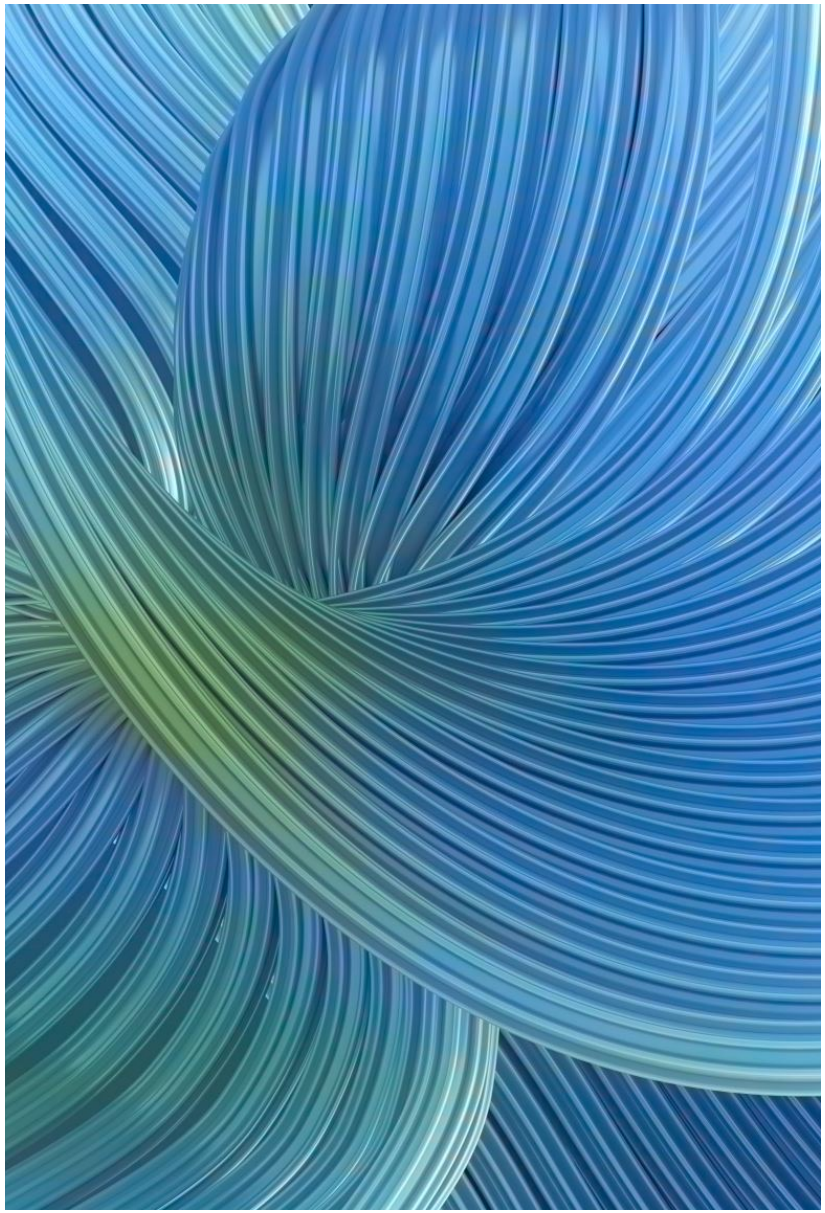
Explainable Artificial Intelligence Framework for Industrial Cybersecurity Threat Detection and Defense

