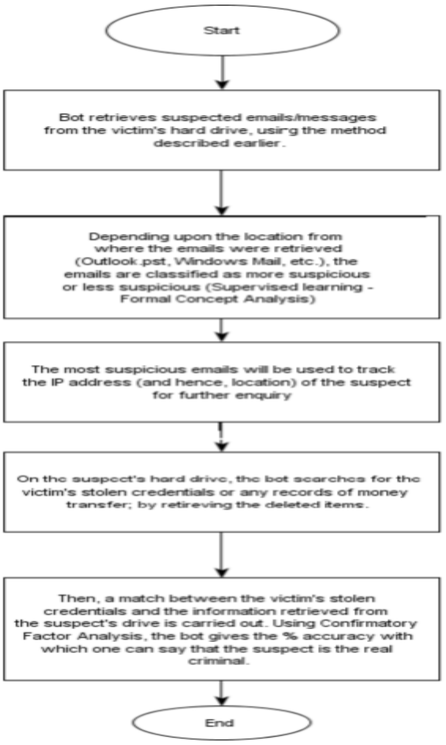

Start
Bot retrieves suspected emails/messages
from the victim's hard drive, usirg the method
described earlier.
Depending upon the location from
where the emails were retrieved
(Outlook.pst, Windows Mail, etc.), the
emails are classified as more suspicious
or less suspicious (Supervised learning -
Formal Concept Analysis)
The most suspicious emails will be used to track
the IP address (and hence, location) of the suspect
for further enquiry
On the suspect's hard drive, the bot searches for the
victim's stolen credentials or any records of money
transfer; by retireving the deleted items.
Then, a match between the victim's stolen
credentials and the information retrieved from
the suspect's drive is carried out. Using Confirmatory
Factor Analysis, the bot gives the % accuracy with
which one can say that the suspect is the real
criminal.
End

**Fig. 3.** **The functioning of the proposed bot is explained in this figure.**

## VI. RESULT AND DISCUSSION

The real problem that threatens millions today is that of cyber fraud and digital malpractices that needs immediate curb. This research paper analyzed how we can do so. First the major two crime detection technique, extraction and identification are pondered upon. All the vital immediate steps to be done for extraction and identification including seizure scan and preservation is formulated. This ensures proper handling of the evidence after the crime is committed and identified.

Next, two methodologies were formulated related to both the victim and the culprit. These included creating a separate copy of the evidence for safekeeping, performing the keyword search using deep neural network and natural language processing by the software used. The initial search for the evidence in suspect's computer, followed by recovery of hard drives with the analytical procedure was formulated next, ensuring the completion of the entire digital forensic process.

Finally, the merging of cyber-security through the usage of bots which would automate the entire process of digital forensic ensured the desired merge that is essential in these times of virtual reality. The functioning of the proposed bot included the additional functionality of tracking the IP address and using the ISP of the suspects and victims seized systems to ensure a well-rounded investigation. This helped to analyze the accuracy of the prediction (as to whether the suspect is the actual criminal or not). The accuracy calculation is done by the bot using the cyber-security technique called as Confirmatory Factor Analysis, in which accuracy is predicted based on past similar cases.

Finally, the merged bot introduced in the digital forensic procedure of analyzing the cyber trends in online financial frauds proved much more efficient and time-saving.

## VII. CONCLUSION

We can conclude that to handle such a large number of finance-related cyber-crimes, AI bots can be trained to predict who has committed the crime. This can be done by embedding the forensics software "Autopsy" within the bot's processor; as well as training the bot (via supervised learning) to classify the emails and predict the accuracy of the results obtained using cyber-security techniques (Formal Concept Analysis and Confirmatory Factor Analysis).

This mechanism will have several advantages. Usage of the digital forensics tool alone does not guarantee the accuracy of the results, and usage of the cyber-security technique alone is a lengthy process. Moreover, the use of bots would save a lot of time and manpower.

The disadvantage of the proposed bot design technique is that it is highly resource-intensive. The Development of AI bots alone requires a lot of technical resources. Providing them further training would incur even greater costs. Plus, as the system is new – it is more prone to glitches, which will be eventually resolved over time.

Thus, if such a bot is successfully designed to investigate the online financial frauds, it would be greatly helpful to the investigating agencies.

## REFERENCES

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
8. Prajval Mohan, Pranav Narayan, Lakshya Sharma, Tejas Jambhale, Simran Koul, "Iterative SARSA: The Modified SARSA Algorithm for Finding the Optimal Path". International Journal of Recent Technology and Engineering (IJRTE). ISSN: 2277-3878, Volume-8 Issue-6, March 2020.
9. Prajval Mohan, Adiksha Sood, Lakshya Sharma, Simran Koul, Simriti Koul, “PC-SWT: A Hybrid Image Fusion Algorithm of Stationary Wavelet Transform and Principal Component Analysis”. International Journal of Engineering and Advanced Technology (IJEAT)’, ISSN: 2249-8958, Volume-9 Issue-5, June 2020.
10. Simran Koul, “Contribution of Artificial Intelligence and Virtual Worlds Towards Development of Super Intelligent AI Agents”

## AUTHORS PROFILE

**Simran Koul** was born in Jammu, India, on 10th November 1998. She completed her senior high school in FAIPS DPS, Ahmadi, Kuwait. She is currently pursuing for her Bachelor’s Degree in Computer Science and Engineering at Vellore Institute of Technology, Vellore, India. She is currently working on projects which involve concepts of Digital Forensics, Robotics, Artificial Intelligence and Natural Language Processing.

**Yash Raj** was born in Tatanagar on 16th November 1997. He completed his Senior High School from DAV Public School, Tatanagar. Yash is currently in his seventh semester with a GPA of 9.22 pursuing his B. Tech in Computer Science and Engineering from Vellore Institute of Technology, Vellore, India. Yash has been the winner of the prominent VIT Hack and has been awarded for the same by the chancellor and board of directors. His areas of interest include MEAN Stack Development, Machine Learning, Artificial Intelligence, Digital Forensics and Database Management. Yash has advanced knowledge of the python programming language and has been certified by Hackerrank. Yash has interned with leading companies like JP Morgan & Co. and Tata Steel Ltd. He is currently working on COVID-19 web portal analysis and forensic cyber-security.

**Simriti Koul** was born in Jammu, India, on 10th November 1998. She completed her senior high school in FAIPS DPS, Ahmadi, Kuwait. She is currently pursuing her Bachelor's Degree in Computer Science Engineering at Vellore Institute of Technology, Vellore, India. She is currently working on projects which involve concepts of Artificial intelligence, Data Analytics, Digital Forensics and Natural Language Processing.