

# On the Adoption of Explainable Deep Learning for Image-Based Network Traffic Classification

Amine Hattak<sup>1,3</sup>, Fabio Martinelli<sup>1</sup>, Francesco Mercaldo<sup>2,1</sup> <sup>a</sup> and Antonella Santone<sup>1</sup>

<sup>1</sup>*Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy*

<sup>2</sup>*University of Molise, Campobasso, Italy*

<sup>3</sup>*La Sapienza, University of Rome, Rome, Italy*

**Keywords:** Internet of Things, Network Traffic Classification, Deep Learning, Network Intrusion Detection, Explainable AI.


**Abstract:** In an era marked by escalating cyber threats, ensuring the security of interconnected devices and networks within the Internet of Things (IoT) landscape is imperative. This paper addresses this pressing concern by delving into network security, focusing on the classification of network traffic through the lens of deep learning techniques. Our study presents a deep learning-based approach customized for network traffic classification in the IoT domain, based on image analysis. Crucially, to enhance the interpretability and the transparency in our model's decisions, we integrate GradCAM (Gradient-weighted Class Activation Mapping), a technique that illuminates the salient regions of input images contributing to the model's predictions. By leveraging GradCAM, we provide deeper insights into the decision-making process, enabling better understanding and trust in our approach. We evaluate the effectiveness of our methodology using the TON\_IoT dataset, consisting of 10 network traces categorized into various vulnerability scenarios and trusted applications. Our findings reveal a remarkable accuracy of 99.1%, demonstrating the potential of our approach in fortifying network security within IoT environments. Moreover, the utilization of GradCAM empowers stakeholders with valuable insights into the inner workings of the model, further enhancing its applicability and trustworthiness.

## 1 INTRODUCTION

The Internet of Things (IoT) has emerged as a revolutionary force, altering our interactions with technology and the world around us. With billions of interconnected devices across multiple domains such as smart homes, healthcare systems, industrial facilities, and transportation networks, IoT has enabled new levels of connectivity, efficiency, and convenience. However, this networked ecosystem provides a fertile field for cyber threats, which have the potential to disrupt and undermine the core fabric of IoT infrastructure.

Cyber-attacks on IoT devices have become more common and sophisticated. These assaults endanger both individual privacy and larger social security by exploiting weaknesses in IoT devices such as weak authentication mechanisms, insecure communication protocols, insufficient software updating, and configuration issues. Such flaws jeopardize the security, integrity, and availability of data transmitted by

IoT devices while also allowing hostile actors to plan large-scale disruptions and attacks. The IoT domain faces a diverse and evolving threat landscape, ranging from distributed denial-of-service (DDoS) attacks leveraging compromised IoT botnets (Kolias et al., 2017) to ransomware campaigns targeting vulnerable smart devices (Yaqoob et al., 2017). Cyber-attacks such as MITM (Li et al., 2017) are also present. Furthermore, successful cyber-attacks against IoT infrastructure can have far-reaching implications, including bodily harm, financial losses, and reputational damage to individuals, businesses, and entire sectors. As IoT expands and integrates into every part of modern life, protecting this networked ecosystem from cyber threats has become critical. To limit the dangers posed by hostile actors, effective cybersecurity strategies must include comprehensive methods for device authentication, data encryption, intrusion detection, and incident response (Vishwakarma 2020 survey). In a study work (Hattak et al., 2023), a unique approach to network traffic classification uses Deep Learning techniques (Zhou et al., 2023; Huang et al., 2023;

<sup>a</sup>  <https://orcid.org/0000-0002-9425-1657>

Huang et al., 2024; Mercaldo et al., 2022) is proposed. This paper describes a novel approach that blends Deep Learning and image-based representations to produce robust and explainable network traffic classification. The authors of the following paper (Moustafa, 2021) present a distributed architecture for testing AI-based security solutions at the edge using TON\_IoT datasets. It focuses on improving security measures in IoT networks using unique architectural techniques. A study (Booij et al., 2021) highlights the necessity of standardizing attributes and attack types in IoT network intrusion datasets such as ToN\_IoT. It emphasizes the importance of heterogeneity in increasing the efficiency of intrusion detection systems in IoT contexts. (Martinelli et al., 2022) suggested an intrusion detection approach for smart grids. Starting with network traffic packet files, they generated a feature vector with 26 distinct features. They used supervised machine learning to create seven different models, with five of them achieving high precision and recall ratings.

