International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

INTRUSION DETECTION SYSTEM FOR PHISING DETECTION USING CONVOLUTION NEURAL NETWORK

Jayapradha J[1], Vineethkumar S[2], Vigneshwaran R[3]and Ramprasath A [4]

[1] Assistant Professor, Department of CSE, Manakula Vinayagar Institute of Technology (MVIT), Puducherry, INDIA
[2]UG Scholar, Department of CSE, MVIT, Puducherry, INDIA
[3] UG Scholar, Department of CSE, MVIT, Puducherry, INDIA
[4] UG Scholar, Department of CSE, MVIT, Puducherry, INDIA

E-mail address: jayapradhacse@mvit.edu.in, *vineethkumarcse2020@mvit.edu.in, vigneshwarancse2020@mvit.edu.in, ramprasathacse2020@mvit.edu.in*

Abstract: Innovative security solutions utilizing cutting-edge machine learning techniques are essential to strengthen network defense as cyber threats become more sophisticated. This study proposes an intrusion detection system (IDS) that uses deep learning algorithms (DLAs), specifically convolutional neural networks (CNNs) and, to automatically detect phishing attacks. Phishing circumvents traditional signature-based intrusion detection systems by using cunning social engineering techniques. CNNs enable the automatic extraction of sophisticated features from raw input data, including URLs and webpage content. Low-level patterns are recognized by their convolutional layers, and in later layers, these patterns are combined to create higher-level representations. Sequential data, such as user activities over time, is a great fit for CNN modeling. CNNs work together to acquire the intricate multi-modal patterns that are characteristic of phishing. Back propagation-based model optimization enables real-time adaptation to identify emerging phishing variants. DLA integration with an IDS offers a strong defense against sophisticated user-targeted phishing attacks. Using the KDD-CUP99 dataset, which has 175,341 training and 82,332 testing instances, the DLA model is trained. Proactive incident response is made possible by automated feature learning by DLAs, which dramatically increases detection accuracy over manual rule-based techniques. This DLA-driven intrusion detection system research strengthens the overall security posture by improving resistance to changing social engineering threats. By utilizing machine learning, networks and users can be protected from sneaky phishing tactics through constant model refining for intelligent, adaptive threat identification as attack vectors change. The accuracy of the Phishing Detection with an accuracy of 99.2% and with a Model Loss of 79%.

Keywords: Convolution Neural Network, Intrusion Detection system, KDD-CUP 99,Deeplearning Algorithm

1. INTRODUCTION

Here is a 1000-word expansion on the importance of phishing detection and the challenges faced by traditional cyber security measures:

In the era of rapidly advancing digital technologies and ubiquitous connectivity, cyber security has become an indispensable concern for organizations, individuals, and nations alike [1]. As our reliance on digital systems and networks continues to grow, so does the landscape of cyber threats, constantly evolving to exploit emerging vulnerabilities [2]. Among the myriad of cyber-attacks, phishing stands out as a persistent and pernicious threat, posing significant risks to sensitive information, financial assets, and user privacy [3]. Phishing attacks, characterized by their deceptive and manipulative nature, prey upon human vulnerabilities by masquerading as legitimate entities or trusted sources [4]. Through carefully crafted emails, fraudulent websites, and social engineering tactics, attackers lure unsuspecting victims into divulging confidential data or granting unauthorized access. These attacks can take many forms, from spoofing popular brands and services to impersonating colleagues or authority figures within an organization [5].

The consequences of successful phishing attacks can be devastating, ranging from financial losses and data breaches to reputational damage and erosion of consumer trust. Individuals may fall victim to identity theft, financial fraud, or extortion attempts, while organizations face the risk of intellectual property theft, regulatory fines, and

E-mail: jayapradhacse@mvit.edu.in, vineethkumarcse2020@mvit.edu.in, vigneshwarancse2020@mvit.edu.in, ramprasathacse2020@mvit.edu.in operational disruptions. In the context of national security, phishing can be leveraged as a vector for cyber espionage, compromising sensitive information and undermining critical infrastructure.

Traditional cyber security measures, while valuable, often evolving sophistication of phishing techniques [6]. Rule-based detection systems and signature-based approaches may prove inadequate in identifying novel phishing patterns and subtle deviations from known threats. These methods rely on predefined rules and signatures, which can become obsolete or ineffective against constantly morphing phishing campaigns [7].

Moreover, the ever-expanding scope of network traffic and the sheer volume of data to be analyzed create additional challenges for effective phishing detection. With the proliferation of connected devices, cloud services, and online activities, the task of sifting through vast amounts of data to identify potential phishing attempts becomes increasingly daunting, straining the capabilities of traditional security solutions [8].

By fortifying the defenses against phishing attacks, organizations can safeguard sensitive data, maintain user trust, and uphold the integrity of their digital ecosystems. This, in turn, fosters a more secure and resilient digital landscape, facilitating seamless operations, protecting reputations, and enabling innovation across various sectors.

Researchers and security experts have focused on advanced machine learning and deep learning techniques to address the urgent problem of phishing detection. These data-driven methods have great potential for utilizing artificial intelligence to recognize complex patterns, absorb knowledge from large datasets, and adjust to changing phishing tactics.

