Unsupervised Online Learning for Long-Term High Sensitivity Seizure Detection

Adelson Chua[1], Michael I. Jordan[1,2], Rikky Muller[1,3]
[1]Department of Electrical Engineering and Computer Sciences,[2]Department of Statistics
University of California, Berkeley, Berkeley, CA
[3]Chan-Zuckerberg Biohub, San Francisco, CA

***Abstract*—Current seizure detection systems rely on machine learning classifiers that are trained offline and subsequently require manual retraining to maintain high detection accuracy over long periods of time. For a true deploy-and-forget implantable seizure detection system, a low power, at-the-edge, online learning algorithm can be employed to dynamically adapt to the neural signal drifts over time. This work proposes SOUL: Stochastic-gradient-descent-based Online Unsupervised Logistic regression classifier, which provides continuous unsupervised online model updates that was initially trained with labels offline. SOUL was tested on two datasets, the CHB-MIT scalp EEG dataset, and a long (>250 hours) human ECoG dataset from the University of Melbourne. SOUL achieves an average cumulative sensitivity of 97.5% and 97.9% for the two datasets respectively, while maintaining <1.2 false alarms per day. When compared with state-of-the-art, a moderate sensitivity improvement of 1-3% is observed on the majority of subjects and a large sensitivity improvement of >12% is observed on three subjects with <1% impact on specificity.**

I. Introduction

Epilepsy is a serious neurological disorder affecting ~1% of the world's population. A relatively new therapy being used for seizure treatment utilizes closed-loop implantable neuromodulators that detect the onset of seizure events from recorded neural signals and trigger neurostimulation to suppress the seizure. The only medically approved device of this kind is called the NeuroPace RNS, which utilizes a small, battery-powered implantable pulse generator surgically implanted in the skull with two electrode leads that are implanted intracranially and/or epicortically. While considerable efficacy has been reported, reducing 66% of seizures by Year 6 [1], frequent visits to a medical professional are required to tune the detection algorithms to compensate for the changing seizure patterns on electrocorticography (ECoG) or scalp electroencephalograpy (EEG), which can also vary from patient to patient. There has been work on seizure prediction and detection systems trained with long-term datasets to capture such variations [2] which reported seizure detection accuracy greater than 90%. However, these are software-only implementations where computational complexity and memory requirements are not a concern. For implantable closed-loop seizure detection systems, both power consumption and long-term accuracy must be considered.

There are several hardware seizure detection implementations in the literature, most of which utilize Support Vector Machine (SVM) based classifiers [3]–[6]. While these systems usually incorporate actual feature calculations on-chip, training and its associated computational complexity are usually completely offloaded to software. The calculated feature weights after offline training are then loaded into the chip which then performs online seizure classification. For long-term datasets with drifting seizure features however, such systems would require external intervention from medical professionals to periodically retrain the classifier for high-accuracy detection over time due to the challenge of getting reliable labels for the data.

Fig. 1 shows our proposed system, which starts with a model that is trained offline, and incorporates an unsupervised online learning scheme during classification to dynamically adapt to the changing EEG/ECoG baseline and seizure patterns, thereby avoiding any offline retraining. To enable incremental updates, stochastic gradient descent (SGD) was utilized on a logistic regression classifier, and the need for labeled seizures is avoided by using the classifier's own prediction as the label. Computational complexity is minimized due to the inherent mathematical simplicity of applying SGD on logistic regression. This makes online learning feasible for a low power integrated circuit implementation. This manuscript focuses on the algorithm design highlighting the framework that we refer to as SOUL (SGD-based Online Unsupervised Logistic regression classifier).

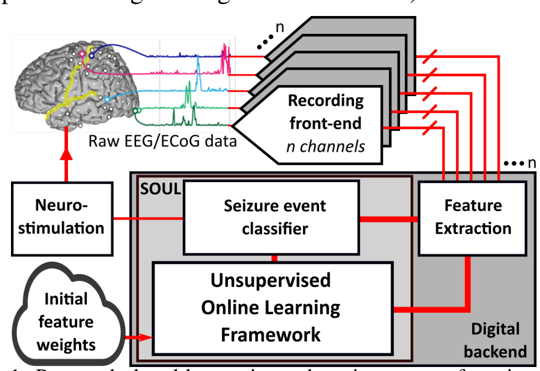

Figure 1: Proposed closed-loop seizure detection system featuring a fully unsupervised online learning framework implemented to dynamically tune the model parameters, initially acquired from offline training, *in situ*.

