**Major Goals of the Project:**

There were three main goals of this project: (1) technology enhancement: close the gap on performance between our state of the art system and clinically acceptable performance, (2) technology hardening: create a real-time system that is capable of being deployed into clinical environments and support clinical testing, (2) 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.

**Accomplishments:**

Major Activities:

We spent the majority of time on this project pursuing goals no. 1 and no. 2 above. Goal no. 1 was particularly critical an overarching goal of the project was to reduce the gap between performance expectations of customers and our best technology. We allocated a PhD student to this task and his work formed the basis of his dissertation.

Goal no. 2 was very important from a commercialization standpoint. Most published research systems are not suitable for clinical applications because they perform multiple passes over the data and have extreme amounts of latency. Performance tends to suffer as latency is reduced. A major differentiation of our work is that we have produced a high performance system that operates with minimal latency and requires modest computing resources to operate in real-time.

The final goal, technology evaluation leading to commercialization, was difficult to achieve due to a variety of reasons. We did conduct the industry’s first widescale technology evaluation to calibrate the performance of the best research systems. That effort is described below. However, our efforts to commercialize the technology through are start up stalled because we were not able to continue funding for our start up company (our Phase II SBIR was not funded). Our efforts to market the technology to leading technology vendors in the field were also not successful. Therefore, at the conclusion of the project, we have made the technology open source so that others can build on our work.

Specific Objectives:

With respect to goal no. 1, we have for a long time established three performance goals for the system: (1) 75% sensitivity, (2) less than 1 false alarm (FA) per 24 hours, and (3) less than 2 seconds of latency. This level of performance was established as a result of discussions with over 100 potential customers over the past 5 years (as part of an NSF I-CORPS project).

Sensitivity degrades very quickly as the FA rate approaches 1/24. hr. This is an extreme operating point not addressed by most machine learning research. Many algorithms are incapable of operating at such a low FA rate. For example, deep learning systems are notorious for the high-level of confidence produced even for incorrect hypotheses (we often joke that they are very sure of themselves). It is difficult to dial down their performance to this very low FA range.

Low latency is critical for clinical deployment of this technology. Neurologists want to be able to medicate patients and see the results instantly. High performing systems that involve multiple passes over the data, such as those entered in the Neureka Challenge described below, are simply not useful in clinical settings because they operate on an entire signal after collection has been completed. The techniques used by these systems to reduce false alarms and remove artifacts are simply not transferable to real-time systems. Therefore, a major objective of our research was to reduce the latency of our best performing system to less than 2 seconds, while operating in real-time on a 1.7 GHz processor.

Finally, with respect to our third goal, we took advantage of a unique opportunity to conduct an industry-wide technology evaluation through the Neureka 2020 Epilepsy Challenge. This was a great opportunity to establish industry-accepted benchmarks for the technology, and to accelerate progress in the field. This challenge is described in detail in:

Shah, V., Obeid, I., Picone, J., Ekladious, G., Iskander, R., & Roy, Y. (2020). Validation of Temporal Scoring Metrics for Automatic Seizure Detection. In I. Obeid, I. Selesnick, & J. Picone (Eds.), *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (pp. 1–5). https://doi.org/10.1109/SPMB50085.2020.9353631.

The web site for the challenge is here: https://neureka-challenge.com/.

We also improved the robustness of the system to variations in recording conditions. Deep learning systems are well-known for lacking robustness to unseen data. We investigated this and demonstrated that our performance results transfer across databases, vendor equipment and recording conditions.

Significant Results:

With respect to the first activity, our best research system, which we refer to as the baseline system for this grant, is fully described in Vinit Shah’s PhD dissertation (Shah, 2021), which was successfully defended in May 2021. This dissertation represents the cumulative work of one of our two students who initiated our work in automatic seizure detection. The central thesis of his work is that separation of the seizure detection problem into a two-phase problem – epileptiform activity detection followed by seizure detection – should improve our ability to detect and localize seizure events. In the first phase, we used a sequential neural network algorithm known as a long short-term memory (LSTM) network to identify channel-specific epileptiform discharges associated with seizures. In the second phase, the feature vector is augmented with posteriors that represent the onset and offset of ictal activities. These augmented features are applied to a multichannel convolutional neural network (CNN) followed by an LSTM network.

