Using a simple decision tree model, we introduce efficient hardware architectures to predict abnormal neurological states in various disorders.

Efficient Feature Extraction and Classification Methods in Neural Interfaces

Mahsa Shoaran, Benyamin A. Haghi, Masoud Farivar, and Azita Emami

Mahsa Shoaran

Benyamin A. Haghi

Masoud Farivar

Azita Emami

Brain disorders such as dementia, epilepsy, migraine, and autism remain largely undertreated, but neural devices are increasingly being used for their treatment. Such devices are designed to interface with the brain, monitor and detect neurological abnormalities, and trigger an appropriate type of therapy such as neuromodulation to restore normal function.

A key challenge to these new treatments is to integrate state-of-the-art signal acquisition techniques, as well as efficient biomarker extraction and

Mahsa Shoaran is an assistant professor in the School of Electrical and Computer Engineering at Cornell University. Benyamin A. Haghi is a graduate student in the Mixed-mode Integrated Circuits and Systems Lab (MICS) at the California Institute of Technology. Masoud Farivar is a research scientist at Google. Azita Emami is the Andrew and Peggy Cherng Professor of Electrical and Medical Engineering, Heritage Medical Research Institute Investigator, and deputy chair of the Engineering and Applied Science Division at Caltech.

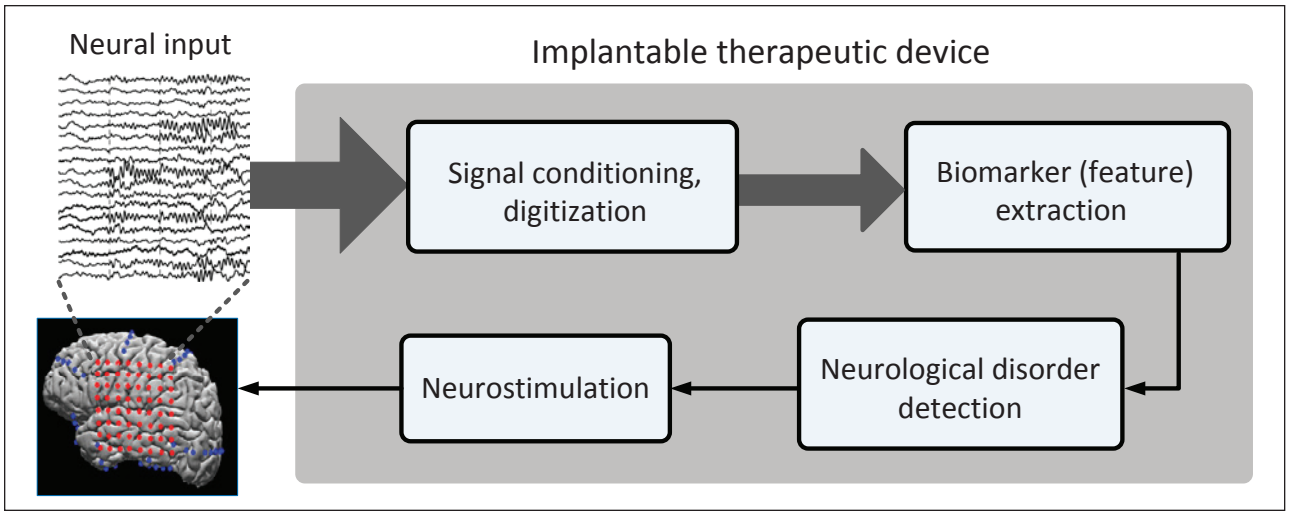


FIGURE 1 General block diagram of a closed-loop therapeutic system for detection and suppression of disabling neurological symptoms.

classification methods to accurately identify symptoms, using low-cost, highly integrated, wireless, and miniaturized devices.

Therapeutic Neural Devices

A general block diagram of a closed-loop neural interface system is shown in figure 1. The neural signals recorded by an array of electrodes (intracranial, scalp, or other types) are initially amplified, filtered, and digitized. A feature extraction processor is then activated to extract the disease-associated biomarkers. Upon abnormality detection, a programmable neural stimulator is triggered to suppress the symptoms of disease (e.g., a seizure, migraine attack, Parkinson's tremor, memory dysfunction) through periodic charge delivery to the tissue.

The abnormality detector device must demonstrate high sensitivity (true positive rate), sufficient specificity (true negative rate), and low latency. It also has to satisfy the safety, portability, and biocompatibility requirements of the human body.