II. Unsupervised Online Learning Classifier

The SOUL framework utilizes logistic regression for classification, which then updates the feature weights online through SGD. SGD is an iterative technique that optimizes the objective function, the set of feature weights, such that classification errors are minimized. The technique takes an estimated step towards optimality by only using a single sample instead of the entire dataset. This leads to a significantly lower computational complexity making it feasible for incremental learning. The feature weight updates take the following form:

$$w_{t+1} = w_t + \eta(y_t - p(w_t, x_t))x_t \tag{1}$$

$$p(w_t, x_t) = \frac{1}{1 + e^{-x_t w_t^T}} \tag{2}$$

The w_t term refers to the current vector of logistic regression weights corresponding to each feature input x_t including the intercept; η is the learning rate of the algorithm, which is set to a constant 0.01 for this system after initial optimizations tuning this parameter translating to a gradual change in model parameters every retraining; y_t is the corresponding label (0 or 1) for the current feature input; and $p(w_t, x_t)$ is the sigmoid function which computes the probability of the current input vector to be either 0 or 1, corresponding to negative or positive seizure event classification respectively. The main benefit of using SGD in conjunction with logistic regression is that the weight update calculations are mathematically straightforward, as shown in (1), and can be done in parallel with minimal hardware cost. The sigmoid function calculation, shown in (2), can also be implemented using a look-up table further reducing computational complexity.

Traditionally, SGD is meant for supervised learning, where an external correct label is provided for every data input. However, for an implantable system operating *in situ*, externally provided labels are not available once deployed. Thus, our approach places SGD within an unsupervised learning paradigm during the online classification phase. This is implemented through bootstrapping, which uses the classifier's predicted probability output to update its own model. The classifier's output probability $p(w_t, x_t)$ is simply rounded to either 0 or 1, and is then treated as a label y_t for SGD to calculate the new set of feature as weights shown in (1).

Cumulative accuracy over time (during online training) is heavily dependent on the initial classifier accuracy after the offline training phase. For example, training on a misclassification can increase the error rate; an undesirable form of positive feedback. To avoid such an occurrence, the model is only updated once a specified confidence threshold is reached. A windowing technique, shown in Fig. 2, is employed such that a series of high-confidence predictions are required to trigger the online model update. The inputs corresponding to these high-confidence predictions would then be used as new data that the SGD algorithm will iterate over for the next model update.

The windowing technique is optimized for the following parameters: a) the window size, which translates to the number of successive seconds of high-confidence classifier predictions before initiating training; and b) the confidence threshold, which corresponds to the minimum classifier output probability that would be considered high-confidence. The parameters are tuned offline and on a patient-specific basis due to EEG/ECoG variability between patients. Achieving good performance offline is essential since final accuracy achieved during testing are dependent on these parameters.

III. Seizure Detection System

A seizure detection system incorporating SOUL is shown in Fig. 3. The system supports 16 channels of EEG/ECoG data, but the algorithm is scalable to any number of channels. The digitized data per channel goes to 4 feature extraction blocks. All 64 calculated features are then input to the classifier.

Two main feature classes are extracted: line length and spectral band powers. These features are commonly used in the implementation of seizure detection systems [7] since they capture the high-amplitude and high-frequency activity characteristic of seizure events. Other features were also considered, such as spectral entropy and time/frequency correlations, but were down-selected due to low weight values after running feature selection algorithms using L1-norm penalization. Tree-based feature selection was also run with these two features being weighted, in terms of feature importance, averaging at least 50% more than the rest. The selected features do not require significant hardware to implement: line length is the accumulated absolute difference between consecutive points; while spectral band power can be approximated by passing the signal through a bandpass FIR filter on a specified frequency range and then performing sum of squares, exploiting Parseval's theorem. This translates well into the low power implantable systems space.

Features are calculated on a sliding window with 0.1 second increments. Spectral band powers within the alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-100 Hz) frequency bands are calculated for each channel.

IV. Classifier Performance Evaluation

The performance of the classifier is tested using an ECoG dataset from the University of Melbourne [2, 8], and the CHB-MIT scalp EEG database. The former features >250-hour recordings on three patients to demonstrate how the online learning scheme of SOUL performs over a long period of time. The latter is a collection of relatively shorter recordings over 24 patients for performance comparisons on a wider population. The CHB-MIT dataset also allows for state-of-the-art comparisons as it is a commonly used dataset to test seizure detection classifiers.

