An Ensemble of Hyperdimensional Classifiers: Hardware-Friendly Short-Latency Seizure Detection With Automatic iEEG Electrode Selection

An Ensemble of Hyperdimensional Classifiers: Hardware-Friendly Short-Latency Seizure Detection with Automatic iEEG Electrode Selection

Alessio Burrello, Simone Benatti, Kaspar Schindler, Luca Benini, Abbas Rahimi

***Abstract*—We propose a new algorithm for detecting epileptic seizures. Our algorithm first extracts three features, namely mean amplitude, line length, and local binary patterns that are fed to an ensemble of classifiers using hyperdimensional (HD) computing. These features are embedded into prototype vectors representing ictal (during seizures) and interictal (between seizures) brain states are constructed. These vectors can be computed at different spatial scales ranging from a single electrode up to many electrodes. This flexibility allows our algorithm to identify the electrodes that discriminate best between ictal and interictal brain states. We assess our algorithm on the SWEC-ETHZ iEEG dataset that includes 99 short-time iEEG seizures recorded with 36 to 100 electrodes from 16 drug-resistant epilepsy patients. Using k-fold cross-validation and all electrodes, our algorithm surpasses state-of-the-art algorithms yielding significantly shorter latency (8.81 s vs. 11.57 s) in seizure onset detection, and higher specificity (97.31% vs. 94.84%) and accuracy (96.85% vs. 95.42%). We can further reduce the latency of our algorithm to 3.74 s by allowing a slightly higher percentage of false alarms (2% specificity loss). Using only the top 10% of the electrodes ranked by our algorithm, we still maintain superior latency, sensitivity, and specificity compared to the other algorithms with all the electrodes. We finally demonstrate the suitability of our algorithm to deployment on low-cost embedded hardware platforms, thanks to its robustness to noise/artifacts affecting the signal, its low computational complexity, and the small memory-footprint on a RISC-V microcontroller.**

***Index Terms*—iEEG, hyperdimensional computing, electrode selection, low-latency seizure detection, local binary patterns, symbolic dynamics, line length, ensemble of classifiers**

I. Introduction

Epilepsy is one of the most severe chronic neurological disorders [1], where one-third of patients continue to suffer from seizures despite pharmacological treatment [2]. For these patients affected by drug-resistant epilepsy, efficient algorithms for seizure detection are needed in particular during pre-surgical monitoring [3].

Intracranial electroencephalography (iEEG) currently provides the best spatial resolution and the highest signal-to-noise ratio (SNR) to record the brain activity in patients with epilepsy [4]. Using iEEG signals, recent machine learning methods have been successful in detecting ictal state (i.e., during seizures) and interictal state (between seizures) [5], [6], [7]. Some attempts are even made to forecast the seizure state [8], [9]. Their primary challenge, however, is to reliably detect seizures from a minimal iEEG recording, thus comprising a low number of seizures. Indeed, given the patient-specific nature of seizure dynamics and the asymmetry inherent in epilepsy [10], all the ML algorithms need a patient-tuning phase on its specific iEEG data. This issue is further exacerbated by the relatively limited amount of data that is generally available during one to three weeks in the epilepsy monitoring unit, where recording time is minimized to reduce the risk of infections and patient discomfort [11]. Therefore learning algorithms that generalize well for novel seizures using only a few ictal examples are needed [12].

Besides, conventional learning methods suffer from further limitations, including: (1) Their high computational complexity, long latency, and large memory footprint render them unsuitable for wearable or implantable devices. (2) Their learning outcome is often a "black box" that cannot be analyzed by an expert neurologist to identify electrodes recording from or close to seizure-generating brain regions. (3) Their training process is slow and iterative.

To address these important limitations, in this work, we propose a hardware-friendly algorithm for epileptic seizure detection, based on our initial one-shot learning algorithm [12] that jointly exploits symbolization and brain-inspired hyperdimensional (HD) computing [13]. The algorithm first maps iEEG signals to discrete symbols using local binary patterns (LBP) [14]. Then, it combines them over time and across electrodes to generate a *transparent* representation in the HD space that encodes the state of interest. To further enhance it, this work makes the following contributions:

