# Preview of Award 1827565 - Annual Project Report



## Cover

| | |
|---|---|
| Federal Agency and Organization Element to Which Report is Submitted: | 4900 |
| Federal Grant or Other Identifying Number Assigned by Agency: | 1827565 |
| Project Title: | PFI-TT: Software for Automated Real-time Electroencephalogram Seizure Detection in Intensive Care Units |
| PD/PI Name: | Iyad Obeid, Principal Investigator<br>Joseph Picone, Co-Principal Investigator |
| Recipient Organization: | Temple University |
| Project/Grant Period: | 08/01/2018 - 01/31/2020 |
| Reporting Period: | 08/01/2018 - 07/31/2019 |
| Submitting Official (if other than PD\PI): | N/A |
| Submission Date: | N/A |
| Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions) | N/A |

---

## Accomplishments

*** What are the major goals of the project?**

There are three main goals of this project: **(1) technology hardening:** create a real-time system that is capable of being deployed into clinical environments and support clinical testing, **(2) technology enhancement:** close the gap on performance between our state of the art system and clinically acceptable performance, and **(3) technology evaluation:** evaluate the system on a previously unseen data set to demonstrate that performance translates to a broad range of clinical operating conditions. The expected impact of these three goals is that the research system developed previously under SBIR funding should be transformed into a commercially viable product.

Regarding the first goal, significant challenges include transitioning to a streaming interface with low latency. Research systems tend to be implemented as non-real-time systems that have the advantage of being able to look deep into the past and far into the future to make intelligent decisions. Such systems, including our research prototype, essentially have infinite latency. Hence, our objective is to match our state-of-the-art performance with a system that has a small amount of latency (on the order of seconds) so that it can be clinically viable. The two modules that will need the most work to accomplish this are feature extraction and postprocessing.

Note that not all applications require this. Many Internet-based applications use client-server architectures where the signal is collected in its entirety on a local device before it is processed. As long as back-end processing occurs relatively quickly, users are fine with this type of implementation. However, EEGs can last as long as 72 hours. Neurologists need results immediately after a significant event, such as a seizure, has occurred. Hence, latency is a major issue in this application.

Regarding the second goal, improving sensitivity while maintaining a very low false alarm rate is very important. False alarms are caused primarily by two type of events: artifacts and slowing. A major challenge here will be to implement artifact reduction as a preprocessor. Another challenge will be to use long-term context to disambiguate slowing from a seizure event.

Regarding the third goal, we planned to do two things: (1) develop a blind evaluation set based on TUH EEG data, and (2) integrate an independent evaluation set based on the Duke University Corpus. We will use these corpora for evaluation purposes only, and not optimize the system based on the results on these corpora. We will demonstrate that performance on TUH EEG translates to previously unseen data.

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities: We have focused on three major activities in the first year of this grant:

**(1) Feature Extraction:** We have taken two approaches to the problem of reducing latency in the feature extraction component of the system. The first approach was to replicate our current feature extraction process with a lean and mean implementation amenable to commercialization. Our standard feature extraction process uses linear frequency cepstral coefficients (LFCCs) and a differential energy term. The LFCCs use a center-aligned window, which adds a small bit of latency to the system (we use a 50% overlap, so this adds 25% of the frame duration in latency, or about 25 ms). The differential energy term uses a 9-frame window to approximate the derivative of the energy. This feature has been shown to be useful for seizure detection. We augment these features with first and second derivatives, following an approach that has been very successful in other applications such as speech recognition. However, since the duration of a frame is 0.1 secs, and we use center-aligned windows that look forward and backward in the data, the feature extraction process adds about one second of delay when you add up the delays introduced by the various elements of the feature extraction process.

We are in the final stages of implementing what is known as a right-aligned window, which means the delay for feature extraction will be zero. Experiments are showing that performance of this system is comparable to the research system. We have completed a full cross-validation suite of experiments yet, but we believe our peliminary results will hold up.

We have also implemented a delayed version of our standard feature extraction system that uses 1 sec of delay. This system supports a streaming interface necessary for real-time operation but otherwise produces the same results as the research system.

Our second approach has been to explore replacing our current feature extraction with an equivalent deep learning system that operates directly on samples of the signal. Our ultimate goal with this approach is to replace feature extraction with an autoencoder. However, thus far we have simply explored adding additional levels of our best Long Short-Term Memory (LSTM) system, and having this system process the same signal samples that are seen by the feature extraction algorithm. Results have been mixed thus far as expected. We are observing a slight degradation in performance when comparing two systems that effectively use only absolute features.

