**Response to the Reviewers**

**Paper No. P035:** Low Latency Real-Time Seizure Detection Using Transfer Deep Learning

The authors would like to convey sincerest appreciation to the organizers, technical members, and reviewers for their comments on the paper followed by its acceptance.

**Summary**:

1. The paper presents a deep learning approach to fast seizure detection using noninvasive EEG signal features, with comparable but improved performance over the authors’ previous best, and improved ROC curve results over the existing state of the art. The existing literature is surveyed well, and the approach and results are clearly communicated. One area for consideration raised by the reviewers is the balance between sensitivity and specificity, especially in a clinical setting: for example, how does a sensitivity of 42% translate to the number of missed events per 24 hours? Some typographical and grammatical errors should be remedied, for example: “We demonstrate the efficacy of this approach We have achieved…” in the abstract.

*Response: There is a relationship between the sensitivity and false positive for every system usually shown in the Receiver Operating Characteristic (ROC) plot. Lower sensitivity has a lower false positive rate and higher sensitivity translates to higher false positivse.*

*The sensitivity percentage is defined by the evaluation metrics. For example, in the overlap metric, detecting an event, independent of its duration, is enough. But for other types of evaluation metrics, the duration may have an impact. In this paper, the validation metrics have been discussed in detail:*

*V. Shah, I. Obeid, J. Picone, G. Ekladious, R. Iskander, and Y. Roy, “Validation of Temporal Scoring Metrics for Automatic Seizure Detection,” in Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2020, pp. 1–5.*

**Reviewer #1:**

1. While the authors defined that their solution is low-latency and close to real-time, there is no concrete reason given for why their solution is needed. It seems like the other solutions discussed in the paper provide similar or sometimes better results. Therefore, the strength of the proposed solution lies with the fact that it can provide results quickly even when only lightweight processing is available. I think the paper would benefit greatly if there were quantifiable comparisons of how much processing power is saved and how fast the seizure detection is compared to the other schemes.

As mentioned above, the paper lacks the clear answer to “Why was this needed?”. Although the authors seem to potentially answer this inherently by presenting their solution.

*Response: All the previous methods that competed in NeurekaTM have infinite processing time because of the heavy pre/post-processing that they had. On the contrary, the proposed system can work in close to real-time with better or close to the current best results. All the competitors’ systems are non-causal, and a timing comparison is not meaningful.*

1. In section II, the authors mention “Neurologists are capable of manually interpreting EEGs with accuracies…”. Have they consulted with a neurologist to see how a human reads and detects the seizure? If so, is it possible to detect a seizure before it starts? Visually looking at the grayscale images that the paper presents, background event vs seizure event is like night and day. Perhaps a discussion could be added to talk about the transition period. Although, it is possible this discussion is a bigger item that is reserved for a future study.

In Section IV, the authors introduce the 3 parameters they use for post processing. While the results are presented in terms of Detection Delay (which consists of BDmin and SDmin), there is no mention of what the ratio between BDmin and SDmin are. Are they equal?

*Response: Due to space limitations, we could not go into more details about many aspects of the system. We plan to submit an expanded version of this paper as a book chapter. BDmin and SDmin are not always equal and they change from 20 to 160 seconds in steps of 10 seconds.*

1. Throughout the paper and specifically in Section V, the results are presented in terms of how low their false alarm rates are. How does this compare to missed detection rate? In a critical-care environment, I would expect to not miss any events that should have been caught even though I might receive an increased number of false alarms. Could you elaborate on the connection between false alarm vs. missed detection rates with your algorithm?

*Response: This topic has been explored in more detail in our other publications. A simple use case we have discussed is the following: in a 12-bed unit, 1 FA per 24 hours per patient means 12 FAs per day, or an average of an FA every 2 hours. This would place an unacceptably high workload on staff in addition to all the other duties they perform, and would seriously cut into a neurologist’s time they can spend with other patients. As a result, clinicians turn off all these automated tools because the FA rates are much higher than 1 FA/24 hours.*

1. Figure 1 might have originally been a powerpoint slide and it seems to have writing but is impossible to read. In fact, most figures in the paper are hard to read as the font on them are too small. Use of color in graphs might also cause issues for readers that use a black-and-white printer.

