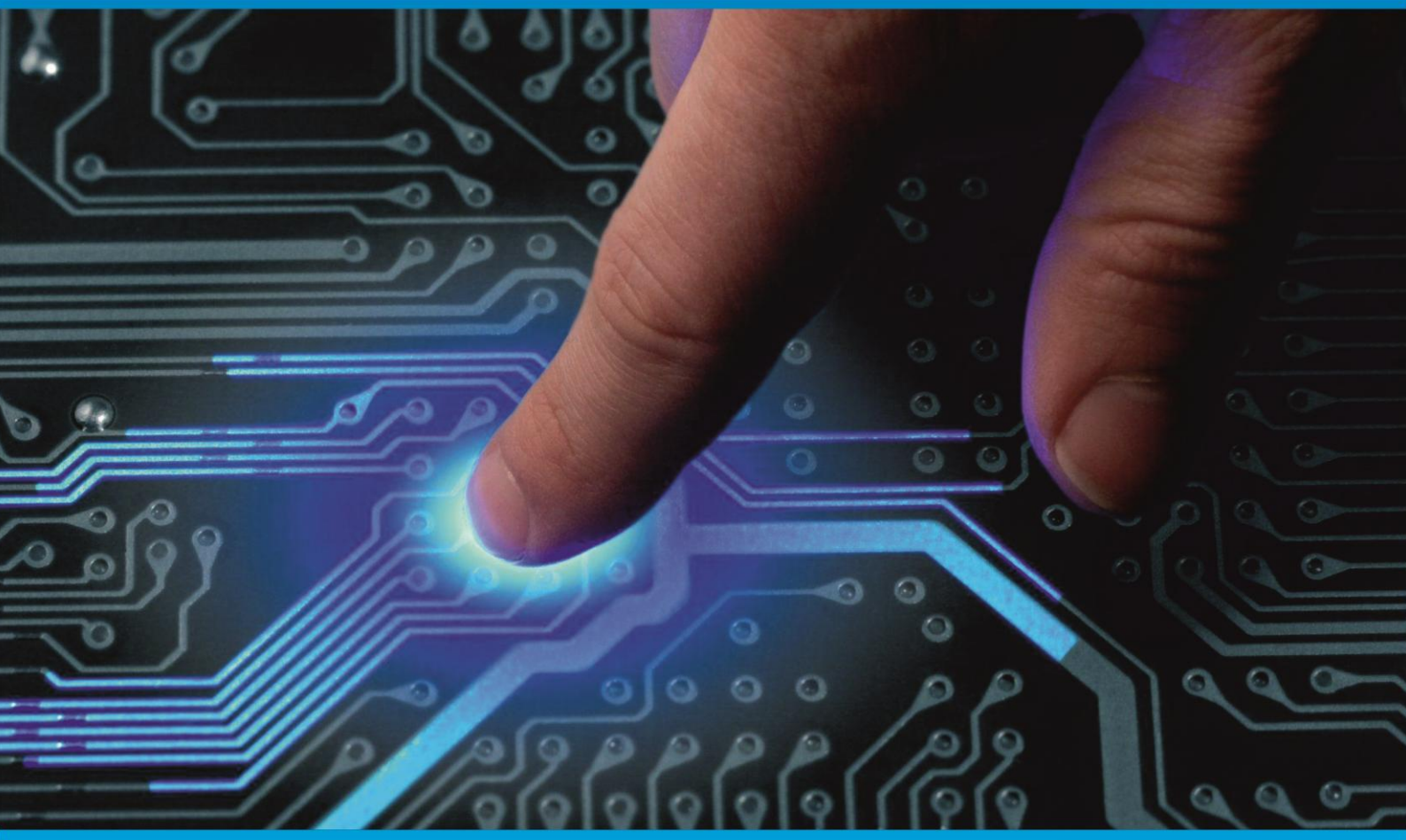
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905197 |

Machine Learning Security: Threats, Countermeasures, and Evaluations

Rohan Shingade, Anuja Kale, Dipali Suryawanshi,Vikas Shinde, Prof.Laxman Deokate

Department of Information Technology, Sinhgad Academy of Engineering, Kondhwa, Pune, India

ABSTRACT: Machine learning is one of the most prevailing techniques in Computer Science, and it has been widely applied in image processing, natural language processing, pattern recognition, cyber security and other fields. Regardless of successful applications of machine learning algorithms in many scenarios, e.g., facial recognition, malware detection, automatic driving and intrusion detection, these algorithms and corresponding training data are vulnerable to a variety of security threats, inducing a significant performance decrease. Hence, it is vital to call for further attention regarding security threats and corresponding defensive technique of machine learning.. This paperaddressed privacy concerns regarding machine learning systems threats by using CNN.

KEYWORDS: Machine learning threats, CNN, security threats, defensive techniques

I. INTRODUCTION

An ML algorithm would use a training set formed of multiple feature vectors, and their associated labels, to build an ML model. This process is called the training or learning phase. When presented with a new test sample, this ML model should give the predicted label (person's name or identifier in face recognition applications). The ability of such an ML model to accurately predict the label is a measure of how well this ML model generalizes to unseen data. It is measured empirically by the test error (generalization error), and it can depend on the quality and quantity of the data used for training the model, what ML algorithm was used to build the model, the selection of ML algorithm hyper parameters (e.g., using cross validation), and even the features' extraction method (if any was required).