In order to detect breaches in the Internet of Things environment, we describe in this work a deep-learning-based approach to solve these security challenges. Through the application of visualization techniques, we are able to transform PCAP format raw collected network traffic files into images, which allows for improved analysis and detection capabilities. Moreover we provide deeper insights into the decision-making process, enabling better understanding and trust in our approach by adopting GradCAM techniques.

## 2 THE METHOD

In this paper, we provide a deep learning model-based approach for classifying network traffic based on images, with an emphasis on the Internet of Things. By incorporating image analysis techniques for the Internet of Things ecosystem, our suggested solution seeks to validate the precision and robustness of intrusion detection systems in IoT through image analysis. We break down our suggested technique into discrete steps, which are shown in Figure 1. This graphical representation illustrates how we envisage assessing deep learning models for network intrusion detection and prediction. The first phase involves gathering samples or a dataset for IoT network traffic. This dataset should include both "malware" and "normal" traces. To guarantee the resilience of later model training and assessment, every sample must be carefully labeled and grouped into unique families. The basis for training and testing the deep learning mod-

els is this carefully chosen dataset.

The study aims to go deeper into network traffic analysis by deploying classification in two different strategies. The first one is based on a binary classification in which we classify the samples as "Normal" or "Malware" disregarding the nature of the original network traffic. We also deployed a more detailed 10-class classification scheme where the samples will be classified based on their corresponding classes, we enlarged paradigm involves classifying network traffic into various families, each representing a different type of activity or communication pattern. By using a multi-class categorization approach, the study hopes to improve the granularity of network traffic analysis, allowing for more precise insights into the heterogeneous nature of network activity. This comprehensive classification scheme not only helps to identify malicious behavior, but it does also provide valuable contextual information about the nature and origins of various network activities, improving the overall effectiveness of intrusion detection and network security measures in the IoT sector. Our strategy is concentrating on the packet headers or metadata and not only the content of the traffic or as it is known by "Payload". This approach is considered robust because of the fact that packet headers include critical information that is useful for network communication and analysis. Following that, we propose the transformation of network data supplied in PCAP format into displayed images, which represents the second step of our suggested approach. This transformation prepares the data for use by deep learning models, allowing for smooth incorporation into training and testing procedures. Notably, image production can be performed in either grayscale or color (RGB) mode to meet individual needs. The next step is to resize the input images to maintain uniformity in dimensions, which is required to reduce the potential loss of information that could compromise the accuracy of deep learning models. These models are then rigorously trained and tested using the selected dataset, with performance metrics compared to the classification job.

The raw network traffic data will be pre-processed to remove any extraneous information or noise that may interfere with the categorization operation. Furthermore, in the image generation phase we will count on both the packet headers and the packet payloads to ensure the durability and flexibility of our method to network intrusion detection in IoT. This method is useful in two scenarios: encrypted and non encrypted network traffic. If the traffic is encrypted, the bytes in the packet payload will be random and indistinguishable from noise. In this instance, it is not possible to generate meaningful images directly from encrypted

payload bytes. However, to generate images from encrypted traffic for viewing reasons, we must adopt a different approach. One such strategy is to concentrate also on the packet headers or metadata rather than the content of the packet only. Packet headers include critical information for network communication and analysis. This information can be described as source and destination addresses, protocols, ports, total length, fragment offset, time to live (TTL), and choices. Following that, we advocate for the transformation of network data supplied in PCAP format into displayed images, which represents the second step of our proposed method. This transformation prepares the data for use by deep learning models, allowing for smooth incorporation into training and testing procedures. Notably, image production can be carried out in either grayscale or color (RGB) modes to accommodate specific requirements. The following action to take is to resize the input photos to maintain uniformity in dimensions, which is required to reduce the potential loss of information that could compromise the accuracy of deep learning models. These models are then rigorously trained and tested using the selected dataset, with performance metrics compared to the classification job.

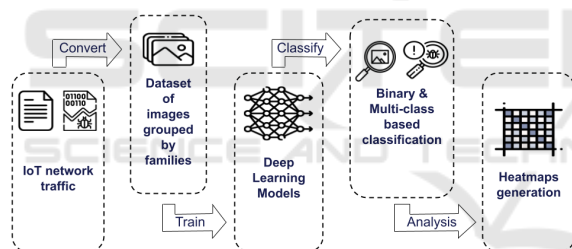


Figure 1: Overall schema of the method.