Table 3 might have a wrong label (mentions Background twice under “Detected”)

Paper seems to have minor grammatical errors and typos. It also seems to be missing keywords in the abstract (if needed)

*Response: We would like to thank the reviewer for pointing out these issues. In our updated paper, we corrected all these problems.*

**Reviewer #2:**

1. This paper presents a method for real-time seizure detection using a CNN and transfer learning. The paper reads well and flows logically. The introduction sufficiently motivates the problem and the presented experiments show good results.

The method of using a weighted loss function to reduce bias is a logical one, but I feel somewhat dangerous in a clinical setting. Does this loss improve both sensitivity and specificity?

*Response: The weighted loss is specially designed to compensate for the unbalanced number of samples in the background and seizure classes. Without these weights, the neural network biases toward the background and ignores the seizure events.*

1. The grayscale images used for classification is stretched from 20x256 to a 256x256 to fit into a ResNet model. I would suggest that the authors try a custom CNN for this problem. Not only do the pretrained weights of ImageNet not transfer well, they also were trained on a set of images with very different structure and statistical properties. I think you will find a simple 3 layer CNN trained from scratch could achieve as good, if not better results than the ResNet 18 pretrained model.

*Response: We have tried several architectures, such as CNN, LSTM, and CNN-LSTM. Because of diverse patterns of seizure and background, small neural networks cannot learn these complex patterns, while the deeper neural networks with a greater number of parameters can learn them. The problem of using larger neural networks with a greater number of parameters is convergence. With a pretrained neural network with preloaded weights, convergence is much faster because the system starts from a reasonable points – presumably closer to an acceptable local optimum.*

1. Minor typo – Table 3, one of the backgrounds should be labeled seizure.

*Response: Thank you for catching this issue. We have corrected the table in the revised paper.*

**Reviewer #3:**

1. That being said, I think they contradict that when they mention the hybrid CNN/LSTM in Evaluation Results since a CNN captures spatial info and LSTM captures temporal info; thus, are they truly the first to focus on jointly modeling both?

*Response: We have actually explored many different permutations of these systems in other work. In our work with CNN/LSTM, we have found a channel-based CNN works best. In such systems, the temporal dimensions of the montage channels are modeled, and they do not include the spatial inter-channel information.*

1. I think it would be helpful for a few citations as this is SPMB and not ICCV or something more with ML; moreover, they mention dropout and batch normalization and actually use batch normalization but never formally define it so it would be helpful to include a citation to some of those regularization methods.

In the same vein, I think it would be helpful for citations to the scaling or normalization techniques such as moving average filters and local scaling.

*Response: Batch normalization was used in the original ResNet18 model which we cited. The moving average technique was not used and local scaling was not taken from any specific references.*

1. I am curious as they randomly selected 1/5 of the background patterns to rebalance the dataset and was unsure if 1/5 was chosen empirically or for some other reason on page 3 in the second paragraph.

*Response: It was selected empirically. We observed that a random selection of 20% background patterns works as well as using all the background. We have clarified this in the manuscript.*

1. I noticed a typo in their abstract "The system can run easily run in real-time using single core CPU"

Also, in the last paragraph of Evaluation Results they write "resnet" and the second paragraph of Evaluation Results they write "Resnet18" and in the methodology Application of Transfer Learning they write "ResNet18". Overall, they are not as focused on consistency in their naming.

*Response: Thank you very much for catching the grammatical issue which we addressed in the revised version. The “resnet” is the name of the proposed system which includes preprocessing, ResNet18, and postprocessing. The core recognition neural network that is embedded into the “resnet” system is ResNet18. We have clarified this in the text.*