The multiphase model was evaluated on a blind evaluation set and was shown to detect segment boundaries within a -second margin of error. Our previous best system, which delivers state-of-the-art performance on this task, correctly detected only 9 segment boundaries. Our multiphase system was also shown to be robust by performing well on two blind evaluation sets. Seizure detection performance on the TU Seizure Detection (TUSZ) Corpus development set is sensitivity with false alarms/ hours (FAs/24 hrs). Performance on the corresponding evaluation set is sensitivity with FAs/ hrs. Performance on a previously unseen corpus, the Duke University Seizure (DUSZ) Corpus is sensitivity with FAs/ hrs. Our previous best system yields sensitivity with FAs/ hrs on the TUSZ development set, sensitivity with 1 FAs/ hrs on the TUSZ evaluation set and sensitivity with FAs/ hrs on DUSZ.

Our second student, Meysam Golmohammadi, who pioneered our work on applying deep learning to seizure detection and was the Chief Technical Officer of our start up, Biosignal Analytics Inc., also successfully defended his dissertation (Golmohammadi, 2021) in June 2021. This dissertation represents a summary of our initial work understanding the problem and addressing it by leveraging our years of experience with other applications such as speech recognition. Automatic analysis of clinical EEGs is a very difficult machine learning problem due to the low fidelity of a scalp EEG signal. Deep learning approaches can be viewed as a broad family of neural network algorithms that use many layers of nonlinear processing units to learn a mapping between inputs and outputs. Deep learning-based systems have generated significant improvements in performance for sequence recognitions tasks for temporal signals such as speech and for image analysis applications that can exploit spatial correlations, and for which large amounts of training data exists. The primary goal of this research was to develop deep learning-based architectures that capture spatial and temporal correlations in an EEG signal. We applied these architectures to the problem of automated seizure detection for adult EEGs. The main contribution of this work is the development of a high-performance automated EEG analysis system based on principles of machine learning and big data that approaches levels of performance required for clinical acceptance of the technology.

With respect to the second goal of developing a real-time system, the process of engineering the system to be amenable to a real-time implementation is discussed in Khalkhali et al. (2021). Scalp electroencephalogram (EEG) signals inherently have a low signal-to-noise ratio due to the way the signal is electrically transduced. Temporal and spatial information must be exploited to achieve accurate detection of seizure events. Most popular approaches to seizure detection using deep learning focus on modeling temporal or spatial information, but do not jointly model this information. We exploit both simultaneously by converting the multichannel signal to a grayscale image and using transfer learning approaches to achieve high performance. The proposed system is trained end-to-end with only very simple pre- and post-processing operations which are computationally lightweight and have low latency, making them conducive to clinical applications that require real-time processing. We demonstrate the efficacy of this approach We have achieved a performance of 42.05% sensitivity with 5.78 false alarm per 24 hours on the development dataset of v1.5.2 of the Temple University Hospital Seizure Detection Corpus. The system can run easily run in real-time using single core CPU, operating at 0.58 xRT on a 1.7 GHz processor in 16 Gbytes of memory with a latency of 300 msec.

With respect to the third major activity, in Golmohammadi’s study, we used the Temple University EEG (TUEG) Corpus, supplemented with data from Duke University and Emory University, to evaluate the performance of these hybrid deep structures. We demonstrated that performance of a system trained only on Temple University Seizure Corpus (TUSZ) data transfers to a blind evaluation set consisting of the Duke University Seizure Corpus (DUSZ) and the Emory University Seizure Corpus (EUSZ). This type of generalization is very important since complex high-dimensional deep learning systems tend to overtrain.

We also investigated the robustness of this system to mismatched conditions (e.g., train on TUSZ, evaluate on EUSZ). We train a model on one of three available datasets and evaluate the trained model on the other two datasets. These datasets are recorded from different hospitals, using a variety of devices and electrodes, under different circumstances and annotated by different neurologists and experts. Therefore, these experiments help us to evaluate the impact of the dataset on our training process and validate our manual annotation process.

Further, we introduce methods to improve generalization and robustness. We analyze performance to gain additional insight into what aspects of the signal are being modeled adequately and where the models fail. The best results for automatic seizure detection achieved in this study are with FA per hours on TUSZ, with FAs on DUSZ, and with 11.26 FAs on EUSZ. We demonstrate that the performance of the deep recurrent convolutional structure presented in this study is statistically comparable to the human performance on the same dataset.