Due to the incremental learning introduced by SOUL, the time series nature of the data must be preserved. Fig. 4 shows how the dataset for each patient is divided into training, validation and holdout. Contrary to random sampling performed

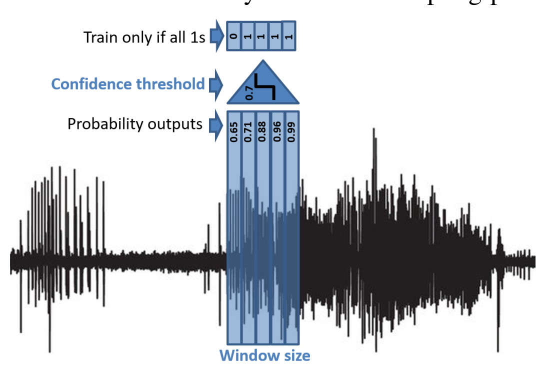

Figure 2: A windowing scheme, tuned using two hyperparameters, is implemented to only train the classifier once a series of high-confidence predictions are generated, allowing for a more robust model over time.

in conventional machine learning approaches, an initial supervised model is created offline followed by classification in a causal manner. It is during the classification phase that SOUL performs incremental updates in an unsupervised fashion to improve detection rates over time.

A. Performance on the long-term ECoG data

The University of Melbourne dataset features ECoG recordings from three human patients who had the lowest seizure prediction performances from the NeuroVista Seizure Advisory System clinical trial [8]. Each recording is more than 250 hours long containing more than 200 seizures per patient.

SOUL was tested on the three patient recordings and compared against an SVM, which is the commonly used seizure state classifier [3]–[6]. For the SVM, training is performed only offline, since model updates for this algorithm have significant impact on computational complexity making it infeasible for the target application. A base logistic regression classifier (which SOUL improves upon) was also included in the comparison as a baseline. Fig. 5 shows a comparison of the algorithms in terms of cumulative sensitivity (true positive/seizure detection rate over time) for Patient 3. The decreasing sensitivities of these conventional offline-only-trained algorithms demonstrate that seizure patterns change over time leading to missed detections. During classification, SOUL tunes the effective threshold for seizure detection, allowing sensitivity to be maintained over time. All algorithms are designed such that specificity (true negative rate) remains >95%, which translates to <1.2 false alarms per day. In a seizure detection application, false alarm rates (*1–specificity*) are of less concern than missed detections (*1–sensitivity*) and are therefore tolerable as long as they are kept below a reasonable level. In this work, <1.2 false alarms per day is maintained for all patients in all algorithms, equivalent to false alarm rates of commercial devices [1].

A comparison of the sensitivity and specificity for all three patients are shown in Fig. 6. SOUL achieves sensitivity that is statistically comparable to the other algorithms in Patients 1 and 2. For Patient 3, SOUL outperformed the SVM by 8.2% after 300 hours, and its offline-only-trained logistic regression counterpart by 16.3%. This is usually indicative of significantly varying seizure event patterns that the static classifier was not able to capture during the offline training phase. The incremental model updates from SOUL shift the classifier threshold during classification, allowing it to track the changing patterns over time. With this dynamic adaptation, later seizures can still be detected, raising classifier sensitivity. This approach can also decrease specificity (which is clearly seen in Patient 3 in Fig. 6) due to SOUL being limited to the threshold-based nature of logistic regression. The average latency for all three patients was 2.6 seconds, which was the median for all algorithms tested.

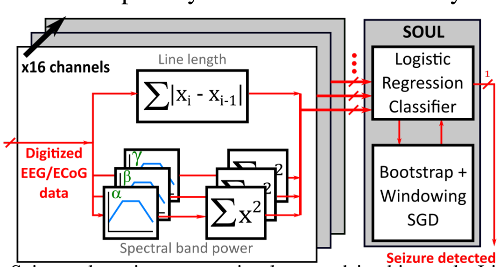

Figure 3: Seizure detection system implemented in this work. Line length and spectral band powers for each of the 16 channels are calculated in parallel which are then fed into SOUL for classification and model update.

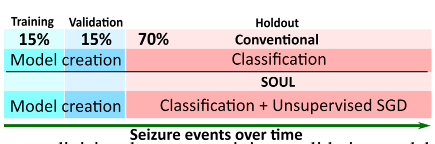

Figure 4: Dataset division between training, validation and holdout while preserving the time series nature of the data. SOUL enables unsupervised model updates during classification on the holdout set.

