ICCP/SL/OP/219

Privacy-preserving cyberbullying detection using federated learning: A comprehensive review of technologies, challenges and solutions

De Silva DSPSU[*], Bhagya RS[2], Magedara TT[3]

Department of Statistics and Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka
[*]*desilva-ec20065@stu.kln.ac.lk*

Background: In the present, cyberbullying has emerged as a more severe threat than traditional bullying and impacts the children's mental health. Centralized detection systems often violate the user's privacy, limitations for scalability, and struggle with heterogeneous data coming from different platforms. This review was aimed to evaluate how federated learning can solve these problems of data privacy, data heterogeneity and scalability in cyberbullying detection.

Method: This review compared and analyzed the current best deep learning models such as Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and transformer-based models like BERT within federated learning frameworks and examine the privacy enhancing techniques including secure aggregation, differential privacy, and blockchain. These methods were examined in their ability to support learning from non-IID (non-independent and identically distributed) and heterogeneous datasets.

Results: Federated learning (FL) based detection systems achieved accuracy ranging from 74% (n=4) to over 90% (n=3) while preserving user privacy. When considering the traditional deep learning models and transformer-based hybrid models, their accuracy and the robustness is high compared to the federated based detection systems But, computation and communication cost is high with lack of privacy. The study also highlighted the ongoing challenges in optimization and operating under resource limited environments.

Conclusions: Federated learning allows the creation of effective, scalable, and privacy-preserving cyberbullying detection tools that work in resource limited environments. Further research should focus on optimizing the hybrid deep learning modes architectures with enhancing the privacy and strengthening the privacy protocols to bring such technologies to make safer digital space for children.

Keywords: Federated learning, cyberbullying, child protection, deep learning.