Hardware-Friendly Seizure Detection with a Boosted Ensemble of Shallow Decision Trees

Mahsa Shoaran, Masoud Farivar, Azita Emami

***Abstract*— Efficient on-chip learning is becoming an essential element of implantable biomedical devices. Despite a substantial literature on automated seizure detection algorithms, hardware-friendly implementation of such techniques is not sufficiently addressed. In this paper, we propose to employ a gradient-boosted ensemble of decision trees to achieve a reasonable trade-off between detection accuracy and implementation cost. Combined with the proposed feature extraction model, we show that these classifiers quickly become competitive with more complex learning models previously proposed for hardware implementation, with only a small number of low-depth ($d < 4$) "shallow" trees. The results are verified on more than 3460 hours of intracranial EEG data including 430 seizures from 27 patients with epilepsy.**

I. Introduction

Given the large population of patients with intractable epilepsy, the automatic detection of seizure onset has sparked great interest among researchers over the past 20 years. In addition to providing a vital seizure alert to the patient, caregiver or a therapeutic device, it significantly eases the task of reviewing and labeling seizure segments in a patient's EEG, a time-intensive task routinely done by neurologists. Implanting a device that performs both detection and closed-loop suppression is the ultimate goal. Today, the Responsive Neurostimulator (RNS) by NeuroPace provides an FDA-approved therapy option to reduce the seizure frequency. However, RNS is bulky, limited in number of channels, and only relies on simple hard thresholding with moderate seizure classification accuracy.

The power and area constraints imposed by implantable devices do not allow the implementation of sophisticated on-chip classification algorithms. Indeed, even the simple arithmetic operations performed in conventional classification methods, such as SVMs [1] and k-nearest neighbor (KNN) algorithms [2] can become very costly with increasing number of recording channels and higher sampling rates. With only simple comparator stages as their building blocks, decision trees (DTs) are a preferable solution to reduce hardware design complexity. Despite all their advantages, decision trees are unfortunately very susceptible to overfitting in seizure detection, particularly due to the high dimensionality of the feature space. This necessitates a careful design.

We present and evaluate a very light seizure detection algorithm using an ensemble of gradient-boosted decision tree classifiers. With the proposed feature extraction steps, we show that these ensembles can compete with more complex learning models proposed for on-chip implementation, with only a small number of low-depth trees. The proposed approach is tested on a large dataset of over 140 days of intracranial EEG data from 27 epileptic patients.

Related Work: [3] has utilized a decision tree spike classification method that interleaves 8 neural channels into one decision tree block. Operating at 50kHz, the proposed system performs spike sorting with negligible power and area per channel. As opposed to spikes that can be classified into multiple shapes, the seizure detection problem is normally simplified into two states of seizure and non-seizure, thus exhibits great potentials for a hardware-optimized implementation using decision trees. In another application [4], a wearable gait monitor using decision tree classifiers achieved roughly identical detection accuracy to support vector machines, while drawing three times less power. It therefore provides a framework for power-efficient detection in wearable systems, by hierarchical activation of sensors through a hierarchical decision tree classifier [4]. The authors in [5] propose a non-linear classifier using Adaboost technique with decision stumps (trees with depth of one) as base classifier, to achieve a low complexity seizure detection system. However, as discussed in Section III, the choice of $d = 3$ achieves a better trade-off between classification performance and implementation complexity.

II. Data Description and Methodology

A. Intracranial EEG Data

In this work, we use the publicly available data from the iEEG portal[1][6], augmented with 8 additional patients from the UPenn and Mayo clinic's seizure detection competition dataset [7], 7 of whom are iEEG recorded at 5kHz and one at 500Hz. The portal includes iEEG recordings at both high and low sampling rates and various types of epilepsy. All patients in the portal with three or more expert marked seizures are included in this analysis. The access IDs of analyzed patients and further details are provided in Table I. In total, more than 3460 hours of data from 27 patients including 430 seizures are processed.