1) We propose an algorithm based on HD computing that consists of three main parts. The first part detects the occurrence of specific signal features, i.e. high amplitude peaks, fast oscillations, periodicity, LBP, line length, and mean amplitude. The extracted features are then mapped into HD space, where vector-space operations are used to combine them over time and across electrodes for encoding a particular state of the brain. During the training phase, the algorithm stores these learned vectors (one for each state per feature) in an associative memory (AM), while, in the inference phase, unseen states are then classified based on the Hamming distance to the learned vectors in the AM. These form an ensemble of HD classifiers. Finally, to increase algorithm robustness, the distance measurements from these vectors are combined in a linear layer followed by simple postprocessing to compute the output label.
2) We present an extensive analysis and comparison with the state-of-the-art (SoA) methods. Compared to LBP+HD [12] and to the SoA methods [7], [15], [5], [12], [9], [6], our new algorithm shortens the latency of seizure onset detec-

A. Burrello and S.Benatti are with the Department of Electrical, Electronic and Information Engineering, University of Bologna, 40136 Bologna, Italy (e-mail: name.surname@unibo.it).

K. Schindler is with the Sleep-Wake-Epilepsy-Center, Department of Neurology, Inselspital, Bern University Hospital, University Bern, 3010 Bern, Switzerland (e-mail: kaspar.schindler@insel.ch).

L. Benini is with the Department of Information Technology and Electrical Engineering at the ETH Zurich, 8092 Zurich, Switzerland, and also with the Department of Electrical, Electronic and Information Engineering, University of Bologna, 40136 Bologna, Italy (e-mail: lbenini@iis.ee.ethz.ch).

A. Rahimi is with IBM Research-Zurich, CH-8803 Zürich, Switzerland (e-mail: abr@zurich.ibm.com).

Support was received from the Hasler Foundation under project 18082, and the EU's Horizon 2020 Research and Innovation Program through the project MNEMOSENE under Grant 780215.

tion (8.81 s vs. 11.57 s) with a higher specificity (97.31% vs. 94.84%) and a slightly lower sensitivity (96.38% vs. 99.77%).

3) Using our algorithm, we demonstrate how specific signal features help to extract clinically useful information. Our algorithm can automatically rank the most discriminating electrodes by measuring the relative distances between the ictal and interictal prototypes generated solely by each electrode. This capability enables reducing the number of electrodes for classification with almost no degradation in accuracy, thereby shortening the execution time, and to focus on essential electrodes for posterior analysis [16]. With only the top-ranked 10% of the electrodes of each patient, we achieve 95.31% mean accuracy, and 9.21 s latency in detection, and still outperform the SoA methods even when those assess information from all electrodes [17].
4) We demonstrate that the proposed algorithm maintains its performance (95.76% accuracy) when it is affected by external noise and artifacts typically occurring when recording (human) biosignals such as white and flicker noise, muscular, and eye-blinking artifacts. Furthermore, due to its bitwise operations, our algorithm can be executed on ultra-low-power RISC-V platforms such as [18]. This is extremely important to enable its deployment on resource-constrained devices, such as wearable or even implantable systems. For instance, we show that energy per classification can be lowered to 18.5 μJ using 4 electrodes. Other works achieve 802 μJ [19] to 840 μJ [20] in general-purpose microcontrollers, and 2.73 μJ [21] to 168.6 μJ [15] in dedicated ASIC implementations.

The rest of the paper is organized as follows. Section II provides an overview of the literature. Section III introduces the used features and HD computing, together with the proposed algorithm. Section V describes the results, the comparisons with the SoA methods, and the details of algorithm performance (operations, energy, and memory requirements). Section VI analyzes the impact of noise, the hardware implementation, and the electrodes ranking. Section VII concludes the paper.

II. Related Work

Automated detection of seizure onset has been widely explored in the last years. The main challenges that a detection algorithm must tackle are related to the performance in terms of accuracy, specificity, sensitivity, and latency, intended as the delay that occurs between the onset of an epileptic seizure and its detection. Indeed, high sensitivity is linked to accurate detection of seizures, while a high specificity is needed to avoid high rates of false alarms and consequent anxiety in the patients. A latency lower than 20 seconds allows to early eliminate symptoms of the seizures [28].

Machine learning [29] and, more recently, deep learning [16] are well-established methods to decode neural signals and identify seizure onsets. Table I gives an overview of principal works in this wide research field.

