Cyber Security Threat Intelligence using Data Mining Techniques and Artificial Intelligence

Shivangi Gupta, A. Sai Sabitha, Ritu Punhani

Abstract: ***Threat intelligence is the procurement of evidence-based knowledge about current or potential threats. The interest of threat intelligence comprises of advancement in efficiency and boosting effectiveness in terms of analytical and prevention capabilities. Cybersecurity represents serious interest for numerous organizations because maximum of them are using Internet-connected data devices which are opening doors for cyber attackers. Outstanding threat intelligence within the cyber sphere requests for the knowledge base of threat information and a thoughtful way to represent this knowledge. This study proposes a clear rationale of significant artificial intelligence (AI) techniques used for recognizing a cyber-attack. Data analysis can be formulated to guide industries and Internet-connected systems such as smartphones or robotic factories on what to do in the appearance of an incident. AI techniques will analyze past incidents and summarize knowledge from experts and will continue to adapt or reform new branches as it reviews from the new incidents. In addition, various data mining approaches used in boosting threat truthfulness in cybersecurity data are also studied. To conclude, we discussed that; AI will robotize the collation of machine-readable external threats and will improve the efficiency and accuracy of the data for each smart organization's specific framework.***

Keywords **:** ***Artificial Intelligence, Cyber Security, Data Security, Intrusion Detection, Internet of Things (IoT), Threat Intelligence.***

I. INTRODUCTION

In today's digital world, every organization is connected to different technologies, working cultures, and processes. This benefits cyber attackers to stroll freely in the working environment. Every organization is likely to be encounter by the stream of attackers and intruders. They target both big organizations as well as small organizations in the public and private sectors. Therefore, unveiling cyber attackers and threat conditions requires cybersecurity defensive tools, processes, and algorithms. Cyber Security refers to the set of algorithms and techniques used to preserve the integrity of nodes, network, and data from damage, attacks and illegal access. Large organizations have petabytes of data including most important and sensitive information, hence it becomes important to protect the data from malicious access and threats. Cyber Security defenses are carried in concern for common attacks (manipulating known vulnerabilities), advanced attacks (abusing complex vulnerabilities) and emerging attacks (new attacks vulnerabilities). The information system has entered into various aspects of the organization such as production, operation, and management departments [3]. These advancements require the need for reliable information security techniques and systems. Cyber Security defenses include employee awareness, identification of intrusions as quickly as possible and analyzing unexpected emerging threat conditions that have never occurred before in the system. Cyber Security demands to reduce false positive and false negative vulnerability conditions.

In spite of various prevention techniques such as installing antivirus, encryption, and firewalls, organizations are still experiencing intrusions at an alarming rate. The most troubled companies that require threat detection are commercial banks, credit card holders, telecommunication sector and many more [8]. Some of the common threats are:

- The disruptive use of information technology services by attackers to fulfill their ideological agenda.
- The unauthorized access of secret information without the permission of the owner.
- The cyberattack leads to compromise with the essential data of communication technology, medical services, and interrupt the e-commerce environment to a large extent.

Since ample amount of data is being generated so there is a need to develop threat intelligence system to identify the pattern of crime or anonymous activities [20]. The cyberattack often comes with a surprise that is why there is a need to recognize patterns of cybercrime to adopt needed patterns posed by newly originating patterns. Cyber Security risks can be reduced in an organization by detecting a cyber threat at its first place.

The main aim of Cyber Security is to safeguard system and data from malicious cyber threats. Cyber Threats takes place in different areas (Refer Fig. 1) and in different forms such as viruses, malware, and information foraging and applications outbreaks. The boost of cybersecurity attacks over the past few years has come up with the requirement for automated threat analysis at each level of the organization or enterprise [18]. Cybersecurity represents significant interest for various organizations because maximum of them are using data devices which are opening doors for cyber attackers. Cyber-attacks are becoming worldly and more organized, so the government has settled to devote more money and time in identifying methodologies to restrain different threats [7]. The financial market is a crucial component of our country's economy.

Revised Manuscript Received on September 25, 2019.
* Correspondence Author
Shivangi Gupta*, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh. Email: shivangi.gpt35@gmail.com
A. Sai Sabitha, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh. Email: assabitha@amity.edu
Ritu Punhani, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh. Email: rpunhani@amity.edu

Guarding financial threats using improved algorithms and correct analysis can help to protect financial crisis and predicting financial trends.

