

# Deep Learning Models for Privacy Risk Assessment in Dynamic Cyber Threat Environments

Suneel Kumar Mogali

Perficient, Inc, USA. E-mail: [suneelmjayshree@gmail.com](mailto:suneelmjayshree@gmail.com)

**Abstract** - Information security management systems and frameworks have embraced traditional risk assessment (RA) methodologies and standards as a cornerstone for secure environments. However, in today's world, where threats are constantly evolving and new vulnerabilities are constantly being found, these approaches encounter numerous challenges. To get around this issue, some have suggested DRA models, which continually and dynamically evaluate risks to an organization's activities in (almost) real time. Connected smart devices, known as the Internet of Things (IoT), have changed the face of modern technology. These advances present new security challenges, but they also bring new opportunities. For intrusion detection systems (IDS), cybersecurity is of the utmost importance. When it comes to protecting Internet of Things (IoT) devices from cyberattacks, Deep Learning has demonstrated encouraging results. Despite intrusion detection systems' (IDS) critical role in protecting sensitive data by detecting and preventing malicious actions, traditional IDS solutions have difficulties when used to the Internet of Things (IoT). This article explores state-of-the-art, Deep Learning-based intrusion detection approaches for Internet of Things security. Recent developments in intrusion detection systems (IDS) for the internet of things (IoT) are reviewed here, with an emphasis on the relevant deep learning algorithms, datasets, attack types, and assessment metrics. This work offers a fresh perspective on managing hazards in system-to-system communication through API calls and helps to tackle this difficulty. Effective threat identification from huge API call datasets is achieved through the introduction of an integrated architecture that integrates deep-learning models, specifically ANN and MLP. In order to improve overall resilience, the detected threats are analyzed to find appropriate mitigations. To ensure that AI models are accessible to all user groups, this work also introduces transparency obligation practices for the whole AI life cycle, beginning with dataset preprocessing and ending with model performance evaluation. These practices include data and methodological transparency as well as SHapley Additive exPlanations (SHAP) analysis. Experiment results showing an average detection accuracy of 88% utilizing the Windows PE Malware API dataset justify the proposed methodology.

**Keywords:** Artificial Intelligence, Deep Learning, Cyber Threat.

## I. INTRODUCTION

The increasing fascination with the use of ICT in a wide range of contexts has made the study of network infrastructure security a major and dynamic field of study. One of the main functions of modern technologies is digital communication. For example, Internet of Things (IoT) devices need secured authentication to guarantee reliable and secure communication between endpoints, as well as a variety of other remote services [1]. The diverse structure of today's networks makes it difficult to maintain security in an ever-evolving and unpredictable setting. Advanced network architectures, like cyber-physical systems, are interconnected webs of smart devices that transmit and receive data. Network traffic is vulnerable to fraud during this process due to factors such as viruses, human mistake, and hackers, which can cause the system to fail, leading to duplicated data and a delay in the planned transmission of the system [2]. Consequently, it is critical to enhance network system security by implementing an attack detection and mitigation model that can guarantee the secrecy, authenticity, and availability of sensitive data.

The ever-changing nature of network systems makes it difficult to predict when new variations or threats may emerge, making it much more difficult to apply a security model in such an environment. Traditional security measures have a hard time spotting these undiscovered dangers. Monitoring the network, protecting it from intrusion, managing incidents, educating and alerting users, and ensuring secure configuration are all necessary components of any security architecture [3]. It takes a lot of time, effort, and research and development to accomplish each feature. It is common practice to classify network assaults as either active or passive. Continuous information extraction without altering the original data at the target device is the hallmark of passive assaults. Conversely, data on the target is frequently altered by active attacks, which either introduce new data or alter the old data. By utilizing an intrusion detection system (IDS) installed in the computer architecture, monitoring operations can identify these types of attacks in a network system.

Effective threat-management methods must be developed through threat analysis, which entails detecting and evaluating possible dangers within a system. As part of its analysis, this procedure takes into account different kinds of data and vulnerabilities [4]. Application, architectural, and device layers are all part of threat management's purview as it seeks to handle these risks. Existing security mechanisms are improved using iterative techniques. In addition, dangers can emerge when there are holes in the system that would be easy for hackers to exploit. It is critical to be aware of the distinction between vulnerabilities and weaknesses. General shortcomings or faults that could cause security difficulties but haven't been exploited yet are called weaknesses. Threat actors can directly exploit vulnerabilities, which are unique, identifiable security flaws [5]. On the other hand, organizations can tackle both existing and new cybersecurity threats head-on if they build their strategies around vulnerabilities.

