Corrosion Management ISSN: 1355-5243

(https://corrosion-management.com/)

Volume 28, Issue 01–June 2018

SCOPUS

AI AND MACHINE LEARNING IN CYBER THREAT DETECTION

1st Rahul Vadisetty Electrical Engineering)

Wayne State University) Detroit, MI, USA rahulvy91@gmail.com

Abstract:

Cybersecurity has gained a lot of attention in today's security issues due to the increased popularity of the Internet of Things, the quick growth of computer networks, and the multitude of important applications. Because of this, identifying different types of cyberattacks or network anomalies and designing an effective intrusion detection system is becoming more and more crucial for modern security. Artificial intelligence, specifically machine learning techniques, can be used to build an intelligent intrusion detection system based on data. To achieve this, we provide in this work an artificial intelligence, Tree machine-learning-based security model, which, after first determining the importance, builds a tree-based generalized artificial intelligence model based on the chosen essential features. Nowadays, the most important problem is cyber threats in machine learning and artificial intelligence. If cyber threats are not detected in time it can cause severe damage to the organization's reputation and revenue and whole networking system in machine learning and artificial intelligence. The demand and need for Cyber security and protection is increasing to deal with the different types of cyber-attacks. To deal with and solve the problems associated with cyber-threats, different organizations, and institutes have implemented several cyber-threat detections.

Index Words: Cyber Security, machine learning, cyber-attacks, Cyber-threat intelligence, Artificial intelligence

1. Introduction:

A. Background:

"Cyber" word refers to the networks that have structured information systems. Cybersecurity deals with the protection of security, integrity, and confidentiality of data provided by organizations, institutions, and individuals. Cyber security ensures the security of data on cyber networks [1]. Ignoring cyber security can lead to high threats like someone with false intentions who can steal user sensitive information like credit card details, which can cause high damage to user trust in an organization as the main purpose of cyber security is to protect the data on the Internet.

In recent years, cybersecurity and its protection have been increasing. Major reasons for cyber-attacks are the Internet of Things, the unbelievable growth of computer systems, and a large number of applications used for either personal or office work by individuals. Cybersecurity has different types of cyber-attacks like denial-of-service attacks, unauthorized access, and computer malware. These types of attacks can often cause lasting damages and financial losses to organizations. For example, in 2017, a ransomware virus led many organizations to huge financial losses of 8 billion dollars. A data breach led the organization to a financial loss of 3.9 million USD. That's why organizations focus on cyber security and protection.

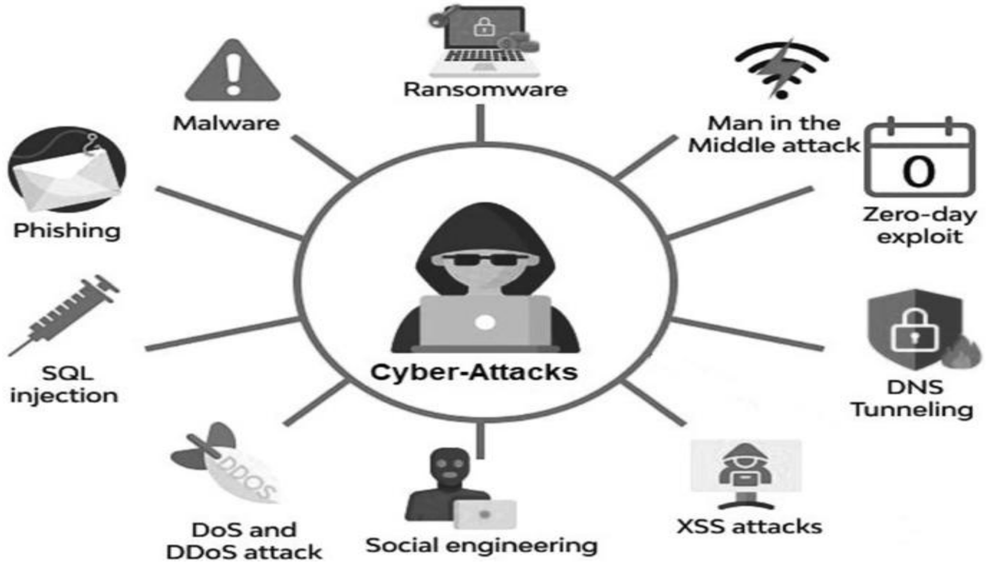

Figure 1Types of Cyber-Attacks [2]