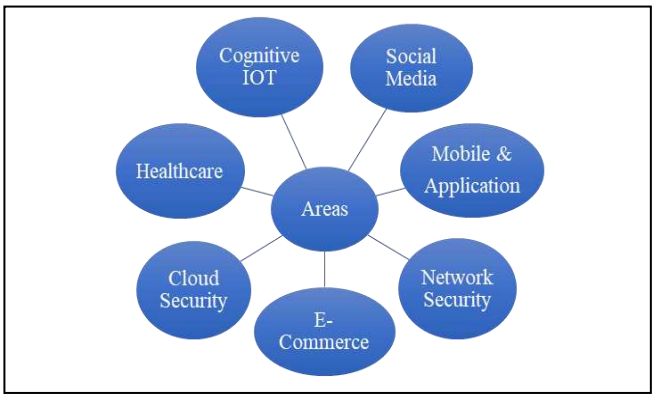

Fig. 1 Research Areas of Cyber Security

Existing techniques for analyzing threat conditions are not sufficient to be implemented on a large scale. Feature extraction and feature selection are the fundamental steps in preprocessing data in order to identify threats using data mining techniques in the area of artificial intelligence, but when it comes to the study of heterogeneous data derived from diverse sources, these tasks appear to be time-consuming and challenging [9]. The biggest issue is big data that is generating structured and unstructured data at an immense rate and creating severe privacy problems [1]. The organization's setup various sensor detection tools and techniques for intrusion detection and network monitoring but handling diverse datasets is a challenging issue. Therefore, the security organizations need to develop a system which can accumulate and correlate diverse datasets that make as a source to help in a longer period.

The objective of this literature review on cybersecurity threat intelligence is as follows:

1) To examine the signs and artificial intelligence techniques identified to determine threat intelligence.
2) To determine the threat truthfulness of abundant amount of data accumulated.
3) To study the patterns of threat detection among cybersecurity issues using statistical analysis.

Based on the first three objectives RQ1, RQ2, RQ3 are discussed.

4) To study the data security challenges faced by industries and measure to overcome those challenges.
5) To analyze the ethical concerns raised by AI algorithms and assessment methods and models developed for the betterment of internet-connected systems.

RQ4, RQ5 and RQ 6 are discussed based on point 4 and 5.

Traditional techniques to cybersecurity were not that efficient. Although if the threats were found, the time it took way too long to recognize [22]. Some attackers modified their approaches to attack but, few modes of operations are usually the same. This regular behavior allows analytical techniques to identify malicious events. Analytics is the discipline of processing huge amount of data with the benefit of mining algorithms and computing devices to generate outcome within a considerable amount of time. Data mining techniques have many operations in security such as analyzing threat to infrastructure services, power grids, auditing and intrusion detection in data storage. Various techniques and algorithms have been investigated to find out malicious users and unusual patterns such as classification, clustering, prediction, fuzzy logic, artificial neural networks, support vector machine and genetic algorithms [15]. Clustering is used to understand patterns and statistical analysis of data based on the collection of similar attributes. Classification can be used to aggregate different cyber attacks and then use the likeliness to detect the threat when it occurs. Prediction can be used to detect hidden future attacks based on the information learned from user behavior.

II. STEMATIC REVIEW

This Before presenting the Systematic Review (SR), Six Research Questions (RQ) were enclosed associated with cybersecurity threat intelligence issues. Analysis in deficiency of existing data mining techniques and artificial intelligence algorithms used to resolve threat issues were considered and studied. The solutions to the RQ framed were inscribed by the review. The questions which do not discuss the inscribed solution in SR are proposed in addition to further research. The research questions in this analysis are examined as follows:

RQ1: Threat Intelligence approaches are very crucial for detecting anomalous actions in the network system of the organizations.

What are the signs and leading data mining approaches and artificial intelligence techniques used for determining threat intelligence?

RQ2: There is an abundant amount of data that exist, but understanding threat truthfulness is nearly unattainable from the data that is accumulated.

How do data mining techniques benefit in boosting threat truthfulness of the cybersecurity data that is accumulated?

RQ3: Case study analysis and certain questions were framed in order to find the threat condition based on discussed data mining techniques.

What are the custom patterns of threat detection across numerous cybersecurity issues using data mining techniques?