In conventional machine learning approaches, such as Support Vector Machine (SVM) [24], Random Forest [25] and Linear Discriminant Analysis [30], the EEG data are preprocessed to extract relevant features both in time and frequency domains. Hence, the feature extracted from the original signal feed the classification algorithm. Random Forest and SVM represent the two most competitive solutions in this category. Random Forests led to the highest rank in the Kaggle competition on the seizure detection [5]. Two different Random Forests coupled with time and frequency features or correlations features achieve 0.981 and 0.984 AUC [5]. In [25], Random Forest is the best performing among KNN, Naive-Bayes, and other classical methods using Empirical Wavelet Transform as preprocessing. Further, Random Forests are widely employed in dedicated hardware implementations, achieving classification energy as low as 41.2 nJ with a slightly lower score (88.1 % specificity, 83.7 % sensitivity) [31], [32].

As another effective classifier, SVM is used in many papers. For instance, the SVM in [23] predicts seizures on 6 channels of intracranial EEG, using the Discrete Wavelet Transform (DWT) on 21 epileptic patients, achieving a sensitivity of 94.46% and a specificity of 95.26% with a false detection rate of 0.58/h, and a low latency of 11.1 s. In [24], the SVM has been used in combination with Local Binary Pattern (LBP) to detect seizures on the single-channel Bonn database. The LBP+SVM requires a 25s window and reaches 99% accuracy with 99.3% and 98.6% sensitivity and specificity respectively, while, in [15], a feature set including spectral features coupled with the SVM is ported on a small chip, achieving 97.7% Sensitivity and 0.185 h^{-1} false alarm rate.

More recently, a remarkable boost in data analysis was provided by the application of deep learning methods, especially based on Convolutional Neural Networks (CNN), Autoencoders, or Long-Short Term memories (LSTM). The main advantage of deep learning algorithms is the ability to automatically learn high-level features from data in time or frequency domain. This eliminates the need for domain expertise in feature extraction. For instance, the work of [6] shows a robust seizure detection based on an LSTM network fed with raw temporal data, which achieves 100% classification accuracy, on a windowed input 16 s long on the Bonn dataset. In a similar vein, CNN-based methods, widely used in image classification, are achieving promising results. To mimic an image-like classification, the CNNs are typically preceded by the Short-Time Fourier Transform (STFT) [16], [27]. In [27], a CNN network with STFT preprocessing is tested using full precision as well as with variable integer quantization, down to 1 bit, reaching up to 0.961 AUC on three datasets with intracranial and scalp electroencephalogram. An autoencoder with the same STFT is used in [16], reaching an accuracy of 94.37%.

Deep learning approaches however require a high-end computational capability for inference and even more for training examples. For these approaches, an embedded implementation often exceeds the power budget available in an implantable or wearable device. On the other hand, classical machine learning approaches are usually tested on individual windows, not reporting latency, hence making it unpractical to evaluate their real-life usage.

The approach we propose in this work combines high performance with low latency in detecting seizures, when compared with the other SoA algorithms (see Table I and Table III for a comparison on the same dataset), and provides a *hardware-friendly* architecture that eases the deployment on real-time resource-constrained chips for wearable/implantable devices. Finally, we report the performance on our dataset on eight algorithms for a fair comparison with our approach. These algorithms are the SoA on different datasets and are either public or shared with us by the authors of the papers.

TABLE I: The state-of-the-art comparison table.
Abbrv.: Sub.: subjects, sEEG: scalp EEG, iEEG: intracranial EEG, SENS: Sensitivity, SPEC: Specificity, FAR: False Alarm Rate, ACC: Accuracy, LAT: Latency, AUC: Area Under Curve.