B. Feature Selection

Based on our initial study on discriminative performance versus hardware complexity of several frequency and time domain features, and the existing literature in [8]-[11], we limited ourself to the following set of features: line-length, time-domain variance, and multiple band powers,

The authors are with the Electrical Engineering Department, California Institute of Technology, Pasadena, CA.
E-mail: {mshoaran, mfarivar, azita}@caltech.edu

[1] www.ieeg.org

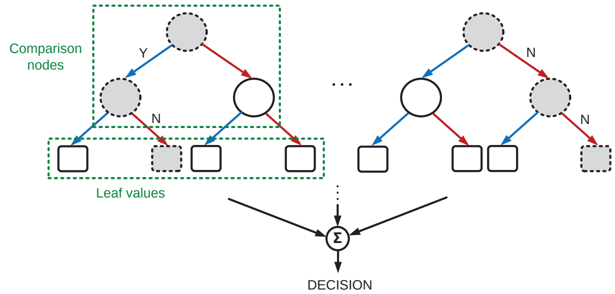

Fig. 1: A general schematic diagram of a boosted ensemble of shallow decision trees proposed for hardware efficient seizure detection (here $d = 2$).

as described in Table II. While studies on EEG signals have emphasized on an epileptic frequency range of below 30Hz [8], [10], the intracranial EEG (iEEG) signals span a wider frequency range, lately shown to go beyond 200Hz for seizure biomarker extraction [12], [13]. These high-frequency oscillations (HFOs) have been previously studied [14] on 36h of iEEG data to evaluate their seizure detection accuracy. The authors have concluded a significant potential of HFOs for seizure detection. In this work, we compare the discriminative performance of various frequency bands, including HFOs, on an extensive iEEG database.

C. Gradient-Boosted Decision Trees

Gradient-boosting [15] is one of the most successful machine learning techniques that exploits Gradient-based optimization and boosting, by adaptively combining many simple models to get an improved predictive performance. Binary split decision trees are commonly used as the "weak" learners. Boosted trees are at the core of the state-of-the-art solutions in a variety of learning domains, given their excellent accuracy, fast computation and operation. The output of a boosted classifier (or regressor) has the additive form of $H(x) = \sum_t \alpha_t h_t(x)$. A general schematic diagram illustrating the components of an ensemble of depth-2 trees is shown in Fig. 1. In this paper, we have employed the XGBoost package [16], a parallelized implementation of Gradient-boosting algorithm. Applying this method to our iEEG dataset, we observed over 100 times improvement in training speed compared to common SVM implementations.

TABLE I: Patient Data and Signal Acquisition Info.

Subj.	iEEG Portal ID	No. Elec.	No. Seiz.	Rec. Dur.	Samp. Rate
1	Study 004-2	56	3	7d 18h	500
2	Study 006	56	5	1d 14h	500
3	Study 040	116	6	2d 23h	5k
4	Study 017	16	9	7d 17h	500
5	Study 011	88	3	3d 12h	500
6	Study 022	56	7	3d 23h	500
7	I001_P034_D01	47	16	1d 8h	5k
8	Study 010	56	3	12d 16h	500
9	Study 023	88	4	2d 5h	500
10	Study 012-1	60	6	3d 7h	500
11	Study 027	48	6	3d 21h	500
12	Study 016	64	7	5d 21h	500
13	Study 031	116	5	6d 19h	500
14	I001_P010_D01	56	10	3d 18h	5k
15	Study 030	64	8	5d 23h	500
16	Study 036	96	4	4d 14h	5k
17	Study 020	56	8	5d 0h	500
18	Study 014	104	15	6d 0h	500
19	Study 021	108	13	6d 11h	500
20	Study 026	96	22	3d 3h	500
21	Study 024	88	19	8d 10h	500
22	Study 028	96	9	1d 16h	500
23	Study 038	88	10	3d 0h	500
24	Study 005	16	151	6d 16h	500
25	Study 012-2	84	28	13d 16h	500
26	Study 019	96	36	5d 16h	500
27	Study 033	128	17	6d 17h	500

III. Classifier Design and Performance Evaluation

Decision trees are very efficient, but also susceptible to overfitting in problems with high feature-space dimensionality. One way to address this is to limit the number of nodes in each tree, i.e., design shallow trees using small number of features. Shorter trees are also more efficient in hardware and equally important, incur less detection delay. Therefore, it is important to carefully select the depth parameter and also to understand the relative predictive value of individual features in prior. Figure. 2 shows the Area Under the Curve (AUC) performance of an ensemble of gradient-boosted trees versus the number of trees for different values of the depth parameter. An important observation is that the detection accuracy is not significantly improved ($< 0.5\%$) with the depth values of 4 and higher. As a simple benchmark, let us consider a boosted ensemble of 5 shallow trees with depth of 3, and compare it to linear SVM, cubic SVM and KNN-3 models, previously proposed in the literature for on-chip classification. Figure. 3 shows the F1-measure performance of these classifiers across different patients. We can see that this benchmark is already competitive with its peers, and that it can outperform with larger ensemble sizes. In our simulations, this benchmark achieved an average seizure detection sensitivity of 98.3%.