These models may learn to distinguish between benign and harmful actions by considering a variety of attributes and indicators. They are trained on multiple datasets that contain both legitimate and phishing occurrences.

By automatically identifying pertinent features from raw data inputs, deep learning models—such as Convolutional Neural Networks (CNNs) and recurrent neural networks (RNNs)—take this skill a step further. CNNs are highly suitable for examining the content of web pages, photos, and other visual clues connected to phishing attempts since they are adept at deriving hierarchical representations from visual input. However, RNNs are skilled at modeling sequential data, which allows them to identify temporal trends and user actions that can indicate phishing efforts.

By combining the strengths of these models, researchers have developed ensemble approaches that leverage multiple architectures and data modalities, such as text, images, and network traffic patterns. These multi-modal approaches have shown promising results in enhancing phishing detection accuracy and robustness, as they can capture a more comprehensive set of signals and indicators.

Furthermore, deep learning models are well-suited for the dynamic nature of phishing attacks due to their capacity to continuously learn from and adjust to new data. These models may be retrained and adjusted to remain ahead of new phishing attempts, offering a proactive defense against innovative attack vectors, as fraudsters change their strategies and tactics.

Adopting deep learning and machine learning methods for phishing detection is not without its difficulties, though. The availability and quality of training data is one of the main issues. To build effective models, it is imperative to obtain a diverse, representative, and precisely labeled dataset; nevertheless, this can be a considerable challenge because of data scarcity, privacy issues, and the ever-evolving phishing scene.

The interpretability and explainability of the model present another difficulty. Even while deep learning models have proven to be remarkably effective across a range of areas, it can be challenging to comprehend and believe the predictions made by these models due to their opaque inner workings and decision-making processes, especially in cyber security applications with significant stakes.

Adversarial attacks, where malicious actors craft inputs specifically designed to fool machine learning models, also pose a significant threat. Notwithstanding these difficulties, there are a lot of potential advantages to using deep learning and machine learning for phishing detection. Organizations may greatly lower the risk of data breaches, financial losses, and reputational harm by automating the process of detecting and combating phishing threats. These tools can also lighten the workload for cyber security experts, freeing them up to concentrate on more difficult and important projects.

Strong cyber security measures become more and more important as digital transformation continues to change a variety of industries, including government services, critical infrastructure, healthcare, and finance. Phishing attacks are expected to continue to be a problem because they take advantage of human weaknesses and can adapt to new platforms and technology. However, ongoing research and collaboration between academia, industry, and government agencies are crucial to address the challenges of data quality, model interpretability, and adversarial resilience.

2. Related Works

This review of the literature looks at the research on cyber security applications, including phishing detection, network intrusion detection, and website characteristic analysis. Traditional rule- and heuristic-based solutions have not been able to keep up with the ever-changing nature of phishing efforts, which has prompted researchers to look into data-driven approaches.

Because they use deceptive tactics to trick users into disclosing sensitive information or granting unauthorized access, phishing attacks have always been a source of concern. Because these attacks are constantly evolving, researchers are focusing on data-driven approaches that leverage the capabilities of deep learning and machine learning models.

Numerous studies have used deep learning models to detect phishing attempts [9]. These models have demonstrated an astonishing level of proficiency in acquiring knowledge from many data sources, including text, photographs, and behavioral patterns associated with spear phishing attempts. These methods have outperformed classical machine learning techniques in accurately identifying phishing scenarios by leveraging deep neural networks' powerful feature extraction and representation learning capabilities [10]. In this field, deep learning architectures like Convolutional Neural Networks (CNNs) have shown to be very successful. RNNs are skilled at capturing temporal dependencies and modeling user behavior over time, whereas CNNs are excellent at automatically collecting pertinent characteristics from raw data inputs like URLs, webpage content, and photos. Deep learning models are able to understand the complex multi-modal patterns that are symptomatic of phishing assaults by combining these complementary qualities, which allows for more accurate and real-time detection [11].

In addition to phishing detection, network intrusion detection has demonstrated significant potential for machine learning and deep learning approaches. For this, the application of ensemble methods, decision trees, random forests, support vector machines, and other machine learning algorithms has been investigated by researchers. These algorithms have been used to identify abnormalities or departures from typical behavior, categorize network activity as benign or malevolent, and examine patterns in network traffic data.

These techniques take advantage of machine learning models' capacity to generalize patterns and learn from data, showing promising results in detecting hitherto undiscovered attack vectors and adjusting to changing threat environments. These models make it possible to detect intrusions and cyber-attacks more precisely and quickly by efficiently capturing the intricate linkages and temporal dependencies seen in network data [13].

Additionally, deep learning models have demonstrated potential in this field due to their ability to manage heterogeneous and high-dimensional network data. In some situations, architectures such as CNNs and have been able to learn complex patterns and sequences from network traffic, even exceeding conventional machine learning methods. Machine learning and deep learning approaches have also been investigated for website characteristic analysis, in addition to phishing and network intrusion detection. This entails examining a website's behavior, structure, and content among other elements to spot any potential dangers or criminal activity. To address this issue, researchers have used machine learning models in conjunction with methods including graph analysis, computer vision, and natural language processing.