In the developing era of AI systems cyber security and data security is the main problem. Cybercrimes depend on three factors which are threats which refer to who the attacker is, vulnerabilities which mean the weakness of the system at which they are attacking, and impacts which relate to the consequences of the attack. If the attacker succeeds in his mission, the company has to compromise on its integrity confidentiality, and availability. A company can suffer many losses from cybercrime such as their image in the market can be damaged [3]

B. Problem Statement:

Organizations have to face many challenges and problems while implementing cyber threat detection like the developing threat landscape. It means that cybercriminals constantly change their techniques and procedures to hack the data which means organizations should update their cyber security defenses to prevent threats like phishing attacks, and data leakage. Organizations are now adopting new technologies like IoT devices and cloud computing which lead to exponential increase of attack surface for cybercriminals. An increase in attack surface makes it more difficult for organizations to identify and monitor threats before any damage is done. Organizations face a major problem of unskilled staff that creates a high challenge for organizations to implement threat detection. Due to unskilled staff, organizations are not able to detect threats and it causes them high disadvantages like loss of data [4].

Budget constraint is another problem for organizations that don't have enough resources for cyber security. As a result, many organizations find it difficult to invest in cybersecurity threats and response solutions. Cybersecurity problems become more difficult due to the advancement of technology like Artificial intelligence and Machine learning even if Artificial intelligence has its advantages it also has disadvantages like artificial intelligence (AI) driven attacks which help the attackers to plan by using Artificial intelligence and machine learning [5]

While using AI systems, the main problem that companies face is data loss. Data loss is an issue due to which the company may have to face very serious consequences. Suppose there is a company, that has its customer's data stored which is important and it is deleted due to any mistake due like accidental deletion or cyber-attack, then the company has to face many issues such as company gets fined, the reputation of the company is damaged and the customer's trust is also lost in the company [6].

2. Literature Review:

A. Introduction:

In this chapter, we have discussed the solutions to related problems and how we can solve the problems by following these solutions. Our goal is to discuss related and modern approaches and their drawbacks to create the theoretical background.

1. Advancements in Artificial Intelligence:

Artificial intelligence has the ability that enable computers to implement a variety of functions like understanding, written language, and analyzing data. In recent years, AI has been the center of attraction due to its innovative benefits like automation which means that AI can perform work independently. For example, AI can help to identify cyber threats by monitoring and observing network traffic. It also has the advantage of reducing human errors, which means that AI can remove manual errors in data processing and other tasks by automation and algorithms. AI is also used to eliminate repetitive tasks which enables humans to work on higher-impact problems. AI believes that once computers have advanced enough algorithms, they will be able to replicate and augment the human mind. Most companies in the world are actively developing different artificial intelligence strategies [7]

AI can easily verify documents, and answer customer questions. AI can process information more quickly than humans. It also provides us with solutions to every problem without wasting any time. The major benefit of AI is that is available every time. We can use AI applications anytime. One of the main reasons to use AI systems for cyber threat detection is that it is always working on their assigned tasks so they can identify threats easily [8].

2. Cyber- Crimes:

Computing Technology and the Internet have a positive impact and brought ease in our lives but they also lead us to some problems such as cyber-crimes like fraud and theft through information technology. As this technology evolves, the number and variety of cybercrime are also increasing as it provides an easy way for criminals [9]. Following are the different cybercrime attacks:

3. Malware attacks:

A malware attack refers to a computer system that is infected with some type of virus. Cybercriminals can use this virus for many purposes like stealing confidential data, and damage to data. For example, a Malware attack was committed in 2017 by the WannaCry ransomware attack which is used by Cybercriminals to take money from victims by holding their data. 230000 computers from 150 countries were affected by the WannaCry ransomware attack which caused a financial loss of 4 billion.

4. Phishing:

In phishing attacks, spam emails, or links are sent to users to hack their data. Phishing messages contain infected attachments and links that ask the user to add their confidential information. For example, in 2018, spam emails were sent to football fans during the World Cup about free trips to Moscow. Many people opened the link and entered their data which caused their data stolen.

5. Distributed DoS attacks:

Cybercriminals use these types of attacks to bring down the system. IoT devices are often lead to DDoS attacks. For example, in 2017, a DDoS attack happened on the UK national lottery website. It prevents the user from playing offline on the website.