Cybersecurity professionals use a variety of threat analysis strategies, including goal-based, asset-based, and risk-based approaches, to catalog, rank, and prioritize possible dangers. When it comes to recognizing and reducing risks, different types of analyses provide different viewpoints and methods. Assessing dangers according to their possible influence on the business and how likely they are to materialize is the main goal of risk-based analysis. Using this method, we may better identify the most pressing risks and devote our resources to mitigating them. By illuminating trends and connections in cyber threats that companies encounter, risk-based cyber threat analysis can help strengthen cybersecurity measures [6]. Since learning algorithms have shown great effectiveness in security and privacy, the goal of [7]'s survey is to have them implemented as IoT NIDS. The survey provides an in-depth evaluation of NIDSs utilizing diverse learning strategies for the Internet of Things, in contrast to previous prominent assessments that concentrate on conventional systems. There has been a categorization of IoT threats and detection methodologies in comparison to more traditional forms of security. Next, we give a comprehensive review of the technologies used to construct NIDS, starting with open-source NIDS, open-source network traffic monitoring tools, and freely available network datasets. These tools can be used by both academics and industry to develop and test cutting-edge NIDS solutions.

According to [8], an extensive evaluation of IDSs should be conducted within the framework of IoT ecosystems. In a similar vein, they have thoroughly examined over 40 significant studies out of 324 foundational papers, reviewed numerous highly developed intrusion detection in the IoT, and assessed and addressed unanswered questions. There are four main types of intrusion detection systems (IDS) based on the

available literature: signature-based, specification-based, anomaly-based, and hybrid. The papers were further split into three groups: centralized, distributed, and hybrid. Additionally, two domains of assessment (theoretical and simulation) are taken into account, along with nine attack types (DoS/DDoS, Sybil, selective forwarding, black hole, sinkhole, jamming, replay, bogus data, wormhole). It also covered the pros and cons of several IDSs. By fixing the issues with current approaches, we can build IDSs that are better in the long run. Intrusion detection systems (IDS) that use machine learning (ML) and deep learning (DL) were reviewed in [9]. Explored the Internet of Things (IoT) and its protocols, as well as its architecture, weaknesses, and attacks at the protocol level. In addition, studies have suggested IDS methodology for the Internet of Things (IoT) or attack detection approaches for the IoT that could be part of an IDS, with an emphasis on the various ML and DL techniques that are available for IoT IDS and how researchers have used them.

## II. LITERATURE REVIEW

Automating processes and enabling solutions to learn on their own, machine learning is the backbone of digital era progress [10]. There is a need to evaluate its security and resilience against adversarial machine learning (AML) assaults since applications such as intelligent firewalls [13], safeguarding autonomous car and IoT systems [12], and spam-filtering systems [11] are examples. In order to compromise security-sensitive apps and compromise their integrity or secrecy, adversaries heavily use machine learning to lower the victim's performance, insert a backdoor, or abuse its privacy.

It is a poisoning attack to compromise integrity by tampering with training datasets or model parameters. These poisoning attacks are already in existence: feature collision, convex polytope, random label flipping, and fast gradient sign technique (FGSM). One kind of evasion attempt is tampering with the testing dataset [14]. At the same time, Clarifai's content moderation classifier is vulnerable to model inversion and inference attacks, which can expose the parameters of the targeted model or extrapolate manipulated data to infer the expected output, allowing an attacker to evaluate and study the model's functional capabilities while compromising its privacy.

Attacks against real-time ML systems have recently been effective, proving the viability of adversarial ML attacks. with an inference accuracy of 86.6% on GPT-3.5 and 50% on GPT-4, assaulted ChatGPT, Claude, and Bard with [15]. Additionally, [16] successfully attacked the commercial Alibaba API 97% of the time. These assaults show how important it is to do thorough research on how to make ML models resilient, with a focus on security-by-design solutions

that aim to make the development process secure and resilient rather than individual models. We have compared and contrasted four adversarial attack types—model inversion, poisoning, evasion, and membership inference—that endanger the development of machine learning models in this literature review.