In this study, we leverage deep learning models to present a novel approach for classifying network traffic based on images, with a particular emphasis on the Internet of Things. By incorporating image analysis techniques specifically designed for the IoT, the suggested solution seeks to improve the precision and robustness of network intrusion detection systems. The process consists of multiple crucial phases, starting with gathering and classifying a dataset of "malware" and "normal" network traffic traces within IoT environment. The methods are investigated is a simple binary classification to distinguishes between "Normal" and "Attack" classes and a complex 10-class classification scheme that provides in-depth insights into a variety of network activities. To ensure robustness and flexibility, pre-processing of raw network traffic data entails eliminating noise and deploying packet headers and payloads that may be applied

to both encrypted and non-encrypted traffic. PCAP-formatted network data is used to create visualized images, which are then resized for uniformity. Ultimately, the efficacy of the suggested methodology is confirmed by training and assessing deep learning models on the carefully selected dataset in order to measure performance against the classification problem.

### 3 EXPERIMENTAL ANALYSIS

This section describes the experiment used to validate the efficacy of the suggested methodology. Initially, we define the (real-world) datasets chosen for study, followed by a presentation of the experimental outcomes.

#### 3.1 Experimental Setup

The experiments were conducted on a workstation equipped with an Intel Core i7 11 the generation processor (2.3 GHz), 16GB of RAM, and an NVIDIA GeForce RTX 3070 GPU. The operating system used was Ubuntu 22.04.2 LTS (GNU/Linux 5.15.133.1-microsoft-standard-WSL2 x86\_64). Python 3.8 along with TensorFlow 2.4 and PyTorch 1.7 libraries were utilized for model training and evaluation. All experiments were run using a freely to use tool for analyzing malware as images (Iadarola et al., 2021).

#### 3.2 Dataset

The TON\_IoT dataset has been utilized in many studies (Moustafa et al., 2020b) (Moustafa, 2019) (Moustafa, 1906) (Ashraf et al., 2021) including studies on cybersecurity and intrusion detection. The article "Analysis of ToN\_IoT, UNW-NB15, and Edge-IIoT Datasets Using DL in Cybersecurity for IoT" was published in the journal Applied Sciences in 2022 (Tareq et al., 2022). The study investigated the ToN\_IoT database thoroughly in order to develop cybersecurity models. According to (Alsaedi et al., 2020), the TON\_IoT Telemetry Dataset represents a new generation of data-driven intrusion detection systems. The TON\_IoT dataset is accessible for academic research and has been used to create realistic botnet datasets for network forensic analytics (Dr Nickolaos Koroniotis, 2021). The TON\_IoT dataset is a cutting-edge collection of datasets intended for testing the usefulness and fidelity of various cybersecurity solutions based on Artificial Intelligence (AI) in the realms of Industry 4.0/Internet of Things (IoT) and Industrial IoT (IIoT). The 'ToN\_IoT' datasets collect data

from several sources, including IoT and IIoT sensor telemetry, operating system datasets (Windows 7 and 10, Ubuntu 14 and 18 TLS) (Moustafa et al., 2020a), and network traffic datasets. The datasets were carefully collected from a realistic large-scale network environment built at the IoT Lab of UNSW Canberra Cyber, School of Engineering and Information Technology (SEIT), UNSW Canberra at the Australian Defence Force Academy (ADFA) (UNSW, 2024). The dataset directories include raw datasets, including IoT/IIoT data logged in log and CSV files from various sensors, such as weather and Modbus sensors, network datasets in PCAP formats from ZEEK (Bro) tool (Author(s), 2020) (Author(s), 2023), and Linux datasets captured through tracing tools on Ubuntu systems. The TON\_IoT dataset is a significant resource for training Machine Learning and Deep Learning algorithms to improve cybersecurity in IoT ecosystems. The TON\_IoT dataset is a valuable resource for academics studying cybersecurity in IoT contexts. It provides a diverse range of data to improve intrusion detection systems and AI-based security solutions. Tables 1 provide a detailed description of Ton\_IoT's network statistics.

Table 1: Statistics of Network Records for Ton\_IoT network dataset.