We have previously explored using samples directly in our research and found that we do not get a measurable improvement in performance. This is counter to what has been observed in applications such as speech and image processing, where they have access to large amounts of data (supervised and unsupervised). We believe our mediocre results are related to the amount of training data available, which is why we continue to develop annotated seizure data. This is also why implementation of an autoencoder strategy is so important – it will allow us to increase the amount of data used to train the system because autoencoders can be trained in an unsupervised fashion. Therefore, this aspect of our work is still in its infancy.

**(2) Artifact Reduction:** Artifacts due to patient movement (and other patient-related behaviors) are known to be a significant, but challenging problem. Most commercial systems perform some form of artifact detection. Artifacts account for about one-third of our errors with our current system.

We have surveyed the literature and explored a number of popular artifact detection systems. Initial implementations of several of these produced no significant improvement in performance. We have looked closely at a family of approaches based on Independent Component Analysis (ICA). The most promising for our application appears to be a technique based on Canonical Correlation Analysis (CCA). Usually, the source signals/variables are separated into two sets of (correlated) variables and CCA is applied between two groups to calculate their correlation weights within and in-between groups. To effectively filter out the signal and avoid outliers, we need to apply this algorithm multiple times for different delay ranges. It is very hard to estimate signals from sources which look like white Gaussian noise, so performance on some portions of the EEG signal is not good. We are trying to incorporate adjacent channel information to work around these issues. There are few things which needs to be automated: (1) selecting target frequencies/lag ranges from the history, (2) determining how many times is the recursion can run before it starts to distort the signal, (3) how to make decisions on preprocessing/filtering the outliers which are single/burst sample spikes, and (4) how to incorporate information from adjacent channels to make a decision on whether the signal estimation is correct. However, our current configuration appears to be promising, and we are beginning to conduct detailed experiments using this method as a preprocessor to feature extraction.

Currently we are also developing deep learning models that detect artifacts. The plan is to postprocess these detected segments using the above artifact reduction algorithm. We have explored a variety of multilayer perceptrons (MLPs) and LSTM networks. A strong consensus on the best architecture has not emerged yet. We are also exploring the use of digital filters to smooth the data and improve the performance of the reduction algorithms. We are having some success filtering muscle artifacts and electrode pop – two very common artifacts that cause our detection systems to fail.

We have also explored the impact of adding additional information to the signal. One of our most vexing problems is that a significant portion of our errors are due to long seizure events – defined as seizures lasting 60 secs or more. These signals display a quasi-periodic structure. We are experimenting with the use of an autocorrelation function as input to the deep learning system to enhance its information about periodicity of the signal. Experiments thus far are inconclusive.

A more interesting attempt that we have implemented involved integrating an old speech coding technique known as Multipulse Linear Predictive Coding (MPLPC). In a nutshell, MPLPC reduces the signal to a combination of a linear filter and a series of orthogonal excitation pulses. These pulses can often clearly depict the periodic nature of the signal. We are experimenting with adding the MPLPC excitation signal to our feature vector as input to our deep learning system. We are in the early stages of experimenting with the best way to do this and the associated impact of this approach. Our initial experiments are promising but inconclusive.

**(3) Evaluation:** We have completed a series of experiments using the Duke University Seizure Detection Corpus (DUSZ) as an evaluation set. This data was collected with completely different EEG hardware and software, as well as being collected at a different hospital. We have shown that performance on this data is comparable to what we are seeing on the TUH EEG development test set. This is extensively documented in one of the publications cited at the end of this report.

We have also introduced a blind evaluation set for TUH EEG as part of a collaboration with IBM Research. IBM approached us about doing an open source Kaggle-style data challenge. To support this effort, we created a blind evaluation set that we are withholding so that it can be used to evaluate technology. The good news is that performance on all three of these evaluation sets (TUH EEG dev test, TUH EEG eval, and DUSZ) are comparable. This is a strong indication that our models are in fact portable and robust.

**Specific Objectives:**

State of the art performance on the seizure detection task we are focused on is a sensitivity of about 40% and a false alarm rate of 7 per 24 hours. Our best system delivers this performance using a combination of LSTMs and heuristic postprocessing. Its latency is considerable because it was not implemented to be a real-time system.

Clinical acceptance requires a sensitivity of 75% at a false alarm (FA) rate of 1 per 24 hours. Sensitivty and false alarm rate are intimiately related. Achieving 1 FA/24 hrs. is an extremely challenging goal. Therefore, our two primary objectives are: (1) to achieve a clinically acceptable level of performance, and (2) decrease latency to less than 10 seconds. Neurologists would like the system to alert them in real time when patients seize. This requires a latency that is on the order of seconds, not hours. A latency less than 10 seconds is more than acceptable in this case because most of the time the neurologists will have to be contacted, which adds considerable delay to the process.