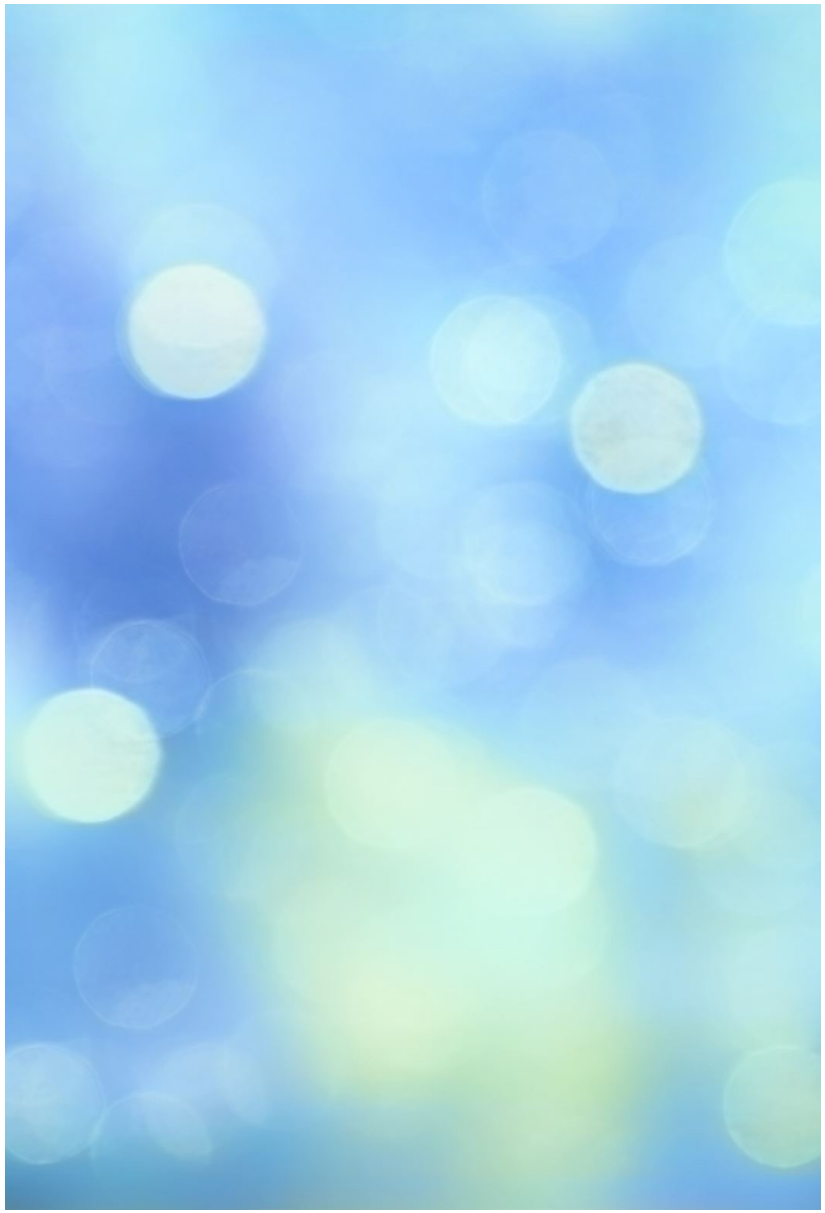
Dimitrios Taketzis, Konstantinos Demertzis, Charalabos Skianis

Department of Information and Communication Systems Engineering (ICSD), University of Aegean