6. Role of Machine learning in cyber threat detection:

Many questions arise about the relationship between ML, DL, and Artificial intelligence (AI). Artificial intelligence helps human intelligence to extend by studying and developing techniques and applications. Artificial intelligence is not human intelligence but from the viewpoint of thinking it also exceeds human intelligence. Machine learning is a subfield of Artificial intelligence that also focuses on prediction-making by using computers. Machine learning techniques are used to perform baseline behavioral profiles for different entities. Deep learning is an innovation in machine learning research. It establishes neural networks and mimics the human brain process to interpret data like images [10].

Machine learning is one of the innovative methods for cybercrime detection. Different techniques of machine learning are used to identify the restrictions and consequences faced by cyber security detection methods. Machine learning techniques are created to learn from data without being programmed. Machine learning techniques are developing in different fields of life like education, medicine, and cybersecurity. Machine learning techniques are used by both attacker and defender sides. Machine learning techniques are used by attackers to pass the defense wall while defenders use these techniques to create defense strategies. Machine learning techniques play an important role in cyber threat detection. Machine learning techniques are skilled in detecting new attacks. The main threat to computer resources is spam messages. Machine learning techniques are used to detect messages as spam.

B. Role of Artificial Intelligence in Cyber Threat Detection:

Data Security has become more important for society as we are dependent on technology. The complexity of cyber threats has increased with the increase of the digital age. Artificial intelligence is used to increase data security. AI also improves security outcomes by detecting anomalies and automating responses to threats before any damage is done. AI has both its strengths and weakness in detecting cyber threats which means that AI excels respond to security threats before wasting any time but it can also be the cause of disruption sometimes AI can determine something as a threat but it might not always be true which leads to disrupt business operations.

C. **Research Objectives**:

The research aims to:

- Discuss the approaches of AI and machine learning in cyber threat detection
- Discuss the usefulness of AI and machine learning in the context of cyber threat detection.

- Suggest consideration for using Artificial intelligence machine study in cyberspace threat detection.

The objectives are mainly aligned to discuss AI and machine learning in the context of cyber threat detection [11]. The objective of the research is to provide techniques and implementation of AI and machine learning in the context of cyber threat detection. The analysis outlines the framework for the practice of AI and machine learning in cyberspace threating remark awareness.

D. Research questions:

The research seeks to answer the following questions:

1. How are AI and machine learning technologies used in cyber threat detection?
2. In what way does this technology contribute to cyber threat detection?
3. What are the strengths and weaknesses of AI and machine learning in the context of cyber threat security detection?
4. Why these technologies are more useful for cyber threat detection?
5. How these techniques can be integrated into a framework of cyber threat detection?

These questions will help to achieve the productive outcomes of the research. These questions will help determine the current situation of machine study and artificial intelligence in cyberspace threat detection. The research will identify the potential strengths and weaknesses of AI and machine learning in cyberspace threat detection.

E. Contribution of the study

The research is crucial for several reasons:

I. **Privacy issues:** The research covers various privacy concerns in AI and machine learning to promote effective cyber threat detection. Privacy concern is the primary concern of the research.

II. **Knowledge contribution:** A study of existing literature provides valuable insights and understanding of AI and machine learning with cyber threat detection.

III. **Security improvement:** Findings of the study assist researchers in improving the security concerns of AI and machine learning technologies for effective implementation in cyber threat detection. These can be used in industries, healthcare, finance and government.

The contribution of this work is centered on expanding existing understanding on AI and machine learning in cyber threat detection. The study examines the role of AI and machine learning in cyber threat identification. This will aid in the development of better and more advanced AI and machine learning technologies that improve the quality of cyber threat detection.

F. Structure of the paper:

The paper is organized as follows:

Chapter 1: Introduction- This will discuss the core concepts of AI, machine learning, and cyber threat detection.

Chapter 2: Literature review - A detailed examination of existing AI and machine learning technologies.

Chapter 3: Methodology- Methods and approaches to evaluate the technologies of AI and machine learning.

Chapter 4: Results and Discussion- An analysis of the findings and their implications.

Chapter 5: Conclusion and future work- A summary of the research and future framework use of technologies.