RQ4: Industries are going through major transformations by providing Internet-connected systems (ICS), smart homes, and robotic factories. However, maintaining the security of these massive and enormous systems that connect large physical environment is challenging.

What are the disparate and data-driven system challenges experienced by ICS and different threat intelligence schemes modeled to overcome the challenges?

RQ5: Artificial intelligence is creating a major scope in diversified fields and providing advantages to society. It includes different algorithms like natural language processing, knowledge learning, and reasoning which are incorporated to put intelligence into the system [21]. But, ethics for writing apt code should be maintained so that no future disaster occurs.

What are the ethical aspects and role of artificial intelligence to ensure that no future calamities occur?

RQ6: Despite data mining techniques accuracy, artificial intelligence algorithms are still evolving and expanding. For betterment and reliability of the internet-connected systems, today researchers are more intended to train machines to perform threat intelligence tasks rather than humans.

What are the various security assessment methods and models developed using artificial intelligence and its future scope?

III. METHODOLOGY OF SEARCH STRATEGIES FOR EXPLORATORY STUDIES

Significant research papers from various data sources were collected by following specified searching strategies. Boolean OR is used to construct search strings from the search terms with similar meanings. Boolean AND is used to concatenate the search terms and restriction of research. Minor search strings as mentioned in the keywords section were used by keeping research question in mind.

A. Information Collection

A Total 200 of papers were extracted during search operation using search strategies as mentioned above. The papers that did not meet the search criteria were excluded for review. A total of 51 most relevant papers and along with their references were selected for review in this paper as given in Table 1. To conduct the latest researches on cybersecurity using data mining and artificial intelligence techniques papers prior to the year 1980 have not taken for study due to limited research findings.

Table1 Research paper collection from research databases

S. no	Source of Databases	No. of search results retrieved	No. of duplicates found	Number of relevant research papers found
1.	IEEE Explore	55	20	24
2.	ACM digital library	35	8	12
3.	Springer link	20	7	09
4.	Other Journals	90	22	06
5.	Total	200	47	60

IV. RESULT AND DISCUSSION

RQ1: What are the signs and leading data mining approaches and artificial intelligence techniques used for determining threat intelligence?

Research Question 1 (RQ1) of the systematic review was addressed to identify various existing data mining techniques and artificial intelligence techniques which have been used in minimizing cyber-attacks. The Part of approaches recently in existence and was implemented to control malicious activities are addressed in Table 2. *Summary of diverse Data Mining and Artificial Intelligence techniques used in early detection of threat condition.*

Data mining technique such as A/B testing is used for e-commerce websites in which comparison is done to identify the changes in the network. Association technique is used in a recommender system, in monitoring logs and user profiling in an e-commerce environment. It is used in identifying the similarity between elements and variables of a large dataset [12]. Association rules are also used in retaining confidentiality, for eg. ARM approach for safe comparison. Naïve Bayesian classification technique was used to predict attack type to help forensic investigator. Genetic algorithms are using natural selection system and are influenced by the human evolution process. They are used in the intrusion detection system and uses genetic programming engine to recognize combination functions [6].

Machine learning algorithms take insight from the data and predict behavior based on known features analyzed from the training dataset. For analyzing information security risk (ISR) in the database management system, a model using the knowledge and fuzzy logic were used for the evaluation of threat condition using previous audits, thereby predicting better results in the evaluation of ISR [2]. It is inferred that Fuzzy Logic, Artificial Neural Networks, Support Vector Machine, Genetic Algorithms, and K-Means clustering are widely used techniques in anomaly detection systems.

RQ2: ***How do data mining techniques benefit in boosting threat truthfulness of the cybersecurity data that is accumulated?***

Research Question 2 (RQ2) of SR was proposed to identify distinct data mining approaches which have been implemented to enhance threat truthfulness in a substantially huge amount of data. With the advent of big data, data mining (DM) techniques have been extensively used to improve models of cybersecurity applications. The part of different DM techniques used for boosting truthfulness of cybersecurity data is discussed below in Table 3.

Summary of various DM techniques used in boosting truthfulness of cybersecurity data and thereby controlling malicious intrusions.

Data mining approach comprises of classification, clustering, forming association rule, random forest and observing anomalies. These techniques were utilized to frame models for cybersecurity applications like spam filtering, virus disclosure, intrusion detection and malicious behavior [4]. Association rule mining and frequent itemset are two prominent and extensively used data analysis techniques for a range of applications [16]. A novel approach was proposed based on ensemble learning for the advancement of various mining algorithms and controlling cyber attacks [5].