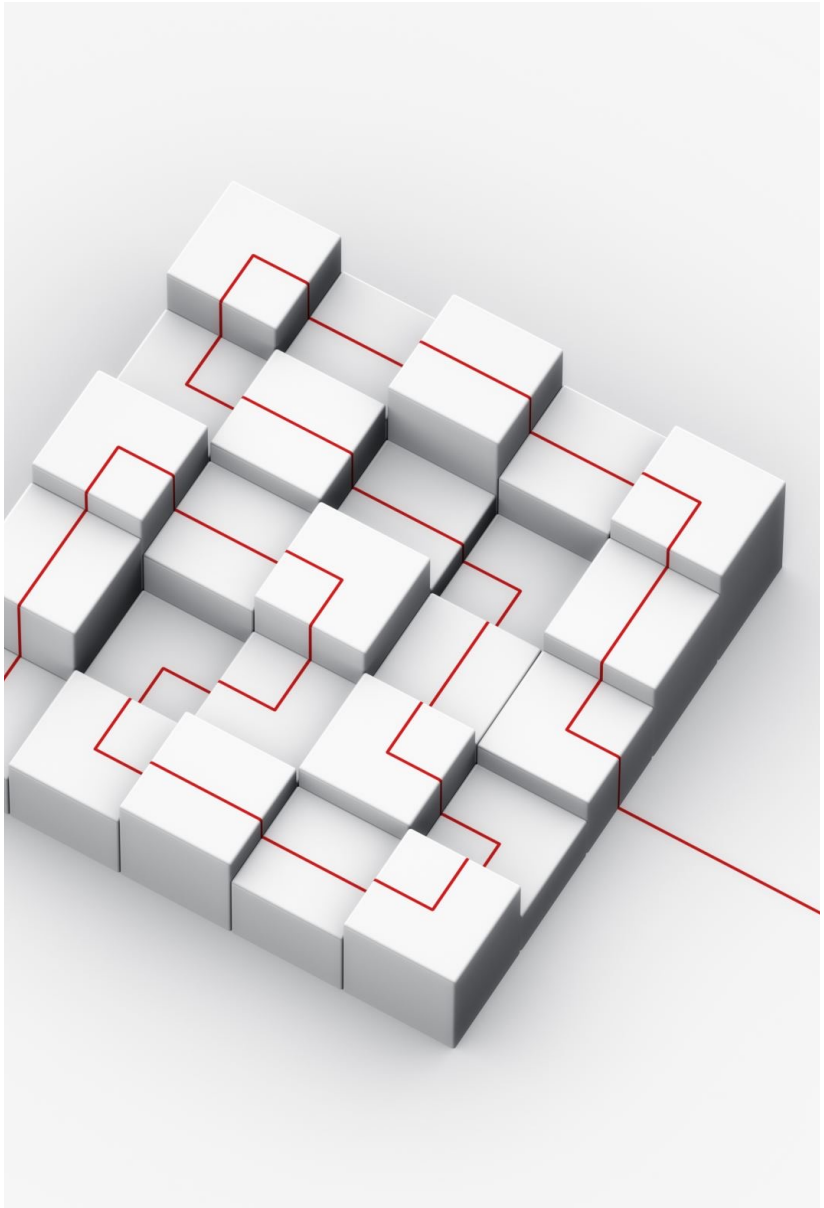
Agenda

- Adoption of AI in Cybersecurity
- The problem
- Our proposal
- Components
- Way ahead
- Our research



Adoption of AI in Cybersecurity

- AI can handle a lot of data
- AI promotes automation
- AI can identify unknown threats (zero-days)
- Reduces detection and response times
- Enhances endpoint protection
- Promotes the timely dissemination of CTI
- …and more



The problem

- Lack of detailed interpretation associated with the use of complex deep learning architectures:
  - can affect the model's performance
  - prevent the frequent adjustment of some critical hyperparameters
  - reduce the algorithm's reliability and the generalization that should characterize the systems in question
- These disadvantages hinder the main stakeholders (e.g., CISO, CEO), from trusting and making meaningful and systematic use of AI-driven cybersecurity systems.

Our proposal

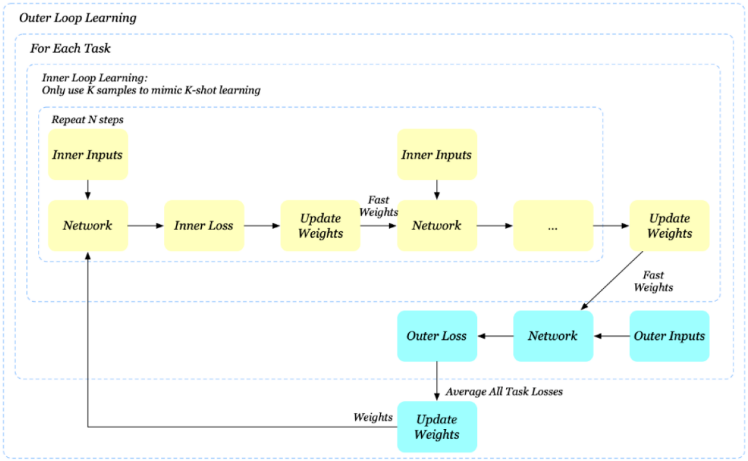
- The construction of AI systems without the need for specialised knowledge
- The system is explainable, interpretable (eg. Why it took a certain decision)
- An explainable artificial intelligence framework for industrial cybersecurity threat detection and defense
- *A* **holistic meta-learning system** that automates selecting and using the most appropriate algorithmic hyperparameters solved by mapping input and output data using the ***Weight Agnostic Neural Networks (WANN) methodology***