Work	Dataset	Electrodes	Signal	Freq.	Preprocessing	Algorithm	Metrics
Khanmohammadi et al. [22]	CHB-MIT, 23 Sub.	23	sEEG	256 Hz	PCA-CSP	ADCD	SENS: 96.0% FAR: 0.12 h^{-1} LAT: 4.21 s
Liu et al. [23]	Freiburg, 21 Sub.	6	iEEG	256 Hz	DWT	SVM	SENS: 94.46% SPEC: 95.26% FAR: 0.58 h^{-1} LAT: 11.1 s
Tiwari et al. [24]	Bonn	1	iEEG	173.6 Hz	LBP	SVM	SENS: 99.3% SPEC: 99.6% ACC: 99.0%
Bhattacharyya et al. [25]	CHB-MIT, 23 Sub.	23	sEEG	256 Hz	EWT	RF	SENS: 97.91% SPEC: 99.57% ACC: 99.41%
Zabihi et al. [26]	CHB-MIT, 23 Sub.	23	sEEG	256 Hz	NLD+LDA	ANN	SENS: 91.15% SPEC: 95.16% ACC: 95.11%
Baldassano et al. [5]	Kaggle, 8 Sub.	16-72	iEEG	400 Hz	Time & Freq Correlation	RF RF	AUC: 0.981 AUC: 0.984
O' Leary et al. [15]	EU dataset, 4 Sub.	up to 122	iEEG	n.a.	Spectral	SVM	SENS:.97.7% FAR: 0.185 h^{-1}
Yuan et al. [16]	CHB-MIT, 23 Sub.	23	sEEG	256 Hz	STFT	Autoencoder	ACC: 94.37%
Truong et al. [27]	Freiburg, 21 Sub. CHB_MIT, 23 Sub. Kaggle, 16-72 Sub.	6 23 16-72	iEEG sEEG iEEG	256 Hz 256 Hz 400 Hz	STFT	CNN	AUC: 0.947 AUC: 0.961 AUC: 0.926
Hussein et al. [6]	Bonn	1	iEEG	173.6 Hz	-	LSTM	SENS: 100.0% SPEC: 100.0%
Our Work	SWEC-ETHZ, 16 Sub.	36-100	iEEG	512 Hz	LBP, Ampl., LL	HD	SENS: 96.38% SPEC: 97.31% ACC: 96.85% LAT: 8.81 s

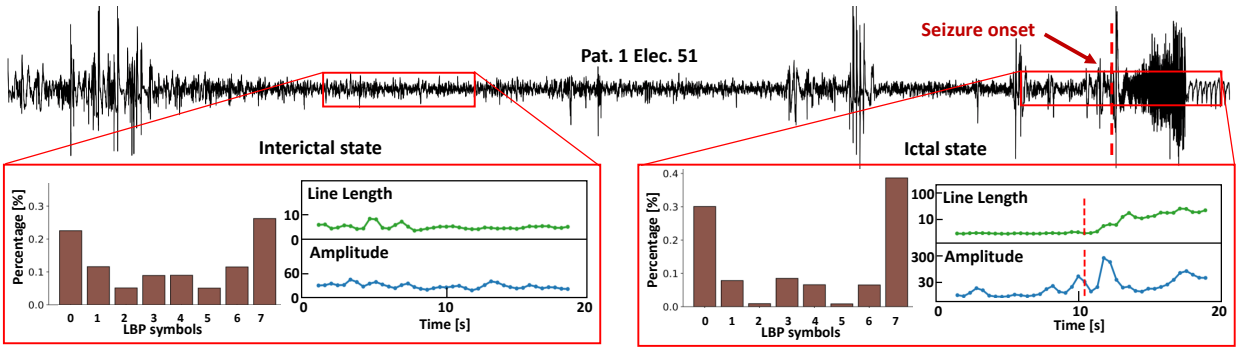

Fig. 1: iEEG signal containing the first seizure from electrode 51 of Patient 1. Red panels show the three features extracted (brown=LBP, green=line length, and blue=amplitude) during interictal and ictal states. The value in the LBP histogram represents the integer corresponding to the string of three concatenated bits.

III. Background

In this section, we describe the three features used in our algorithm: line length (Section III-A1), mean amplitude (Section III-A2), and local binary patterns (Section III-A3). All the features are computed over time windows of 1 s with 0.5 s overlap. This temporal interval demonstrated the best performance throughout an extensive experiment if compared to wider windows or larger shifts – lower overlap (window $\in$ [1.0 s, 3.0 s], shift $\in$ [0.5 s, 3.0 s]). On the other hand, increasing the shift proportionally reduces energy consumption. Furthermore, we present a method for embedding them homogeneously into an HD space that promotes simple learning and shortens the latency of seizure detection.