Figure. 4 summarizes the overall performance of examined features across patients. In order to obtain a more realistic estimation of accuracy under various measurement conditions, we have not used any pre-processing techniques. The performance could be further boosted by artifact removal, as some datasets (e.g. patient 3 and 18) are contaminated by high-frequency artifacts that particularly overlap with HFO band. Line-length stands out as the best *single* discriminative

TABLE II: Evaluated Features

Feature	Description
Line-Length (LLN)	$\frac{1}{d}\sum_d \lvert x[n] - x[n-1] \rvert$, d = window length
Power (POW)	Total spectral power
Variance (VAR)	$\frac{1}{d}\sum_d (x[n]-\mu)^2$ where $\mu = \frac{1}{d}\sum_d (x[n])$
Delta Power (δ)	Spectral power in 1-4Hz
Theta Power (θ)	Spectral power in 4-8Hz
Alpha Power (α)	Spectral power in 8-13Hz
Beta Power (β)	Spectral power in 13-30Hz
Gamma Power (γ)	Spectral power in 30-80Hz
Ripple Power (Ripple)	Spectral power in 80-200Hz
Fast Ripple Power (FR)	Spectral power in 200-250Hz @ SR = 500Hz, Spectral power in 200-600Hz @ SR = 5kHz
HFO Power (HFO)	Spectral power in 80-250Hz @ SR = 500Hz, Spectral power in 80-600Hz @ SR = 5kHz

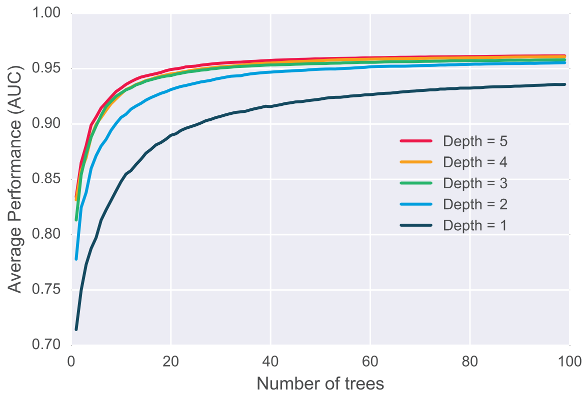

Fig. 2: The overall classification performance at various depths versus number of trees.

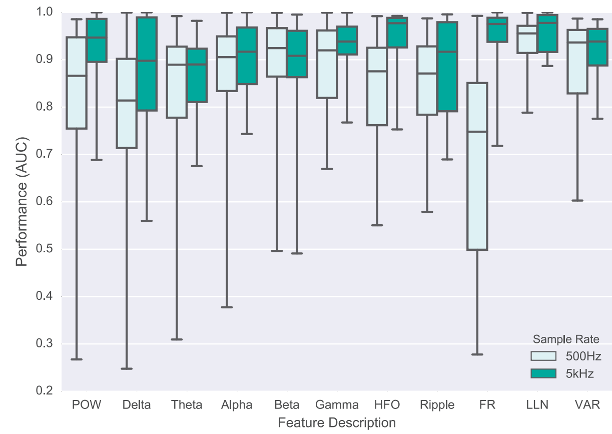

Fig. 4: Feature importance for patients with two different sampling rates of 500Hz and 5kHz.

feature, in confirmation with the results reported in [8] and being used as a gold standard in [17]. It captures both low-amplitude fast and high-amplitude slow activities during the course of a seizure. As shown in Fig. 4, the optimal frequency range that exhibits the most discriminating epileptiform activity is patient-dependent, but in majority of patients sampled at a sufficiently high rate of 5k, it has a clear shift from Berger bands (delta, theta, alpha, beta) towards gamma, fast ripple, and more specifically, the HFOs.

As discussed in [14], HFOs may be missing in some cases and hard to capture at low sampling rates. In addition, it is somewhat challenging to capture them due to presence of artifacts, their low amplitude and duration, and rare occurrence. However, their potential in early detection of seizure onset is promising, a factor of great importance in seizure control devices. Inspired by the early works on exploration and analysis of HFOs [13], several researchers are therefore seeking to alleviate these challenges by developing automatic detection methods [18] to ease the use of HFOs in clinical routine. Applying circuit techniques to suppress the effect of artifacts and improve the signal-to-noise ratio in measurements may help to successfully capture them in future devices.