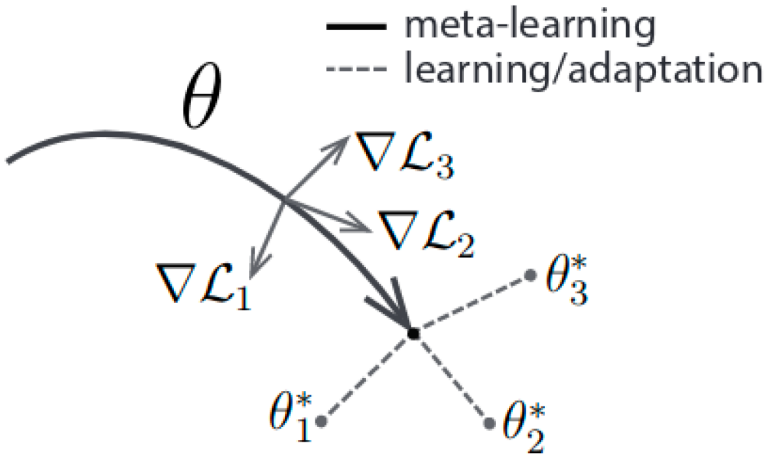
MetaLearning

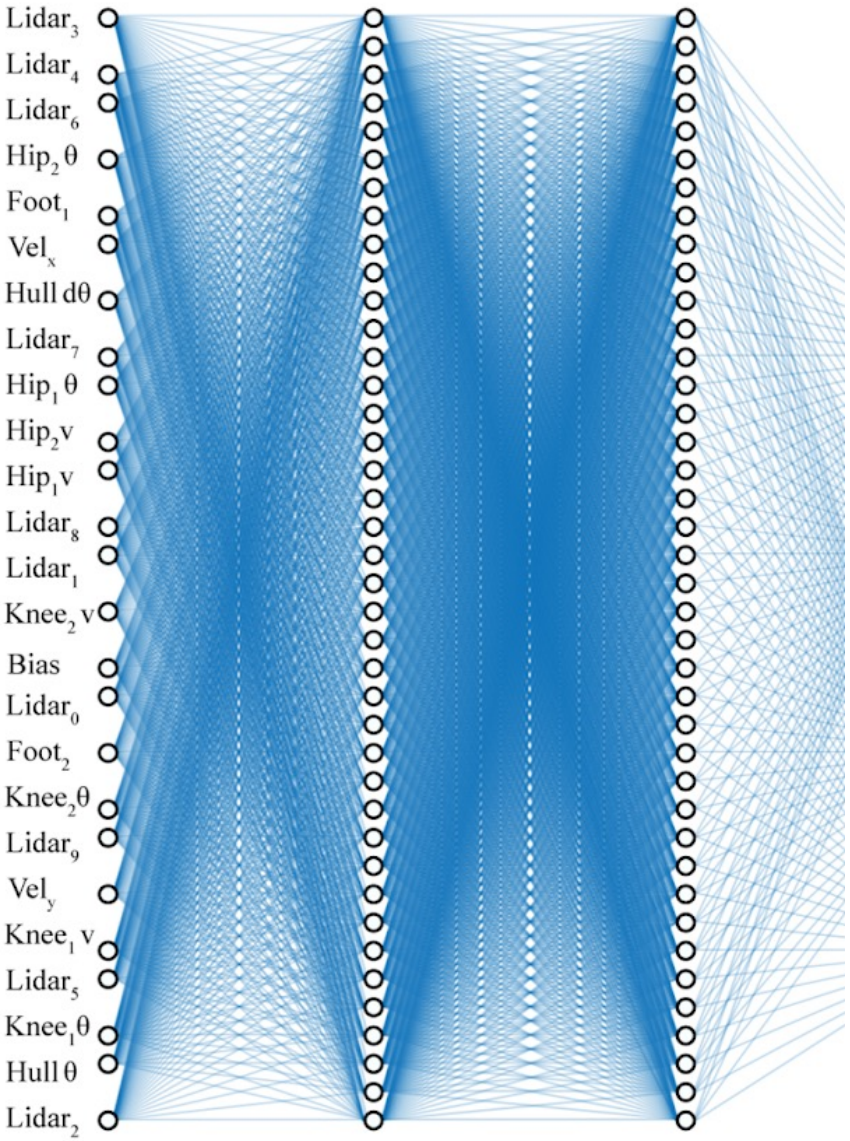
- MetaLearning is a novel holistic approach, which automates and solves the problem of the specialized use of machine learning algorithms.
- It aims for the use of automatic machine learning to learn the most appropriate algorithms and hyperparameters that optimally solve the problem
- A meta-learning system combines three requirements:
  - It must include a learning subsystem
  - Experience has to be gained by utilizing the knowledge extracted from metadata related to the dataset under process or from previous learning tasks that have been completed in similar or different fields
  - Learning bias must be chosen dynamically