CTI highlights on the creation of multilayered threat intelligence system. Search robot was developed that analyze the false activities of internet resources by self-learning patterns based on the workings of neural networks. The goal of pattern generation technique is to train the system to learn newly generated threats and generate an alert against new processes whether they are safe or not. Information foraging (IF) addresses issues for volume and velocity of data generated by stabilizing human intuition with automation. This addresses that information foraging is helpful in expansion of tools to anticipate cyber intrusion using publically attainable data.

Table 2 Research Work in determining threat using data mining and artificial intelligence techniques.

S.no.	Authors	Year	Application	Techniques Used
1.	Sisiaridis and Markowitch	(2018)	Minimizing Data Complexity in Feature Extraction and Feature Selection for massive datasets.	Machine Learning Approach implemented in Apache Spark using its python API
2.	Basallo et al.	(2018)	Information Security Risk Assesment in Database Management Systems	Using Artificial Intelligence techniques, a model based on Fuzzy logic was developed.
3.	Ghimeş and Patriciu	(2017)	Discovering patterns and malicious activities of users in Big data	Neural Network Models were proposed using machine learning algorithms
4.	Jayasingh et al.	(2016)	The challenges faced by an analyst in fraud detection, network forensics, data privacy issues, and data provenance problems	Naïve Bayesian classification to predict fraud and minimize data privacy issues
5.	Kumar et al.	(2016)	Identification of Cyber Threats in the computing world to enhance revenue generation and cost-cutting	K - Means Clustering technique is used to group similar data relevant to different attributes.
6.	Veeramachaneni et al.	(2016)	Building artificial intelligence solution to recognize and defend against malicious and unseen attacks	Artificial intelligence approach to build an outlier detection system to gather feedback from the system for security analysis

Table 3 Research Work in determining DM techniques used in boosting truthfulness of cybersecurity data and controlling threat in cybersecurity applications.

S.no.	Authors	Year	Application	Techniques Used
1.	Cao and Wang	2017	To increase the robustness of Data Mining models used in Cyber Security applications	Weighted Random Forest and Cluster-Based Random Forest approaches to enhance the robustness of data
2.	Deng et al.	2017	To improve the performance of content filtering in cyber applications to control threat	Content filtering function mining algorithm to boost efficiency in determining threat conditions
3.	Khan et al.	2017	Identification of Denial of Service Attack in the organizations	Pattern Recognition technique in log files using data mining.
4.	Li et al.	2016	Privacy-preserving mining for various outsourced databases that allow sharing of data without negotiating data privacy	Association rule mining and frequent itemset mining techniques were used for developing encryption scheme
5.	Hochbaum and Baumann	2016	The sparse projection was used for massively large datasets which substantially reduced testing time with minimal effect on accuracy	Sparse computation was projected on data mining algorithms such as KNN, SVM, graph-based learning techniques
6.	Ng and Banik 2015	2015	Intrusion detection system to identify unauthorized patterns and security breaches in the system	Anomaly detection and signature database using data mining techniques were used

Weighted Random Forest (WRF) technique and Cluster-based Weighted Random Forest approach used to reform the robustness of random forest [14]. These approaches used bagging to boost uncertainty of the models. In addition, the sparse computation that generates a similarity matrix based on a pair of objects that share the same neighborhood was projected efficiently for massively large datasets. Further, sparsification was performed on the k-nearest neighbor algorithm ,graph-based technique and support vector machine [11].

This technique resulted in the reduction of testing time with minimal effect on accuracy. With the advent of usage of internet, criminal activities have also accelerated; therefore the solution was proposed to detect unauthorized activity based on anomaly detection and database signature using data mining techniques [17]. This was inferred that intrusion detection system helped to recognize security threat in the system.

