Please accept this submission in response to your invitation:

On behalf of the National Institutes of Health (NIH) BD2K program, we are inviting you to participate in the upcoming annual International Society for Computational Biology meeting, ISMB 2018 in Chicago, IL. BD2K will be hosting several exciting sessions on July 7th-8th, highlighting talks from NIH officials around biomedical data science, BD2K projects, and training programs. Example sessions will include:

* BD2K Power Tools: Faster, Cheaper, Better
* Building the FAIR Data Ecosystem from the Ground Up
* Advancing Biomedical Sciences through Machine Learning

**Affiliated BD2K Project:** Automatic discovery and processing of EEG cohorts from clinical records

**BD2K Grant Number:** U01HG008468

**Letter of Support:** PIs Obeid and Picone are co-PIs on this project and co-authors on this submission.

**Abstract:**

**Curriculum Learning Based on Sample Selection Using Posterior Probabilities**

*Saeedeh Ziyabari, Iyad Obeid and Joseph Picone*

The Neural Engineering Data Consortium

Temple University

Philadelphia, Pennsylvania, 19122, USA

Email: {saeedeh, iobeid, picone}@temple.edu

Training of deep learning algorithms is highly dependent on the order of training samples. Various forms of curriculum learning have been proposed to reduce the sensitivity of the training process. The general concept behind curriculum learning is to use easy samples first and gradually introduce more complex samples. Identifying the difficulty level of samples is a major challenge. We propose a new data selection strategy based on using a less sensitive algorithm that excels at automatic segmentation to triage samples, rank the data based on posteriors generated in this first pass, and then proceed with training a more complex deep learning system using this derived ordering of the data. We use a hybrid hidden Markov model / Stacked denoising Autoencoder based system for the first pass, and a more powerful system based on a Convolutional Neural Network and a Long Short-Term Memory Network for the second pass. We demonstrate this strategy on a seizure detection task based on the TUH EEG Seizure Corpus. Our system produced a sensitivity 32.13% with 10 false alarm per 24 hours, which is very close to our overall best performance, yet is a robust process that can be easily applied to new tasks.

Keywords: electroencephalogram (EEG), curriculum learning, seizure detection, deep learning