To document the lack of DRA methods adapted to the intricate requirements of an ICS, [17] zeroed in on quantitative RA approaches in ICSs. They examined the primary technologies employed by each RA approach and came to the conclusion that DRA is an essential tool for improving network safety. The authors of [18] sought to highlight the importance of DRA in the field of information systems. Despite the availability of several standards, they still see RA more as a periodic than a real-time process, they claimed. The classic RA has given way to DRA in recent years. DRA uses attack trees to define potential attack patterns on the information system and input data from databases like the National Vulnerability Database (NVD) to keep the system up-to-date.

### Enabling the Evolutionary Approach in Cybersecurity

The evolutionary approach to cybersecurity firmly adheres to the principles of capability development and preventive protection mechanisms. This requires shifting from traditional, static security methods to more dynamic forms of cyber protection [19]. Organizations can enhance their security posture and adaptability to emerging threats by adopting this evolutionary approach. A methodological strategy based on an evolutionary model can be employed to ward off cyber assaults. Using lessons learned from previous security breaches, this model illustrates how contemporary cyber-physical systems can adapt to new threats. Cybersecurity that takes an evolutionary approach allows businesses to strengthen their defenses against cyberattacks by evolving and adapting to new threats. Among these endeavors is the cultivation of threat anticipation, assault response preparation, and incident recovery speed.

On top of that, businesses can improve their proactive cybersecurity skills over time by using an evolutionary strategy. Building a strong security culture, increasing security awareness across the company, and creating training programs for staff are all part of this. As a result, the evolutionary strategy helps both defend against and adapt to cyber assaults [20].

The evolving strategy also facilitates cross-stakeholder and partner collaboration in the fight against cybercrime. This includes sharing details regarding current cybersecurity risks, effective methods for reducing those risks, and new research

and advances in the area. Collaborating allows organizations to better detect, assess, and resolve threats.

When it comes to cybersecurity, the evolutionary method offers a complete and adaptable framework that may help enterprises tackle the problems of the digital age. To better withstand cyber assaults and future threats, companies should adopt a mentality of capability growth and implement preventative protection measures [21]. Figure 1 depicts our cybersecurity strategy for the modern digital age, which is based on the resilience paradigm. Several steps are involved in this approach:



Figure 1: Dynamic security approach

#### 1. Explore Digital Skills:

To tackle cybersecurity concerns, one must first determine what digital skills are necessary. To build strong cybersecurity, it is necessary to collect data regarding technical capabilities, open procedures, and linked human aspects.

#### 2. Analyze Threats and Risks:

Get started by cataloguing the digital abilities needed to counteract cyber dangers. Building strong cybersecurity requires gathering knowledge regarding technical capabilities, transparent processes, and interrelated human aspects.

#### 3. Develop Threat Response:

Come up with a strategy for dealing with threats that includes safeguards, identification, and action. Plan ways to restore systems from incidents and counterattacks quickly and efficiently.

#### 4. Integrate Security and Resilience:

Organizational strategy should incorporate cybersecurity and resilience. To achieve attack resilience and adaptation,

build a framework that integrates technology, processes, and human factors.

**5. Enhance Readiness and Flexibility:**

Strengthen preparedness and adaptability to deal with cyber threats. Create and implement tried-and-true incident response plans; can quickly restore systems and data. Be flexible in the face of evolving dangers by learning from your mistakes.

**6. Promote Participation and Collaboration:**

Inspire the many players in the cybersecurity ecosystem to work together and take an active role. Motivate enterprises to take part in resilience and security activities by offering them financial and non-financial incentives. This includes vendors, business partners, end-users, and other groups.

The foundation of this strategy is a comprehensive view of cybersecurity that incorporates human factors, procedures, and technology to build resilience against threats. Businesses can improve their defenses against cyber threats by encouraging teamwork and concentrating on resilience and security integration. Critical components like threat intelligence and capability monitoring enable evolving strategies. Continual monitoring of user actions, system records, and network traffic is essential for organizations [22]. New risks, vulnerabilities, and attack trends can be discovered by businesses through the collection and analysis of pertinent data. Proactive decision-making and the deployment of adaptable defenses are based on this knowledge.