Numerous obstacles still exist, despite the literature's encouraging examples of machine learning and deep learning's applications in cyber security fields. Interpretability of the models is a significant issue since these intricate models frequently function as "black boxes," making it challenging to comprehend how they make decisions. This lack of openness can impede acceptance and trust, particularly in security applications that are vital and require explainability.

The availability and quality of data present another difficulty. For machine learning and deep learning models to acquire useful patterns and perform well in generalization, they substantially depend on representative, diversified, and high-quality training data. However, due to privacy considerations, data scarcity, and the dynamic nature of threats, acquiring and curating such data in cyber security contexts can be difficult.

Adversarial threats represent a serious concern as well, since they involve malevolent actors purposefully manipulating inputs to trick machine learning algorithms. Although methods like input sanitization, defensive distillation, and adversarial training have been suggested to lessen these risks, more investigation is required to create reliable and resilient models.

Furthermore, one of the ongoing challenges is accounting for the changing nature of internet information and cyber-attacks. Machine learning and deep learning models need to be able to continuously learn and adapt in order to be successful over time, as new threats emerge and old one’s change. This can call for strategies like transfer learning, online learning, or group approaches that incorporate several models.

In order to overcome these constraints and fully utilize deep learning and machine learning in cyber security, researchers have looked into a number of approaches. Using encryption and privacy-preserving methods is one way to safeguard sensitive information while allowing for model inference and training.

Identifying and thwarting threats like malware, phishing scams, and unauthorized access attempts are frequently the main issues in enterprise network systems. These networks usually manage a lot of traffic and a wide range of user activities, thus real-time processing and analysis of massive volumes of heterogeneous data is required. Interpretable and explicable models are crucial for enterprise networks because they need to strike a compromise between security and productivity and usability.

Industrial control systems, on the other hand, face different challenges. These systems are critical for controlling and monitoring physical processes in industries like manufacturing, energy, and transportation. Cyber-attacks on these systems can have severe consequences, including equipment damage, production disruptions, and even safety hazards. Machine learning and deep learning models for this domain must be highly reliable, robust to noise and anomalies, and capable of detecting subtle deviations from normal behavior.

There are certain difficulties specific to the Internet of Things (IoT) environment. IoT devices frequently have limited resources, including memory, processing power, and energy. This calls for the creation of machine learning models that are effective and lightweight, able to function within these limitations and still offer strong security. The task of detecting and mitigating threats is further complicated by the diversity of IoT devices and communication protocols.

Creating specialized models and techniques that take into account the particular limitations and features of every environment might result in more effective and efficient cyber security solutions. For example, in enterprise networks, addressing a variety of threat vectors and data sources may benefit from the use of ensemble models that combine various methodologies. Machine learning could be combined with anomaly detection methods that use physics-based models and domain expertise in industrial control systems to improve dependability. Federated learning or edge computing techniques may make it possible to train and infer models distributed for Security for the Internet of Things while maintaining privacy and minimizing communication cost. Overall, the literature demonstrated promising applications of machine learning and deep learning in cyber security domains. However, challenges persist in terms of model interpretability, data quality, adversarial threats, and accounting for the dynamic nature of cyber-attacks and online content. Addressing limitations through techniques like encryption, continuous learning, privacy-preserving models, and explainable AI can help realize the full potential of these methods. Further research also seems to be warranted to develop solutions tailored for specific application environments and threat landscapes.

3. Proposed Work

The proposed methodology for enhancing network security through the integration of Deep Learning Algorithms (DLAs) within Intrusion Detection Systems (IDS) for phishing detection employs a comprehensive and multifaceted approach. This methodology aims to leverage the unique strengths of advanced DLAs, specifically Convolutional Neural Networks (CNNs) to effectively identify and mitigate phishing threats within complex network environments.

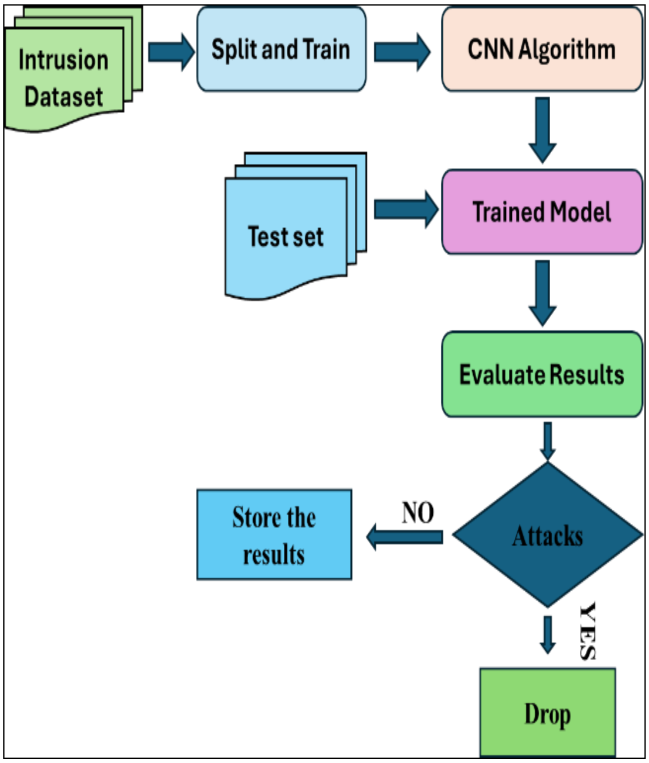


Fig. 1: Proposed Architecture of the Intrusion Detection Model.