A. Feature Extraction

1) Line Length: Line length, also known as total variation and—up to a scaling factor—identical to absolute slope [33], is the total length of the iEEG signal in a given time window. Line length captures well large signal amplitudes and high-frequency oscillation, which both are typical characteristics of epileptiform iEEG signals while being linear in time of execution with the number of channels. Importantly, line length outperforms other often used features such as zero-crossings [34]. Line length can be efficiently computed by accumulating over time the absolute difference of consecutive iEEG sampling points, thereby requiring only the last two EEG samples:

$$ll = \frac{1}{N}\sum_{t=1}^{N}(|EEG_t - EEG_{t-1}|). \quad (1)$$

Figure 1 shows the difference of line length between interictal (left) and ictal (right) states. As shown, growing values of line length clearly indicate the seizure onset.

2) Mean Amplitude: Mean amplitude is defined as the mean value of the peaks of a time series within a window. Since high amplitude slow oscillations are recorded at the onset of

some seizures [5], this feature is effective for detection. Other similar features also reflect the same signal characteristic, e.g., the energy of the signal or the area detection tool of [35], however demanding significantly greater computational efforts. Moreover, as for line length, we do not need to store all the EEG samples of the time window. Figure 1 shows that mean amplitude can also discriminate between ictal and interictal states.

Algorithm 1 Mean amplitude extraction

Input: window's dimension: N, $EEG_1, \ldots, EEG_N$.
mean $= 0$, npeaks $= 0$
for $t \leftarrow 3, \ldots, N$ **do**
  if $EEG_{t-2} < EEG_{t-1}$ and $EEG_{t-1} > EEG_t$ **then**
    mean $+=$ EEG_{t-1}
    npeaks $+=$ 1
  end if
end for
mean $=$ mean $/$ npeaks
Output: $mean$

3) Local Binary Pattern (LBP): Local binary patterns (LBP) belong to a class of data-analysis methods known as symbolization, which consists of transforming raw continuous time series into a finite alphabet of symbols. LBP can efficiently map iEEG samples into symbols, using bit strings that represent the sign of the first derivative of the iEEG signal, i.e., one-dimensional LBP [14], [36]. Note that a LBP code reflects the short-term relation between consecutive samples of a temporal series, namely if their derivative is positive or negative. Prior studies [7] have demonstrated that the distribution of LBPs is more informative than their sequences. Algorithm 2 shows how to compute the LBP symbols for a time window.

Algorithm 2 LBP extraction

Input: $\dim_{window}$: N, $\dim_{LBP}$: l, $EEG_1, \ldots, EEG_N$.
for $t \leftarrow l, \ldots, N$ **do**
  $\delta = EEG_{t-l+1:t} - EEG_{t-l:t-1}$
  $\text{LBP}_t = concat(\ int(\delta > 0))$
end for
Output: LBP

Figure 1 shows the difference between the LBP histogram of the interictal (left) and ictal (right) states. In the interictal state, the LBP codes are distributed more equally, while in the ictal state, there are two LBP codes with relatively higher probability than the other codes; this is mainly due to the slow and asymmetric oscillations during seizures [37], [38].

B. Pattern Recognition Algorithm: Hyperdimensional (HD) Computing

Inspired by the very large size of the brain's circuits, hyperdimensional (HD) computing [13] is all about computing with random vectors of very high dimensionality (d), e.g. $d \geq 1000$. In this formalism, the vectors can be further combined or compared, with a well-defined set of vector-space operations implementing the task of interest. This computational paradigm provides a novel perspective on data representations and associated operations with unique features in terms of robustness [39], [40], speed of learning [41], [42], [12], and energy efficiency [43], for a wide range of biosignal applications, e.g., EMG, EEG, or in general, ExG processing [41], [43], [44] .

We consider d-dimensional binary random vectors of equally probable 1s and 0s, i.e., dense binary elements of $\{0,1\}^d$ known as binary spatter code [45], where random vectors are nearly orthogonals [46]. HD computing begins by generating a vector space, composed of orthogonal hypervectors, to remap given data into hypervectors. In our seizure detection scenario, LBP codes and input channels are stored in an *item memory* (IM), while their values are mapped via a set of hypervectors correlated with their amplitude, namely *continuous* item memory (CIM) [41]. In this continuous vector space of CIM, orthogonal endpoint vectors are generated for the minimum and maximum input levels. Vectors for intermediate levels are then generated by linear interpolation between these endpoints so that the Hamming distance of vectors corresponds to the closeness of levels.

The output vector of the encoder is then fed into an associative memory (AM) for training and inference. Interested readers can refer to [17] for more details.