**III. METHODOLOGY**

Two primary sequential steps make up the suggested method, as depicted in Figure 2: threat identification pertaining to APIs and threat management. Multiple phases make up each phase, offering a clear and organized technique. In the first stage, deep-learning models are utilized to detect threats through APIs. Data preparation, model results, and transparency obligation procedures are the four main areas covered by this stage. Using deep learning in the first step is beneficial since it can easily process large amounts of API data. Predicting patterns, optimizing resource allocation, and anticipating maintenance needs are all made easy with this type of analysis. System responsiveness and proactive management and optimization of operations based on predictive analytics are both made possible by the integration of deep-learning algorithms into API frameworks. Phase two of our technique involves analyzing the essential traits that were generated in phase one.

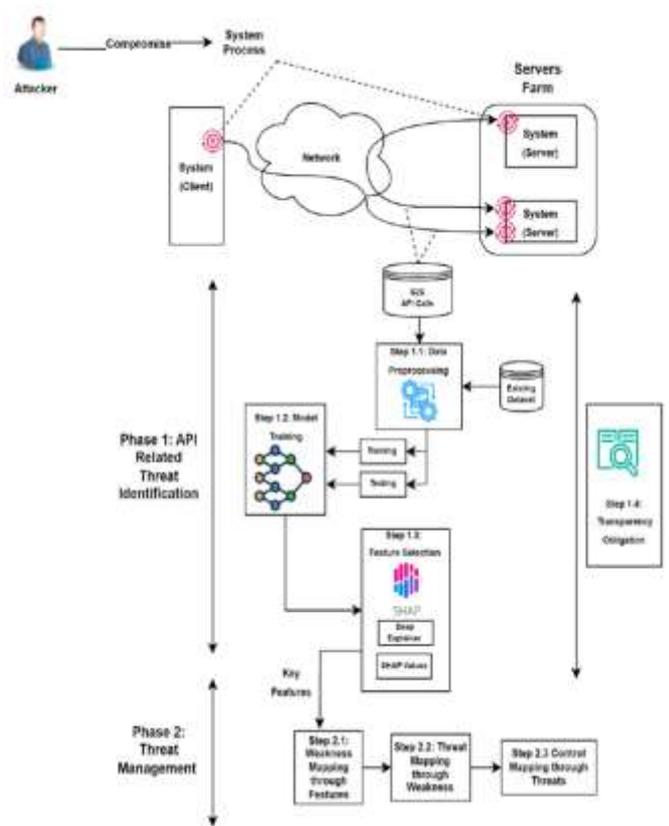


Figure 2: Proposed architecture (source: Nihala Basheer)

Mapping vulnerabilities via features, threats through features, and controls through threats are the three processes that make up the threat assessment and management phase, which is the second phase. In order to identify and comprehend vulnerabilities, the important aspects from the prior stage are evaluated with open-source security intelligence frameworks that are both commonly used and publicly available. The next step is to associate the vulnerabilities with the dangers they provide; this will allow us to identify potential attack vectors and threat actors. Finally, the identified threats are mitigated by determining appropriate control measures. In order to make API-driven systems more secure and resilient, this method advocates for comprehensive threat management approaches.

The two primary steps of this diagrammatic architecture for detecting and controlling security risks to APIs are:

Phase 1: API-Related Threat Identification

**1. System Compromise by an Attacker:**

- The attacker compromises a system process, which can occur at the client or server level.
- The compromised system interacts with APIs over a network connecting the client and server systems.

**2. Data Preprocessing (Step 1.1):**

- Existing datasets are preprocessed to prepare for analysis.
- The processed data is split into training and testing sets.

**3. Model Training (Step 1.2):**

- A machine learning model is trained using the preprocessed data to identify threat patterns or anomalies.

**4. Feature Selection (Step 1.3):**

- Techniques like SHAP (SHapley Additive exPlanations) and deep explainers are used to determine key features influencing the model's decisions.

**5. Transparency Obligation (Step 1.4):**

- Outputs of the analysis are documented to ensure transparency and accountability in threat identification.

Phase 2: Threat Management

**1. Weakness Mapping Through Features (Step 2.1):**

- Key features identified in Phase 1 are mapped to specific system weaknesses.