An Example of Neuroengineering Treatment: Epilepsy

The emerging field of neuroengineering uses engineering technologies to investigate and treat neurological diseases. Epilepsy has been one of the primary targets, along with movement disorders, stroke, chronic pain, affective disorders, and paralysis (Stacey and Litt 2008).

Approximately one third of epileptic patients exhibit seizures that are not controlled by medications. Neuromodulation offers a new avenue of treatment for intractable epilepsy.

Over decades, research on epilepsy has led to fundamental understandings of brain function, with strong implications for other neurological disorders. In addition, because of the severity of refractory epilepsy and the need for surgery, human tissue and epileptic EEG datasets are largely available. Most therapeutic neural interfaces reported in the literature have therefore focused on extracting epileptic biomarkers for automated seizure detection (Shoaran et al. 2015; Shoeb et al. 2004; Verma et al. 2010).

The spectral energy of neural channels in multiple frequency bands as well as various time and frequency domain features have been used as potential seizure biomarkers. To improve the power and area efficiency in multichannel systems, a spatial filtering technique was proposed to precede the seizure detection unit (Shoaran et al. 2016b). But in most devices the classification of neural features is performed either remotely or by means of moderately accurate thresholding techniques.

For one patient-specific support vector machine (SVM) classifier (implemented by Yoo et al. 2013), the classification processor contributes to a significant portion of chip area and power. To improve the accuracy of detection, resource-efficient on-chip learning is becoming an essential element of next-generation implantable and wearable diagnostic devices.

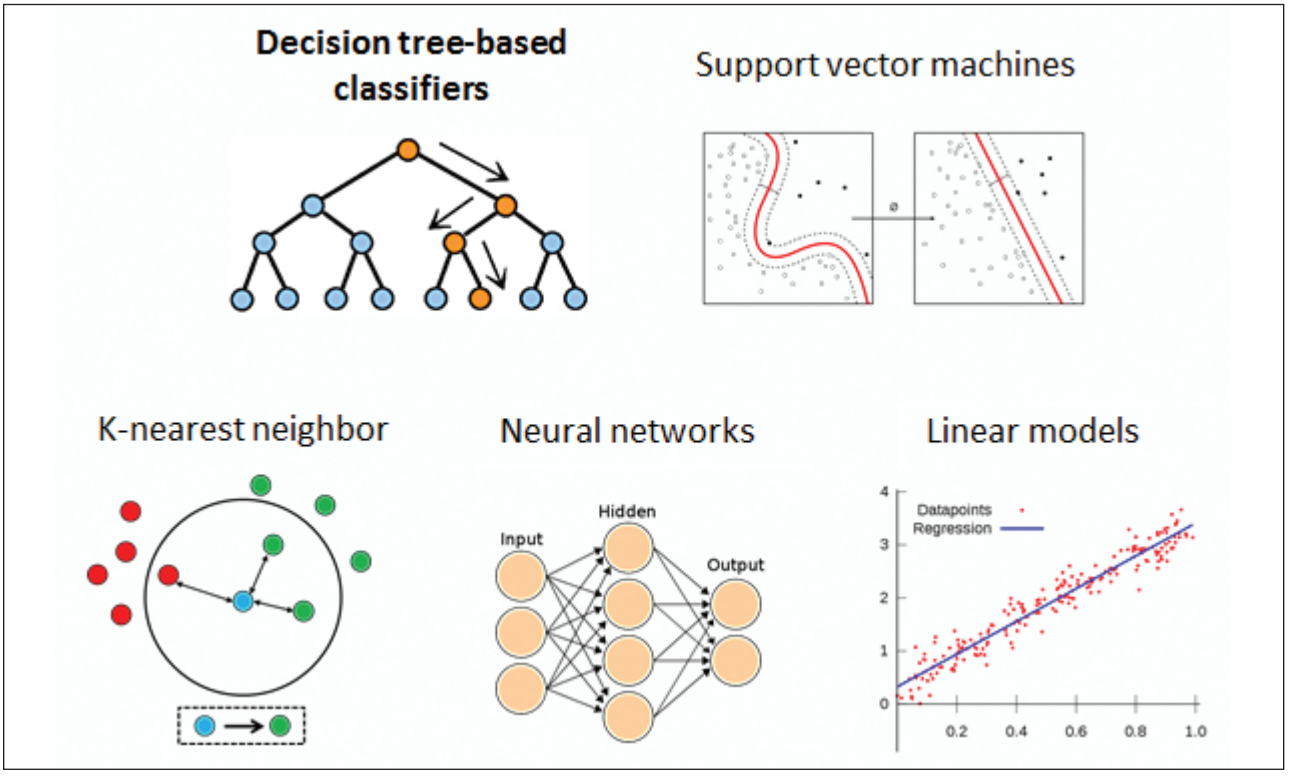


FIGURE 2 Schematic of common learning models as potential candidates for hardware implementation.