Type	No of rows
Backdoor	508116
DDoS	6165008
DOS	375328
Injection	452659
MITM	1052
Password	1718568
Ransomware	72805
Scanning	7140161
XSS	2108944
Normal	796380

In this study, we employed networked datasets in PCAP format, which were acquired using the Zeek tool by the IoT Lab of UNSW Canberra. Following pre-processing and picture transformation, we divide the data into three sets: training, validation, and test, with an 80:10:10 split ratio. The test set for both datasets (the binary and multi-class) contains 2,764 samples, whilst the training set contains 24,806 samples, which are further divided into 22,053 for training and the remaining 2,753 for validation. To maintain a balanced distribution, samples from each family are spread evenly among the sets.

### 3.3 Image Generation

Converting collected network data in PCAP format to an image allows us to visualize network traffic data in a more intuitive and human-readable format, while also giving valuable input for deep learning model training. This method improves the efficiency and effectiveness of network traffic pattern analysis and interpretation by discovering trends and abnormalities in network traffic that are not immediately visible in raw data. The method normally begins by scanning the binary data of the PCAP file and extracting the packet header, which contains information such as source and destination IP addresses, port numbers, protocol types, packet sizes, and so on. After that, calculating the image size, rearranging the data to make a 2D array, and finally, we will end up with visualised images that can be used for training deep learning models for pattern extraction and classification tasks. The generated images through this process will be used as input for training various deep-learning models. The ability of deep learning models to automatically extract features and patterns from images makes them well-suited for analyzing IoT network traffic data. Furthermore, the ability to create a large dataset of network traffic images will be used to train and evaluate different deep learning architectures, such as models that are known in the literature (MobileNet, VGG16, etc). The main function "process\_pcap(pcap\_path)" is defined to handle packet processing. It creates an empty list "current\_bytes" to collect bytes from packets, and then iterates through each packet in the PCAP file with "rd\_pcap(pcap\_path)". The raw bytes from each packet are extracted and appended to current\_bytes if they exceed the image's maximum size. It generates an image from the accumulated bytes using the create\_image() function, and it constructs a filename depending on the packet summary and the number of bytes accumulated. After processing all packets, it saves the image to the specified output location, resets current\_bytes to an empty list, and begins accumulating bytes for the next picture. If there are any remaining bytes in current\_bytes, it generates an image and saves it using the same method as before.

Finally, images are created from the accumulated bytes of packets, ensuring that no data is lost in the process. Each image filename includes the packet summary and the size of the gathered bytes, providing context for the data used to create the image. Figure 2 shows the final output from PCAP to picture. It is evident that diverse network traffic produces different images with distinct properties for each family.

**Data:** PCAP

**Result:** Generating PNG images  
initialization;

```

while Packets are not finished do
  Extract raw bytes from the packet and
  append them to current_bytes;
  if accumulated bytes exceed maximum
  size allowed for an image then
    Create an image from accumulated
    bytes using create_image()
    function Construct a filename based
    on packet summary and the number
    of accumulated bytes;
    Save the image to the specified output
    directory;
    Reset current_bytes to an empty list
    to start;
    accumulating bytes for the next
    image;
  else
    there are remaining bytes in
    current_bytes Create an image and
    save it using the same procedure as
    above;
  end
end

```

Algorithm 1: From PCAP to PNG Image.

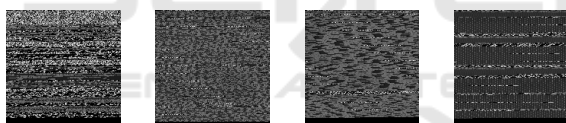


Figure 2: A visualization analysis of some network traffic samples that are belonging to different classes of Ton\_IoT dataset.

### 3.4 Results and Discussion

#### 3.4.1 Experiments

Table 2 presents the hyperparameters used for various deep learning (DL) models in the context of the research work. These parameters are critical as they directly influence the performance and efficiency of the models in processing and analyzing network traffic data.

The models encompassed in this study include CNN (Convolutional Neural Network), MobileNet, ResNet50, and VGG16, each distinguished by unique architectures and capabilities suited for various tasks within network traffic analysis and intrusion detection. Standardized input dimensions of 224x224 pixels with three RGB color channels are maintained across all models to ensure consistency and compa-

rability in the experimental setup. The pivotal parameters of epochs and batch size, crucial for the training process of DL models, involve a configuration of 40 epochs and a batch size of 32 across all models, striking a balance between computational efficiency and model convergence. Additionally, the depth of DL models, denoted by the number of layers, significantly impacts their ability to extract intricate patterns and features from input data.