RQ3: ***What are the custom patterns of threat detection across numerous cybersecurity issues using data mining techniques?***

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Merchant_id	Average_transa	Transaction_a	is_declined	Total_declines_per_day	isForeignTransa	isHighRiskCountry	Daily_chargeback	6_month_avg_(	6-month_chbk	is_fradnt	is_fradulent	
2	3160040998	100	3000	N	5	Y	1	0	0	0	Y	1	
3	3160040998	100	4300	N	5	Y	1	0	0	0	Y	1	
4	3160041896	185.5	4823	Y	5	N	0	0	0	0	Y	1	
5	3160141996	185.5	5008.5	Y	8	N	0	0	0	0	Y	1	
6	3160241992	500	26000	N	0	Y	1	800	677.2	6	Y	1	
7	3160241992	500	27000	N	0	Y	1	800	677.2	6	Y	1	
8	3160272997	262.5	11287.5	N	0	N	0	900	345.5	7	Y	1	
9	3162041996	185.5	11130	Y	20	N	0	0	0	0	Y	1	
10	3162041996	185.5	6121.5	Y	20	N	0	0	0	0	Y	1	
11	3162041996	185.5	7049	Y	20	N	0	0	0	0	Y	1	
12	3356298138	166.788473	4836.865717	N	0	N	0	721	229	9	Y	1	
13	3359162473	444.9970144	21804.85371	N	0	Y	1	0	0	0	Y	1	
14	3359690891	152.451565	4116.192255	N	0	Y	1	865	375	8	Y	1	
15	3364840542	36.91948763	2141.330283	N	5	Y	1	0	0	0	Y	1	
16	3365355395	806.1795426	23379.20674	N	0	N	0	816	811	5	Y	1	
17	3369900897	257.0911668	10283.64667	N	4	Y	0	0	0	0	Y	1	

Fig. 2 Screenshot of Dataset "creditcardcsvpresent.csv"

Research Question 3 (RQ3) of the systematic review was framed to analyze the significance of data mining techniques used to find out the relationship between data mining techniques and threat detection.

With the help of a case study, this study intends to find out the patterns and relationship between attributes of the security data.

The case study is intended to define threat detection strategy The dataset used in this case study is Abstract data set for Credit card fraud detection (creditcardcsvpresent.csv) (Refer Fig. 2) from Kaggle.com by Vinayak Joshi [13]. The dataset summarizes the usage of 3076 customers and 11 attributes. The attributes are described in Table 4.

Table 4 List of Attributes of Dataset "Abstract data set for Credit card fraud detection" (creditcardcsvpresent.csv)

Attributes	Description
Merchant_id	Id of the merchant
Average Amount/transaction/day	Transaction performed per day by each merchant
Transaction_amount	Transaction amount
Is declined	Transaction declined or not
Total Number of declines/day	Total number of transactions declined per day
isForeignTransaction	Transaction performed is foreign transaction or not
isHighRiskCountry	Transaction performed is in high-risk country or not
Daily_chargeback_avg_amt	Daily chargeback average amount
6_month_avg_chbk_amt	6 months average chargeback amount
6-month_chbk_freq	6 months chargeback frequency
isFradulent	Transaction performed is fraudulent or not

Using python, a statistical test was performed to find out the following relationships. So that based on this analysis, the threat can be interpreted.

(i) Is there any relationship between Total Number of declines_per_day and is_fradulent attribute?

(ii) Is there any relationship between isHighRiskCountry and is_fradulent attributes?

Case1: One value is categorical and other is numerical so, ANOVA test is performed to find the relationship between the two attributes. Let us consider Ho (null hypothesis): There is no relationship between the two attributes. By statistical test, statistics value is greater than pvalue (statistics > pvalue) (Refer Fig. 3). Therefore, the null hypothesis is

rejected and it is inferred that relationship exists between attributes Total Number of declines_per_day and is_fraudlent attributes.

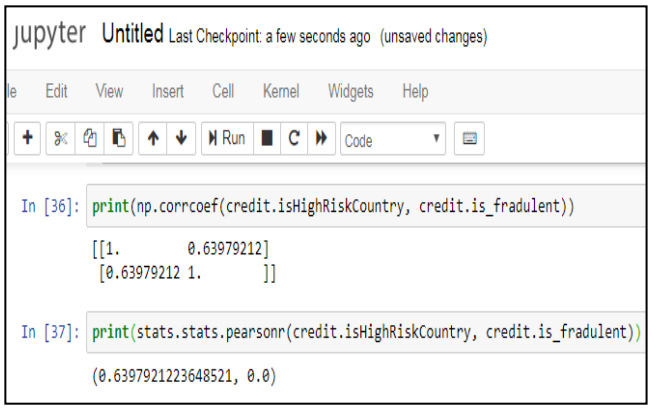

Fig. 3 ANOVA TEST for Case 1(Total Number of declines_per_day and is_fradulent attribute)

Case2: Since both the values are numerical so Correlation test is performed to find the relationship between the two attributes. Let us consider Ho (null hypothesis): There is no relationship between the two attributes. By statistical test, values are close to 0 (If values are close to 0 then no relationship and if values are close to 1 then there is relationship) (Refer Fig 4), therefore no relationship between the two attributes. The null hypothesis is correct.