Neurologists would also like to be able dose patients and see results in real-time. This requires a much smaller latency – on the order of seconds. This is why we have explored zero latency approaches. We have made good progress at reducing the latency of our system, but accurate and powerful postprocessing will require some amount of latency.

**Significant Results:**

We have demonstrated that we can do extraction of absolute features with minimal loss in performance. Differentiation of features, however, will require several frames of delay. We have made good progress towards our goal of less than 1 second of delay for feature extraction.

We have also demonstrated we can integrate feature extraction into the deep learning system with minimal degradation in performance. This is work in progress, but preliminary experiments are encouraging.

We have also shown that our models are robust and transfer across domains. This was a major concern since most of our EEG data comes from Temple University Hospital. However, portability of these models appears good since clinical EEGs are normally collected under controlled conditions.

Preliminary results on an ICA-based artifact reduction approach appear promising, but need a much more comprehensive evaluation.

**Key outcomes or Other achievements:**

We were approached by IBM a couple of years ago to collaborate on a Kaggle-style open source challenge. We agreed to participate under the condition that this be a community-wide open source activity. They agreed and were planning to host this using IBM cloud computing resources. Unfortunately, it has taken them much longer than expected to release this challenge to the public.

However, they did complete the internal challenge recently. Approximately 100 teams within IBM participated. The best performing systems, as we predicted, suffered from high false alarm rates and did not approach the level of performance that we have demonstrated (and shared with them). Therefore, we feel some amount of validation that our industry-leading technology is representative of what state of the art deep learning systems can deliver. The IBM internal challenge will be discussed in an upcoming publication that the IBM team is developing under our guidance.

*** What opportunities for training and professional development has the project provided?**

The students involved in the project have presented their work at an IEEE-sponsored technical conference – the IEEE Signal Processing in Medicine and Biology Symposium. They have also participated in the development of several publications. We have several journals papers under review as well. We spent a significant amount of time training our students how to publish and present their work. Journal papers typically go through 10 to 20 revisions while conference papers typically go through 5 to 10 revisions. The PIs are very committed to mentoring students with respect to technical communications.

In terms of technical development, the students have developed a combination of important practical skills including: (1) design and implementation of machine learning experimental paradigms using big data, (2) development of software using a strict software engineering paradigm that includes revision control, documentation standards, programming guidelines, regression testing and standard interface specifications, and (3) implementation of real-time streaming interfaces. The students have also dramatically improved their C++ and Python programming skills.

*** How have the results been disseminated to communities of interest?**

We have been very active in disseminating data (see *https://www.isip.piconepress.com/projects/tuh_eeg/*). We recently augmented our web site to allow anonymous rsyncing of our data resources. Previously users had to either send us a disk drive via snail mail or use a web-based tool like wget. Rsync makes it much easier for users to download our data incrementally over slow-speed links. We have placed all of our corpora, including the TUH EEG training and development test sets, on this server so they are openly available. We currently have over 2,000 subscribers of these resources.

We also actively support these resources. We average approximately three customer interactions per week in which we provide detailed problem-solving for users. Several commercial and non-profit entities have also made use of these resources and solicited our support in exploiting our resources.

We are also in the process of publishing or have published several important papers, including a comprehensive book chapter and journal paper that describe our technology in great detail.

*** What do you plan to do during the next reporting period to accomplish the goals?**

Our focus on technology development for the remainder of the project includes: (1) evaluation of new artifact detection technology that we have developed, (2) optimization of the real-time version of the system to decrease latency while maintaining performance, and (3) integration of additional information sources to improve our accuracy at the detection of long-term seizure events.

We continue to monitor the deep learning literature and adapt emerging algorithms to our application. Increasing the robustness of the training process so that results are reproducible remains an important goal. Making the system less sensitive to changes in the software base (e.g., upgrades to packages such as TensorFlow) is also important.

We also need to complete integration of the real-time system into our demonstration environment so that we can showcase the technology in the best possible light. This involves integrating a streaming interface so that we can visualize the detection of seizure events in real-time.

Finally, we plan to re-engage potential customers for the technology once we can effectively demonstrate the real-time system.

