

An Integrated Trustworthy Detection and Classification of Cyber-Physical Attacks in the Presence of Disturbances Using Morphological Image Processing and Explainable AI

Ahmad Abu Nassar¹, Matthew Oinonen¹, Walid Morsi¹

¹Ontario Tech University

Smart Digital Substations (SDSs), are cyber-physical systems (CPSs) that rely on communication networks to exchange information among physical devices, making such CPSs vulnerable to cybersecurity threats. The problem of detecting and classifying attacks in SDSs has been traditionally studied by considering machine learning as a closed box with no interpretation of the decisions made, which has led to untrustworthy algorithms. The attack detection becomes more challenging in the presence of disturbances, as certain types of attacks may exhibit similar behavior to some disturbance events. Furthermore, some attacks may concurrently occur in the presence of disturbances, which may increase the misclassification rate. This paper presents a novel trustworthy approach for detecting and classifying attacks considering the simultaneous occurrence of disturbances in SDSs. This study uses Explainable Artificial Intelligence (XAI) to provide interpretability of the classification decisions using the cyber and physical features in SDSs. This method applies a series of processes, including the use of the Bartlett observation window and morphological image processing, to enhance the learning of the Convolutional Neural Network (CNN) to better extract the hidden features relevant to the attacks and the disturbances when applying the Continuous Wavelet Transform. The proposed approach achieved detection and classification accuracies of 99.37% and 98.44%, while reducing the computational time by 90%, due to the incorporation of a hardware acceleration of multiple graphics processing units (GPUs).