**Summary of Recent Progress Grant Number:** 1622765

*"STTR Phase I:  Real-time Automatic Analysis of*

*Electroencephalograms in an Intensive Care Environment*

*Using Deep Learning."*

***BioSignal Analytics, Inc***

**PIs:**  Meysam Golmohammadi, Joseph Picone

**Date:** 12/01/2016

**Progress Specific to Step 1 – Data Preparation:**

Using the TUH EEG Corpus, the world’s largest open-source clinical EEG corpus, we implemented a semi-automated strategy to label seizures by: (1) EEG reports were parsed using natural language processing techniques to locate sessions most likely to contain seizures. (2) Two seizure detection tools (Persyst and AutoEEG) were used to identify sessions with seizures. (3) Sessions, where both tools agreed with high confidence, were studied and divided into comprehensive training and evaluation subsets. (4) These subsets were manually annotated by a group of experts based on ACNS guideline.

In this process, in the first step, we provided a training and evaluation dataset that are partially annotated by our experts at Temple University. This gave us an opportunity to develop experiments for seizure detection task. Then in the second step, a golden evaluation dataset, related to 50 patients with seizures, is collected and then annotated by a group of experts at Temple University. The statistics of the evaluation dataset is presented in Table 1. Currently we are in the process of collecting and annotating more data for training dataset.

Additionally, to deal with the problem of inter-rater agreement, it is necessary that the evaluation dataset get annotated by neurologists out of Temple University. We have now identified about 20 neurologists who could potentially can annotate for us; some of them are still in negotiation for further details before the final agreement. The plan is to make sure for each and every file, we allocate three neurologists for marking EEG files to deal with inter-rater reliability analysis. Therefore in the first stage, the golden evaluation dataset will be divided into roughly 7 portions, of each portion comprises 7 to 8 patients. Each portion's data is given to 3 different neurologists for annotation. If more neurologists can be identified later on, the splitting ways can be further increased to reduce the burden per person.

Table 1. An overview of the seizure detection dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **Description** | **Files** | **Sessions** | **Patients** |
| TUH\_EEG | 38,408 | 16,982 | 10,868 |
| Evaluation Dataset | 1052 | 233 | 50 |

**Progress Specific to Step 2 – Baseline Technology Evaluation:**

Using partially annotated dataset from TUH-EEG, we adjusted and transformed our event detection system into a seizure detection system. The current system is a hybrid HMM/deep learning system in which the channel independent model is implemented using HMM technology and a stacked denoising autoencoder (SdA) deep learning is used to postprocess the output of channel-independent event detector. The results of a real-time seizure detection system based on this technology was amazing which delivered 92% detection accuracy with 9% false alarms using NEDC evaluation methodology.

During our research, we found that unfortunately neurology community is suffering from the lack of a standard and comprehensive evaluation methodology for seizure detection task. For example, in MIT, the classification was evaluated based on seizure and non-seizure records and files. In UPenn, the evaluation was done by the false negative and false positive rates, as well as the latency of each seizure. However, in speech recognition community, National Institute of Standards and Technology (NIST), created an evaluation infrastructure called as Spoken Term Detection (STD). Based on NIST STD tools, we are developing a standard evaluation methodology for EEG classification tasks including seizure detection.

In another effort, using both of commercial seizure detection tool (Persyst) and our hybrid HMM/deep learning system, we ran experiments to find seizures in the evaluation dataset with 30 patients, to compare two systems using NIST STD evaluation methodology. The results are presented in Table 2.

In summery all the parts of step 2 are implemented successfully including statistical language mode and active learning. Since we are collecting and annotating more and more data, we need to optimize these systems and experiments using new evaluation and training dataset in future.

**Progress Specific to Step 3 – Algorithm Enhancement:**

Regarding the algorithm enhancement, currently we are trying to replace the hybrid HMM/SdA system with a new deep learning network, called as long short-term memory (LSTM) networks. LSTM networks are used because they are capable of learning long-term dependencies, which is crucial for smoothing. LSTMs have outperformed alternatives such as standard Recurrent Neural Networks (RNNs) and Hidden Markov Models (HMMs) in problems of sequential nature, such as speech recognition. Our latest experiments showed us that, for EEG event classification, a one layer LSTM performed better than a three layer SdA network.

Table 2. A comparison between Persyst and AutoEEG

using evaluation dataset for 30 patients based on NIST STD metric.

|  |  |  |
| --- | --- | --- |
| **Description** | **Persyst** | **AutoEEG** |
| Sensitivity | 47.0% | 38.0% |
| Specificity |  80.6% | 78.7% |
| ATWV | +0.0732 | -0.693 |
| Seizure FAs/24 hrs | 44 | 108 |
| Total FAs/24 hrs | 97 | 220 |

**Progress Specific to Step 4 – Assessment:**

**Progress Specific to Fundraising:**

**Progress Specific to Marketing:**

**Progress Specific to finding collaborators for Phase 2:**