**2. Threat Mapping Through Weakness (Step 2.2):**

- System weaknesses are then analyzed to identify potential threats they could enable.

**3. Control Mapping Through Threats (Step 2.3):**

- Mitigation controls are applied to address and manage identified threats effectively.

Overview of the Workflow:

- The process begins with the attacker compromising a system.
- API call logs and datasets are analyzed through the framework to identify weaknesses and threats.
- Machine learning models and feature explainability methods are used to enhance threat detection and transparency.
- The findings guide organizations in mapping vulnerabilities to threats and applying targeted controls.

This end-to-end pipeline highlights a structured approach to securing APIs from potential attacks.

**IV. RESULTS AND DISCUSSION**

**4.1 Data Preprocessing**

By preparing the selected data, this first phase seeks to kick off the deep-learning life cycle. In order to make sure the data is suitable for training and evaluating models, we processed it according to the selected dataset. Figure 3 shows the distribution of classes before to SMOTE, and it is clear that goodware (0) and malware (1) are significantly outnumbered. There are 582 instances in the goodware class, making up about 57% of the dataset, and 439 instances in the malware class, making up about 43% of the dataset. The fact that the goodware bar is much larger than the malware bar suggests that there are more goodware samples in the dataset to begin with.

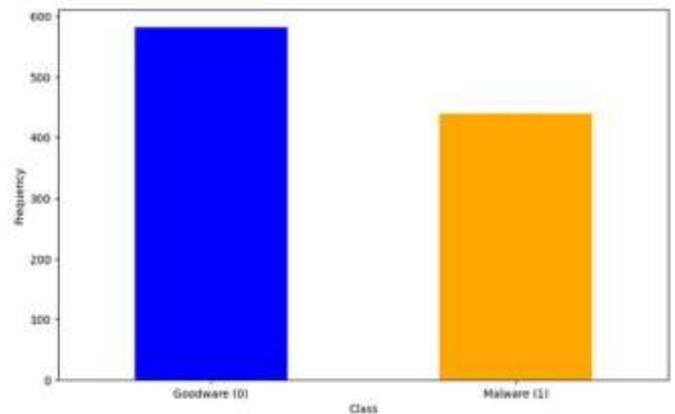


Figure 3: Class distribution before SMOTE

After SMOTE was applied, the class distribution is shown in Figure 4. The technique has been employed to rectify the imbalance by augmenting the malware class's sample size to parity with the goodware class's sample size. As a result, there are 582 instances of each class of malware and goodware, making the two bars of equal height. Machine learning models trained on this data may perform better as a result of the class imbalance being eliminated.

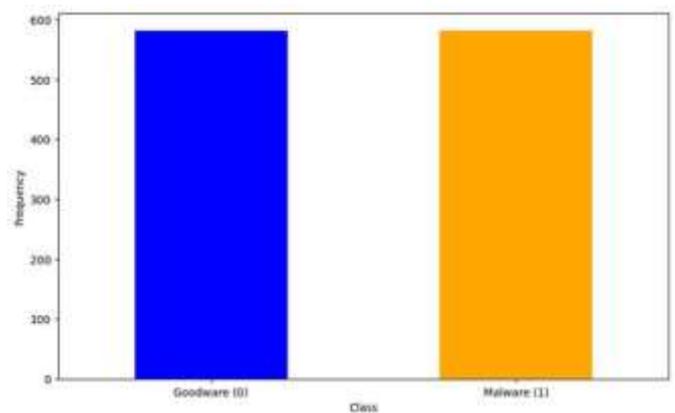


Figure 4: Class distribution after SMOTE

## 4.2 Model Training

Model performance evaluation follows data processing; it involves testing the trained model using a variety of measures to see how well it can predict outcomes from unseen data. When evaluating models for classification tasks, these metrics are crucial. An impressive 91% overall accuracy shows that the ANN model is reliable in predicting results across all test conditions. This model demonstrates a high level of accuracy in identifying positive class labels, with a precision of 93%. With an impressive recall rate of 89%, it clearly captures a significant portion of real positive experiences. The 91% balanced F1 score indicates that the model is very trustworthy since it efficiently maintains a balance between recall and precision. In contrast, the MLP model demonstrates an accuracy of 88%, which is marginally lower.