MetaLearning

- Demertzis, K.; Iliadis, L. GeoAI: A Model-Agnostic Meta-Ensemble Zero-Shot Learning Method for Hyperspectral Image Analysis and Classification. *Algorithms* **2020**, *13*, 61. https://doi.org/10.3390/a13030061







Weight Agnostic Neural Networks - WANN

- An evolving strategy in neural network development techniques that can perform a specialized task regardless of the weights of the connections in the building blocks of the neural network

Shapley methodology

- Its a very effective way of generating explanations from game theory and specifically a cooperative game.
- The connection of Shapley values with the problem of explaining the architectural structures of WANN is made by considering WANN as a collaborative game whose players are the characteristics of the data set.
- The payoff/gain of the players of a cooperative game is given by a real function that gives values to sets of players.

Shapley methodology

Lloyd Shapley

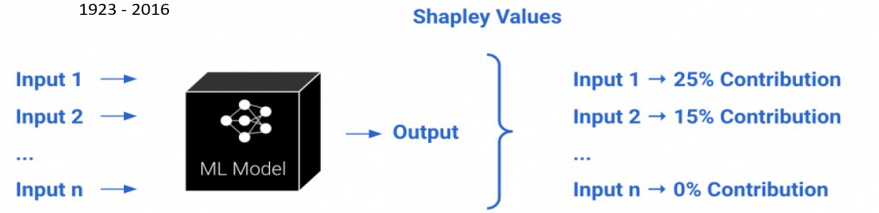
1923 - 2016

Shapley Value

Given a characteristic function game $G = (N, v)$ the Shapley value of a player $i \in N$ is given by:

$$\phi_i(G) = \frac{1}{|N|!} \sum_{\pi \in \Pi_N} \Delta_\pi^G(i)$$

$$\varphi_i(v) = \frac{1}{\text{number of players}} \sum_{\text{coalitions excluding } i} \frac{\text{marginal contribution of } i \text{ to coalition}}{\text{number of coalitions excluding } i \text{ of this size}}$$



Connection between Shapley and WANN

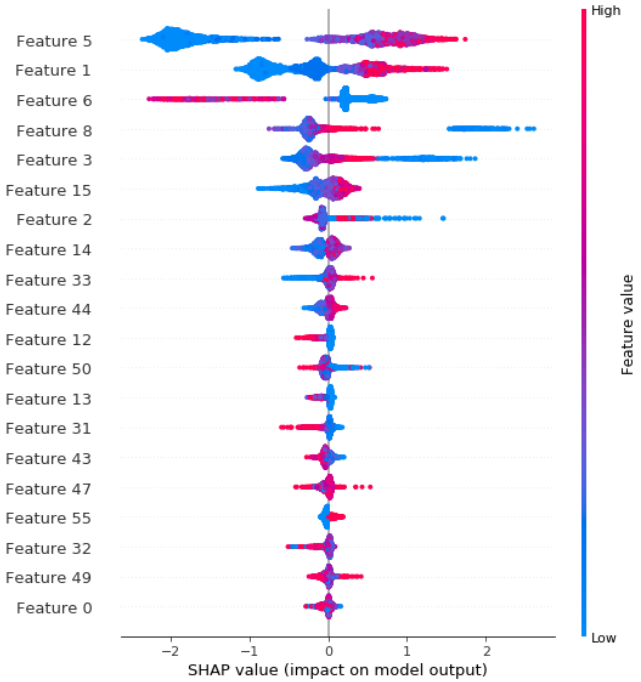
We consider the problem of WANN architectural structures as a cooperative game.