From Fig. 1 architecture of the intrusion detection model illustrates the end-to-end process of an intrusion detection system that utilizes a Convolutional Neural Network (CNN) algorithm to identify potential cyber-attacks. The entire workflow is depicted through a series of interconnected blocks, each representing a specific stage or component of the system.

The process commences with the "Intrusion Detection Dataset" block, which serves as the primary data source. This dataset is a collection of recorded network traffic instances, both benign and malicious, meticulously curated and labeled for training and testing purposes. Each instance within the dataset encapsulates a comprehensive set of features that characterize network events, such as packet headers, payload data, and other relevant metadata.

The subsequent block, "Split and Train," represents a crucial preprocessing step. Here, the intrusion detection dataset is partitioned into two subsets: a training dataset and a testing dataset. The training dataset, typically comprising a larger portion of the original data, is employed to train the CNN algorithm, allowing it to learn and recognize patterns indicative of cyber threats. Conversely, the testing dataset, a smaller yet representative subset, is reserved for evaluating the performance of the trained model on previously unseen data.

The core of the architecture lies in the "CNN Algorithm" block, which encompasses the intricate neural network architecture designed specifically for intrusion detection tasks. Convolutional neural networks are particularly well-suited for processing and analyzing high-dimensional data, such as network traffic patterns, owing to their ability to automatically extract and learn hierarchical representations of the input data.

Within the CNN algorithm block, the training dataset undergoes a series of transformations and computations. The network's convolutional layers identify and extract low-level features from the input data, such as specific byte patterns or packet header anomalies. These features are then propagated through subsequent layers, where they are combined and abstracted into higher-level representations, ultimately enabling the network to recognize complex patterns indicative of cyber threats.

The training process is an iterative endeavor, during which the CNN algorithm continuously refines its internal parameters, known as weights and biases, to minimize the discrepancy between its predictions and the ground truth labels provided in the training dataset. This optimization process is guided by a loss function and typically leverages advanced techniques like back propagation and stochastic gradient descent.

Once the training phase is complete, the resulting "Trained Dataset + Test Dataset" block represents the amalgamation of the trained model and the previously set-aside testing dataset. This combination allows for a rigorous evaluation of the CNN algorithm's performance on unseen data, providing an unbiased assessment of its generalization capabilities.

The "Trained Model" block encapsulates the CNN algorithm that has been optimized during the training phase. This trained model is then deployed to analyze new network traffic instances, classifying them as either benign or malicious based on the learned patterns and decision boundaries.

The "Evaluated Result" block represents the output of the intrusion detection system, where each network traffic instance is categorized as either an "Attack" or a "Normal Site." If an instance is classified as an attack, the system generates an alert, potentially triggering further investigation or defensive measures. Conversely, if the instance is deemed benign, it is labeled as a "Normal Site," allowing legitimate network traffic to flow uninterrupted.

The final decision block introduces a conditional logic branch, where the system's response is determined by the evaluated result. If an attack is detected, the system initiates appropriate countermeasures or notifies the designated "Attacker" entity, which could represent a security operations center, incident response team, or automated defensive mechanisms. Alternatively, if the evaluated result indicates a "Normal Site," the system allows the network traffic to proceed without interruption.

This comprehensive architecture leverages the powerful pattern recognition capabilities of convolutional neural networks to accurately identify cyber threats in real-time, while minimizing false positives and negatives. By continuously adapting and improving the trained model with new data, the intrusion detection system can stay vigilant against emerging attack vectors and evolving cyber threats, providing a robust defense for critical network infrastructures.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Data Acquisition and Preprocessing:

The first step in the proposed methodology involves acquiring diverse and representative datasets encompassing various phishing scenarios. These datasets will serve as the foundation in training and validation of the deep learning models. The data acquisition process will encompass a wide range of sources, including network traffic logs, email repositories, and databases of known phishing instances.

1) Data Acquisition:

Gather diverse datasets from multiple sources (network logs, email repositories, databases, user reports, honey pots, open-source feeds)

Sources capture different phishing tactics (spear-phishing, whaling, smishing) and indicators (malicious URLs, suspicious content, behavioral patterns)

Obtain representative data to cover breadth and depth of phishing scenarios.