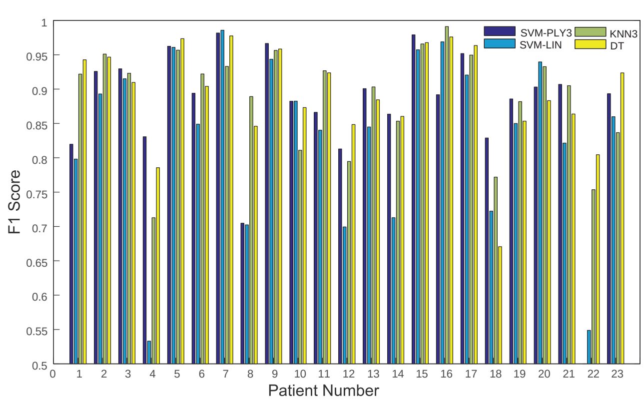

Fig. 3: Comparison of predictive ability (F1 Scores) of three different classification methods with an ensemble of five trees of depth $d = 3$.

The results presented above encourage a patient-specific training step to set the frequency passband of the feature extraction filter, in order to get the desired performance for every patient. While the physical implementation of all spectral power features, then selection and elimination of failed ones may cause significant hardware cost, the circuit-level tuning of a band-pass filter is much more practical. Upon training for each patient, the bandwidth may be fixed, as the dominant range of rhythmic seizure activity for each person is nearly consistent over time [10].

IV. Hardware-Friendly Classification

As our feature importance studies showed, two features prove to be dominant: line-length and a single spectral power specific to each patient. Furthermore as shown in Fig. 2, very little improvement in performance is achieved by using trees with a depth of 4 and above. These findings can lead to proper design solutions to implement hardware-efficient decision trees, as discussed below.

A. Mixed-Signal Decision Tree Topology

In addition to choosing an inherently simple classifier such as DT, further hardware saving could be achieved by performing an initial detection in analog domain. As opposed to the fully digital approach in [19], we suggest to build a mixed-signal DT classifier by combining the light analog pre-detectors with more complex digital features. The proposed architecture is shown in Fig. 5. Since the final decision in majority of cases during the operation of device is equal to NO (i.e., seizures are rare events), the power consumption can be significantly reduced by performing an initial ultra-low-power and sensitive analog detection within each channel (e.g. line-length and a tunable bandpass filter) and keeping the digital circuitry off during this phase. Once this step is completed, those channels with a “YES” or “UNCERTAIN” state are further processed in digital domain. This technique

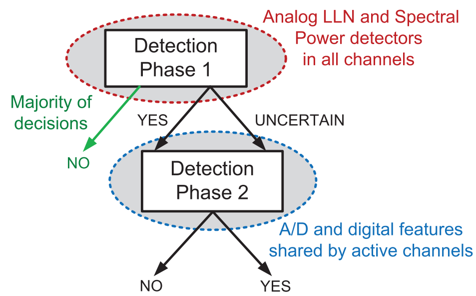

Fig. 5: An efficient implementation of DT classifier in two steps: analog per channel and shared digital.

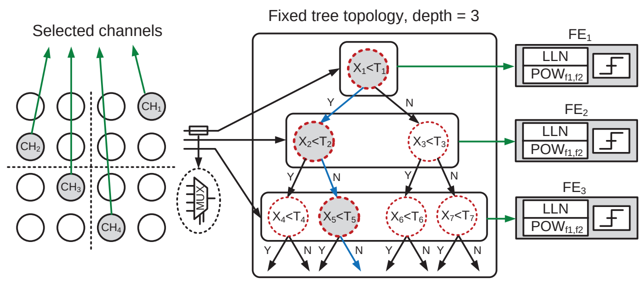

Fig. 6: A tree of depth 3 assigned to the selected channels during training. The low and high cut-off frequencies of the filter (f_1 and f_2) are externally trained and set per patient.

alleviates the conventional overhead of digitization and high-complexity digital feature extraction inside the channels.

B. Optimal Channel Allocation upon Learning from Data

A critical challenge of online seizure detection using an implantable device is that the seizure detection algorithm and corresponding circuit architecture has to be chosen and implemented in advance. Using switching techniques and multiplexing, however, provide some degree of flexibility in allocation of physically implemented blocks to selected channels. To partially alleviate this problem, a generic decision tree architecture with a reasonable depth and complexity can be implemented on chip. During each comparison step, only the feature value of the channel appearing in the active node and path of tree is needed, as shown in Fig. 6. The rest of array can be switched off to save power. The channels can be multiplexed either across the entire array, or chosen among a selected subset of channels which are dominant decision makers during training. Interestingly, only $D \times N$ feature extraction blocks are required, with D being the depth of tree and N being the number of trees. The drawback is that the depth of tree will affect the detection latency. Alternatively, since the final decision of each tree is made upon completing the decisions in prior levels, one single feature extraction block (analog or digital) can be sequentially used per tree, resulting in significant hardware saving.