Machine Learning in Neural Devices: Scalability Challenges

Conventional classification techniques such as SVMs, k-nearest neighbors (KNNs), and neural networks (illustrated in figure 2) are hardware intensive and require high processing power and large memory units to perform complex computations on chip.

Numerous studies show that a large number of acquisition channels are required to obtain an accurate representation of brain activity, and that the therapeutic potential of neural devices is limited at low spatio-temporal resolution. It is expected that future interfaces will integrate thousands of channels at relatively high sampling rates, making it crucial to operate at extremely low power. The device must also be very small to minimize implantation challenges.

Despite a substantial literature on machine learning, hardware-friendly implementation of such techniques is not sufficiently addressed. Indeed, even the simple arithmetic operations performed in conventional classification methods can become very costly with an increasing number of channels.

Finally, filter banks and, in general, feature extraction units can be hardware intensive, particularly at higher frequencies associated with intracranial EEG. Extensive system-level design improvement is needed to meet the requirements of an implantable device while preserving high-resolution recording capability.

Decision Tree–Based Classifiers

We present and evaluate a seizure detection algorithm using an ensemble of decision tree (DT) classifiers. The general schematic of a single decision tree is shown in figure 2.

With only simple comparators as their core building blocks, DT classifiers are a preferable solution to reduce hardware design complexity. Using a gradient-boosted ensemble of decision trees, we achieve a reasonable tradeoff between detection accuracy and implementation cost.

Gradient boosting (Friedman 2001), one of the most successful machine learning techniques, adaptively combines many simple models to get an improved predictive performance. Binary split decision trees are

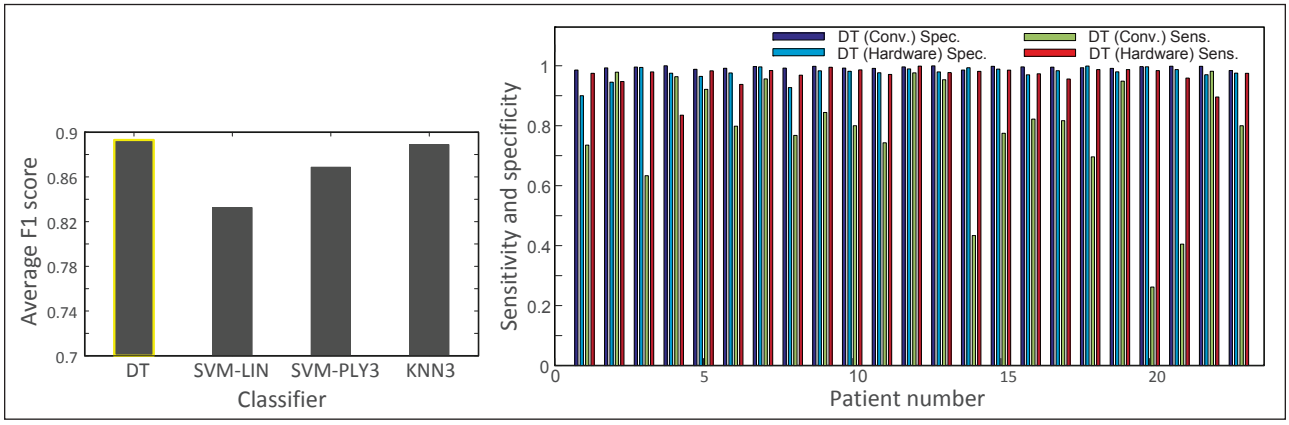


FIGURE 3 Comparison of predictive ability of different classification methods with an ensemble of 8 decision trees (DT) of depth 3 (left), and the classification performance of the asynchronous hardware model compared to a conventional (conv.) DT (right). KNN = K-nearest neighbor; LIN = linear; PLY3 = polynomial kernel of order 3; SVM = support vector machine.

commonly used as the "weak" learners. Boosted trees are at the core of state-of-the-art solutions in a variety of learning domains because of their accuracy and fast computation and operation.