---

## Products

### Books

### Book Chapters

Golmohammadi, M. Shah, Vinit Obeid, Iyad Picone, Joseph (2019). Deep Learning Approaches for Automatic Analysis of EEGs. *Deep Learning: Algorithms and Applications 1st.* Obeid, I. Picone, J.. Springer-Verlag. New York, New York, USA. .

Status = ACCEPTED; Acknowledgement of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: TBD.

**Inventions**

**Journals or Juried Conference Papers**
**View all journal publications currently available in the NSF Public Access Repository for this award.**

The results in the NSF Public Access Repository will include a comprehensive listing of all journal publications recorded to date that are associated with this award.

**Licenses**

**Other Conference Presentations / Papers**
Shah, Vinit Anstotz, Ryan Obeid, Iyad Picone, Joseph (2018). *Adapting an Automatic Speech Recognition System to Event Classification of Electroencephalograms*. Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB). Philadelphia, Pennsylvania, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

**Other Products**
*Databases.*

(20181206) NEDC TUH EEG Artifact Corpus (v1.0.0): This is our first release of the TUH EEG Artifact Corpus. This corpus was developed to aid in EEG event classification such as seizure detection algorithms. This corpus is a subset of the TUH EEG Corpus and contains files with 5 different types of artifacts: (1) eye movements (EYEM), (2) chewing (CHEW), (3) shivering (SHIV), (4) electrode pop, electrode static, and lead artifacts (ELPP), and (5) muscle artifacts (MUSC).

https://www.isip.piconepress.com/projects/tuh_eeg/downloads/tuh_eeg_artifact/v1.0.0/

*Databases.*

NEDC TUH EEG v1.2.0: an open source EEG Corpus

https://www.isip.piconepress.com/projects/tuh_eeg/downloads/tuh_eeg/v1.2.0/

*Databases.*

The TUH EEG Seizure Detection Corpus (v1.5.0 and v1.6.0):

https://www.isip.piconepress.com/projects/tuh_eeg/downloads/tuh_eeg_seizure/v1.5.0/

https://www.isip.piconepress.com/projects/tuh_eeg/downloads/tuh_eeg_seizure/v1.6.0/

This is our open source manually annotated seizure detection corpus that can be used to train and evaluate seizure detection systems.

*Software or Netware.*

A demonstration program that illustrates how to correctly access EEG data from an EDF file using montages and channel labels:

(20190716) NEDC Python Streaming Software (v1.0.0): The first release of a self-contained Python script tha demonstrates how to correctly read EDF files.

https://www.isip.piconepress.com/projects/tuh_eeg/downloads/nedc_pystream/

*Software or Netware.*

Standardized Scoring Software

https://www.isip.piconepress.com/projects/tuh_eeg/downloads/nedc_eval_eeg/

**Other Publications**

Ferrell, Sean Jakielaszek, Luke Elseify, Tarek Picone, Joseph (2019). *The Temple University Hospital EEG Corpus: Annotation File Formats*. A technical report that describes how to access our annotation data.. Status = OTHER; Acknowledgement of Federal Support = Yes

Ferrell, Sean Mathew, Vineetha Ahsan, Tameem Picone, Joseph (2019). *The Temple University Hospital EEG Corpus: Electrode Location and Channel Labels*. A technical report that describes how to access our EEG data. Available at https://www.isip.piconepress.com/publications/reports/2019/tuh_eeg/electrodes/.. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

**Patents**

**Technologies or Techniques**

**Thesis/Dissertations**

**Websites**

## Participants/Organizations

**What individuals have worked on the project?**

| Name | Most Senior Project Role | Nearest Person Month Worked |
|---|---|---|
| Obeid, Iyad | PD/PI | 1 |
| Picone, Joseph | Co PD/PI | 1 |
| Shah, Vinit | Graduate Student (research assistant) | 3 |
| Shawki, Nabila | Graduate Student (research assistant) | 3 |
| Cap, Thao | Undergraduate Student | 1 |

**Full details of individuals who have worked on the project:**

**Iyad Obeid**
**Email:** iobeid@temple.edu
**Most Senior Project Role:** PD/PI
**Nearest Person Month Worked:** 1

**Contribution to the Project:** Provides software engineering and bioengineering expertise.

**Funding Support:** No other support.

**International Collaboration:** No
**International Travel:** No

---

**Joseph Picone**
**Email:** joseph.picone@gmail.com
**Most Senior Project Role:** Co PD/PI
**Nearest Person Month Worked:** 1

**Contribution to the Project:** Provides machine learning and signal processing expertise.

**Funding Support:** No other funding.

**International Collaboration:** No
**International Travel:** No

**Vinit Shah**
**Email:** tug14467@temple.edu
**Most Senior Project Role:** Graduate Student (research assistant)
**Nearest Person Month Worked:** 3