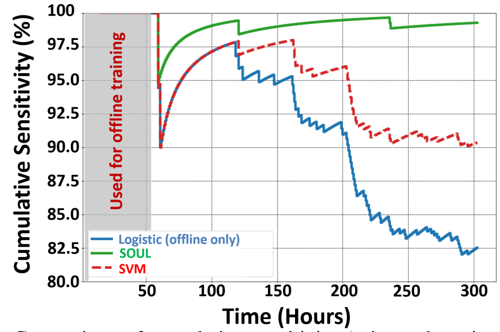

Figure 5: Comparison of cumulative sensitivity (seizure detection rate over time) for SOUL, offline-only-trained logistic regression and SVM for Patient 3 of the University of Melbourne dataset. For SOUL, sensitivity was maintained >95%, with final value 8.2% better than the SVM.

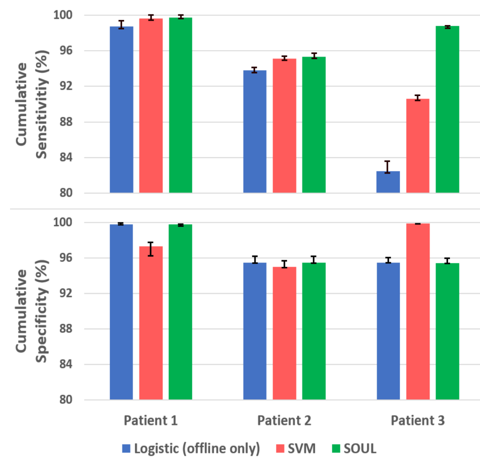

Figure 6: Final sensitivity and specificity values (higher is better) at the end of testing period; error bars indicate max/min values for the last 24 hours.

B. Performance on the CHB-MIT dataset

The CHB-MIT dataset consists of scalp EEG recordings from 24 pediatric subjects with intractable seizures. Across all subjects, the mean recording time was 41 hours (min: 19 hours, max: 156 hours, median: 33 hours). Correspondingly, the mean number of recorded seizure events per subject was 7.6 (min: 3, max: 27, median: 6).

Using SOUL, all 24 subjects resulted in improved performance when compared with the state-of-the-art. The majority (87%) of subjects resulted in a moderate (~1-3%) improvement in sensitivity and specificity when compared to [3, 4], which reported these values per patient. For three subjects (6, 8, and 18), a significant benefit can be observed. Relative to [3, 4], the average sensitivity for these subjects improved by >12%, while the average specificity differed by <1%. Without online training, the cumulative sensitivity for these subjects decreased over time due to missed detections. Similar to Patient 3 in the dataset discussed in Section A, this may be due to varying seizure event patterns over time that can be tracked by SOUL and updated in real time. For the majority of subjects, the moderate improvement may be due to more stable seizure patterns, or it may be due to the limited number of seizure events per patient in the data, which restrict the number of training and inference datapoints.

The advantage of SOUL is clearly seen in Fig. 7, a plot of the cumulative classifier sensitivity on the EEG recording from Subject 6, who has the longest recording among the subjects identified earlier. The base offline-only-trained logistic regression classifier performs the worst due to its threshold-based nature being unable to track varying patterns on long-term data. With SOUL, the initial model parameters can be updated online allowing it to adapt to signal changes leading to increased sensitivity. The final sensitivity achieved was >14% higher than an SVM classifier, which reached the same cumulative sensitivity as [4] on the same subject.

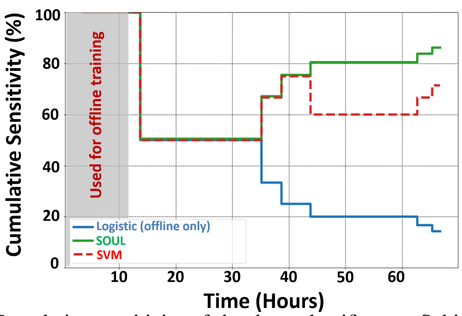

Figure 7: Cumulative sensitivity of the three classifiers on Subject 6 of the CHB-MIT database. Subject 6 has the longest recording and most number of seizures among subjects which SOUL sees a significant benefit.

A comparison of SOUL vs. state-of-the-art implementations that use the same dataset is shown in Fig. 8. While SOUL favors longer datasets to take advantage of incremental learning, the algorithm was still able to achieve a mean sensitivity of 97.5%, which is higher compared to other works. There is a relative ~1% penalty on specificity which is a very small trade-off to achieve higher detection rates. SOUL was also able to achieve the lowest detection latency, which is mainly attributed to the algorithm's ability to track the slow-changing EEG baseline due to online learning.