G. Summary:

This chapter covers a detailed explanation of the existing literature related to artificial intelligence and machine learning. The research covers topics like artificial intelligence and machine learning benefits and strengths and weaknesses in cyber threat detection. The research addresses the weaknesses and provides valuable solutions and techniques to overcome the problems. It will also discuss the importance of artificial intelligence and unit study in cyber threat awareness.

3. Methodology

A. Introduction:

This chapter provides a detailed methodological approach to an effective understanding of AI and machine learning. It provides a research method that measures the efficiency of AI and machine learning in the context of cyber threat detection. The approaches for collecting data help to analyze the effectiveness of collected data. The chapter also highlights the reliability and validity of collected data. It also considers various topics that can be linked to the ethical considerations of the study. Ethical consideration provides valuable insights into the presence of ethical norms during the study. This will assist in the perfect evaluation of AI and machine learning technologies in cyber threat detection.

B. Research Design:

The research conducts both qualitative and quantitative methods to evaluate the effectiveness of Al and machine learning in cyber threat detection. The research design always rules out from research questions. Both studies provide an efficient outlook of the assessment in the cyber threat detection context.

C. Qualitative analysis:

The study conducts qualitative analysis for the analysis of text to get a deeper understanding of the AI and machine learning technologies with context to cyber threat detection. Textual analysis such as a literature review assists the researcher in studying previous studies of the author to examine the conclusion and results of their finding [12]. The analysis of existing studies helps the research to study the cyber threat issues related to artificial intelligence and machine learning. The study focuses on examining key patterns and words in the previous studies for better understanding. The research used Google Scholar and authentic websites from secondary sources. The articles and websites used in the research are selected based on title and abstract. The researcher also reviewed the reference list of these articles to study in cyberspace AI and machine study threat detection.

D. Quantitative Method:

The research uses qualitative methods to examine a deep understanding of AI and machine research in cyberspace threat detection. Experiments were conducted which are also referred to as experimentalism to analyze various AI and machine learning qualities [13]. These assessments may include the validity and reliability of the data techniques conceivably, and the versatility and confidentiality of the artificial intelligence and machine learning in cyber threat detection. Therefore, the quantitative analysis aims at providing effective solutions for effective cyber threat detection implementation through AI and machine learning.

E. Data collection methods:

Gathering of data is a critical process in any form of research. Data collection method based on the research questions. Collection of data was gathered through primary and secondary sources for effective outcomes.

1. Literature review:

Data gathered from previous literature reviews assist the research in gaining an understanding of AI and machine learning in cyber threat detection from existing sources. These sources are gathered from academic journals, peer reviews and research papers [14]. White papers and company reports are used to evaluate the measures and current trends of AI and machine learning in cyber threat detection.

2. Expert interviews:

Effective interviews were conducted with people who are experts in the field of AI and machine learning integration in cyber threat detection. These interviews provide effective insights into the strengths and weaknesses of AI and machine learning in cyber threat detection [15]. Professionals from academics, industries, finance, and government are included in these interviews.

3. Experimental data:

Experimental data was applied to gain statistical analysis of the collected data. This will help to produce better outcomes of the result.

F. Data analysis techniques:

Data analysis techniques are used to identify and examine the effectiveness of collected data whether it is useful or not. Qualitative and Quantitative Methods will be analyzed to ensure effectiveness.

1. Qualitative analysis:

The research examines the data for the completion of the research. Qualitative analysis is used as a secondary source in the research for the data collection. For the analysis of qualitative data thematic analysis was used to analyze the data in the research [16]. The author used some of these methods to analyze the data. Narrative analysis was also used to convert the interviews case studies and experimental setup into narrations [17]. Data analysis techniques enable the researcher in the proper and effective narration of the research. This helps the researcher to narrate the data for the research.

2: Quantitative analysis:

Quantitative data was obtained from statistical results and this is used to analyze data. With the help of graphs and charts, these techniques are effectively analyzed for better results. Statistical analysis makes a wise assessment of the experimental setup [18].

F. **Experimental setup:**

Experimental setup is the quantitative technique used in the research. Various topics like applying techniques of AI and machine learning and identification of techniques in artificial intelligence and machine learning in cyber threat detection assist the researcher in effectively deriving results of the experimental setup.

G. **Evaluation metrics:**

The efficiency and strength of artificial intelligence and tool research are assessed using some of the factors:

1. Accuracy: The most important factor that will determine the accuracy of forecasting and current trends of AI and machine learning in cyber threat detection.