Finally, in Golmohammadi’s work, we also developed some effective visualization tools to understand exactly what the network is learning. We believe these tools will be relevant to a larger class of deep learning systems.

**References**

Golmohammadi, M. (2021). *Deep Architectures for Spatio-Temporal Sequence Recognition With Applications in Automatic Seizure Detection* [Temple University]. https://doi.org/https://www.isip.piconepress.com/publications/phd\_dissertations/2021/seizure\_detection/

Khalkhali, V., Shawki, N., Shah, V., Golmohammadi, M., Obeid, I., & Picone, J. (2021). Low Latency Real-Time Seizure Detection Using Transfer Deep Learning. In I. Obeid, I. Selesnick, & J. Picone (Eds.), *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (pp. 1–7). IEEE. https://doi.org/10.1109/SPMB52430.2021.9672285

Shah, V. (2021). *Improved Segmentation for Automated Seizure Detection Using Channel-Dependent Posteriors* [Temple University]. *https://www.isip.piconepress.com/publications/phd\_dissertations/2021/seizure\_segmentation/*

Key Outcomes:

In Spring 2020, we had a unique opportunity to coordinate an industry-wide challenge on seizure detection. We collaborated with several partners and designed the Neureka 2020 Epilepsy Challenge. Initially, 19 sites participated. We provided the data sets and evaluation methodology. The results, as mentioned before, are available from this web site: *https://neureka-challenge.com/*. It established an industry accepted benchmark for the performance of the technology. However, the best performing systems were not easily translated to real-time technology and involved complex, multi-pass systems. The previously mentioned real-time system we have developed compares favorably with the best systems but is much more amenable to real-time implementation (which we provide). The Neureka Challenge was a unique opportunity to baseline our technology, presumably making it more attractive for commercialization.

We have continued to support participants in this evaluation. We routinely run evaluations for researchers on the blind evaluation data set to support their publications as a community service. We are also going to release the blind evaluation data as part of TUSZ v1.5.3 (expected in Spring 2022).

Additional information:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| [(Download)](https://www.research.gov/rppr-web/downloadAttachment?docID=1200898) | roc.pdf | Receiver Operating Characteristics (ROC) comparing the three significant EEG seizure detection systems we have developed. | Joseph Picone | 09/30/2021 |
| [(Download)](https://www.research.gov/rppr-web/downloadAttachment?docID=1200902) | table.pdf | A comparison of our EEG seizure detection system to several competitive systems that participated in our open source challenge. Our system is the only one of these that can be implemented in real-time and used in clinical applications. |  |  |

to collect feedback on how acceptable the current level of performance would be toclinicians.Significant Results:

Regarding EEG performance, we established our best result to date on the TU SeizureDetection Corpus (TUSZ) - 41.60% sensitivity with 5.63 FAs/24 hrs. The real-time versionof this system gives performance of 45.59% sensitivity with 12.24 FAs/24 hrs. This iscompetitive with the best research systems presented at the open source NeurekaEpilepsy Challenge and is the only one of these systems that operate in real-time with lowlatency.We have also released v1.5.2 of the TUSZ Corpus and will release v1.5.3 by the end ofthe 2021 calendar year (we are running behind schedule on this due to some unexpectedstaffing issues). v1.5.3 will include a new eval, dev, and training set definition. Allannotations in v1.5.2 have been carefully reviewed by a new team of annotators.

Regarding generalization and robustness, a system that achieves 45.59% sensitivity with12.24 FAs/24 hrs on the TUSZ dev set achieves 45.91% sensitivity with 11.86 FAs/24 hrson the Duke University Seizure Corpus (DUSZ), and 62.56% sensitivity with 11.26 FAs/24hrs on the Emory University SEizure Corpus (EUSZ). The performance of our deeprecurrent convolutional structure is statistically comparable to the human performance onthe same dataset. The system was only trained on TUSZ and did not need to be retrainedto deliver good performance on the other datasets.Key outcomes or Otherachievements:

We have released v1.5.2 of the TUSZ Corpus and will release v1.5.3 by the end of thecalendar year. v1.5.3 contains a significant increase in the amount of training data. Wealso completed a review of all of the /dev and /eval data to make sure the annotationswere accurate.