ISSN 2278-3091
Volume 10, No.3, May - June 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse861032021.pdf
https://doi.org/10.30534/ijatcse/2021/871032021

Using ML and Data-Mining Techniques in Automatic Vulnerability Software Discovery

Imdad Ali Shah, Samina Rajper,
Department of Computer Science Shah Abdul Latif University Khairpur. Sindh, Pakistan
imdad.shah@salu.edu.pk, Samina.rajper@salu.edu.pk
Noor ZamanJhanjhi
School of Computing Science and EngineeringSCE, Taylor's University, Malaysia
NoorZaman.Jhanjhi@taylors.edu.my

ABSTRACT

Today's age is Machine Learning (ML) and Data-Mining (DM) Techniques, as both techniques play a significant role in measuring vulnerability prediction accuracy. In the field of computer security, vulnerability is a fault that might be exploited as a risk artist that performs unlawful activities inside computer security. The attackers have several different fitting tools and they are taking advantage to operate software illegally and are using it for getting self-profit. Additionally, that helps to expose and identify the violence external. Weakness management remains a repeating exercise to identify, remediating, and justifying weaknesses. These exercises normally send software faults in computing security. The meaning of using weakness with the same risk might go to misperception. It is possible to have a major effect because of possible stability and the window of weakness presented a risk hole in the software and required to fruitfully finish and smoothly operate. A security room has to be set up (zero-day invaders). Software Security Faults stand serious among unavoidable complications in the realm of computer risk. In this study, we have provided a comprehensive review of three book chapters, more than a hundred research articles papers, and several associated papers of different work that have been studied within the capacity of SVA and discovery applying ML and data-mining techniques. The earlier work has been thoroughly read and an adequately comprehensive summary has been provided in table-1. ML techniques that can professionally handle these attacks and we expect the net result of this survey article to help in designing the new detection model for identifying the above-mentioned attacks.

Keywords: Privacy vulnerability management, security, machine learning, and cybersecurity

1. INTRODUCTION

Our age is computer software (CS) and its usage everywhere. Day to day, increasing, present life remains on various types of software. [1]Cybersecurity activities and rules are the main reasons for the problems and profits of a probabilistic nature. Together approaches could not avoid handling with danger and possibility. The ethics risk issues have been discussed compressively in two consequentialist approaches, namely deontological and contractual approaches to risk, and appraised the problems through special cases. [2] Has focused business ethical stakeholders toward reply to them and responsibilities have been assessed using an approach of care-based stakeholders including business's ethical responsibilities. [3] Has given a theoretical investigation of ethical hacking, containing history and trying to provide a systematic sorting and further study to identify ethical hackers with hackers that are bound to a code of ethics benefiting business-friendly value. Finally concludes by advancing a practical greatest repetition approach for indicating ethical hacking, which helps to reach the final decision. Finally, present the challenges in the area of Automatic Vulnerability Software Discovery are discussed, future scenarios and emerging guidelines are planned.

Our contributionto this study, we provide a comprehensive review of three book chapters, hundred research papers, and several associated papers of different work in the area of software vulnerability analysis and discovery applying Machining Learning and data mining techniques. We also analyze research trends and research focus on ML techniques that can professionally handle these attacks and we expect the net result of this survey article to help designing the new detection model for identifying the above-mentioned attacks.

Table 1: Explains the survey of the chapter's topic (s) of ethical issues

Year	Survey of the chapter's topic (s)	Ethical Issue	Conflicting
2020	**[1]**	**Emerging technology and identify two approaches.**	**Protecting information**
2020	**[2]**	**Repairing find vulnerabilities inside providing a time frame**	**Protection business data**
2020	**[3]**	**Bound toward a code of ethics benefiting business-friendly values**	**Validity**

2. COMPREHENSIVE DETAILED OF MORAL NATIONAL SECURITY

i. Cyber Terrorism has claimed that cyber-crime remains normally go it to comprise regarding accessing computer except holder's allow, maximum possibility regarding consent of one to access computer security, destroying computer info. Cyber-terrorism fundamentally involves doing these same activities to advance single ideological or political conclusions. There is a double relation between the Internet and Terrorism. Initially, the Internet has been raised as a forum and a terrorist group aimed at individual terrorists, to spread violence and hate messages among the people, as well as conversation with one another and their supporters.

ii. Cyber-EspionageCyber spying remains to use electronic abilities to unlawfully collect data. The info technology (IT) revolution has improved the path to govt run, for whole states. The unbalance threat added by cyber-attacks and the medium errors of a cyber gap are thoughtful. The security threat to all nations. While successes about cyber spying to that rule counterintelligence and administration create short response signals which additional thoughtful cyber-attacks about the dangerous substructures stand the trouble of time.

iii. Cyber Awareness to make great awareness regarding cyber-security threats and weaknesses and on civilization has happened energetically, it watches absently in humanity. If we matched headship which administrations about states' effort toward emerging. Through increasing awareness, corporate and individual users might know in what way to act now in the online realm and defend them by characteristic danger.

iv. Consciousness happenings become & run change supply approaches to touch wide viewers. Main outside issues of long current security breach dangers and events, new risks, updates of security rules and approach.

v. Profiling people are approached, behaved in the assured path because they have features that need to reach an exact profile and are associated with definite other qualities. The profile is to be used by security agencies or police to catch terrorists or offenders, through airport security, organizations to target exact customers and banks determining to provide loan facilities to a needy person and granted loan on their profile serve security purposes [3],[4],[5],[6],[7].

3. PAPER CONTRIBUTION

We have done an integrated deep study on the vulnerability of Software Discovery. Several presenting threat analysis techniques algorithmsare available and create hard for the practitioners to tell them regarding the choosing the fitting approaches. Our review work will open new doors for the researchers and help to reach an exact solution to the emerging technology of automatic software vulnerability. Further, this review help to choose the data analysis techniques algorithm, the primary object of this review article to give strategies for the detect knowledge holes for the forthcoming research guidelines.

a. SUMMARYOF SURVEYS RESEARCH PAPERS ON ML and DMT in AVSD

[8] Has appraised that every year huge figures of security weaknesses have been found in the production of software. The publicly stated CVE records/exposed within the propriety code. A wealth of OP code is presented for examination and the chance to study the forms of viruses that might lead to SC from the information. In a data-driven method to weaken discovery, applying ML, particularly specifically, has been applied to C and C++ platforms. The data compiled a large dataset of hundreds of thousands of open-source functions labeled with the outputs of static analysis. The dataset compared the application of deep neural network models with methods applied to artifacts from the build process and found that source-based models perform better. [9] Has described that the figure of software vulnerabilities is increasing each year, they are discovered internally in proprietary code. Hence these weaknesses are serious risks of the exploit, may not compromise the result of the system and information can leak. Further study appraised that C and C++ open-source code present a large-scale function level for vulnerability detection using ML. This worked on developing a fast and scalable

vulnerability detection tool on deep feature representation learning and that worked directly interprets lexed source code. The study results are a promising approach for automated software vulnerability detection. [10]Has studied that the fuzzing techniques are to be used for finding bugs by executing the software with a large number of abnormal inputs and focused on how to improve the code coverage. It is inefficient the vulnerable code only takes a tiny fraction of the entire code. This study focused more on design and implements a vulnerability-oriented evolutionary fuzzing prototype namely V Fuzz, on an object to find bugs efficiently and quickly in a limited time. [11]Has studied that many engineering tasks greatly trust the classification/regression handcrafted software features, just like detection, software requirements, vulnerability discovery, malware detection, and code review. This study appraised more than previous solutions usually directly using handcrafted features. This leads to suboptimal results due to their lack of powerful representations of the handcrafted features and focused to adopt the effort aware just in time software defect prediction (JIT-SDP), which has a typical handcrafted based task and exploit new possible solution. [12]Has introduced an approach for predicting the cumulative number of software vulnerabilities and it is more accurate than vulnerability discovery models in most cases. The author approaches using a neural network model (NNM) to model the nonlinearities related to vulnerability disclosure. While nine common VDMs used interest to compare their prediction capabilities with the researcher's approach. As the study shows that NNMs are accurate predictions of software vulnerabilities. [13]Has studied that open-source software systems are available on the internet today and studies focused on automatic label software code. While software code may be labeled with a set of keywords, the study referred to these keywords as software labels. Further the main object of this study is to provide a quick view of the software code vocabulary. [14]Has appraised that automated web application penetration emerged as a development. As the computer assigned the task of penetrating web application security with the penetration testing technique due to the computer's relevant program reduces the time, resources, and cost required for assessing a web application security. [15] Has appraised the computer's security investigation and its results would benefit from the information about the characteristics of the human attacker behind security incidents. As the current security mechanisms gave attention to the characteristics of the attack, rather than the attacker and focused on the attacker behavior analysis due to it beinga challenging problem. [16] Has worked to identify the vulnerabilities from the internet sites with WCMS applications and remedies to be applied. The important feature of the WCMS is the ability to perform automated, dynamic, and fast vulnerability scans of the WCMS and the attached plugins on a large scale with in-built ethical respect. [17]Has worked on the many Android app security analysis tools to detect many of the known vulnerabilities. This study appraised advantages for the security measures of password authentication and every type of authentication technology and application service input from the keyboard. [18] studied worked to identify the gaps in hacking practices and developed their activity plan for self-protection against such cyber-attacks. [19]Has worked on an Al-based Penetration System using reinforcement learning (RL), on interest to learn to reproduce average and complex Penetration activities.[20] Studied on the multifaceted aspect of IoT networks & applications, based on real-life use cases, and focused on the difficulties engendered in mounting an ADE from both software system Engineering and network convergence perspectives. [21]Studied the performance of several ML models compared to predict attacks and anomalies, on the IoT systems accurately. [22] Has studied conducting a systematic and in-depth survey of the ML & DL-based resource management mechanisms in cellular wireless and IoT networks. [23] Has discovered unstructured raw data contains a rich source of exact information on how the huge volume of data may be leveraged to create cyber intelligent situational awareness to mitigate advanced cyber threats. [24]Has studied continuously changing network behavior and rapid growth of attacks made which were generated through static and dynamic approaches.