Table 3 compares machine learning models' performance on test sets for binary (2 classes) and multi-class (10 classes) classification tasks. The models tested were Convolutional Neural Network (CNN), MobileNet, ResNet50, and VGG16. The performance metrics reported include accuracy (Acc), precision (Prec), recall (Rec), F1-score (F1), and area under the receiver operating characteristic curve (AUC) for the binary classification task (2 classes) and similarly, for the multi-class classification task (10 classes).

The results show that for all models, the binary classification task outperforms the multi-class classification job. This difference is expected because binary classification tasks are fundamentally simpler than multi-class classification problems. CNN routinely outperforms other models in both binary and multi-class classification tasks, with accuracy, precision, recall, F1-score, and AUC values more than 0.9. This shows that CNN is good at identifying underlying patterns in data and making accurate predictions. MobileNet, ResNet50, and VGG16 also perform well on both classification tasks, with accuracy values ranging from about 0.9 to 0.95 for binary classification and around 0.9 to 0.94 for multi-class classification. These results indicate that these models are capable of achieving high levels of accuracy and reliability in classifying images across different categories.

Overall, the results reported show that deep learning models, particularly CNN, are successful at classifying network traffic based on images. These findings can help guide the selection of appropriate machine learning models for comparable categorization problems in practical applications.

#### 3.4.2 Explainability

Following the flow of our work discussed in section 2 after the training and testing phases, we will generate heatmaps for the test set of our dataset. For this reason 3 different heatmaps were convocated for providing explainability of decision macking for the best performing model (according to section 3.4.1 Gradcam (Selvaraju et al., 2016), Gradcam++ (Jamil et al., 2023) and scorecam (Wang et al., 2020). The three separate CAM techniques were used to determine

Table 2: The hyperparameters of different DL models.

Model	CNN	MobileNet	ResNet50	VGG16
Input image/vector size	224x224x3			
Epochs and Batch size	40 - 32			
Number of layers	13	29	50	16

Table 3: Comparison between the results of different models on the test sets (binary and 10 classes classification).

Mode	2 classes					10 classes					
	Model	Acc	Prec	Rec	F1	Auc	Acc	Prec	Rec	F1	Auc
CNN	0.991	0.991	0.991	0.991	0.991	0.993	0.899	0.904	0.896	0.9	0.981
MobileNet	0.916	0.917	0.913	0.915	0.915	0.975	0.916	0.919	0.915	0.917	0.977
ResNet50	0.952	0.954	0.951	0.952	0.952	0.99	0.948	0.949	0.947	0.948	0.989
VGG16	0.945	0.95	0.944	0.947	0.947	0.991	0.948	0.95	0.946	0.948	0.994

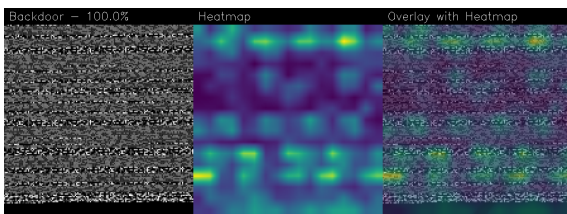


Figure 3: An example of a sample correctly classified as Backdoor class plot after applying the Grad-CAM algorithm.

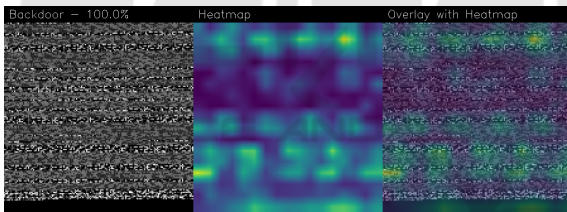


Figure 4: An example of a sample correctly classified as Backdoor class plot after applying the Grad-CAM++ algorithm.

