Privacy Preserving Machine Learning with EEG Data

Martine De Cock[1,2], Rafael Dowsley[3], Nick McKinney[1],
Anderson C. A. Nascimento[1], Dongrui Wu[4]

[1] University of Washington Tacoma
[2] Ghent University
[3] Aarhus University
[4] DataNova LLC

Machine learning and big data are revolutionizing many areas of research; the era of big data breaches also makes sensitive data especially vulnerable to exploitation by malicious actors. Electroencephalogram (EEG) data are feature rich and highly valuable in the neurosciences. EEG data is by its intimate nature so rich with personal information, often known only to the subconscious, that researchers can extract information beyond their professed scope. Under the right conditions, a malicious actor can extract private information such as passwords, ATM PINs, and many other pieces of private information. As both researchers and industry move to integrate brain-machine interfaces into our daily lives, the implications for data loss, especially in the face of malicious actors, is troubling.

In this work, we address this problem by proposing, to the best of our knowledge, the first specific solution for privately training predictive models based on EEG data. More particularly, we propose a privacy preserving system to detect driver drowsiness based on EEG data. Our solution is based on secure multi-party computation (MPC). More specifically, it is based on the secret sharing approach for secure multi-party computation: all the sensitive information is not given to any single party; instead, each participant receives a share of the information that looks completely random. And the information can only be recovered if all the participants reveal their shares. The usefulness of this approach comes from the fact that it is possible for the participants to collaborate in order to perform computations on the shared values without leaking any information.

In contrast to generic techniques available in the literature for solving general secure MPC problems, our solutions are tailored for the specific kind of data in question (EEG), present efficient communication and computation complexities, and work efficiently even when the number of participants in the protocol are as high as 14. We are not aware of any other implementation of MPC in the literature that works with this number of participants. As a side result, we also propose a generic framework for MPC in Java that extends the results to an arbitrary number of players.