b.SUMMARY OF OTHER ASSOCIATED RESEARCH PAPERS

The study focused on an important research problem about automatic detection of a software vulnerability, while the exact solution of human's expert difficult types. Further, the author used deep learning-based for vulnerability discovery and deep learning is suitable to deal with problems. As the author took some guiding principles to apply deep learning to vulnerability discovery and find representations of software programs that are suitable for deep learning [25]. The researcher used code gadgets to symbolize platforms and transformed them into paths. The author designed the implementation of a bottomless LB weakness recovery structure which is called Weakness Bottomless Pecker and this Vulnerability is practically missed by other vulnerability Discovery Systems. The author experimented with other authors proposed, some preliminary guiding principles for using deep learning on interest to discover a vulnerability, further the author's worked on these principles due to these are sufficient. These principles centered on answering three fundamental questions (a) what is the appropriate granularity for bottomless LB weakness discovery? (b) What is the appropriate granularity for deep learning-based vulnerability detection? (c) How to select the exact neural system aimed at weak discovery? The researcher

proposed that every year a huge safety weakness has been exposed in manufacturing software, publicly appraised usually weakness and experiences recorded/ revealed inside the propriety code. While the affluence of open-source code presents aimed at the examination and chance to study the patterns of bugs that might be central to security weaknesses from records. The author proposed a data-driven method to weak discovery applying ML, exactly practical to C and C++ programs. Further, the author compiled a huge dataset of thousands of open-source functions labeled through the results of static analysis. The researcher compared the application of deep neural network models through attitudes applied to artifacts from the creative practice and searching that source-based models execute superiorly. The author used a data-driven methodology to weaken exposure applying ML. Further, the researcher also compared the application of a deep neural network model and has implemented machine learning models for the detection of bugs that might be central to security weaknesses in the C/C++ code. Hence these weaknesses are effective risks of exploitation, may not compromise in the result of the system, and information can leak. The author appraised that C and C++ open-source code present to grow huge measure function level for weakness detection applying ML. The researcher worked on developing a quick and scalable weakness exposure tool on bottomless types of learning signs and that worked directly interprets lexed source code. While the results of the research are a commented method aimed at automated software weakness discovery. The researcher has established a quick and scalable weakness discovery tool built on deep type's illustration learning regarding openly interpreting the lexed source code. The further author has established the probability of applied ML to identify software weaknesses directly from source code. The researcher proposed that as a maximum broadly employed method used to catch weaknesses in practice software platforms. The author presented a structure to examine weaknesses exposed from the current compositional investigation tool & allocate CVSS3 (Common Weakness Counting Structure v3.0) scores to them. Further, researchers trained simply based on machine learning models. The researcher appraised a structure aimed at routinely predicting the sternness of vulnerable function which is informed by a compositional symbolic execution tool. As used as a logical method. Further, the author picked data from NVD regarding weaknesses that were informed earlier through CVSS3 sternness scores aimed at C programs. The researcher proposed that the fuzzing techniques are to be used for searching bugs through implementing the software through a huge number of irregular involvements and focused on what steps have been made to recover code handling. While this is ineffective as the weak code single obtains a little part of the whole code. The author focused to form & used weakness concerned with evolutionary fuzzing prototype namely V Fuzz, on the object to find bugs efficiently and quickly in a limited time. The author designed and used V-Fuzz, a vulnerability-oriented evolutionary fuzzing structure, through joining the weakness prediction with evaluating V-Fuzz. The researcher proposed an important problem and attracted much attention regarding automatically detecting software vulnerabilities, while weakness finders still can't obtain weakness exposure ability. Further, the author presented Weakness Bottomless Learning-based locator (VulDeeLocator). The researcher proposed that many engineering tasks that greatly trust the classification/regression handcrafted software features, just like detection, software requirements, vulnerability discovery, malware detection, and code review. Further, the author appraised those earlier explanations regarding tasks usually directly practice the handcrafted types. The researcher introduced a method aimed at predicting the collective quantity of software weaknesses and it is additional exact than weakness detection models in most cases. The author approaches using a neural network model (NNM) to model the nonlinearities related to weakness revelation. While nine usually VDMs applied on interest to match their prediction ability through the researcher method. The researcher proposed that there are exposed source software structures available on requirement today and the author focused on automatic sign software code. The researcher proposed that the information system mostly builds on the weak supervision process in the present days and active safety made measures earlier to stop the attacks. The researcher proposed that automatically advanced web application penetration emerged as developed. While the computer gave the task of penetrating (WAS) through the penetration checking method due to the computer's relevant program, it reduced the time, resources, and cost essential for assessing a (WAS). The researcher proposed the state of the art of black-box (WA) scanner is scientifically studied and examines the methods aimed at detecting (WA) weakness in an unseen checking atmosphere. Further, the author designs sophisticated algorithms for interpreting and understanding the semantics of under-test web applications with an absence of source codes. The researcher proposed to present CASEI, complete a system that removes information about cyber-security events from the text, and populates a semantic model on interest to integration into a knowledge graph of cybersecurity data. Further author's study covered both cyber-attacks and vulnerability-related events. The researcher proposed a new cybersecurity event removal task and provided definitions of five event types. Further, the study targeted works significant responsibilities in an incident detection system. The researcher developed CASIE and information elimination structure trained using the annotated data and studied their output and components to link event arguments to Wikidata entities and for computing reference and sequence relations between events. The researcher proposed to work on the

computer's security investigation and its results will benefit through the material regarding the appearances of the human attacker behind safety incidents. While the current security devices attention to the appearance of the violence, somewhat after the attacker. Further, it focused on the attacker behavior analysis due to it being a challenging problem [26],[27],[28],[29],[30],[31],[32],[33],[34],[35],[36],[37],[38],[39],[40],[41],[42],[43],[44],[45],[46],[47],[48],[49],[50].

b. SOFTWARE WEAKNESS ANALYSIS AND DISCOVERY

In the perspective of software security, weaknesses remain exact flaws. Everyone watches through the above-cited explanations, as we can see from the above-cited definitions, several important relations were applied to express software weaknesses. Explain the relations and choose the maximum appropriate. The object stands, whether this detects exploited reason destruction(www.sypsnopsy.com).

c. RELIABILITY, WHOLENESS, AND UNDESIRABILITY

Platform malfunction examination is the problem of deciding to provide a platform with identified confidence weaknesses. Further, we explanation as:

i. Undecidable:according to the Compatibility idea & computational difficulty idea an undecidable problem is a decision problem aimed at proving unmanageable to build procedure to continuously centralize to give exact answer no or yes.

ii. Turing test: This test was established by Alan Turing in 1950 and it can show the machine's ability to be equal or unlike any human. Denial proposes that a human assessor justice NL exchanges amid human and machine formed in instruction to produce a human response. Assessor sees that amid partners in the exchange remains a machine and that contributors are unique. Discussion is incomplete to text channels such as computer keyboards & screens, hence this result didn't depend on the machine's capability to using words to speak. Uncertainty the assessor couldn't dependably check the machine, they inquired about the machine to permit a test. While test results depend on machines to answer the questions properly, only the human obtained the answers they essential. Figure-1.