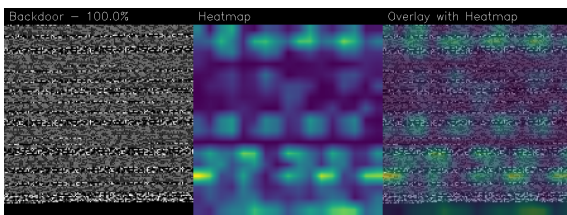


Figure 5: An example of a sample correctly classified as Backdoor class plot after applying the Score-CAM algorithm.

which portions of the photos were most distinctive for classification using the CNN architecture-trained model. As a result of using CAMs, we acquired a large number of .png images with the activation validation stages map overlaid on the original image. In particular, the heatmaps obtained were made up of

three separate colors: blue, yellow, and green. Areas overshadowed with blue show the model’s indifference. Instead, the model is drawn to the yellow areas of the image. Green spaces, on the other hand, are utilized to identify central locations. In most instances, all CAM methods have highlighted the same areas, indicating that they are the most relevant parts of the image for categorization. The results obtained after the CAMs execution allow us to certify that all of the algorithms can identify plots belonging to "Backdoor" class with 100% accuracy.

#### 4 CONCLUSIONS

In this paper, we suggested a graphical network analysis technique for the IoT ecosystem with the goal of performing multi-class classification to automatically classify between benign and malicious network traces in multi-class scenario. In detail, we propose extracting packet headers that contain information about the source and destination addresses, protocols, ports, and other relevant metadata in the case of encrypted traffic, and representing these network traces in the form of images that will be used as input for several deep-learning models to detect the application that generated the specific network trace. The highest score achieved in our results was 99.1% by CNN model in binary classification scenario and 94.8% by ResNet50 and VGG16 in terms of accuracy in the multi-class classification task. as well as we applied three different Class Activation Mapping algorithms: Grad-CAM, Grad-CAM++ and Score-CAM to provide a better understanding of the decision making for classification. As future work, we want to consider more contemporary deep-learning algorithms such as the Vision Transformers (ViT), We intend to examine the robustness of the proposed technique, as well as its resilience in real-world

circumstances.

## ACKNOWLEDGEMENTS

This work has been partially supported by EU DUCA, EU CyberSecPro, SYNAPSE, PTR 22-24 P2.01 (Cybersecurity) and SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the EU - NextGenerationEU projects, by MUR - REASONING: foRmal mEthods for computAtional analySis for diagnOsis and progNosis in imagING - PRIN, e-DAI (Digital ecosystem for integrated analysis of heterogeneous health data related to high-impact diseases: innovative model of care and research), Health Operational Plan, FSC 2014-2020, PRIN-MUR-Ministry of Health, the National Plan for NRRP Complementary Investments D<sup>3</sup> 4 Health: Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care, Progetto MolisCTe, Ministero delle Imprese e del Made in Italy, Italy, CUP: D33B22000060001 and FORESEEN: FORmal mEthodS for attack dEtECTION in autonomous driving systems CUP N.P2022WYAEW.

This work has been carried out within the Italian National Doctorate on Artificial Intelligence run by the Sapienza University of Rome in collaboration with the Institute of Informatics and Telematics (IIT), the National Research Council of Italy (CNR).