The ANN and RNN ROC curves are shown in Figure 5. With an Area Under the Curve (AUC) score of 0.94 for the ANN model and an AUC of 0.96 for the RNN model, both models demonstrate strong performance in classification. The steep slopes at the top left of both curves, which originate from the origin, indicate high TPRs relative to FPRs. With their curves positioned high above the diagonal line, it is clear that both models perform substantially better.

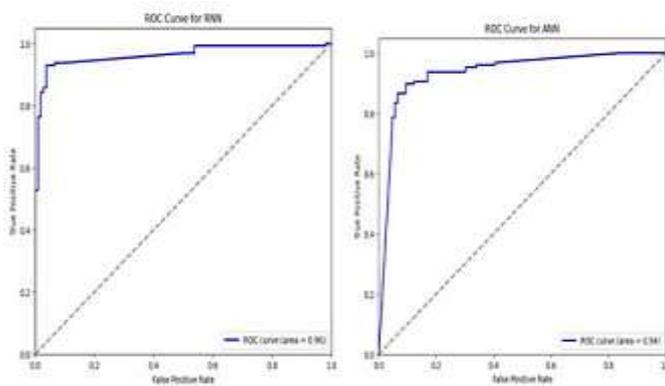


Figure 5: ROC curve of both the models

## V. CONCLUSIONS

The extensive use of application programming interfaces (APIs) for system-to-system communication has raised several security problems, chief among them being threats in API calls. The sheer number of API-related requests and the variety of threat types make it difficult to resolve these issues. In order to address this difficulty, this work uses a deep-learning-based method for threat detection and a security knowledge-based method for threat analysis and management. Highlighting the significance of mapping system-to-system API call functions that may be utilized through the extraction of critical threat-related attributes using deep-learning models, the suggested technique presents a well-structured threat

identification and management strategy. Our suggested strategy not only improves decision-making and transparency but also achieves a high accuracy of over 85% for both ANN and MLP models in threat detection. This is achieved by effectively surpassing previous methodologies through the integration of XAI methods. This makes sure that any user group, irrespective of their level of technical knowledge, can comprehend which qualities or attributes are helpful for threat detection. In addition, our approach is made more resilient and applicable by focusing on using open-source intelligence, which increases the likelihood of successful practical application. This paper also makes a new contribution by recommending practices that promote transparency. We can earn the trust of all kinds of users—developers, application vendors, suppliers, and end users—by keeping and sharing transparent, accurate, and easily accessible information regarding the datasets, decision-making, and model performance. Enhancing the total effectiveness of the threat-management process, the transparency duty of the entire AI lifecycle not only fulfills regulatory obligations but also develops a culture of openness and accountability. Finally, we intend to fix the present issues and improve our methodology in the future, even though our suggested strategy greatly improves the capacity to identify and handle API-based risks.

## REFERENCES

- [1] Jones, C.L.; Bridges, R.A.; Huffer, K.M.T.; Goodall, J.R. Towards a Relation Extraction Framework for Cyber-Security Concepts. In ACM International Conference Proceeding Series; Association for Computing Machinery: New York, NY, USA, 2015.
- [2] Jeimy, J.; Cano, M. FLEXI—A Conceptual Model for Enterprise Cyber Resilience. *Procedia Comput. Sci.* 2023, 219, 11–19.
- [3] Wallis, T.; Dorey, P. Implementing Partnerships in Energy Supply Chain Cybersecurity Resilience. *Energies* 2023, 16, 1868.
- [4] Lubis, M.; Lubis, A.R. Designing Secured Cafe Network with Security Awareness Domain and Resource (SADAR) by Simulation using Cisco Packet Tracer. In ACM International Conference Proceeding Series; Association for Computing Machinery: New York, NY, USA, 2022; pp. 233–238.
- [5] Bemthuis, R.; Iacob, M.-E.; Havinga, P. A Design of the Resilient Enterprise: A Reference Architecture for Emergent Behaviors Control. *Sensors* 2020, 20, 66–72.
- [6] Lubis, M.; Rahman, N.A.; Alam, P.F. Marketing Strategies Design for Crowd sourcing Application in Indonesia. In ACM International Conference Proceeding Series; Association for Computing Machinery: New York, NY, USA, 2021; pp. 25–31.