Therefore, the threat can be recognized based on the Total number of declines happening. But high-risk country factor cannot help in determining whether the transaction will be fraudulent or not.

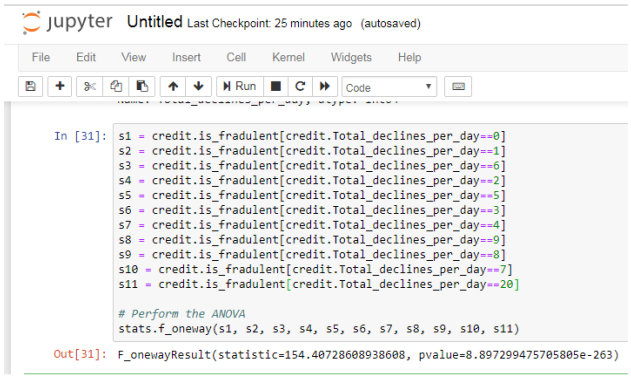

Fig. 4 Correlation TEST for Case 2 (isHighRiskCountry and is_fradulent attributes)

RQ4: ***What are the disparate and data-driven system challenges experienced by Internet Connected Systems (ICS) and different threat intelligence schemes modeled to overcome the challenges?***

Research Question 4 (RQ4) of the systematic review was proposed to determine the challenges experienced by the smart system (ICS) and various intelligence techniques developed to control system breaches and identifying threat conditions.

Summary of disparate and data-driven system challenges experienced by massive and enormous systems that are connecting large physical environment and intelligence schemes modeled to overcome these challenges.

Industry and manufacturing revolution in recent times have provided advancement to ICS like automated factories, Internet of Things (IoT), smart cities, smart homes, and robotic factories, but designing a protected architecture faces major challenges. The challenges recognized like, handling security over heterogeneous massive sources, distinct standards for manufacturing and technology requirement. One of the big challenges in present threat detection techniques stems from those in transforming heterogeneous data sources gathered from software, platforms, and Fog and Cloud computing systems. New threat intelligence scheme was proposed based on two components: smart management element and threat intelligence element. Smart management module handled diverse data sources needed to interact with industrial systems. Threat intelligence scheme helped in identifying anomalous movements against both physical and network systems. Another challenge is implementing Machine learning for cybersecurity is not an easy task because of the fast and rapidly changing threat outlook. To this issue, fractal-based cognitive neural network methodology improving classification problem of neural networks was proposed to detect and distinguish malicious samples.

RQ5: ***What are the ethical aspects and role of artificial intelligence to ensure that no future calamities occur?***

The Research Question 5 (RQ5) of the systematic review was framed to analyze the esthetical concern in the development of machine intelligence and how artificial intelligence algorithms are helping to create apt rules in order to resemble the thought process of a human.

Summary of motivation, expectations in the advancement of machine intelligence and various ethical aspects and role of artificial intelligence in new emerging artificial intelligence scope.

Artificial intelligence is the understanding shown by machines and software. It helps to determine the problem in a similar fashion like humans [19]. The presence of AI is invariably expanding in cyberworld because of its tireless performance of tasks. The main motivation of intelligence machines is that decisions made are more logical and appropriate because of the absence of emotions. On the other hand, humans make decisions after taking everything in mind of his/her emotions. Some ethical decisions are being encountered such as to support or oppose the development of lethal autonomous weapons systems (LAWS). Ethical issues are developing as we are giving more importance and power to robots. Ethical issues can be dealt with by writing appropriate code and testing issues properly. A hybrid system among AI and humans is an intelligent solution to implement machine intelligence [8]. Knowledge Engineering (KE) reviews the structure of a decision to recognize how a conclusion is attained. It is widely in practice by engineers like they are integrating KE in decision support software in order to gain the ability to identify a face or parse what a person says for meaning. Sooner, the area of knowledge engineering will help in creating systems that could solve problems better than humans.