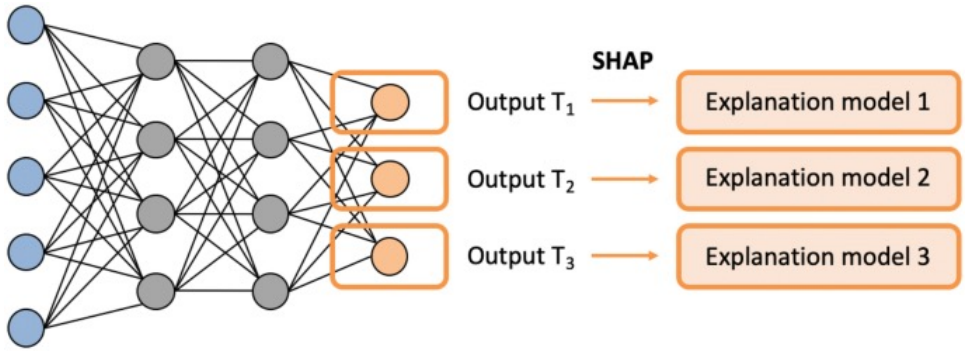
The players are the characteristics of the dataset.

The profit function is the neural network model under consideration, and the model predicts the corresponding profits.

The Shapley values show the contribution of each feature and therefore the explanation of why the model made a specific decision.