To teach an algorithm how to recognize objects in images, we use a specific type of Artificial Neural Network: a Convolutional Neural Network (CNN).It is a special type of Neural Network widely used for recognize the image.

Machine learning has been pervasively used in a wide range of applications due to its technical breakthroughs in recent years. It has demonstrated significant success in dealing with various complex problems, and shows capabilities close to humans or even beyond humans. However, recent studies show that machine learning models are vulnerable to various attacks, which will compromise the security of the models themselves and the application systems. Moreover, such attacks are stealthy due to the unexplained nature of the deep learning models. In this paper, security issue of machine learning is systematically analyzed.

The present-day community accesses advanced technologies, both hardware, and software, at an unprecedented pace in possibly every imaginable field. However, this has resulted in a whole new range of threats in terms of privacy and security. Therefore, there is a demanding need to address the security and privacy perspective of different types of cyber threats which are increasing at a drastic pace with unknown malware. According to a special report, out of seven billion population in the world, about six billion rely on mobile phones or other smart gadgets for banking, shopping, financing, healthcare, internet-of-things (IoT), blockchain applications, posts on social media and for professional information and updates. Therefore, during downloading of the applications on smart devices, there is a strong chance of data leakage and theft. Apart from that, malware is also triggered by corrupt system routines, unauthorized network access to resources and gather sensitive information. To cope

up with these issues, many anti-virus tools, intrusion detection systems, defenders, and latestfirewalls with updated security patches are available.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905197 |

II. LITERATURE SURVEY

Machine learning security is an active research topic and remains an open problem. In [1], a comprehensive survey on machine learning security covers the whole lifecycle of machine learning systems with respect to five major types of attacks and their corresponding countermeasures is presented. A general conclusion is that the threats are real, and new security threats are constantly emerging. For example, studies show that there is a transferability in adversarial examples, which means adversarial examples can generalize well between different machine learning models. It is shown that poisoning examples can also generalize well across different learning models. The transferability can be used to launch attacks in black-box scenarios effectively. Due to the unexplained nature of machine learning models, the essential reasons for these attacks, i.e., is the adversarial example a bug or an intrinsic property of the model, need to be further studied. This paper can hopefully provide comprehensive guidelines for designing secure, robust and private machine learning systems.
The survey, "A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations", [2] aims to present a taxonomy of contemporary insider types, access, level, motivation, insider profiling, effect security property, and methods used by attackers to conduct attacks and a review of notable recent works on insider threat detection, which covers the analyzed behaviors, machine-learning techniques, dataset, detection methodology, and evaluation metrics. Several real cases of insider threats have been analyzed to provide statistical information about insiders. In addition, this survey highlights the challenges faced by other researchers and provides recommendations to minimize obstacles.

Machine learning is one of the most prevailing techniques in Computer Science, and it has been widely applied in image processing, natural language processing, pattern recognition, cybersecurity and other fields. Regardless of successful applications of machine learning algorithms in many scenarios, e.g., facial recognition, malware detection, automatic driving and intrusion detection, these algorithms and corresponding training data are vulnerable to a variety of security threats, inducing a significant performance decrease. Hence, it is vital to call for further attention regarding security threats and corresponding defensive techniques of machine learning, which motivates a comprehensive survey in [3].

With the approaching of the big data age, privacy protection is becoming a unavoidable hurdle in front of us. Motivated by this, we surveyed the major milestones in privacy study up to date from different perspectives, aiming to pave a reliable ground for interested readers to explore this exciting, emerging, and promising field. SHUI YU [4] summarized the outputs of privacy study in different research principles and communities. In particular, he presented the mathematical effort of the related privacy models and frameworks

For privacy concerns to be addressed adequately in today's machine learning systems, the knowledge gap between the machine learning and privacy communities must be bridged. "Privacy Preserving Machine Learning: Threats and Solutions", aims to provide an introduction to the intersection of both fields with special emphasis on the techniques used to protect the data[5].

Machine- and deep-learning algorithms are adopted in many application domains, but in the cyber security field, they are affected by several open issues. In [6], T. Minárik et al consider adversarial attacks where the machine-learning model is compromised to induce an output favourable to the attacker. Literature on this subject is still immature, and most documented examples of adversarial attacks against security systems consider only few algorithms and few application areas.

