ABSTRACT

Darknet traffic classification is an essential cybersecurity issue that will be addressed in this study due to anonymity, class imbalance, and unavailability of labeled data in Darknet datasets. The objective of paper is a Federated Semi-Supervised Learning model to address these issues, with federated learning's privacy features with semi-supervised techniques for efficient use of both labeled and unlabeled data. The two experiments using the CIC-Darknet2020 dataset as the approach for this study. The first experiment employed the FedAvg aggregation strategy with 10 clients for 10 rounds and achieved a global accuracy of 93.55%. The following experiment compared FedAvg, FedMedian, FedTrimmedMean, and FedKrum using 5 clients for 5 rounds, among which FedMedian achieved the highest accuracy of 88.88%. The study succeeded in taking advantage of pseudo-labeling for enhanced performance promotion, data privacy preservation, and could potentially find everyday application in cyber security, yet one domain which requires future improvement is class imbalance.

Keywords
Darknet Traffic Classification, Federated Learning, Semi-Supervised Learning, Class Imbalance, Pseudo-Labeling, Neural Networks, Cybersecurity, Privacy Preservation