We use HD computing for learning and classification of ictal vs. interictal states in this paper. We apply four main steps after the feature extraction: 1) embedding discrete and analog features to the HD space via a combination of IM and CIM; 2) combining the mapped vectors with the multiplication/addition operations inside an *encoder* to produce a new vector representing the state of interest; 3) adding those vectors from the same state to create a *prototype* vector representing the entire category; 4) finally comparing the prototype vectors with a query vector to assign the query vector to the best matching category (i.e., inference/classification).

IV. Methods

In this section, we present the main contribution of the paper. We show how three different features can be used by HD computing to learn seizures from few examples of iEEG recordings efficiently. We extend previous HD algorithms in biosignal analysis [12], [43] by including more feature extractors in its structure.

A. Mixture of Features and Binary HD Computing

An ensemble of classifiers is formed, each operating on a feature that is combined by a linear layer. First, three features are extracted from the iEEG signal of each electrode. Then d-dimensional binary representations are constructed to capture the evolution of the features across all electrodes and over time. Finally, a linear layer is trained with the output of the HD classifiers and followed by simple postprocessing. The three main blocks of our algorithm are shown in Figure 2 (feature extraction+HD computing), and Figure 3 (final linear layer and postprocessing). This approach is computationally efficient and enables fast learning from one or a few examples of seizure recordings per patient. Our dataset (Section V) is already filtered and downsampled. Thus no additional preprocessing is applied. Figure 2 shows the three different feature extractors applied to the iEEG input signals. As described in Section III-A, both amplitude and line length are parameter-free. On the other hand, we assess different values of LBP code lengths, and we fix $l = 3$: the accuracy is comparable for all values of l in the range $[3, 10]$. We have chosen the minimum value of l to reduce the number of resulting symbols (2^l). The duration of the window, namely the number of samples on which we compute the features, is as short as