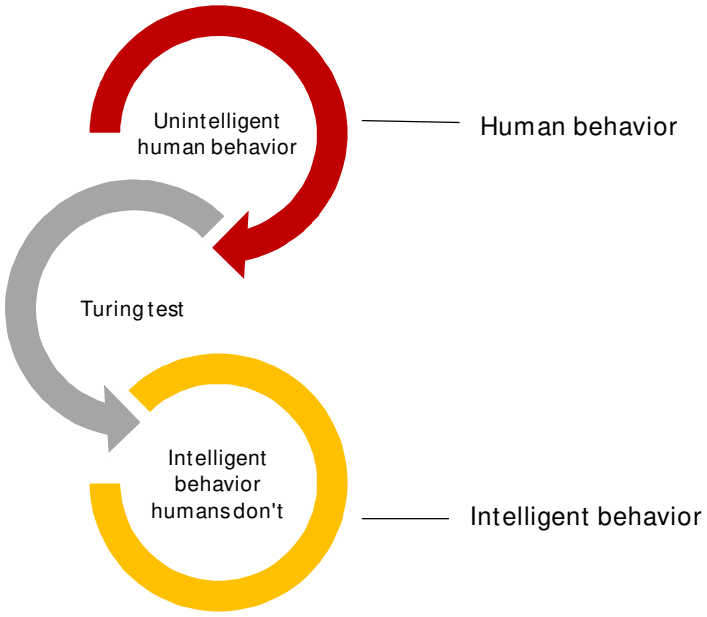

Figure-1 Explain Turing Testing

iii. Halting Problem: HP is decisive from an explanation of an arbitrary computer platform and enters whether the platform would run forever.

iv. Rice's Theorem: Altogether non-trivial (NT), semantic properties of platforms were undecidable. A property is NT if it is neither correct for every comparable function nor incorrect. Built regarding the pending of

Turing's HP & RT, this might show that several platform examination difficulties were pending in the overall event. Program safety examination remains the problem to choose where the program was set covers identified safety weakness (as per security policy) or not [50]. Grounded in the frame of resolving Turing's HP & Rice's Theorem, it might express that several program investigation troubles were chosen in the common event, have chosen experts was regarding comprehensive & thorough results to the problem didn't exist. The logic of mathematics, an evidence method is found is comprehensive if no unacceptable argument might sanction from the method. An evidence method is comprehensive if altogether correct the argument might be sanctioned by the method. As cited previously such a comprehensive method is known as non-existent. In additional weakness investigation, an extra in practice beneficial method is a program weakness detection (or safety reporting) method. Indifference to weakness investigation method that authorizes or rejects the safety of assigned program (binary output) program weakness detection method reports extra comprehensive info (type, location, etc.) aimed at every weakness detection assigned in the program [51],[52],[53],[54],[55],[56],[57],[58],[59],[60],[61].

4. TRADITIONAL METHODS

TM refers to the traditional method of instruction that is frequently used in lecture mode. Using this teaching methodology is to be focused on the textbook, the teacher-dominant, based on the exams. The key here is to specifically learn, memorize and reproduce the principles and theories. Although, the un-guaranteed environment of the trouble of software weakness investigation, detection & an excess of methods studied and planned from experts mutually in the academic public, the software industry because of serious status regarding the trouble. The projected methods remain completely unavoidably estimated answers, they altogether moreover deficiency broadness. Therefore, altogether research steps try to plan a better method matched to earlier works, concerning the exact phase of the practice of software weakness examination and discovery, for example, discovery correctness, runtime effectiveness, and vulnerability coverage. Changed methods to decrease program safety weaknesses need extensive review, including platform vulnerability examination and detection procedures were printed amidand were presented by Shahriar &Zulkemine. As altogether program examination methods have divided into three main types:

v. Static Analysis:SA so-called Static Code Analysis (SCA), is a practice of computer program debugging. The assigned program was examined based on its source. These methods are applied for generalizing concepts to examine the belongings of a program,therefore SAA greatest might complete false vulnerabilities may report and the greatest is no lost weaknesses. The less reported untrue weaknesses, extra accurate generalization. In the drill, a trade-off builds amid examination accuracy and computational effectiveness. Further, it is an automatic tool that might support developers and programmers transporting out SA. Code verification through visual inspection only, without the help of automatic tools, is occasionally named program comprehension or program understanding.

vi. Dynamic Analysis:DA is a program checking and evaluating data planning in real-time. The assigned program was examined from implementing it with input data & checking during process attitude. Using this way, sets of check-cases aimed at an examination of the properties of a program, frequently significantly promising entries & runtime conditions, therefore dynamic analysis systems (DAS) cannot analyze the whole program's behavior. Hence, DAS may finish (i,e not report false vulnerability & approve all security programs). Hence analysis methods, the condition aimed at functioning runtime environment of the given program and it is possible lengthy-time & extraordinary charges compulsory aimed at the handling of altogether entry checking cases never examining huge, complex software. However, SA methods significantly use in the software industry.

vii. Hybrid Analysis: HA is a progressive and advanced security tool that provides detailed information about supporting files that upload to the service. While it needs to understand some deeply of Windows and Program code to understand the advanced parts of the examination. The Assign program examines through a combination's DA and SAT.Itis promising regarding misreading built on earlier debates regarding static and DA methods that hybrid analysis methods can sound and complete. Unfortunately, this is not true, and while HA methods might profit through the compensations of mutually static and DA, these effects by limits of mutual methods. This is not true, unfortunately; the advantages of both static and dynamic analysis can take benefit through

hybrid analysis approaches. There are many differences among the VDA.

viii. Software Penetration Testing:SPT is a software safety checking method, passed out through a group of safety specialists. Penetration Testing should try to exploit weaknesses and security vulnerabilities throughout the environment, try to penetrate both at the key application and network level. PT is to determine whether unauthorized access to important systems and files can achieve. There is a large amount of misunderstanding in the industry about the differences between vulnerability scanning and PT, as two expressions usually interchange. However, their implications and meaning are very different. Vulnerabilities assessment simple reports and identifiers noted vulnerabilities, while PTs (pen test) try to exploit the vulnerabilities to determine whether unapproved access or other malicious activity is possible Figure-2.

Figure-2 Explain STS

ix. Fuzz-Testing (Fuzzing):FT is called proper checking, anywhere a great designed entry record stands properly changed & fed toward the program in checking huge although checking for disappointments. Further, Fuzzing is a functionality testing method and software security that feeds randomly built input to the system and looks for signs that a failure in response to that input occurred. Fuzzing may help attackers discover sure assumptions made about user input in the system.

x. Static Data-Flow Analysis:SDFA isidentified Tainted Data-flow Analysis, Static program examination method wherever unreliable data through input sources mark for example taint & flow toward complex program declarations identified through sinks track as a possible sign of weakness. Further, DFA is a type of static analysis. The object of SA is to reason about program behavior at compile-time, before ever running the program[62],[63],[64],[65],[66],[67],[68],[69],[70],[71], [72],[73].

5. MACHINE LEARNING AND DATA MINING TECHNOLOGY

There are several processes of work that applies method through the study of artificial intelligence and data science to identify the problem of software weakness examination and detection. It attracted a class of methods ignored regarding the review of Zulkernine, hence it got a growing focus on the research group in these years. MLT through the study of AI proved useful applying exercise aimed at several other application fields Russell. MLT might generally divide into three approaches Figure-3& 4.

SupervisedLearning: SL is an assignment of learning a task that gives a path for input and output grounded on pattern input-out twosomes. The function has been obtained from inferring training label data which comprising of training set examples. The Learning Algorithm is required to the undetected location from training data for an accurate approach [74],[75].

i. Un-Supervised Learning:Un-Supervised Learning is a techniqueof Machine Learning procedure used to attract results through datasets consisting of input data without label responses. The most common Un-Supervised Learning method is cluster analysis, which uses exploratory data analysis to find hidden grouping in data or patterns [76],[77].

ii. ReinforcementLearning:Reinforcement Learning is an approach of Machine Learning, it's concerned with how software agents ought to take actions in an environment to maximize the notion of cumulative reward. RL differs from supervised learning in not needing labeled input/output pairs to be presented, and in not needing sub-optimal actions to be explicitly corrected. Instead, the focus is on finding a balance between the exploration of uncharted territory and the exploitation of current knowledge [78],[79],[80].