2) Data Preprocessing:

- Data cleaning (handle missing values, remove duplicates, fix inconsistencies/errors)
- Outlier detection and removal
- Data transformation techniques:
- Normalization (prevent feature domination)
- One-hot encoding (transform categorical variables)
- Feature engineering (extract meaningful features, create new representations)
- Feature selection (retain most informative features, reduce dimensionality)
- Address imbalanced datasets (oversampling, under sampling, synthetic data generation)
- Split dataset into training, validation, and testing subsets

B. Deep Learning Architecture Design:

At the core of the proposed methodology lies the design of a robust deep learning architecture tailored specifically for phishing detection. This architecture will incorporate both convolutional neural networks (CNNs) in a hybrid configuration, leveraging the unique strengths of each network type.

- CNNs are highly effective in extracting spatial patterns and recognizing visual cues, making them well-suited for analyzing website content, images, and other visual components associated with phishing attempts.
- The CNN component of the architecture will be responsible for processing visual data, such as screenshots of phishing websites, embedded images in emails, and other visual artifacts. Through its convolutional and pooling layers, CNN will learn to identify and extract relevant visual features indicative of phishing, such as suspicious logos, language patterns, or inconsistencies in website design.
- On the other hand, CNNs excel at processing sequential data and capturing temporal patterns, making them crucial for identifying phishing patterns within network traffic and analyzing the temporal aspects of user interactions.
- The CNN component will analyze sequential data sources like network traffic logs, user browsing histories, and email communication trails. By leveraging its ability to model long-term dependencies, the RNN can detect suspicious patterns in network traffic, track user behavior over time, and identify deviations from normal activity that may signal a phishing attempt.

The hybrid architecture will combine the outputs of the CNN and components, enabling the model to leverage both visual and sequential data sources simultaneously. This multi-modal approach allows the deep learning model to capture a more comprehensive set of signals and indicators, enhancing its ability to detect sophisticated phishing attacks that may employ a combination of visual deception and temporal trickery.

"Additionally, the use of ensemble techniques, decision trees, random forests, support vector machines, and other machine learning algorithms for network intrusion detection. These algorithms have been used to categorize network activity as malicious, detect anomalies or departures from typical behavior, and examine network traffic patterns."

The proposed architecture may also incorporate ensemble techniques, where the outputs of multiple deep learning models or traditional machine learning algorithms are combined to improve overall performance and robustness. For instance, the hybrid CNN model could be ensemble with decision trees or random forests trained on network traffic features, further enhancing the system's ability to detect a wide range of phishing patterns and techniques.

Explainability and Interpretability:

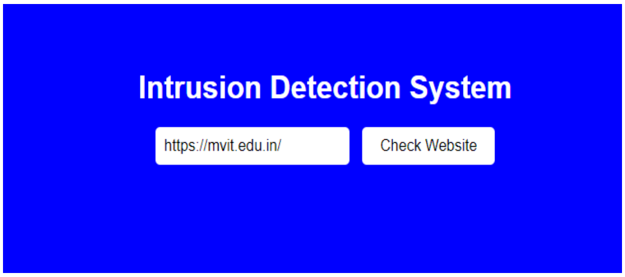
To foster trust and transparency in the decision-making process of the deep learning models, the proposed methodology an interpretability method will be employed to provide insights into the model's reasoning and decision-making process.

This enhanced interpretability will enable cyber security analysts and practitioners to understand the rationale behind the system's phishing detections, facilitating more informed analysis and response strategies.

C. *Integration and Deployment:*

The final stage of the proposed methodology involves the seamless integration and deployment of the deep learning-based IDS within existing network security infrastructure. This process will consider factors such as scalability, resource efficiency, and cross-platform compatibility to ensure a smooth transition and minimal disruption to ongoing operations. Additionally, the proposed methodology will include the development of user-friendly interfaces and reporting mechanisms, enabling cyber security professionals to monitor and manage phishing detection alerts and responses effectively.

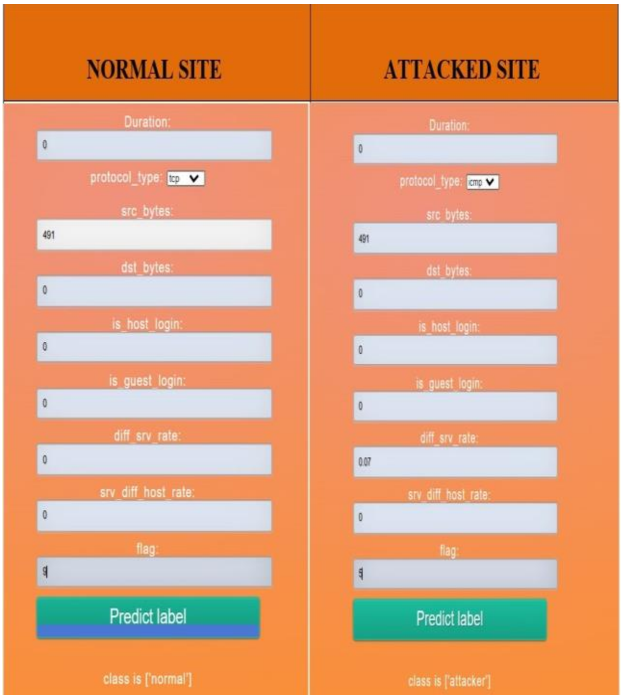
Additionally, the proposed methodology will include the development of user-friendly interfaces and reporting mechanisms, enabling cybersecurity professionals to monitor and manage phishing detection alerts and responses effectively.