2. Efficiency: Efficiency in this context refers to time management and inspection of technologies.

3. Scalability: Scalability is used to solve the problems of the technologies.

4. Reliability and validity: Reliability and validity are also the major concerns of the research. For this concern, tests are conducted multiple times and if they generate the same value every time, this indication shows that they are reliable and valid [19].

5. Confusion matrix:

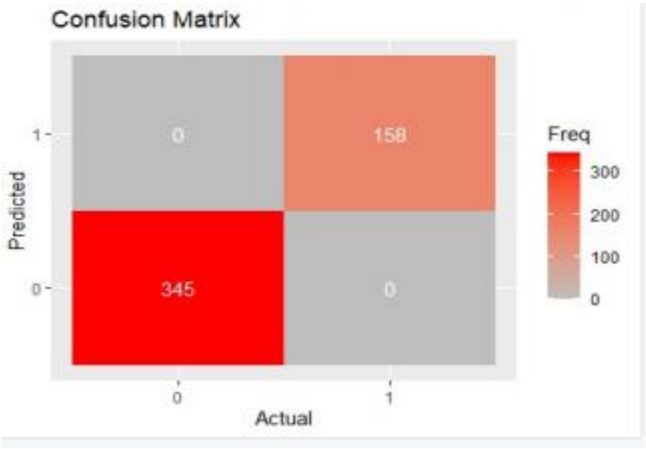

Figure 2. Confusion MATRIX

The study effectively demonstrates AI and machine learning in cyber threat detection. The results of the confusion matrix and ROC curves AI and machine learning in cyber threat detection. During the experiment, the accuracy and efficiency of the model were significantly high. Data protection can process huge amounts of data faster than data quality and analyse the detection of cyber threats. Additionally, they can detect vulnerabilities and predict threats with unprecedented accuracy.

6. ROC Curves:

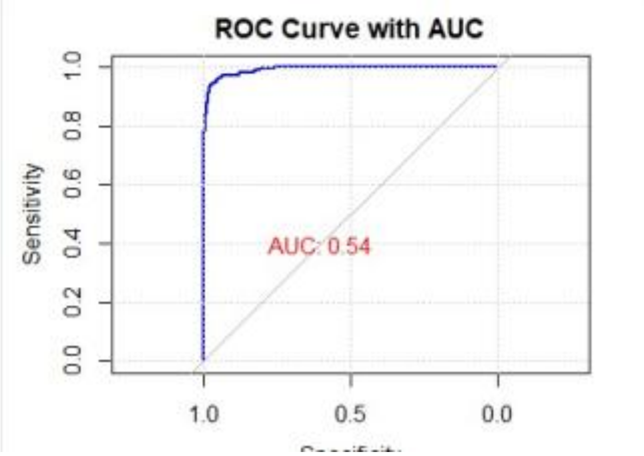

Figure 3. ROC Curve.

The ROC curve presented in Figure 3 shows the actual positive rate between AI and machine learning in cyber threat detection. The true positive rate is plotted as a function of the false positive rate for different cut-off points. A test with perfect discrimination has a ROC curve that passes through the upper left corner; the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

H. Ethical Considerations:

1. Privacy: Data gathered from sources will be safe and protected as data protection is the main concern of the research. If any disruption happens in the data protection, the researcher should inform it in the research. Data will only be used for the information purposes. It should not be considered for unethical purposes [20]. When a survey is conducted, the participants that are involved in the data collection are assured and promised to keep their data private and their privacy will not be disturbed during the whole scenario. The privacy of the participants will not be threatened and the privacy of the data collection should not be affected.

2. Confidentiality: The researcher should provide information regarding the confidentiality of the data Protective measures are adopted by the researcher to maintain the confidentiality of the article's data [21].

3. Conflict: The researcher should avoid those factors that can create conflicts in the research. Research obtained from the existing literature will be done smoothly and efficiently to achieve productive outcomes. The study will be conducted effectively by considering the factors to avoid conflicts in the research.

4. Consent:

Participants are voluntary and free to respond to the study at any time without any hurry or obstacles. Participants were informed that their privacy would not be harmed. The participants were first fully informed of the information and data that was provided.