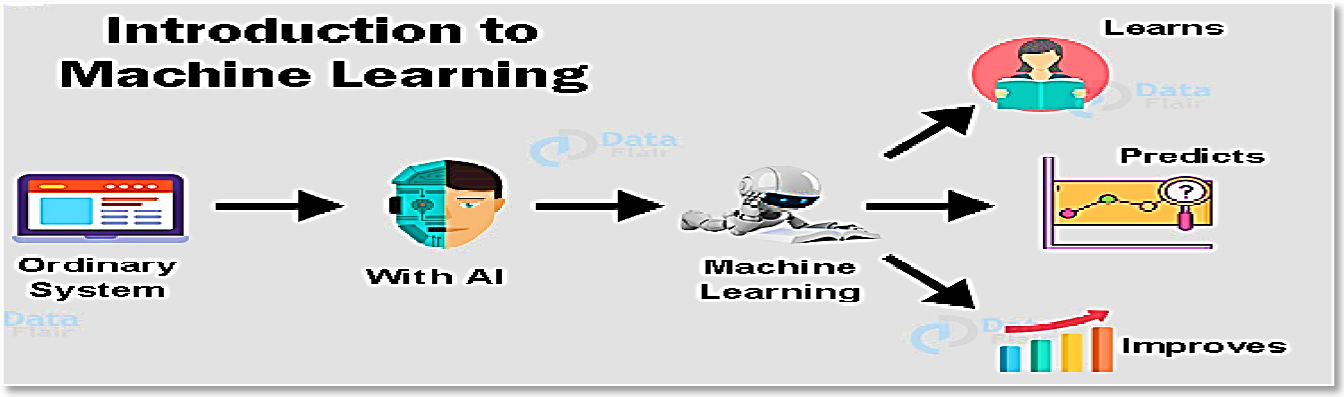

Figure-3 shows the introduction of ML [77]

Figure-4 shows the process of Data Mining Technology [80]

6. VULNERABILITY PREDICTION BASED ON SOFTWARE MODEL

We review the primary type of techniques that vulnerability prediction models that consume statistical analysis, ML, DM & methods to foresee weak software artifacts (object orient classes, source code records, and binary sections) grounded on usually software engineering metrics. The key thoughts of these techniques borrow through the ground of software value & dependability assurance in the study of SE due to their inadequate sources aimed at software verification & testing, which demands a supervisory model regarding

allowing additional effective software checking ideas. Vulnerability Prediction Models are a priority for security checks and for repairing pitfalls. Increasing safe software remains a time-consuming & difficult movement. The key cause of self-doubt was a vulnerability in the software. Hence prediction of software weakness plays a significant character in software engineering, particularly in web application growth. A software vulnerability prediction model estimates whether a software unit remains weak or not [81], [82], [83]Figure-5& 6.

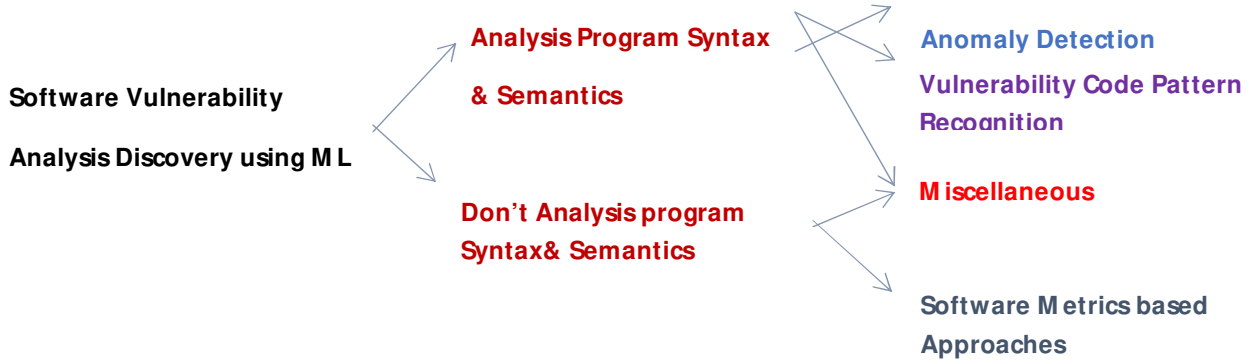

Figure-5shows the concept of software vulnerability

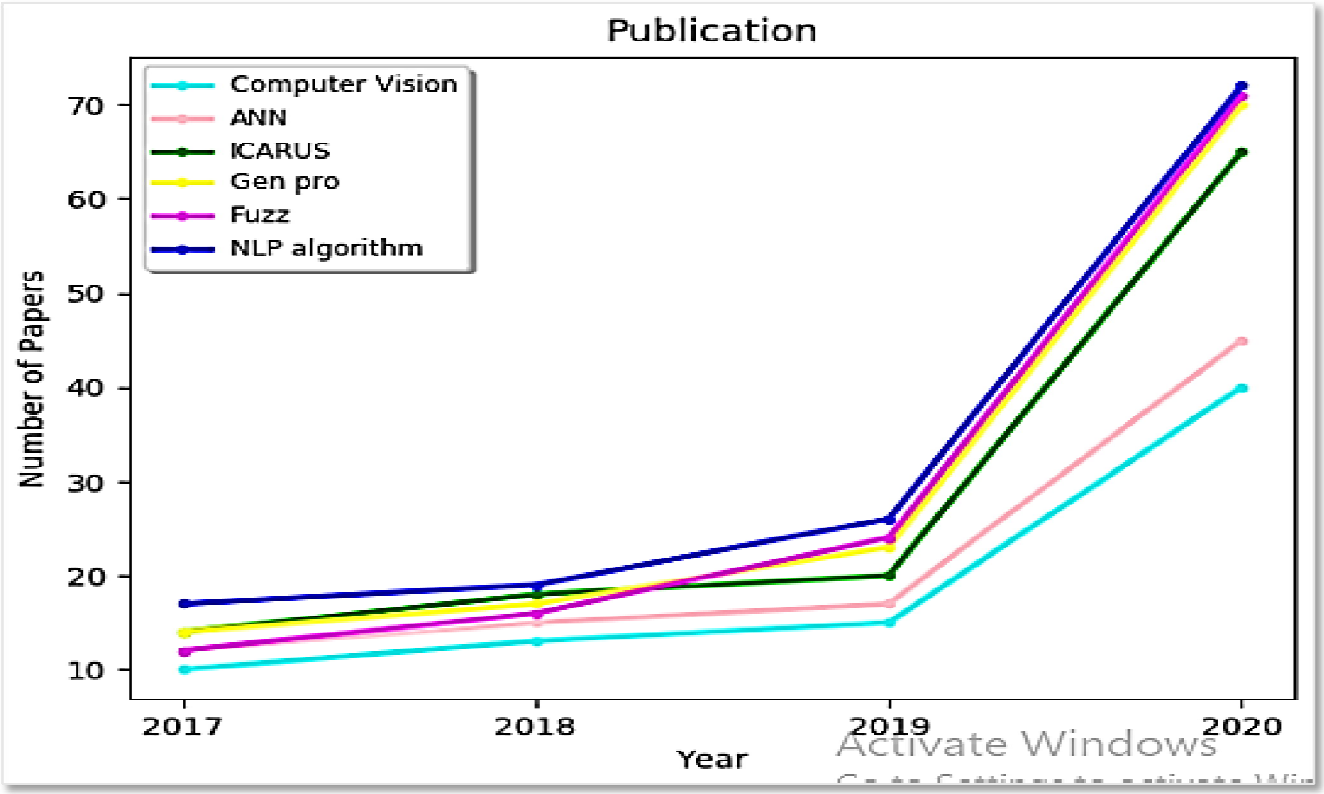

Figure-6 shows the MLT

7. EXPLANATIONS OF EMERGING ON A DIFFERENT ERA

[84] Study to assess the capability of IVA toward catch CVEs linked to a software product applying this automatically allocated CPE. Trouble remains regarding CPE vocabulary didn't comprise CPEs aimed at altogether products. It remains a tough assignment to search accurate CPEs used software info pick by a computer. Few cases info on a software product pointedly differed by these CPEs. [85,86] has recommended a system in which the method takes elements from both collaborative filters and attacks path discovery approaches to recognize attack pathways and forecast assaults. This study presented the performance assessment from the violence diagram generation procedure. Examined the performance & possibility of the planned violence pathways generation method toward recognized and compute altogether the potential attacked forms; we used the PSR SC subprocess of the "vehicles transport service. This was a portion of the given cable along with series of assets about that service only. The PSR subprocess goals to explain the interfaces among Port Authority, It was working through PCS which was electronic ground and connected various structures functioned from a variety of organization & level

procedure of section and logistic practices via only a requested plan. It has observed approximately 180 cyber properties (145 software properties and 35 hardware properties) by various product appearances and practical terms (vendor, product chains) various associates have shown weaknesses that were essential to the backing provision of the procedure. [87]Fuzzes techniques are used aimed at searching bugs via implementing software through a huge sum of irregular inputs and focused on in what way toward recovery code reporting. Further, the study focused to form and implements weaknesses concerned with an evolutionary fuzzing prototype namely V Fuzz, on an object to search bugs professionally and early within the due period. This study pursued a method to transformed binary program functions into numerical vectors. Furthermore, the vectors must able to carry enough information for future training. Towards this, chosen to leverage the Attributed Control Flow Graph (ACFG) was representing the binary function**.** The author introduced the key sections and workflow of V-Fuzz. As the above picture appraised the design of V-Fuzz, this shows Vulnerability-Oriented Evolutionary Fuzzer. [88] has worked on an Al-based Penetration system using the RL, on interest toward learning repeat regular and hard Penetrationhappenings.[89],[90],[91],[92],[93],[94],[95],[96],[97],[89],[99],[100],[101],[102],[103],[104],[105],[106],[107],[108] Figure-7