A significant benefit of SOUL lies in its low memory requirements, which can translate to low power operation when implemented in hardware. Compared to an SVM which requires an array of support vectors for classification, SOUL only requires a single vector of feature weights. The SVM classifiers used in [3, 4] required 64 kB of memory to store the support vectors, while the SOUL architecture shown in Fig. 3 would only require <200 bytes, which is a >300x decrease. Most of the power savings would come from the lower memory requirement, as the computations for the online-learning-capable SOUL roughly translate to an equivalent amount of logic compared to a hardware-optimized offline-only-trained SVM classifier.

V. Summary

This work presents SOUL, a logistic-regression-based classifier that continuously updates an initial offline-trained model

		JSSC 2013[3]	JBHI 2016[4]	BIOCAS 2018[5]	TBCAS 2018[6]	This Work
Classifier		Linear SVM	2 Linear SVM with arbiter	SVM	Non-linear SVM	SOUL (Logistic + SGD)
Training method		Offline Supervised	Offline Supervised	Offline Supervised	**Online** Supervised	**Offline + Online Unsupervised**
Classifier memory size		64 kB	64 kB	-	-	<200 B
Sensitivity	All subjects	82.7%	95.7%	91.1%	96.8%	97.5%
	Subjects 6, 8, 18	79.3%	81%	-	-	93%
Specificity	All subjects	95.5%	98.0%	90.1%	99.2%	98.2%
	Subjects 6, 8, 18	95.6%	94.3%	-	-	95.3%
Detection latency		<2 s	3.4 s	-	-	1.6 s

Figure 8: Comparison of SOUL versus the state-of-the-art seizure detection implementations using the CHB-MIT dataset.

with an unsupervised online learning scheme using a combination of stochastic gradient descent and bootstrapping. The classifier was tested on two human patient datasets. On long-term >250-hour ECoG data for three adult subjects, the proposed classifier maintains >95% sensitivity and specificity. Compared to offline-only-trained algorithms, SOUL achieved up to 8.2% higher cumulative sensitivity when compared an SVM classifier, while maintaining <1.2 false alarms per day. On pediatric EEG data, SOUL demonstrated a performance benefit for every patient, and achieved the highest average sensitivity of 97.5% when compared with state-of-the-art classifiers. Significant performance improvement was observed for a subset of patients, whose average sensitivity increased by >12% with <1% impact on specificity. SOUL achieved the shortest latency when compared with state-of-the-art. Implementing SOUL in a hardware system enables a >300x reduction in memory requirements compared to an SVM, which can translate to low power operation suitable for implantable seizure detection systems.

Acknowledgements

This research was funded by NSF CAREER, the Wagner Foundation, and the Army Research Office under contract/grant number W911NF-16-1-0368. The authors thank the sponsors of the Berkeley Wireless Research Center, and the Berkeley Research Computing program. Thanks to Dr. Dean Freestone and Dr. Mark Cook for providing the patient datasets, and to Paul Rigge, Yonatan Nozik, and Ryan Kaveh for technical discussion.

References

[1] "Summary of safety and effectiveness data: NeuroPace RNS System." https://www.accessdata.fda.gov/cdrh_docs/pdf10/p100026b.pdf.

[2] L. Kuhlmann *et al.*, "Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG," *Brain*, vol. 141, pp. 2619–2630, 08 2018.

[3] J. Yoo, "An 8-Channel Scalable EEG Acquisition SoC With Patient-Specific Seizure Classification and Recording Processor," *JSSC*, vol. 48, pp. 214–228, Jan 2013.

[4] C. Zhang *et al.*, "Design and Implementation of an On-Chip Patient-Specific Closed-Loop Seizure Onset and Termination Detection System," *JBHI*, vol. 20, pp. 996–1007, July 2016.

[5] L. Tang *et al.*, "Automatic Artifact Reduction Based on MEMD- for Seizure Prediction," in *BioCAS 2018*, pp. 1–4, Oct 2018.

[6] L. Feng *et al.*, "VLSI Design of SVM-Based Seizure Detection System With On-Chip Learning Capability," *TBCAS*, vol. 12, pp. 171–181, 2018.

[7] L. Logesparan, A. J. Casson, and E. Rodriguez-Villegas, "Optimal features for online seizure detection," *Medical & Biological Engineering & Computing*, vol. 50, pp. 659–669, Jul 2012.

[8] M. Cook, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study.," *Lancet Neurology*, vol. 12, pp. 563–71, 06 2013.