**Contribution to the Project:** Lead algorithm designer. Artifact reduction. Deep learning architectures.

**Funding Support:** No additional funding. Transferred from another externally funded project for Summer 2019.

**International Collaboration:** No
**International Travel:** No

---

**Nabila Shawki**
**Email:** tuk02200@temple.edu
**Most Senior Project Role:** Graduate Student (research assistant)
**Nearest Person Month Worked:** 3

**Contribution to the Project:** Real-time implementation. Feature extraction development using augmented information.

**Funding Support:** Worked 9 months as a department teaching assistant prior to Summer 2019.

**International Collaboration:** No
**International Travel:** No

---

**Thao Cap**
**Email:** tuj64267@temple.edu
**Most Senior Project Role:** Undergraduate Student
**Nearest Person Month Worked:** 1

**Contribution to the Project:** Software engineer. Supports the development of some of software and infrastructure.

**Funding Support:** Nothing other than this grant.

**International Collaboration:** No
**International Travel:** No

---

**What other organizations have been involved as partners?**
Nothing to report.

**What other collaborators or contacts have been involved?**

As mentioned, IBM Research has been tangentially involved in this project:

Subhrajit Roy <subhrajit.roy@au1.ibm.com>, IBM Research (Australia)

Stefan Harrer <sharrer@au1.ibm.com>, IBM Research (Australia)

---

## Impacts

**What is the impact on the development of the principal discipline(s) of the project?**

The software tools, particularly the scoring software) and data we are providing are an attempt to bring discipline to the evaluation process in this field. It is important that commercial systems evaluate in a common framework so that scientifically valid comparisons can be made. We are laying this groundwork so that our commercial techology can be fairly evaluated by potential investors, and so that state of the art performance is well understood. We are partnering with several organizations to promote this concept across the international community. Thus far, one windfall from this is that we are very certain our technology is competitive with the very best commercially available systems.

**What is the impact on other disciplines?**

Seizure detection is a difficult problem. It is exposing some of the weaknesses/challenges in deep learning technology. Hopefully this will motivate researchers to tackle more fundamental problems such as reproducibility, data imbalance, and low false alarm rates.

**What is the impact on the development of human resources?**

We believe our software provides an excellent model for how software should be designed and implemented. For example, our recently released Python software to access EDF files is an excellent tutorial on how to write flexible data-driven software. This was developed to help a number of our constituents who were having trouble accessing data.

**What is the impact on physical resources that form infrastructure?**

Our state of the start small computing cluster, Neuronix, continues to be an excellent model for how researchers can build cost-effective clusters that are capable of providing state of the art price/performance ratios for big data applications. It is very competitive with Amazon AWS because it supports vast amounts of data storage inexpensively.

**What is the impact on institutional resources that form infrastructure?**

The data developed in this project is actually being used as test data for a graduate-level machine learning class (https://www.isip.piconepress.com/courses/temple/ece_8527/). We also use a number of our experimental results to demonstrate key concepts and challenges in machine learning.

**What is the impact on information resources that form infrastructure?**
Nothing to report.

**What is the impact on technology transfer?**

Of course, the entire scope of this project is focused on commercialization of our EEG technology. The technology being developed flows directly to our startup company: Biosignal Analytics, Inc. The technology will be demonstrated to leading technology companies in the EEG market space.

**What is the impact on society beyond science and technology?**
Nothing to report.

---

## Changes/Problems

**Changes in approach and reason for change**
Nothing to report.

**Actual or Anticipated problems or delays and actions or plans to resolve them**

It is important to note that we had two senior PhD graduate students and one MS-level graduate student lined up to work on this project when it was proposed. However, prior to the initiation of the project, all departed unexpectedly for lucrative jobs in industry. This caused delays because we had to re-staff the project. We are now on course, however, and have our two most senior PhD students working on the project. As a result, we will be requested a no-cost extension at the end of the current period of performance.

**Changes that have a significant impact on expenditures**

As described above, the unexpected changes in staffing resulted in slower than expected expenditures. Hence, we hope we can extend the period of performance of the project.

**Significant changes in use or care of human subjects**
Nothing to report.

**Significant changes in use or care of vertebrate animals**
Nothing to report.

**Significant changes in use or care of biohazards**
Nothing to report.

## Special Requirements

**Responses to any special reporting requirements specified in the award terms and conditions, as well as any award specific reporting requirements.**
Nothing to report.