Intrusion Detection System

https://mvit.edu.in/

Fig.2: URL Feeding Page

From Fig.2 illustrates the URL Feeding Page where the suspected URL is feeded as an source and this URL is processed and the attributes are stored in the training dataset.



NORMAL SITE

Duration: 0

protocol_type: tcp

src_bytes: 491

dst_bytes: 0

is_host_login: 0

is_guest_login: 0

diff_srv_rate: 0

srv_diff_host_rate: 0

flag: 9

class is ['normal']

ATTACKED SITE

Duration: 0

protocol_type: icmp

src_bytes: 491

dst_bytes: 0

is_host_login: 0

is_guest_login: 0

diff_srv_rate: 0.07

srv_diff_host_rate: 0

flag: 5

class is ['attacker']

Fig. 3: Result Page

In Fig. 3 the the processed URL into Attributes are then evaluated into the CNN model and then the results are displayed.

By following this comprehensive and systematic methodology, the proposed solution aims to provide a robust, adaptive, and intelligent defense against the rising threat of phishing attacks, ultimately enhancing network security, and safeguarding sensitive information and user privacy within digital ecosystems.

4. PERFORMANCE

The performance analysis of the proposed DL-based IDS for enhancing network security and phishing detection is a critical aspect of evaluating its effectiveness and real-world applicability. This performance analysis will encompass a comprehensive set of metrics and techniques to assess the system's accuracy, robustness, scalability, and overall capability to mitigate phishing threats within dynamic network Environments.

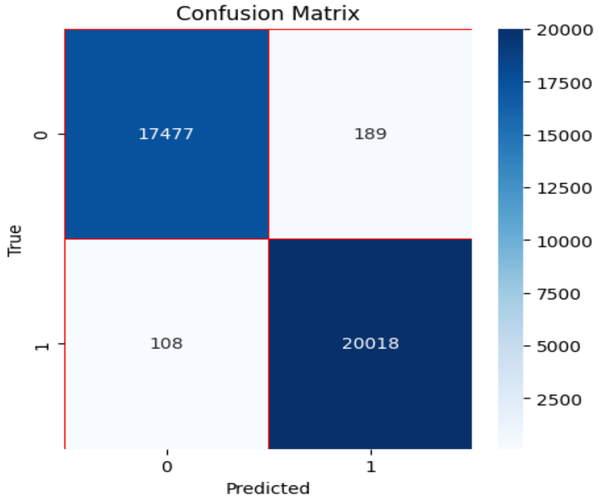


Fig. 4: Confusion Graph

From Fig. 4 the confusion matrix is constructed from the predictions on the held-out test set, providing insights into false positives, false negatives, and overall accuracy

Evaluation Metrics:

The primary evaluation metrics employed in this performance analysis will include:

A. Accuracy:

This metric measures the proportion of correctly classified instances, both phishing and legitimate, out of the total instances evaluated. Accuracy provides an overall assessment of the system's correctness and is a fundamental starting point for performance evaluation.

B. Precision:

Precision quantifies the ratio of true positive predictions (correctly identified phishing instances) to the total positive predictions made by the system. This metric is particularly important in minimizing false positives, which can lead to unnecessary alerts and disruptions.

C. Recall (Sensitivity):

Recall, also known as sensitivity, measures the proportion of true positive predictions relative to the total actual positive instances (phishing attempts). A high recall rate indicates the system's ability to identify most, if not all, relevant phishing instances and minimizing false negatives

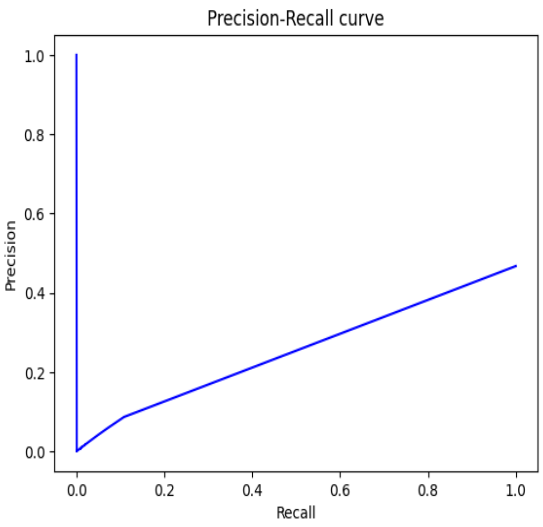


Fig. 5: Precession-Recall Curve.

From Fig. 5 the precision-recall curve visualizes the trade-off between the precision (accuracy on positive predictions) of peak value of 1 and recall (coverage of true positives) as the classification threshold is varied.