CHINTHAPALLI SUDHEER et al [7] provide a user-centered computer learning system that affects large data from various security logs, awareness information, and inspector intelligence. This method provides complete configuration and solution for dangerous user detection for the Enterprise System Operating Center. Select machine learning methods in the SOC product environment, evaluate efficiency, IO, host and users to create user-centric features. . Even with simple mechanical learning algorithms, we prove that the learning system can understand more insights from the rankings with the most unbalanced and limited labels. More than 20% of the neurological model of modeling is 5 times that of the current rule-based system.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905197 |

Pattern classification systems are commonly used in adversarial applications, like biometric authentication, network intrusion detection, and spam filtering, in which data can be purposely manipulated by humans to undermine their operation. As this adversarial scenario is not taken into account by classical design methods, pattern classification systems may exhibit vulnerabilities, whose exploitation may severely affect their performance, and consequently limit their practical utility. Several works have addressed the problem of designing robust classifiers against these threats, although mainly focusing on specific applications and kinds of attacks. In this paper, we address one of the main open issues: evaluating at design phase the security of pattern classifiers, namely, the performance degradation under potential attacks they may incur during operation. Asharani V et al proposed a framework for empirical evaluation of classifier security that formalizes and generalizes the main ideas proposed in the literature [8].

Adversarial attacks pose a serious threat to the success of deep learning in practice. This fact has recently lead to a large influx of contributions in this direction. The article "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey", presented the first comprehensive survey on adversarial attacks on deep learning in Computer Vision. NaveedAkhtar and AjmalMianreview the works that design adversarial attacks, analyze the existence of such attacks and propose defenses against them. To emphasize that adversarial attacks are possible in practical conditions, and separately review the contributions that evaluate adversarial attacks in the real-world scenarios. Finally, drawing on the reviewed literature, we provide a broader outlook of this research direction [9].

III. PROPOSED SYSTEM

Insider threat has become a widely accepted issue and one of the major challenges in cyber security. This phenomenon indicates that threats require special detection systems, methods, and tools, which entail the ability to facilitate accurate and fast detection of a malicious insider.

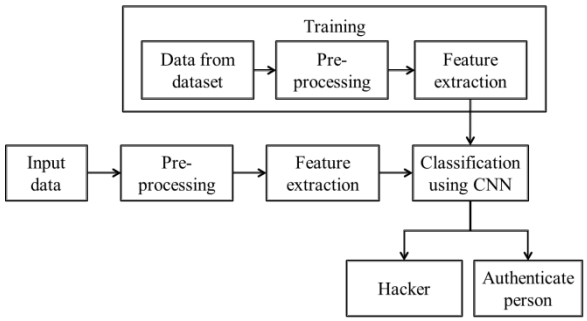


Fig architecture of proposed system

Training: dataset is in the form of excel sheet. Data from datasheet is pre-processed to remove any redundant data. Important feature are extracted from pre-processed data which then fed for classifier for comparison. Data input from user is pre-processed and then features are extracted. Extracted features are similar to extracted on dataset. Classifier as name suggest classifies data from user into genuine or hacker by comparing extracted features from dataset and input. CNN (Convolutional Neural Network) is used as classifier.

IV. CONCLUSION

As machine learning is becoming widely used in many practical applications including but not limited to image processing, natural language processing, pattern recognition, computer vision, intrusion detection, malware identification and autonomous driving, protecting the security of machine learning at both training and inferring phases becomes an urgent need. In this paper, we have presented a system based on CNN to findinsider threat.

REFERENCES

[1] MingfuXueChengxiang Yuan, Heyi Wu, Yushu Zhang, Weiqiang Liu, "Machine Learning Security: Threats, Countermeasures, and Evaluations", Digital Object Identifier 10.1109/ACCESS.2020.2987435

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905197 |

[2] Nasser Al-Mhiqani, Rabiah Ahmad, Z. ZainalAbidin, WarusiaYassin, Aslinda Hassan, KarrarHameedAbdulkareem, NabeelSalih Ali ZahriYunos "A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations" Appl. Sci. 2020, 10, 5208; doi:10.3390/app10155208 www.mdpi.com/journal/applsci

[3] Liu, Q., Li, P., Zhao, W., Cai, W., Yu, S., & Leung, V. C. M. (2018). "A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View", IEEE Access, 6, 12103–12117. doi:10.1109/access.2018.2805680

[4] SHUI YU "Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data" 2169-3536 2016 IEEE VOL 4, 2016 Digital Object Identifier 10.1109/ACCESS.2016.2577036

[5] Mohammad Al-Rubaie J. Morris Chang, "Privacy Preserving Machine Learning: Threats and Solutions", © 2018 IEEE Accepted for publication in IEEE Security and Privacy Magazine 1

[6] T. Minárik, S. Alatalu, S. Biondi, M. Signoretti, I. Tolga, G. Visky, "Addressing Adversarial Attacks Against Security Systems Based on Machine Learning" 2019 11th International Conference on Cyber Conflict: Silent Battle (Eds.) 2019 © NATO CCD COE Publications, Tallinn Complexity International Journal (CIJ) Volume 24, Issue 01, March 2020 Impact Factor (2020): 5.6 ISSN: 1320-0682 http://cij.org.in/Currentvolumeissue2401.aspx 200

[7] ChinthapalliSudheer, M S VenugopalaRao, "User-Centric Meachine Learning Framework For Cyber Security Operation Center" MSVGOPALARAO@GMAIL.COM

[8] Asharani V et al, "Security Evaluation of Pattern Classifiers in Adversarial Environments", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April- 2015, pg. 768-774

[9] NaveedAkhtar and AjmalMian "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey" arXiv:1801.00553v3 [cs.CV] 26 Feb 2018

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details