Combined with an efficient feature extraction model, we show that, with only a small number of low-depth "shallow" trees, the boosted classifiers quickly become competitive with more complex learning models (Shoaran et al. 2016a). These ensembles of axis-parallel DT classifiers are excellent candidates for on-chip integration, eliminating the multiplication operation and offering significant reductions in power and chip area.

Performance Evaluation and Hardware Design

As a benchmark, we compare a boosted ensemble of 8 trees with a depth of 3 to linear SVM, cubic SVM, and KNN-3 models proposed for on-chip classification, using the following features: line length, time-domain variance, and multiple band powers. The proposed approach is tested on a large dataset of over 140 days of intracranial EEG data from 23 epileptic patients.

Figure 3 (left) shows the average F1 measure of classifiers. This benchmark is already competitive with its peers, and can outperform using larger ensemble sizes. It achieves an average seizure detection sensitivity of 98.3 percent.

Decision trees are very efficient, but also susceptible to overfitting in problems with high feature space dimensionality. To address this, we limit the number of nodes in each tree—that is, we design shallow trees with a small number of features. These shorter trees are also more efficient in hardware and, equally important, incur less detection delay. In our simulations, the detection accuracy is not significantly improved (<0.5 percent) with DT depth values of 4 or more.

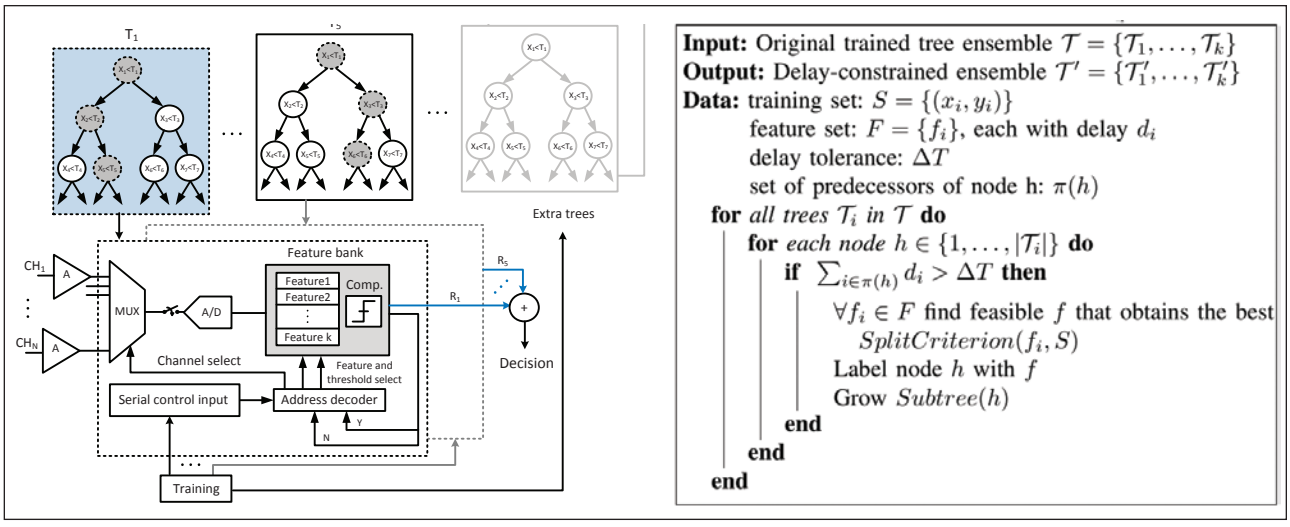
Proposed Decision Tree Architecture

We propose the architecture shown in figure 4 (left) to implement ensembles of decision trees. At each comparison step, only the features appearing in the active nodes of trees are needed; the rest of the recording array can be switched off to save power.

Because the final decision is made upon completing decisions at prior levels, a single feature extraction unit can be sequentially used per tree. This results in a significant hardware saving, in contrast to SVM, which requires all features from the entire array.

For example, the memory required to classify 32-channel neural data with 8 trees (a maximum depth of 3 and threshold resolution of 8 bits) is as low as 100 bytes, while SVM and KNN-based arrays would need over 500 kB of memory. Depending on the specific patient and the difficulty of the detection task, additional "supportive" trees can be used to further boost the classification accuracy.