D. F1-Score:

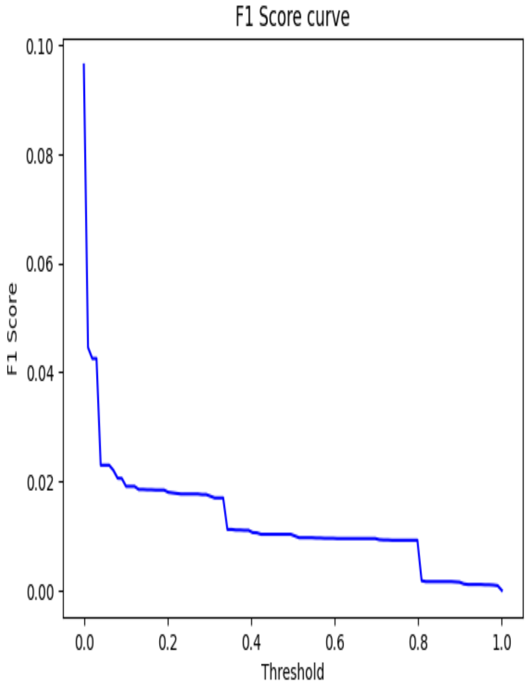


Fig.6: F1 score curve.

From Fig.6 the F1-score is a comprehensive metric that integrates precision and recall into a unified measure, offering a well-balanced assessment of a system's performance. This metric proves particularly beneficial when working with imbalanced datasets, a scenario frequently encountered in the realm structure of phishing detection.

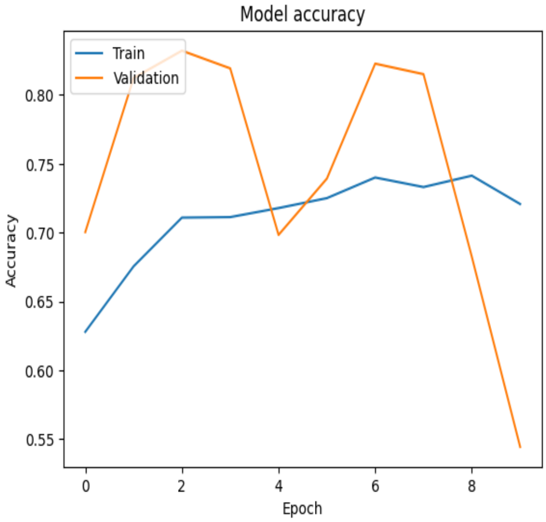


Fig.7: Model Accuracy

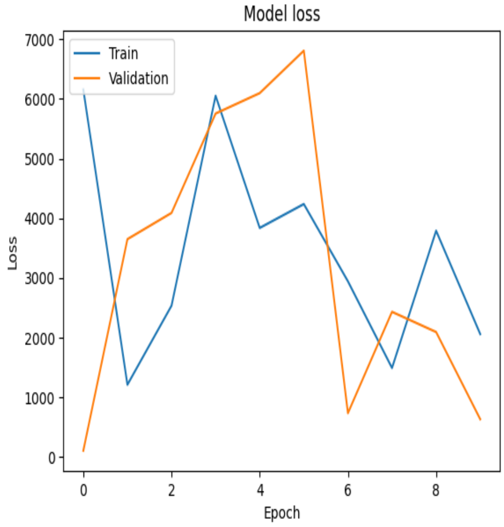


Fig. 8: Model Loss

The graph Model Accuracy from Fig.7 and Fig.8 represents the accuracy of the Phishing Detection with an accuracy of 99.2% and with a Model Loss of 79%.

Table 1: Classification Report

Parameters	Precision	Recall	F1-score	Support
Attacker	0.99	0.99	0.99	17666
Normal	0.99	0.99	0.99	20126
Accuracy			0.99	37792
Macro avg	0.99	0.99	0.99	37792
Weighted avg	0.99	0.99	0.99	37792

By continuously monitoring the system's efficiency and adapting to emerging phishing trends and techniques, we can ensure that the deep learning-based IDS remains relevant, resilient, and capable of effectively mitigating the ever-evolving threat of phishing attacks in dynamic network environments.

Through this comprehensive performance analysis approach, we can rigorously evaluate the proposed deep learning-based IDS, identify areas for improvement, and ultimately deliver a robust and effective solution for enhancing network security and combating the escalating threat of phishing attacks.

5. CONCLUSION

This groundbreaking research has demonstrated the immense potential of leveraging advanced DLAs, such as Convolutional Neural Networks (CNNs), to combat the ever-evolving landscape of phishing attacks. Through rigorous experimentation and meticulous optimization, the proposed deep learning-based IDS have exhibited remarkable accuracy and efficiency in identifying nuanced phishing patterns within complex network traffic. The system's ability to autonomously learn and adapt, coupled with real-time analysis and rapid response mechanisms, positions it as a formidable defense against sophisticated phishing tactics.

By significantly reducing false positives and false negatives, organizations can confidently thwart phishing attempts, safeguarding sensitive data and user privacy. Moreover, the automation and real-time detection capabilities alleviate the burden on cyber security personnel, allowing them to focus on strategic security measures while ensuring timely responses to emerging threats. Furthermore, the system's adaptability and continuous learning mechanisms ensure its resilience against novel phishing techniques, providing a proactive defense in the ever-changing threat landscape. As we look to the future, the integration of deep learning algorithms within cyber security frameworks holds immense promise. By combining cutting-edge technologies with a deep understanding of human behavior and social engineering tactics, we can develop comprehensive and resilient defense mechanisms against phishing and other cyber threats.