RQ6: ***What are the various security assessment methods and models developed using artificial intelligence and its future scope?***

Artificial intelligence opens potential capabilities and will bring a new range of experiences. Research Question 6 (RQ6) of the systematic review was framed to study various techniques and models evolved for overcoming the resistance in emerging security intelligence.

Summary of models and their applications developed for identification of Cyber threats using AI and its future outlook.

For emerging and ongoing new techniques used by intruders, it is important to develop models that are most reliable for most vulnerable situations occurring. Cross-stack sensor framework is one such development for achieving attackers' information and monitoring their activities (Araujo et al. 2018). It provides defenders with new potent tools and methodologies for new occurring deceptions. Cyber Threat Intelligence models enable cyber defenders to scrutinize their threat intelligence capabilities. It says that attacks must be properly recognized before performing the safeguarding efforts.

V. CONCLUSION

This study concludes that ongoing cyber defenses are mainly perimeter-based, and are often framed on known statistical patterns to search threat condition. Therefore, these approaches are limiting visibility into emerging cyber threats. Multiple intensely embedded and threat sensing techniques can minimize the impact of thriving attack campaigns and can act as an obstacle against future attacks.

Threat intelligence sharing allows entities to interchange patterns of danger with each other, in the form of indicators, for threat analysis and occurrence response. The main drawback is that none of them covers all of the significant data and information required for efficient threat intelligence. Knowledge from publically available data should be gathered and formally represented to assist in the progress of advanced reasoning. Despite data mining techniques available there is a need to use subjective knowledge to train models or artificial neural networks in the system to identify and resist unknown threats.

It is inferred from this study that threat intelligence requires analyzing and sharing of threat data and information in an efficient way but, sharing of information requires common protocols, standard format, common representation and understanding of basic terminologies. Therefore, artificial intelligence is the solution approach to this need. This review includes statistical techniques to test threat intelligence based on data mining techniques. Further, the scope of the paper can be extended using artificial intelligence algorithms using data mining techniques to recognize unknown patterns of threats.