The proposed architecture faces a practical challenge of designing decision trees under application-specific delay constraints. Given any DT ensemble $\tau = \{\tau_1,\ldots,\tau_k\}$ obtained from our original method, we need to ensure that each tree τ_i satisfies the delay constraint: $\sum_{i\in\pi(h)} d_i \le \Delta T$, where d_i is the time required to compute feature f_i, ΔT is the maximum tolerable detection delay, and $\pi(h)$ is the set of all predecessors of node h. We propose a "greedy" algorithm to solve this



Input: Original trained tree ensemble $\mathcal{T} = \{\mathcal{T}_1, \ldots, \mathcal{T}_k\}$
Output: Delay-constrained ensemble $\mathcal{T}' = \{\mathcal{T}'_1, \ldots, \mathcal{T}'_k\}$
Data: training set: $S = \{(x_i, y_i)\}$
feature set: $F = \{f_i\}$, each with delay d_i
delay tolerance: ΔT
set of predecessors of node h: $\pi(h)$
for *all trees* $\mathcal{T}_i$ *in* $\mathcal{T}$ **do**
  for *each node* $h \in \{1, \ldots, |\mathcal{T}_i|\}$ **do**
    if $\sum_{i \in \pi(h)} d_i > \Delta T$ **then**
      $\forall f_i \in F$ find feasible f that obtains the best $SplitCriterion(f_i, S)$
      Label node h with f
      Grow $Subtree(h)$
    end
  end
end

FIGURE 4 Hardware-level architecture for an ensemble of decision tree classifier with primary and supportive trees (left) and a greedy training algorithm to meet the delay constraints (right). A = amplifier; A/D = analog to digital converter; CH = channel; Comp. = comparator; k, N = number of features and channels; MUX = multiplexer; R = result.

practical constraint by building trees that satisfy the delay requirement, as illustrated in figure 4 (right).

However, this algorithm may result in a suboptimal solution. We therefore investigate a novel asynchronous model to learn from neural data streams, the results of which are shown in figure 3 (right). In this model, the trees are built with features that maximize accuracy regardless of their computational delay. Based on averaged results of completed trees and previous results of incomplete trees, decisions are frequently updated (over 0.5-sec intervals) to avoid long latencies and maximize sensitivity. Once completed, longer trees contribute to decisions at future time steps.

Conclusions

Based on a simple yet sufficiently accurate (98.3 percent) decision tree model, we introduce efficient hardware architectures and related training algorithms to predict the abnormal neurological states in various disorders, such as epilepsy, Parkinson's disease, and migraine. Such classifiers may allow the full integration of processing circuitry with the sensor array in various resource-constrained biomedical applications.

References

Friedman JH. 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29(5):1189–1232.

Shoaran M, Pollo C, Schindler K, Schmid A. 2015. A fully-integrated IC with 0.85μW/channel consumption for epileptic iEEG detection. IEEE Transactions on Circuits and Systems II: Express Briefs (TCAS-II) 62(2):114–118.

Shoaran M, Farivar M, Emami A. 2016. Hardware-friendly seizure detection with a boosted ensemble of shallow decision trees. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), August 16–20, Orlando.

Shoaran M, Shahshahani M, Farivar M, Almajano J, Shahshahani A, Schmid A, Bragin A, Leblebici Y, Emami A. 2016. A 16-channel 1.1mm^2 implantable seizure control SoC with sub-μW/channel consumption and closed-loop stimulation in 0.18μm CMOS. Proceedings of the IEEE Symposium on VLSI Circuits (VLSIC), June 13–17, Honolulu.

Shoeb A, Edwards H, Connolly J, Bourgeois B, Treves ST, Guttag J. 2004. Patient-specific seizure onset detection. Epilepsy and Behavior 5(4):483–498.

Stacey WC, Litt B. 2008. Technology insight: Neuroengineering and epilepsy—Designing devices for seizure control. Nature Clinical Practice Neurology 4(4):190–201.

Verma N, Shoeb A, Bohorquez J, Dawson J, Guttag J, Chandrakasan AP. 2010. A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system. IEEE Journal of Solid-State Circuits 45:804–816.

Yoo J, Yan L, El-Damak D, Altaf MAB, Shoeb A, Chandrakasan AP. 2013. An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor. IEEE Journal of Solid-State Circuits 48:214–228.