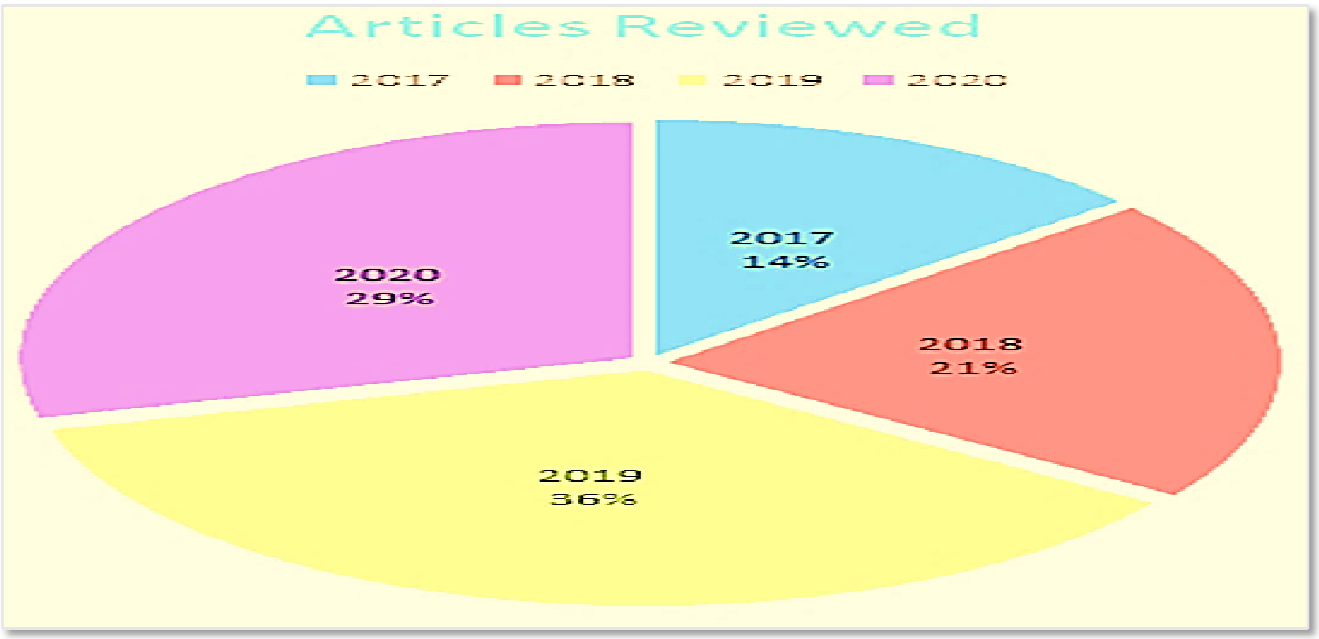

Figure-7Percentage of publication years of articles reviewed

8. CONCLUSION AND FUTURE WORK

In this study, we present a review of three chapters, more than one hundred research papers, and other several associated papers of different work were done, which have been studied in the capacity of SV analysis and discovery applying ML and DMT. We extensively have been reviewed earlier work that has applied DM and ML procedures for software weakness investigation and discovery. The earlier work has been studied which is presented in Table-1. Additionally, comprehensive summaries of survey papers have been provided. The researcher was taking the comprehension of main stages; methods and assessment methodology in each stage and the summary of table-1 and summary of survey papers have been argued the successes and limits of each step by way of exposed spaces forthcoming research to every type. The chief determined through these classified re-scenarios of earlier research providing a planned scenario of the success and inadequacies of earlier findings for possible researchers in the emerging ground. Application of DM & ML methods aimed at SVA and discovery gives several chances, and we pondered several exposed complications and exposed research areas for upcoming work in each step. Further, the earlier methods are still immature public investigation in the areas regarding SVA and discovery using ML and data mining techniques. We also analyze research trends and research focus, finally concluding with research challenges and hoping for research guidelines for ML & DMTs in AVSD. Future work, as research gaps have been mentioned in table-1. Additionally, I will work on one of the holes which have been mentioned in the summary of book chapters table-1 and summariesof surveypapers.

REFERENCES

[1]Martin, C. Dianne. "TAKING THE HIGH ROAD White hat, black hat: the ethics of cybersecurity." *ACM Inroads* 8.1 (2017): 33-35.https://dl.acm.org/doi/fullHtml/10.1145/3043955

[2] Martin, C. Dianne. "TAKING THE HIGH ROAD White hat, black hat: the ethics of cybersecurity." *ACM Inroads* 8.1 (2017): 33-35. https://dl.acm.org/doi/fullHtml/10.1145/3043955

[3] Martin, C. Dianne. "TAKING THE HIGH ROAD White hat, black hat: the ethics of cybersecurity." *ACM Inroads* 8.1 (2017): 33-35.https://dl.acm.org/doi/fullHtml/10.1145/3043955

[4] Harer, Jacob A., et al. "Automated software vulnerability detection with machine learning." *arXiv preprint arXiv:1803.04497* (2018). https://arxiv.org/abs/1803.04497

[5] Russell, Rebecca, et al. "Automated vulnerability detection in source code using deep representation learning." 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018. https://ieeexplore.ieee.org/abstract/document/8614145

[6] Hiller, Janine S. "The benefit corporation and corporate social responsibility." *Journal of Business Ethics* 118.2 (2013): 287-301. https://link.springer.com/article/10.1007/s10551-012-1580-3

[7] Li, Zhen, et al. "VulDeeLocator: A Deep Learning-based Fine-grained Vulnerability Detector." *arXiv preprint arXiv:2001.02350* (2020).https://arxiv.org/abs/2001.02350

[8] Harer, Jacob A., et al. "Automated software vulnerability detection with machine learning." *arXiv preprint arXiv:1803.04497* (2018). https://arxiv.org/abs/1803.04497

[9]Russell, Rebecca, et al. "Automated vulnerability detection in source code using deep representation learning." *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018. https://ieeexplore.ieee.org/abstract/document/8614145

[10] Li, Yuwei, et al. "V-fuzz: Vulnerability-oriented evolutionary fuzzing." *arXiv preprint arXiv:1901.01142* (2019). https://arxiv.org/abs/1901.01142

[11] Russell, Rebecca, et al. "Automated vulnerability detection in source code using deep representation learning." *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018. https://ieeexplore.ieee.org/abstract/document/8614145

[12] Movahedi, Yazdan, Michel Cukier, and IlirGashi. "Vulnerability prediction capability: A comparison between vulnerability discovery models and neural network models." *Computers & Security* 87(2019):101596.https://www.sciencedirect.com/science/article/pii/S016740481930158

[13] Al-Msie'deen, Ra'Fat. "Automatic labeling of the object-oriented source code: The lotus approach." *arXiv preprint arXiv:1803.00048* (2018). https://arxiv.org/abs/1803.00048

[14] Seng, Lim Kah, NorafidaIthnin, and Syed ZainudeenMohdShaid. "Automating Penetration Testing Within an Ambiguous Testing Environment." *International Journal of Innovative Computing* 8.3 (2018). https://ijic.utm.my/index.php/ijic/article/view/180

[15] Mallikarjunan, K. Narasimha, S. Mercy Shalinie, and G. Preetha. "Real Time Attacker Behavior Pattern Discovery and Profiling Using Fuzzy Rules." *Journal of Internet Technology* 19.5 (2018): 1567-1575. https://jit.ndhu.edu.tw/article/view/1777

[16] Cigoj, Primoz, and BorkaJermanBlazic. "An Intelligent and Automated WCMS Vulnerability-Discovery Tool: The Current State of the Web." *IEEE Access* 7 (2019): 175466-175473. https://ieeexplore.ieee.org/abstract/document/8922605

[17] Lee, Kyungroul, and KangbinYim. "Cybersecurity threats based on machine learning-based offensive technique for password authentication." *Applied Sciences* 10.4 (2020): 1286. https://www.mdpi.com/2076-3417/10/4/1286