- [7] Pieters, W.; Hadžiosmanović, D.; Dechesne, F. Cyber Security as Social Experiment. In ACM International Conference Proceeding Series; Association for Computing Machinery: New York, NY, USA, 2014; pp. 15–24.
- [8] Lubis, M.; Fathoni, M.; Lubis, A.R. New Product Development Architectural Framework for Sustainability and Innovation within Telecommunication Industry. In ACM International Conference Proceeding Series; Association for Computing Machinery: New York, NY, USA, 2020; pp. 145–150.
- [9] Grigaliūnas, Š.; Brūzgienė, R.; Venčkauskas, A. The Method for Identifying the Scope of Cyberattack Stages in Relation to Their Impact on Cyber-Sustainability Control over a System. *Electronics* 2023, 12, 591.
- [10] Carías, J.F.; Labaka, L.; Sarriegi, J.M.; Hernantes, J. Defining a Cyber Resilience Investment Strategy in an Industrial Internet of Things Context. *Sensors* 2019, 19, 138.
- [11] Kupsch, J.A.; Miller, B.P.; Heymann, E.; César, E. First principles vulnerability assessment. In Proceedings of the 2010 ACM Workshop on Cloud Computing Security Workshop, Chicago, IL, USA, 8 October 2010; pp. 87–92.
- [12] Garcia-Perez, A.; Cegarra-Navarro, J.G.; Sallos, M.P.; Martinez-Caro, E.; Chinnaswamy, A. Resilience in healthcare systems: Cyber security and digital transformation. *Technovation* 2023, 121, 102583.
- [13] Ademujimi, T.; Prabhu, V. Digital Twin for Training Bayesian Networks for Fault Diagnostics of Manufacturing Systems. *Sensors* 2022, 22, 1430.
- [14] AlMajali, A.; Viswanathan, A.; Neuman, C. Resilience Evaluation of Demand Response as Spinning Reserve under Cyber-Physical Threats. *Electronics* 2017, 6, 2.
- [15] Linkov, I.; Ligo, A.; Stoddard, K.; Perez, B.; Strelzoffx, A.; Bellini, E.; Kott, A. Cyber Efficiency and Cyber Resilience. *Commun. ACM* 2023, 66, 33–37.
- [16] Hausken, K. Cyber resilience in firms, organizations and societies. *Internet Things* 2020, 11, 100204.
- [17] Van Haastrecht, M.; Golpur, G.; Tzismadia, G.; Kab, R.; Priboi, C.; David, D.; Răcățăian, A.; Baumgartner, L.; Fricker, S.; Ruiz, J.F.; et al. A Shared Cyber Threat Intelligence Solution for SMEs. *Electronics* 2021, 10, 2913.
- [18] Rizwan, K.; Ahmad, M.; Habib, M.A. Cyber Automated Network Resilience Defensive Approach against Malware Images. In ACM International Conference Proceeding Series; Association for Computing Machinery: New York, NY, USA, 2022; pp. 237–242.
- [19] Kotenko, I.; Izrailov, K.; Buinevich, M.; Saenko, I.; Shorey, R. Modeling the Development of Energy Network Software, Taking into Account the Detection and Elimination of Vulnerabilities. *Energies* 2023, 16, 5111.
- [20] Estay, D.A.S.; Sahay, R.; Barfod, M.B.; Jensen, C.D. A systematic review of cyber-resilience assessment frameworks. *Comput. Secur.* 2020, 97, 101996.
- [21] Blay, K.B.; Yeomans, S.; Demian, P.; Murguia, D. The Information Resilience Framework. *J. Data Inf. Qual.* 2020, 12, 1–25.
- [22] Jones, S.L.; Collins, E.I.M.; Levordashka, A.; Muir, K.; Joinson, A. What is ‘cyber security’?: Differential language of cyber security across the lifespan. In Proceedings of the Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; p. LBW0269.

**Citation of this Article:**

Suneel Kumar Mogali, “Deep Learning Models for Privacy Risk Assessment in Dynamic Cyber Threat Environments” Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 5, Issue 2, pp 109-115, February 2021. Article DOI <https://doi.org/10.47001/IRJIET/2021.502016>

\*\*\*\*\*