V. Conclusion

Hardware-efficient seizure detection becomes increasingly important in systems with hundreds of recording electrodes, a future trend in neuroscience and neuroengineering. Towardes this goal, we studied the performance of gradient-boosted ensemble of low-depth decision trees with a selected subset of features on a large iEEG database. We show that the proposed solution performs comparatively well against previously reported learning models for hardware implementation, with only a handful of trees of depth three.

References

[1] J. Yoo, L. Yan , D. E. Damak , M. B. Altaf , A. Shoeb, A. Chandrakasan, "An 8-channel scalable EEG acquisition soc with patient-specific seizure classification and recording processor," *IEEE J. Solid-State Circuits*, vol. 48, pp. 214-228, 2013.

[2] A. Page, C. Sagedy, E. Smith, N. Attaran, T. Oates, T. Mohsenin, "A flexible multichannel EEG feature extractor and classifier for seizure detection," *IEEE Trans. Circuits Sys. II: Express Briefs*, vol. 62, pp. 109-113, 2015.

[3] Y. Yang, C. Boling, A. J. Mason, "Power-area efficient VLSI implementation of decision tree based spike classification for neural recording implants," *Biomedical Circuits and Systems Conference (BioCAS)*, pp. 380-383, 2014.

[4] A. Benbasat and J. Paradiso, "A Framework for the Automated Generation of Power-Efficient Classifiers for Embedded Sensor Nodes," *ACM SenSys*, 2007.

[5] M. Ayinala and K. K. Parhi, "Low complexity algorithm for seizure prediction using Adaboost," *Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1061-1064, 2012.

[6] J. B. Wagenaar, G. A. Worrell, Z. Ives et al., 'Collaborating and sharing data in epilepsy research," *J Clin Neurophysiol.*, vol. 32, pp. 235-239, 2015.

[7] Available online at `www.kaggle.com/c/seizure-detection`

[8] L. Logesparan, A. J. Casson, E. Rodriguez-Ville, "Optimal features for online seizure detection," *Med Biol Eng Comput.*, vol. 50, pp. 659-669, 2012.

[9] R. Esteller, J. Echauz, T. Tcheng, B. Litt, B. Ples, "Line length: an efficient feature for seizure onset detection," *Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1707-1710, 2001.

[10] A. Shoeb and J. Guttag, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment," *Proc. Int. Conf. Machine Learning*, Jun. 2010.

[11] M. Shoaran, C. Pollo, K. Schindler, A. Schmid, "A Fully-Integrated IC with 0.85-μW/Channel Consumption for Epileptic iEEG Detection," *IEEE Trans. Circuits Sys. II: Express Briefs*, vol. 62, pp. 114-118, 2015.

[12] A. Bragin, J. Engel, C. L. Wilson, I. Fried, G. Buzski, "High-frequency oscillations in human brain," *Hippocampus*, vol. 9, pp. 137-142, 1999.

[13] R. J. Staba, C. L. Wilson, A. Bragin, I. Fried, J. Engel , "Quantitative analysis of high-frequency oscillations (80-500 Hz) recorded in human epileptic hippocampus and entorhinal cortex," *J Neurophysiol.*, vol. 88, pp. 1743-1752, 2002.

[14] L. Ayoubian, H. Lacoma, J. Gotman, "Automatic seizure detection in SEEG using high frequency activities in wavelet domain," *Med Eng Phys.*, vol. 35, no. 3, pp. 319-328, 2013.

[15] J. H. Friedman, "Greedy function approximation: a gradient boosting machine", *Annals of Statistics*, pp.1189-1232, 2001.

[16] T. Chen, T. He, "xgboost: eXtreme Gradient Boosting," *R package version 0.4-2*, 2015.

[17] K. Schindler, H. Leung, C. E. Elger, K. Lehnertz, "Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG," *Brain*, vol. 130, pp. 65-77, 2007.

[18] S. V. Gliske, Z. T. Irwin, K. A. Davis, K. Sahaya, C. Chestek, W. C. Stacey, "Universal automated high frequency oscillation detector for real-time, long term EEG," *Clin Neurophysiol.*, vol. 127, pp. 1057-1066, 2016.

[19] E. Shih, J. Guttag, "Reducing Energy Consumption of Multi-channel Mobile Medical Monitoring Algorithms," *in Proc. Int. Workshop Sys. Netw. Supp. for Healthcare and Assisted Living Env.*, 2008.