[18] Ranganath, Venkatesh-Prasad, and Joydeep Mitra. "Are free Android app security analysis tools effective in detecting known vulnerabilities?." *Empirical Software Engineering* 25.1 (2020): 178-219. https://link.springer.com/article/10.1007%2Fs10664-019-09749-y

[19] Ghanem, Mohamed C., and Thomas M. Chen. "Reinforcement learning for efficient network penetration testing." *Information* 11.1 (2020): 6. https://www.mdpi.com/2078-2489/11/1/6

[20] Hasan, Mahmudul, et al. "Attack and anomaly detection in IoT sensors in IoT sites using machine

learning approaches." *Internet of Things* 7 (2019): 100059.https://www.sciencedirect.com/science/article/pii/S2542660519300241

[21] Shafiq, D. A., Jhanjhi, N. Z., Abdullah, A., &Alzain, M. A. (2021). A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications. *IEEE Access*, *9*, 41731-41744.https://ieeexplore.ieee.org/abstract/document/9374987

[22] Almusaylim, Z. A., Alhumam, A., Mansoor, W., Chatterjee, P., &Jhanjhi, N. Z. (2020). Detection and Mitigation of RPL Rank and Version Number Attacks in Smart Internet of Things.https://www.preprints.org/manuscript/202007.0476/v1

[23] Vinayakumar, R., et al. "ScaleNet: scalable and hybrid framework for cyber threat situational awareness based on DNS, URL, and email data analysis." *Journal of Cyber Security andMobility* 8.2 (2019): 189-240. https://www.riverpublishers.com/journal_read_html_article.php?j=JCSM/8/2/3

[24] Vinayakumar, R., et al. "Deep learning approach for intelligent intrusion detection system." *IEEE Access* 7 (2019): 41525-41550. https://ieeexplore.ieee.org/abstract/document/8681044

[25] Muzammal, S. M., Murugesan, R. K., Jhanjhi, N. Z., & Jung, L. T. (2020, October). SMTrust: Proposing Trust-Based Secure Routing Protocol for RPL Attacks for IoT Applications. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 305-310). IEEE.https://ieeexplore.ieee.org/abstract/document/9247818

[26] Humayun, M., Jhanjhi, N. Z., Hamid, B., & Ahmed, G. (2020). Emerging smart logistics and transportation using IoT and blockchain. *IEEE Internet of Things Magazine*, *3*(2), 58-62.https://ieeexplore.ieee.org/abstract/document/9125435

[27] Alamri, Malak, N. Z. Jhanjhi, and Mamoona Humayun. "Blockchain for Internet of Things (IoT) Research Issues Challenges & Future Directions: A Review." *Int. J. Comput. Sci. Netw. Secur* 19 (2019): 244-258.https://expert.taylors.edu.my/file/rems/publication/109566_6018_1.pdf

[28] He, Ya-Ling, et al. "Review of the solar flux distribution in concentrated solar power: non-uniform features, challenges, and solutions." *Applied Thermal Engineering* 149 (2019): 448-474. https://www.sciencedirect.com/science/article/abs/pii/S1359431118360769

[29] Stevens, Rock, et al. "Summoning demons: The pursuit of exploitable bugs in machine learning." *arXiv preprint arXiv:1701.04739* (2017). https://arxiv.org/abs/1701.04739

[30] Sanguino, Luis Alberto Benthin, and Rafael Uetz. "Software vulnerability analysis using CPE and CVE." *arXiv preprint arXiv:1705.05347* (2017). https://arxiv.org/abs/1705.05347

[31] Dam, HoaKhanh, et al. "Automatic feature learning for predicting vulnerable software components." *IEEE Transactions on Software Engineering* (2018). https://ieeexplore.ieee.org/abstract/document/8540022

[32] Maghrabi, Louai&Pfluegel, Eckhard& al-Fagih, Luluwah& Graf, Roman &Settanni, Giuseppe &Skopik, Florian. (2017). Improved software vulnerability patching techniques using CVSS and game theory. 1-6. 10.1109/CyberSecPODS.2017.8074856. https://ieeexplore.ieee.org/abstract/document/8074856

33] Li, Hongbin, and Li-An Zhou. "Political turnover and economic performance: the incentive role of personnel control in China." *Journal of public economics* 89.9-10 (2005):1743-1762. https://www.sciencedirect.com/science/article/abs/pii/S0047272704001355

[34] Singh, Umesh & Joshi, Chanchala& Singh, Suyash. (2017). Zero day Attacks Defense Technique for Protecting System against Unknown Vulnerabilities.https://www.sciencedirect.com/science/article/abs/pii/S0047272704001355

[35] Wasylkowski, Andrzej, Andreas Zeller, and Christian Lindig. "Detecting object usage anomalies." *Proceedings of the the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering*. 2007. https://www.sciencedirect.com/science/article/abs/pii/S0047272704001355

[36] Singh, U. K., Joshi, C., & Singh, S. K. (2016). ZDAR system: defending against the unknown. *International Journal of Computer Science and Mobile Computing*, *5*(12), 143-149.https://link.springer.com/chapter/10.1007/978-3-642-54525-2_27

[37]Ponnusamy, V., Jhanjhi, N. Z., & Humayun, M. (2020). Fostering Public-Private Partnership: Between Governments and Technologists in Developing National Cybersecurity Framework. In *Employing Recent Technologies for Improved Digital Governance* (pp. 237-255). IGI Global.https://www.igi-global.com/chapter/fostering-public-private-partnership/245984

[38] C. Chung, C. Chen, W. Shih, T. Lin, R. Yeh and I. Wang, "Automated machine learning for Internet of Things," *2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, Taipei, 2017 https://ieeexplore.ieee.org/abstract/document/7991112

[39] Le Roux, J. L., et al. "OSPF protocol extensions for path computation element (PCE) discovery." *RFC5088, January* (2008).https://www.hjp.at/doc/rfc/rfc5088.html

[40] S. VenkatramuluAssociate Professor Computer Science and Engineering Department, Kakatiya Institute of Technology and Science, Warangal, Telangana, 506015, India. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) https://www.ripublication.com/ijaer17/ijaerv12n24_10.pdf

[41] Cigoj, Primoz &Jerman, Borka. (2019). An Intelligent and Automated WCMS Vulnerability-Discovery Tool: The Current State of the Web. IEEE Access. 7. 1-1. 10.1109/ACCESS.2019.2957573. https://ieeexplore.ieee.org/abstract/document/8922605

[42]Alferidah, D. K., &Jhanjhi, N. Z. (2020). A Review on Security and Privacy Issues and Challenges in Internet of Things. *International Journal of Computer Science and Network Security IJCSNS*, *20*(4), 263-286.https://expert.taylors.edu.my/file/rems/publication/109566_7213_1.pdf

[43] Ranganath, Venkatesh-Prasad, and Joydeep Mitra. "Are free android app security analysis tools effective in detecting known vulnerabilities?." *Empirical Software Engineering* (2019): 1-42. https://link.springer.com/article/10.1007%2Fs10664-019-09749-y

[44] Imran, M., Faisal, M., & Islam, N. (2019, November). Problems and Vulnerabilities of Ethical Hacking in Pakistan. In *2019 Second International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT)* (pp. 1-6). IEEE.https://ieeexplore.ieee.org/abstract/document/8955459

[45] Ghanem, Mohamed C., and Thomas M. Chen. "Reinforcement Learning for Efficient Network Penetration Testing." *Information* 11.1 (2020): 6.https://www.mdpi.com/2078-2489/11/1/6

[46] Yu, Miao, et al. "A Survey of Security Vulnerability Analysis, Discovery, Detection, and Mitigation on IoT Devices." *Future Internet* 12.2 (2020): 27.https://www.mdpi.com/1999-5903/12/2/27

[47] Ali, Bako, and Ali Ismail Awad. "Cyber and physical security vulnerability assessment for IoT-based smart homes." *Sensors* 18.3 (2018): 817.https://www.mdpi.com/1424-8220/18/3/817

[48] Mohamudally, Nawaz, and MahejabeenPeermamode-Mohaboob. "Building an anomaly detection engine (ADE) for Iot smart applications." *Procedia computer science* 134 (2018): 10-17.https://www.sciencedirect.com/science/article/pii/S1877050918311013

[49] Hasan, Mahmudul, et al. "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches." *Internet of Things* 7 (2019): 100059.https://www.sciencedirect.com/science/article/pii/S2542660519300241

[50] Hasan, M., Islam, M. M., Zarif, M. I. I., & Hashem, M. M. A. (2019). Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things*, *7*, 100059.https://www.sciencedirect.com/science/article/pii/S2542660519300241