5. Free

Another important ethical consideration was that the participants who were involved in the survey were free to respond. The people who participated in the survey were not pressured and restricted. They were free and independent to respond to the survey.

6. Proper Citation (Credit):

In this paper, proper citation is given to give credit to the authors. Proper citation is necessary to avoid any ethical issues in the completion of the research. Proper information regarding previous authors assists the researcher in effectively conducting the research.

I. Conclusion:

The article provides a valuable understanding of AI and machine learning in cyberspace threat awareness. This study demonstrates that results conduct useful information that can be used by

Organizations for effective implications for efficient combination of AI and machine learning in cyber threat detection. It uses mixed methodology and analyzes the data through thematic and statistical analysis. Therefore, the research is useful for effective Integration of AI and machine learning for cyber threat detection.

IV. Results and discussion:

A. Introduction:

In this chapter, the author assesses the aims that the researcher tends to find. In this chapter, it was highlighted what has been done, how it is done, and how it is useful in the future. It demonstrates the efforts used in the research for validation of the data. This chapter contains the results of the findings conducted in the article. It provides valuable insights to describe what the study has achieved in this article. The results and discussion chapter of the study focuses on the experiment and simulation of the collected data. This section presents the results of the research and details related to AI and machine learning in cyber threat detection.

B. Data processing

The type of data used in the research is taken from the Kaggle dataset. In the Kaggle dataset, the data obtained is interpreted in the ROC curve graph and confusion matrix. The data set obtained from the data processing data will be further categorized into training and test data. This data set will help to perfectly evaluate the performance of the findings of AI and machine learning in cyber threat detection.

C. Data set description:

The research used data for the analysis is the Kaggle dataset which is a classic dataset in anomaly detection The Kaggle dataset contains extensive material related to AI and machine learning in cyber threat detection. This dataset centers on various components to study AI and machine learning in cyber threat detection. The data set has been taken by Kaggle and then there is a description stated in the graphs of the ROC curve graph and confusion matrix as shown in figs and these graphs are interpreted.

D. Data normalization:

Data normalization is used to improve the learning and training of all the images in the research of AI and machine research in cyberspace threat detection. The normalization step is crucial because it scales and arranges all the values of the article. In the research, the normalization of data is one of the crucial steps before processing collected data from the database. It aimed to ensure the contribution of training of the model equally. The range is used to manage the measurement of variables on different scales within the Kaggle dataset.

D. Standardization:

All the steps used in the model of the Kaggle dataset, are for the accuracy and consistency of the research. These methods enhance the convergence of the model and make it faster than before. All the unit features are scaled for standardization.

1. Results

The research focuses on the level of accuracy of the model which is used to test the model that was the Kaggle dataset model. The test results show AI and machine learning in cyber threat detection. It can be seen that preparation and testing accuracy directly affect the effectiveness of the model. This reflects how anomaly helps in the detection of cyber threats within AI and machine learning. The results also show that the limitations of the model warrant expectations for the dataset used in the review.

2. Discussion:

These results help the author to evaluate Artificial Intelligence and machine learning in cyberspace threating mark detection. These results also support the author to solve problems like privacy by using the results. The results achieved with accurate formation enable the author to complete the research reliably and validly.

V. Conclusion and Future Work:

In this paper, the data set was used to provide effective results and that data set was taken from the Kagal model dataset. The results provide valuable methods and models to consider in the data quality assurance AI and machine learning in cyber threat detection.

A. Conclusion:

The article provides a valuable understanding of AI and machine study in cyber threat detection. The research used the Kaggle dataset to conclude effective results. The study demonstrates that results conduct useful information that can be used by the organization for effective implications of AI and machine learning in cyber threat detection. The training test and confusion model provides effective information about the classes that help in operating more accurate results. The researcher used a mixed method for better understanding that generates productive outcomes. The main findings of the results provide effective solutions to the problem statement. Overall, the research was useful and its implications can be used in organizations and workplaces shortly. The research evaluates AI and machine learning in cyber threat detection. The research focuses on learning AI and machines with AI using Kaggle datasets. The research shows that artificial consciousness plays an important role in AI and machine learning in cyber threat detection. ROC curves and matrixes show AI and machine learning in cyber threat detection. AI Systems are effective methods of studying abnormal behaviour within the network system.