## REFERENCES

- Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A., and Anwar, A. (2020). Ton\_iiot telemetry dataset: A new generation dataset of iiot and iot for data-driven intrusion detection systems. *Ieee Access*, 8:165130–165150.
- Ashraf, J., Keshk, M., Moustafa, N., Abdel-Basset, M., Khurshid, H., Bakhshi, A. D., and Mostafa, R. R. (2021). Iotbot-ids: A novel statistical learning-enabled botnet detection framework for protecting networks of smart cities. *Sustainable Cities and Society*, 72:103041.
- Author(s) (2020). zeek-osquery: Host-network correlation for ... *arXiv preprint arXiv:2002.04547*.
- Author(s) (2023). Introducing uwf-zeekdata22: A comprehensive network traffic ... *Journal Name*, 8(1):18.
- Booij, T. M., Chiscop, I., Meeuwissen, E., Moustafa, N., and Den Hartog, F. T. (2021). Ton\_iiot: The role of heterogeneity and the need for standardization of features and attack types in iiot network intrusion datasets. *IEEE Internet of Things Journal*, 9(1):485–496.
- Dr Nickolaos Koroniotis, D. N. M. (2021). The bot-iiot dataset. <https://research.unsw.edu.au/projects/bot-iiot-dataset>.
- Hattak, A., Iadarola, G., Martinelli, F., Mercaldo, F., Santone, A., et al. (2023). A method for robust and explainable image-based network traffic classification with deep learning. In *Proceedings of the 20th International Conference on Security and Cryptography*, pages 385–393.
- Huang, P., Xiao, H., He, P., Li, C., Guo, X., Tian, S., Feng, P., Chen, H., Sun, Y., Mercaldo, F., et al. (2024). La-vit: A network with transformers constrained by learned-parameter-free attention for interpretable grading in a new laryngeal histopathology image dataset. *IEEE Journal of Biomedical and Health Informatics*.
- Huang, P., Zhou, X., He, P., Feng, P., Tian, S., Sun, Y., Mercaldo, F., Santone, A., Qin, J., and Xiao, H. (2023). Interpretable laryngeal tumor grading of histopathological images via depth domain adaptive network with integration gradient cam and priori experience-guided attention. *Computers in Biology and Medicine*, 154:106447.
- Iadarola, G., Casolare, R., Martinelli, F., Mercaldo, F., Peluso, C., and Santone, A. (2021). A semi-automated explainability-driven approach for malware analysis through deep learning. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
- Jamil, M. S., Banik, S. P., Rahaman, G. A., and Saha, S. (2023). Advanced gradcam++: Improved visual explanations of cnn decisions in diabetic retinopathy. In *Computer Vision and Image Analysis for Industry 4.0*, pages 64–75. Chapman and Hall/CRC.
- Kolias, C., Kambourakis, G., Stavrou, A., and Voas, J. (2017). Ddos in the iot: Mirai and other botnets. *Computer*, 50(7):80–84.
- Li, C., Qin, Z., Novak, E., and Li, Q. (2017). Securing sdn infrastructure of iiot-fog networks from mitm attacks. *IEEE Internet of Things Journal*, 4(5):1156–1164.
- Martinelli, F., Mercaldo, F., and Santone, A. (2022). A method for intrusion detection in smart grid. *Procedia Computer Science*, 207:327–334.
- Mercaldo, F., Zhou, X., Huang, P., Martinelli, F., and Santone, A. (2022). Machine learning for uterine cervix screening. In *2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 71–74. IEEE.
- Moustafa, N. (1906). A systemic iiot-fog-cloud architecture for big-data analytics and cyber security systems: A review of fog computing. *arXiv 2019. arXiv preprint arXiv:1906.01055*.
- Moustafa, N. (2019). New generations of internet of things datasets for cybersecurity applications based machine learning: Ton\_iiot datasets. In *Proceedings of the eResearch Australasia Conference, Brisbane, Australia*, pages 21–25.
- Moustafa, N. (2021). A new distributed architecture for evaluating ai-based security systems at the edge: Network ton\_iiot datasets. *Sustainable Cities and Society*, 72:102994.
- Moustafa, N., Ahmed, M., and Ahmed, S. (2020a). Data analytics-enabled intrusion detection: Evaluations of

- ton\_iot linux datasets. In *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 727–735. IEEE.
- Moustafa, N., Keshky, M., Debiez, E., and Janicke, H. (2020b). Federated ton\_iot windows datasets for evaluating ai-based security applications. In *2020 IEEE 19th international conference on trust, security and privacy in computing and communications (TrustCom)*, pages 848–855. IEEE.
- Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D. (2016). Grad-cam: Why did you say that? *arXiv preprint arXiv:1611.07450*.
- Tareq, I., Elbagoury, B. M., El-Regaily, S., and El-Horbaty, E.-S. M. (2022). Analysis of ton-iot, unw-nb15, and edge-iiot datasets using dl in cybersecurity for iot. *Applied Sciences*, 12(19).
- UNSW (2024). School of engineering and technology. <http://www.unsw.edu.au/canberra/about-us/our-schools/engineering-technology>.
- Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25.
- Yaqoob, I., Ahmed, E., ur Rehman, M. H., Ahmed, A. I. A., Al-garadi, M. A., Imran, M., and Guizani, M. (2017). The rise of ransomware and emerging security challenges in the internet of things. *Computer Networks*, 129:444–458.
- Zhou, X., Tang, C., Huang, P., Tian, S., Mercaldo, F., and Santone, A. (2023). Asi-dbnet: an adaptive sparse interactive resnet-vision transformer dual-branch network for the grading of brain cancer histopathological images. *Interdisciplinary Sciences: Computational Life Sciences*, 15(1):15–31.