[51] Arko, A. R., Khan, S. H., Preety, A., & Biswas, M. H. (2019). *Anomaly detection In IoT using machine learning algorithms* (Doctoral dissertation, Brac University).http://dspace.bracu.ac.bd/xmlui/handle/10361/12776

[52] Hussain, F., Hassan, S. A., Hussain, R., & Hossain, E. (2020). Machine learning for resource management in cellular and IoT networks: Potentials, current solutions, and open challenges. *IEEE Communications Surveys & Tutorials.* https://ieeexplore.ieee.org/abstract/document/8951180

[53] Grammatikis, Panagiotis I. Radoglou, Panagiotis G. Sarigiannidis, and Ioannis D. Moscholios. "Securing the Internet of Things: challenges, threats and solutions." *Internet of Things* 5 (2019): 41-70. https://www.sciencedirect.com/science/article/pii/S2542660518301161

[54] Vinayakumar, R., et al. "ScaleNet: scalable and hybrid framework for cyber threat situational awareness based on DNS, URL, and email data analysis." *Journal of Cyber Security and Mobility* 8.2 (2019): 189-240.https://www.riverpublishers.com/journal_read_html_article.php?j=JCSM/8/2/3

[55] Vinayakumar, R., et al. "Deep learning approach for intelligent intrusion detection system." *IEEE Access* 7 (2019): 41525-41550.https://ieeexplore.ieee.org/abstract/document/8681044

[56] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Robust

intelligent malware detection using deep learning. *IEEE Access*, *7*, 46717-46738.https://ieeexplore.ieee.org/abstract/document/8681127

[57] Mohan, Vysakh S., et al. "SPOOF net: syntactic patterns for identification of ominous online factors." 2018 IEEE Security and Privacy Workshops (SPW). IEEE, 2018.https://ieeexplore.ieee.org/abstract/document/8424657

[58] Vigneswaran, K. Rahul, et al. "Evaluating shallow and deep neural networks for network intrusion detection systems in cyber security." *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2018.https://ieeexplore.ieee.org/abstract/document/8494096

[59] Vinayakumar, R., PrabaharanPoornachandran, and K. P. Soman. "Scalable framework for cyber threat situational awareness based on domain name systems data analysis." *Big data in engineering applications*. Springer, Singapore, 2018. 113-142.https://link.springer.com/chapter/10.1007/978-981-10-8476-8_6

[60] Vinayakumar, R., et al. "Detecting Android malware using long short-term memory (LSTM)." *Journal of Intelligent & Fuzzy Systems* 34.3 (2018): 1277-1288.https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs169424

[61] Vinayakumar, R., K. P. Soman, and PrabaharanPoornachandran. "Evaluating deep learning approaches to characterize and classify malicious URL's." *Journal of Intelligent & Fuzzy Systems* 34.3 (2018): 1333-1343.https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs169429

[62] Vinayakumar, R., et al. "Evaluating deep learning approaches to characterize and classify the DGAs at scale." *Journal of.*https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs169423

[63] Lu, J., Bi, J., Shang, C., Yue, C., Morillo, R., Ware, S., Kamath, J., Bamis, A. and Russell, A.: Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning, *Proc. ACM onInteractive, Mobile, Wearable and Ubiquitous Technologies*, Vol.2, pp.1–21 (2018). https://dl.acm.org/doi/abs/10.1145/3191753

[64] Saeb, S., Lattie, E., Kording, K. and Mohr, D.: Mobile phone detection of semantic location and its relationship to depression and anxiety, *JMIR mHealth and uHealth*, Vol.5, No.8, p.e112 (2017). https://mhealth.jmir.org/2017/8/e112/

[65] Sabatelli, M., Osmani, V., Mayora, O., Gr¨unerbl, A. and Lukowicz, P.: Correlation of significant places with self-reported state of bipolar disorder patients, *Proc. 4th International Conference on WirelessMobile Communication and Healthcare* (*MobiHealth*), pp.116–119 (2015). https://ieeexplore.ieee.org/abstract/document/7015923

[66] Huang, Y., Xiong, H., Leach, K., Zhang, Y. and Barnes, L.: Assessing social anxiety using GPS trajectories and point-of-interest data, *Proc. 2016 ACM International Joint Conference on Pervasive andUbiquitous Computing* (*UbiComp): Adjunct* (2016). https://dl.acm.org/doi/abs/10.1145/2971648.2971761

[67] Boukhechba, M., Daros, A., Fua, K., Chow, P., Teachman, B. and Barnes, L.: DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones, *Smart Health,CHASE 2018 Special Issue*, Vol.9-10, pp.192–203 (2018). https://www.sciencedirect.com/science/article/abs/pii/S2352648318300400

[68] Mehrotra, A. and Musolesi, M.: Using autoencoders to automatically extract mobility features for predicting depressive states, *Proc.ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol.2, pp.1–20 (2018). https://dl.acm.org/doi/abs/10.1145/3264937

[69] Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M. andWeidt, S.: Mobile sensing and support for people with depression: A pilot trial in the wild, *JMIR MhealthUhealth*, Vol.4, p.e111 (2016). https://mhealth.jmir.org/2016/3/e111/

[70] St¨utz, T., Kowar, T., Kager, M., Tiefengrabner, M., Stuppner, M., Blechert, J., Wilhelm, F. and Ginzinger, S.: Smartphone based stress prediction, *Proc. International Conference on User Modeling, Adaptation,and Personalization*, pp.240–251 (2015). https://link.springer.com/chapter/10.1007/978-3-319-20267-9_20

[71]Almusaylim, Z. A., Alhumam, A., &Jhanjhi, N. Z. (2020). Proposing a secure RPL based internet of things routing protocol: a review. *Ad Hoc Networks*, *101*, 102096.https://www.sciencedirect.com/science/article/abs/pii/S1570870519308388

[72] Tron, T., Resheff, Y., Bazhmin, M., Peled, A. and Weinshall, D.: Real-time schizophrenia monitoring using wearable motion sensitive devices, *Proc. 7th EAI International Conference on Wireless*

MobileCommunication and Healthcare (*MobiHealth*) (2017). https://link.springer.com/chapter/10.1007/978-3-319-98551-0_28

[73] Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M. and Riper, H.: Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: An explorative study, *Journalof Medical Internet Research*, Vol.18, No.3, p.e72 (2016). https://www.jmir.org/2016/3/e72/

[74] Rabbi, M., Ali, S., Choudhury, T. and Berke, E.: Passive and In-Situ assessment of mental and physical well-being using mobile sensors, *Proc. 2011 ACM International Joint Conference on Pervasive andUbiquitous Computing* (*UbiComp*), pp.385–394 (2011). https://dl.acm.org/doi/abs/10.1145/2030112.2030164

[75] Ferdous, R., Osmani, V. and Mayora-Ibarra, O.: Smartphone app usage as a predictor of perceived stress levels at workplace, *Proc. 9th International Conference on Pervasive Computing Technologies forHealthcare*(*PervasiveHealth*), pp.225–228 (2015). https://ieeexplore.ieee.org/abstract/document/7349403

[75] Mehrotra, A., Hendley, R. and Musolesi, M.: Towards multi-modal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction, *Proc. 2016 ACM International JointConference on Pervasive and Ubiquitous Computing* (*UbiComp*)*:Adjunct*, pp.1132–1138 (2016). https://dl.acm.org/doi/abs/10.1145/2968219.2968299

[76] Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S., McInnis, M., Ajilore, O., Nelson, P., Ryan, K. and Leow, A.: Predicting mood disturbance severity with mobile phone keystroke metadata: The biaffect digital phenotyping study, *Journal of Medical Internet Research*, Vol.20, No.7, p.e241 (2018). https://www.jmir.org/2018/7/e241/

[77] Faurholt-Jepsen, M., Vinberg, M., Frost, M., Christensen, E., Bardram, J. and Kessing, L.: Smartphone data as an electronic biomarker of illness activity in bipolar disorder, *European Psychiatry*, Vol.17, No.7, pp.28–31 (2015). https://onlinelibrary.wiley.com/doi/abs/10.1111/bdi.12332

[78] Buddi, P., Prasad, V.V.R., Sunitha, K.V.N., Reddy, N.C.S. and Anil, C.H.: DetectStress: A novel stress detection system based on smartphone and wireless physical activity tracker, *Proc. 1st InternationalConference on Artificial Intelligence and Cognitive Computing* (2018). https://link.springer.com/chapter/10.1007/978-981-13-1580-0_7