B. Future work:

The study was conducted to analyze AI and machine learning technologies in cyber threat detection. The research conducts data very useful but it can be more informative in the research if it conducts several values. More than one value conduction can derive more effective results. The addition of more models adds more productiveness to the results. The research doesn't use real-world data sets as these are not easy to use but they operate more accurate information on the findings. Real base data sets allow the viewing of participants involved in the data with the differential association. In the future, more accurate and extensive information will provide more comprehensive results about the function of AIand machine research in cyberspace threating mark detection.

Bibliography

[1] R. v. Solms and J. v. Niekerk, "From information security to cyber security," *Science direct*, vol. 38, pp. 97-102, 2013.

[2] A. Bendovschi, "Cyber-Attacks – Trends, Patterns and Security Countermeasures," *Procedia Economics and Finance*, vol. 28, pp. 24-31, 2015.

[3] E. A. Fischer, "Cybersecurity Issues and Challenges: In Brief," *Congressional Research Service*, 12 august 2016.

[4] G. N. Reddy and G. U. Reddy, "A Study Of Cyber Security Challenges And Its Emerging Trends On Latest Technologies," *arXiv*, p. 5, 8 February 2014.

[5] A. Sinha, T. H. Nguyen, D. Kar, M. Brown, M. Tambe and A. X. Jiang, "From physical security to cybersecurity," *Journal of Cybersecurity*, vol. 1, no. 1, pp. 19-35, 2015.

[6] Namasudra, "Cloud computing: A new era," *Journal of Fundamental and Applied Sciences*, vol. 10, 29 may 2017.

[7] S. Jahanzaib and A. Tarique, "Artificial Intelligence and its Role in Near Future," *JOURNAL OF LATEX CLASS FILES*, vol. 14, 2015.

[8] N. Wirkuttis and H. Klein, "Artificial Intelligence in Cybersecurity," *Artificial_Intelligence_in_Cybersecurity-libre.pdf*, vol. 1, january 2017.

[9] D. Selma, C. Huseyin and A. Mustafa, "APPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNIQUES TO COMBATING CYBER CRIMES: A REVIEW," *International Journal of Artificial Intelligence & Applications (IJAIA)*, vol. 6, 2015.

[10] Y. XIN, L. KONG, Z. LIU, Y. CHEN, Y. LI, H. ZHU, M. GAO, H. HOU and C. WANG, "Machine Learning and Deep Learning Methods for Cybersecurity," *IEEExplore*, vol. 6, 2017.

[11] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEE Explore*, vol. 18, no. 2, pp. 1153-1176, 2015.

[12] C. Belsey, "Textual analysis as a research method," *Research methods for English studies*, vol. 2, pp. 160-178, 2013.

[13] Chatterji, Aaron, Findley, Michael, Jense, N. M, Meier, Stephan and D. Nielson, "Field experiments in strategy research," *Econstar*, vol. 37, no. 1, pp. 16-132, 2016.

[14] A. Denney and R. Tewksbury, "How to write a literature review," *Tandfo*, vol. 24, no. 2, pp. 218-234, 2013.

[15] M. Muskat, D. A. Blackman and B. Muskat, "Mixed methods: Combining expert interviews, cross-impact analysis and scenario development," *The Electronic Journal of Business Research Methods*, vol. 10, no. 1, pp. 09-12, 2012.

[16] V. Clarke and V. Braun, "Thematic analysis," *Tandfo*, vol. 12, no. 3, pp. 297-298, 2017.

[17] L. McAlpine, "Why might you use narrative methodology? A story about narrative," *Eesti Haridusteaduste Ajakiri. Estonian Journal of Education*, vol. 4, no. 1, pp. 32-57, 2016.

[18] Z. Ali and B. Bala, "Basic statistical tools in research and data analysis," *Indian journal of anaesthesia*, vol. 60, no. 9, pp. 662-669, 2016.

[19] M. M. Mohamad, N. L. Sulaiman, L. C. Sern and K. M. Salleh, "Measuring the validity and reliability of research instruments," *Science direct*, vol. 204, pp. 164-171, 2015.

[20] M. Abid, N. Iynkaran, X. Yong and H. Guang, "Protection of big data privacy," *IEEE Explore*, vol. 4, pp. 1821-1834, 2016.

[21] P. Ethicist, "Simplifying the complexity of confidentiality in research," *SAGE Publications*, vol. 10, no. 1, pp. 100-102, 2015.