REFERENCES

1. Alguliyev, R., & Imamverdiyev, Y. (2014, October). Big data: big promises for information security. In *Application of Information and Communication Technologies (AICT), 2014 IEEE 8th International Conference on* (pp. 1-4). IEEE.
2. Basallo, Y. A., Senti, V. E., & Sanchez, N. M. (2018). Artificial intelligence techniques for information security risk assessment. *IEEE Latin America Transactions*, *16*(3), 897-901.
3. Bastos, M. R., & Martini, J. S. C. (2015, October). A model-free voltage stability security assessment method using artificial intelligence. In *Innovative Smart Grid Technologies Latin America (ISGT LATAM), 2015 IEEE PES* (pp. 679-682). IEEE.
4. Cao, N., & Wang, Y. (2017, July). A Novel Approach to Improve Robustness of Data Mining Models Used in Cyber Security Applications. In *Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference on* (Vol. 2, pp. 297-300). IEEE.
5. Deng, S., Yuan, C., Yang, J., & Zhou, A. (2017). Distributed Mining for Content Filtering Function Based on Simulated Annealing and Gene Expression Programming in Active Distribution Network. *IEEE Access*, *5*, 2319-2328.
6. Folino, G., Pisani, F. S., & Sabatino, P. (2016, July). An incremental ensemble evolved by using genetic programming to efficiently detect drifts in cyber security datasets. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion* (pp. 1103-1110). ACM.
7. Ghimeş, A. M., & Patriciu, V. V. (2017, June). Neural network models in big data analytics and cyber security. In *Electronics, Computers and Artificial Intelligence (ECAI), 2017 9th International Conference on* (pp. 1-6). IEEE.
8. Gupta, S., & Chowdhary, S. K. (2017, September). Authentication through electrocardiogram signals based on emotions-a step towards ATM security. In *2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 440-442). IEEE.
9. Gupta, S., & Dhir, S. (2016, February). Issues, Challenges and Estimation Process for Secure Web Application Development. In *Computational Intelligence & Communication Technology (CICT), 2016 Second International Conference on* (pp. 219-222). IEEE.
10. Gupta, S., & Sabitha, A. S. (2018). Deciphering the attributes of student retention in massive open online courses using data mining techniques. *Education and Information Technologies*, 1-22.
11. Hochbaum, D. S., & Baumann, P. (2016). Sparse computation for large-scale data mining. *IEEE Transactions on Big Data*, *2*(2), 151-174.
12. Jayasingh, B. B., Patra, M. R., & Mahesh, D. B. (2016, December). Security issues and challenges of big data analytics and visualization. In *Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on* (pp. 204-208). IEEE.
13. Kaggle (2018). Abstract data set for Credit card fraud detection, creditcardcsvpresent.csv, by VinayakJoshi, Link: https://www.kaggle.com/shubhamjoshi2130of/abstract-data-set-for-credit-card-fraud-detection#creditcardcsvpresent.csv
14. Khan, M. A., Pradhan, S. K., & Fatima, H. (2017, March). Applying Data Mining techniques in Cyber Crimes. In *Anti-Cyber Crimes (ICACC), 2017 2nd International Conference on*(pp. 213-216). IEEE.
15. Kumar, N., Kharkwal, N., Kohli, R., & Choudhary, S. (2016, February). Ethical aspects and future of artificial intelligence. In *Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016 International Conference on* (pp. 111-114). IEEE.
16. Li, L., Lu, R., Choo, K. K. R., Datta, A., & Shao, J. (2016). Privacy-preserving-outsourced association rule mining on vertically partitioned databases. *IEEE Transactions on Information Forensics and Security*, *11*(8), 1847-1861.
17. Ng, J., Joshi, D., & Banik, S. M. (2015, April). Applying data mining techniques to intrusion detection. In *2015 12th International Conference on Information Technology-New Generations (ITNG)* (pp. 800-801). IEEE.
18. Sisiaridis, D., & Markowitch, O. (2018, April). Reducing Data Complexity in Feature Extraction and Feature Selection for Big Data Security Analytics. In *Data Intelligence and Security (ICDIS), 2018 1st International Conference on* (pp. 43-48). IEEE.
19. Vattapparamban, E., Güvenç, İ., Yurekli, A. İ., Akkaya, K., & Uluağaç, S. (2016, September). Drones for smart cities: Issues in cybersecurity, privacy, and public safety. In *Wireless Communications and Mobile computing Conference (IWCMC), 2016 International* (pp. 216-221). IEEE.
20. Veeramachaneni, K., Arnaldo, I., Korrapati, V., Bassias, C., & Li, K. (2016, April). AI^ 2: training a big data machine to defend. In *Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2016 IEEE 2nd International Conference on* (pp. 49-54). IEEE.
21. Vijay, V. C., Lees, M., Chima, P., Chapman, C., & Raju, P. (2015, March). Knowledge based educational framework for enhancing practical skills in engineering distance learners. In *Global Engineering Education Conference (EDUCON), 2015 IEEE* (pp. 124-131). IEEE.
22. Yavanoglu, O., & Aydos, M. (2017, December). A review on cyber security datasets for machine learning algorithms. In *Big Data (Big Data), 2017 IEEE International Conference on* (pp. 2186-2193). IEEE.

AUTHORS PROFILE

Ms Shivangi Gupta is a student and pursuing Bachelors degree in Information Technology. She is currently student of Department of Information Technology, Amity School of Engineering and Technology, Amity University, Noida. She has published research papers in Journal and conferences. She has recent published manuscript "Deciphering attributes of student retention in massive open online courses using Data Mining Techniques" in "*Educational and Information Technology"* journal. She has published in *Computational "Intelligence & Communication Technology", IEEE*. She has published in "*International Conference on Reliability, Infocom Technologies and Optimization", IEEE*. She has handled android and networking projects. Her area of interests are Android Development, Data Mining, Networking and Web Technologies.

Dr A. Sai Sabitha is an Associate Professor in Computer Science and Engineering. She is currently heading Department of Information Technology, Amity University, Noida. She has published several research papers in conferences & journals. She has over 17 years of Academic and industry experiences. She has handled various B.Tech & M. Tech Projects. Her area of interests are e-Learning, Knowledge Management, Data Mining, Artificial Intelligence & Web Technologies.

Ms Ritu Punhani assistance professor in Information Technology in Amity University, she handled many responsibilities. She was c*onducted corporate training for South African High Commission to train their employees the concept of software project management. Earned "YOUNG RESEARCHER" Award for research proposal "ISMS Process Maturity Standardization for Data Mining Industry", held at Central University of Rajasthan in collaboration with IIT Kharagpur, National School of Leadership (NSL) Pune and International Association of Research Scholars (IARS).* Published 21 research papers in field of information technology and write few books for MBA and MCA courses.