[79] Eyben, F.,W¨ollmer, M. and Schuller, B.W.: Opensmile: The munich versatile and fast open-source audio feature extractor, *Proc. ACMMultimedia*(2010). https://dl.acm.org/doi/abs/10.1145/1873951.1874246

[80] Tacconi, D., Mayora, O., Lukowicz, P., Arnrich, B., Setz, C., Tr¨oster, G. and Haring, C.: Activity and emotion recognition to support early diagnosis of psychiatric diseases, *Proc. 2nd International Conferenceon Pervasive Computing Technologies for Healthcare* (*PervasiveHealth*), pp.100–102 (2008). https://ieeexplore.ieee.org/abstract/document/4571041/

[81] Guidi, A., Salvi, S., Ottaviano, M., Gentili, C., Bertschy, G., De Rossi, D., Scilingo, E. and Vanello, N.: Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: System performance evaluation and case study, *Sensors*, Vol.15, pp.28070–28087 (2015). https://www.mdpi.com/1424-8220/15/11/28070

[82] Tanaka, H., phd, Taira, K., Arakawa, M., Masuda, A., Yamamoto, Y., Komoda, Y., Kadegaru, H. and Shirakawa, S.: An examination of sleep health, lifestyle and mental health in junor high school students, *Psychiatry and Clinical Neurosciences*, Vol.56, pp.235–236 (2002). https://onlinelibrary.wiley.com/doi/full/10.1046/j.1440-1819.2002.00997.x

[83] Abdullah, S., Matthews, M., Murnane, E., Gay, G. and Choudhury, T.: Towards circadian computing: "Early to bed and early to rise" makes some of us unhealthy and sleep deprived, *Proc. 2014 ACMInternational Joint Conference on Pervasive and Ubiquitous Computing* (*UbiComp*), pp.673–684 (2014). https://dl.acm.org/doi/abs/10.1145/2632048.2632100

[84] Cho, K.: Chronic 'jet lag' produces temporal lobe atrophy and spatial cognitive deficits, *Nature Neuroscience*, Vol.4, pp.567–568 (2001). https://www.nature.com/articles/nn0601_567

[85] Staples, P., Torous, J., Barnett, I., Carlson, K., Sandoval, L., Keshavan, M. and Onnela, J.-P.: A comparison of passive and active estimates of sleep in a cohort with schizophrenia, *npjSchizophrenia*,Vol.3, No.37 (2017). https://www.nature.com/articles/s41537-017-0038-0

[86] Freedman, D., Pisani, R. and Purves, R.: *Statistics*, W.W. Norton (1998). https://amstat.tandfonline.com/doi/full/10.1080/10691898.2002.11910665

[87] Lu, J., Bi, J., Shang, C., Yue, C., Morillo, R., Ware, S., Kamath, J., Bamis, A. and Russell, A.: Joint modeling of heterogeneous sensing data for depression assessment

via multi-task learning, *Proc. ACM onInteractive, Mobile, Wearable and Ubiquitous Technologies*, Vol.2, pp.1–21 (2018). https://dl.acm.org/doi/abs/10.1145/3191753

[88] Hastie, T., Tibshirani, R. and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springerseries in statistics, Springer (2009). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.158.8831&rep=rep1&type=pdf

[89] Salzberg, S.L.: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., *Machine Learning*, Vol.16, No.3, pp.235–240 (1994). https://link.springer.com/content/pdf/10.1007/BF00993309.pdf

[90] Breiman, L.: Random forests, *Machine Learning*, Vol.45, No.1, pp.5–32 (2001). Tanaka, H., phd, Taira, K., Arakawa, M., Masuda, A., Yamamoto, Y., Komoda, Y., Kadegaru, H. and Shirakawa, S.: An examination of sleep health, lifestyle and mental health in junor high school students, *Psychiatry and Clinical Neurosciences*, Vol.56, pp.235–236 (2002). https://onlinelibrary.wiley.com/doi/full/10.1046/j.1440-1819.2002.00997.x

[91] Cover, T. and Hart, P.: Nearest neighbor pattern classification, *IEEE Trans. Information Theory*, Vol.13, pp.21–27 (1967). https://ieeexplore.ieee.org/abstract/document/1053964

[92] Tibshirani, R.: Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B* (*Methodological*), Vol.58, No.1, pp.267–288 (1996). https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x

[93] Ng, S. K., Krishnan, T., & McLachlan, G. J. (2012). The EM algorithm. In *Handbook of computational statistics* (pp. 139-172). Springer, Berlin, Heidelberg.https://link.springer.com/chapter/10.1007/978-3-642-21551-3_6

[94] Duda, R., Hart, P. and Stork, D.: *Pattern classification*,Wiley (2012). https://www.eejournal.ktu.lt/index.php/elt/article/view/1816

[95] Sun, J., Lu, J., Xu, T. and Bi, J.: Multi-view sparse co-clustering via proximal alternating linearized minimization, *Proc. 32nd InternationalConference on Machine Learning* (*ICML*), pp.757–766 (2015). http://proceedings.mlr.press/v37/sunb15.pdf

[96] Ozgur, C., Colliau, T., Rogers, G., Hughes, Z. and Bennie, E.: Mat- Lab vs. Python vs. R, *Journal of Data Science*, Vol.15, pp.355–372 (2017). https://www.jstage.jst.go.jp/article/ipsjjip/28/0/28_16/_article/-char/ja/

[97] Kromme, J.: Python & R vs. SPSS & SAS (2017), available from _https://www.r-bloggers.com/python-r-vs-spss-sas/_.

[98] Flach, P.A. and Lachiche, N.: Confirmation-Guided Discovery of First-Order Rules with Tertius, *Machine Learning*, Vol.42, No.1, pp.61–95 (2001). https://link.springer.com/article/10.1023/A:1007656703224

[99] Lu, J., Bi, J., Shang, C., Yue, C., Morillo, R., Ware, S., Kamath, J., Bamis, A. and Russell, A.: Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning, *Proc. ACM onInteractive, Mobile, Wearable and Ubiquitous Technologies*, Vol.2, pp.1–21 (2018). https://dl.acm.org/doi/abs/10.1145/3191753

[100]Li, Zhen & Zou, Deqing& Xu, Shouhuai&Ou, Xinyu&Jin, Hai & Wang, Sujuan& Deng, Zhijun & Zhong, Yuyi. (2018). VulDeePecker: A Deep Learning-Based System for Vulnerability Detection. 10.14722/ndss.2018.23165. https://www.eecis.udel.edu/~hnw/paper/ndss18.pdf

[101]Dhamodaran, Sasikala, and Archana Balmoor. "Future Trends of the Healthcare Data Predictive Analytics using Soft Computing Techniques in Data Science." *CVR Journal of Science and Technology* 16.1 (2019): 89-96.http://cvr.ac.in/ojs/index.php/cvracin/article/view/422

[102] Moholth, Ole Christian, RadmilaJuric, and Karoline Moholth McClenaghan. "Detecting Cyber Security Vulnerabilities through Reactive Programming." *Proceedings of the 52nd Hawaii International Conference on System Sciences*. 2019. https://scholarspace.manoa.hawaii.edu/handle/10125/60157

[103] Sangkaran, T., Abdullah, A., &Jhanjhi, N. Z. (2020). Criminal Community Detection Based on Isomorphic Subgraph Analytics. *Open Computer Science*, *10*(1), 164-174.https://www.degruyter.com/view/journals/comp/10/1/article-p164.xml

[104] Alferidah, D. K., &Jhanjhi, N. Z. (2020). A Review on Security and Privacy Issues and Challenges in Internet of Things. *IJCSNS*, *20*(4), https://expert.taylors.edu.my/file/rems/publication/109566_7213_1.pdf

[105]Almusaylim, Z. A., Alhumam, A., Mansoor, W., Chatterjee, P., &Jhanjhi, N. Z. (2020). Detection and Mitigation of RPL Rank and Version Number Attacks in Smart Internet of Things.https://www.preprints.org/manuscript/202007.0476/v1

[106] Diller, G. P., Kempny, A., Babu-Narayan, S. V., Henrichs, M., Brida, M., Uebing, A., ... &Dimopoulos, K. (2019). Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. European heart journal, 40(13), 1069-1077.https://discovery.ucl.ac.uk/id/eprint/100766 28/

[107]Linthicum, K. P., Schafer, K. M., & Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. Behavioral sciences & the law, 37(3), 214-222.https://onlinelibrary.wiley.com/doi/full/10.1002/bsl.2392

[108]Bisaso, K. R., Anguzu, G. T., Karungi, S. A., Kiragga, A., &Castelnuovo, B. (2017). A survey of machine learning applications in HIV clinical research and care. Computers in biology and medicine, 91, 366-371.https://www.sciencedirect.com/science/article/abs/pii/S001048251730361X