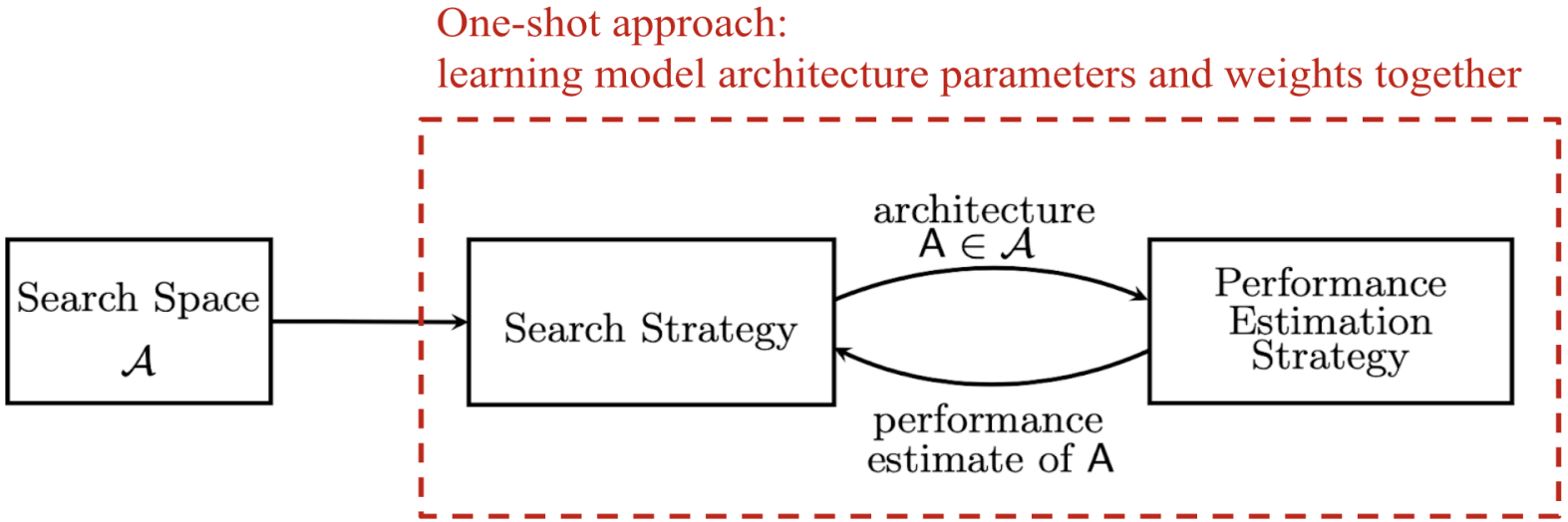
Connection between Shapley and WANN





Way ahead

- Our methodology is a universal system that can adapt to virtually every environment for cybersecurity purposes
- Eg. How to correlate internal organization data (logs, network traffic, events, etc.) with external data or intel (CTI, security repositories, etc.)
- Automated optimization of the appropriate parameters of method pre-training to achieve an even more efficient, accurate, and faster process (fine-tuning)
- Explainable personalized federated learning model to cyber-attacks identification



Our research

- Darknet Traffic Big-Data Analysis and Network Management for Real-Time Automating of the Malicious Intent Detection Process by a Weight Agnostic Neural Networks Framework, Published: 25 March 2021, **https://doi.org/10.3390/electronics10070781**
- Tsiknas, K.; Taketzis, D.; Demertzis, K.; Skianis, C. Cyber Threats to Industrial IoT: A Survey on Attacks and Countermeasures. IoT 2021, 2, 163-186, Published: 07 March 2021 **https://doi.org/10.3390/iot2010009**
- Demertzis, K.; Tsiknas, K.; Taketzis, D.; Skoutas, D.N.; Skianis, C.; Iliadis, L.; Zoiros, K.E. Communication Network Standards for Smart Grid Infrastructures. Network 2021, 1, 132-145. Published: 03 August 2021 **https://doi.org/10.3390/network1020009**
- Our research is ongoing!

References

- Demertzis, K.; Tsiknas, K.; Takezis, D.; Skianis, C.; Iliadis, L. Darknet Traffic Big-Data Analysis and Network Management for Real-Time Automating of the Malicious Intent Detection Process by a Weight Agnostic Neural Networks Framework. Electronics2021, 10, 781. https://doi.org/10.3390/electronics10070781
- Kerschke, P.; Hoos, H.H.; Neumann, F.; Trautmann, H. Automated Algorithm Selection: Survey and Perspectives. Evol. Comput. 2019, 27, 3–45.
- Xu, Z.; Cao, L.; Chen, X. Learning to Learn: Hierarchical Meta-Critic Networks. IEEE Access 2019, 7, 57069–57077.
- Dyrmishi, S.; Elshawi, R.; Sakr, S. A Decision Support Framework for AutoML Systems: A Meta-Learning Approach. In Proceedings of the 2019 International Conference on Data Mining Workshops (ICDMW), Beijing, China, 8–11 November 2019; pp. 97–106.
- Makmal, A.; Melnikov, A.A.; Dunjko, V.; Briegel, H.J. Meta-learning within Projective Simulation. IEEE Access 2016, 4, 2110–2122.
- Demertzis, K.; Iliadis, L. GeoAI: A Model-Agnostic Meta-Ensemble Zero-Shot Learning Method for Hyperspectral Image Analysis and Classification. Algorithms 2020, 13, 61.
- WLee, S.; Bartlett, P.L.; Williamson, R.C. Efficient agnostic learning of neural networks with bounded fan-in. IEEE Trans. Inf. Theory 1996, 42, 2118–2132
- Zhang, K.; Wang, Q.; Liu, X.; Giles, C.L. Shapley Homology: Topological Analysis of Sample Influence for Neural Networks. Neural Comput. 2020, 32, 1355–1378. [CrossRef]
- Zhang, L.; Gao, Z. The Shapley value of convex compound stochastic cooperative game. In Proceedings of the 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Zhengzhou, China, 8–10 August 2011; pp. 1608–1611.

Thank you!

For further contact :

cskianis@aegean.gr