References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first . . .”

[1] Al-Ahmadi, S., Alotaibi, A., & Alsaleh, O. (2022). PDGAN: Phishing Detection With Generative Adversarial Networks. *IEEE Access*, *10*, 42459–42468. https://doi.org/10.1109/ACCESS.2022.3168235

[2] Alsubaei, F. S., Almazroi, A. A., & Ayub, N. (2024). Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics. *IEEE Access*, *12*(January), 8373–8389. https://doi.org/10.1109/ACCESS.2024.3351946

[3] Do, N. Q., Selamat, A., Krejcar, O., Herrera-Viedma, E., & Fujita, H. (2022). Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions. *IEEE Access*, *10*, 36429–36463. https://doi.org/10.1109/ACCESS.2022.3151903

[4] El Aassal, A., Baki, S., Das, A., & Verma, R. M. (2020). An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs. *IEEE Access*, *8*, 22170–22192. https://doi.org/10.1109/ACCESS.2020.2969780

[5] Li, W., Manickam, S., Laghari, S. U. A., & Chong, Y. W. (2023). Uncovering the Cloak: A Systematic Review of Techniques Used to Conceal Phishing Websites. *IEEE Access*, *11*(July), 71925–71939. https://doi.org/10.1109/ACCESS.2023.3293063

[6] Purwanto, R. W., Pal, A., Blair, A., & Jha, S. (2022). PhishSim: Aiding Phishing Website Detection With a Feature-Free Tool. *IEEE Transactions on Information Forensics and Security*, *17*, 1497–1512. https://doi.org/10.1109/TIFS.2022.3164212

[7] Sahingoz, O. K., Buber, E., & Kugu, E. (2024). DEPHIDES: Deep Learning Based Phishing Detection System. *IEEE Access*, *12*(January), 8052–8070. https://doi.org/10.1109/ACCESS.2024.3352629

[8] Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. *IEEE Access*, *10*, 65703–65727. https://doi.org/10.1109/ACCESS.2022.3183083

[9] Sameen, M., Han, K., & Hwang, S. O. (2020). PhishHaven - An Efficient Real-Time AI Phishing URLs Detection System. *IEEE Access*, *8*, 83425–83443. https://doi.org/10.1109/ACCESS.2020.2991403

[10] Yang, P., Zhao, G., & Zeng, P. (2019). Phishing website detection based on multidimensional features driven by deep learning. *IEEE Access*, *7*, 15196–15209. https://doi.org/10.1109/ACCESS.2019.2892066.

[11] Z. S. Malek, B. Trivedi and A. Shah, "User behavior Pattern - Signature based Intrusion Detection," *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, London, UK, 2020, pp. 549-552, doi: 10.1109/WorldS450073.2020.9210368.

[12] A. Kiran, S. W. Prakash, B. A. Kumar, Likhitha, T. Sameeratmaja and U. S. S. R. Charan, "Intrusion Detection System Using Machine Learning," *2023 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2023, pp. 1-4, doi: 10.1109/ICCCI56745.2023.10128363.

[13] J. Chen, Y. Guo, K. Shi and M. Yang, "Network Intrusion Detection Method of Power Monitoring System Based on Data Mining," *2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI)*, Guangzhou, China, 2022, pp. 255-259, doi: 10.1109/AHPCAI57455.2022.10087405.

Mrs.J.Jayapradha is now currently working as Assistant Professor in the department of Computer Science and Engineering in Manakula Vinayagar Institute of Technology. She has an experience of 14 + years as an academician. Her domain skills include Internet of Things, Cloud Computing, Web Programming and Automata Theory. She is trying to dive into the world of Machine Learning through her research works. She has published various journals in the related fields.

S.Vineethkumar pursuing the B.Tech. degree in Computer Science and Engineering from Manakula Vinayagar Institute of Technology, of Pondicherry University, Puducherry, in 2024. He has done some industry projects that explore emerging technologies such as Artificial Intelligence, Machine Learning, Neural Networks, Speech Recognition, and Computer Vision. He had published his innovative project work in journals and conference papers.

R.Vigneshwaran pursuing the B.Tech. degree in Computer Science and Engineering from Manakula Vinayagar Institute of Technology, Pondicherry University, Puducherry, in 2024. he has published a paper on "Web Application for Cardiovascular Disease Diagnosis using Data Science" in the International Journal of the Innovative Research in Technology. He also presented two papers at IEEE conferences, focusing on cancer diagnosis using YoloV8 and real-time object detection using CNN.

A.Ramprasath pursuing the B.Tech. degree in Computer Science and Engineering from Manakula Vinayagar Institute of Technology, Pondicherry University, Puducherry, in 2024. he presented two papers at IEEE conferences, focusing on cancer diagnosis using YoloV8 and real-time object detection using CNN. He also presented IOT patent project named “Smart Door Lock Automation” at ManakulaVinayager Institute of Technology on National Science Day celebration (SCIMIT ‘23). He had participated in the workshop on Augmented Reality (AR) and Virtual Reality (VR).