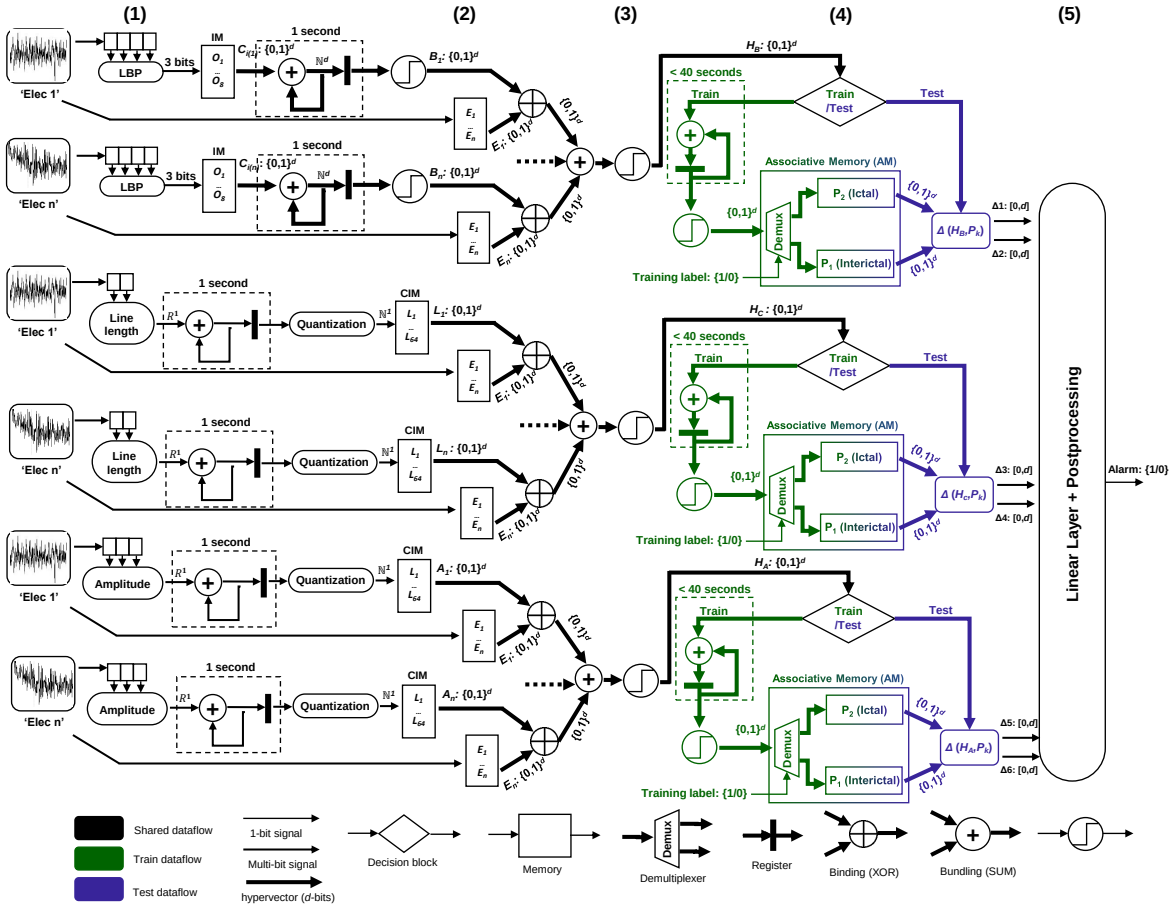

Fig. 2: The processing chain of our algorithm: (1) LBP, line length, and mean amplitude feature extractors; (2) Embedding of features in the HD space; (3) HD operations to compute H_A, H_B, H_C vectors that encode the state of interest; (4) Learning and inference using three associative memories; (5) Combining Hamming distances using a linear layer followed by postprocessing. Linear layers training is based on the AM outputs (see Fig. 3 for its details).

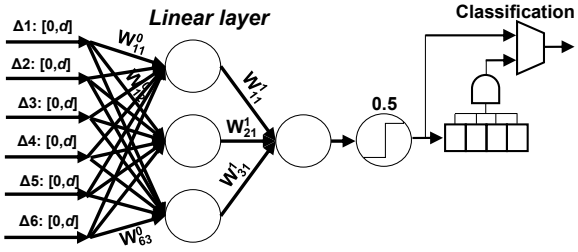

Fig. 3: Linear layer and postprocessing used in our algorithm. Three neurons are trained to combine the distances from the HD classifiers followed by a four samples checker for classification.

1 s to minimize the latency of classification while ensuring a low computational cost.

B. Learning and Classification using HD computing

We use HD computing for fast learning and classifying from the three extracted features. For every feature, we train a HD model containing an associative memory with two prototypes. This has been done by first embedding the features in the HD space that is tailored to each specific feature, whereas the structure of associative memory is the same for every feature.

HD computing first embeds the features in the HD space via the IM or CIM, by assigning a binary vector of dimension $d = 10,000$ to different entities. For the first two features, i.e., line length and mean amplitude, our embedding starts by quantizing their continuous values to 64 discrete levels with equal intervals. Then, a CIM assigns a vector to each different level such that orthogonal endpoint vectors are generated for the minimum and maximum levels in the range. Other vectors for intermediate levels are then generated by linear interpolation between these endpoints so that the similarity of vectors corresponds to the closeness of levels. Figure 2 details the extraction of features and their embedding and shows how the H vector is computed per feature and used for training and computing the distances.

To combine these mapped vectors across all electrodes without interference, we generate a spatial record (H), in which the *name* of each electrode is a field, and the feature is the *value*. To generate these fields, the IM maps the name of n electrodes to orthogonal binary vectors: $E_1 \perp E_2 \ldots \perp E_n$. This orthogonal mapping eases binding ($\oplus$) the name electrodes ($E_j \mid j \in [1,n]$) to their corresponding feature values. For instance, for the LBP features, it leads to: $H_B = [(B_1 \oplus E_1) + \cdots + (B_n \oplus E_n)]$. Similarly, for the line length features: $H_C = [\sum_{j=1}^{n} L_{j(i)} \oplus E_j] \mid i \in [1,64]$; as well as for the amplitude features: $H_A = [\sum_{j=1}^{n} A_{j(i)} \oplus E_j] \mid i \in [1,64]$. All three H_A, H_B, and H_C vectors are generated in parallel

every 0.5 seconds, and together encode the state of the brain at a specific time. Each vector is used to train a related associative memory (AM) containing two *prototype* vectors representing ictal and interictal states. For instance, all H_A vectors computed over an interictal window of 40 s are bundled to be stored in the related AM as a binary prototype vector representing the interictal state. This bundling operation (i.e., addition followed by normalization) can be applied online for new training examples to apply incremental updates in the prototype vector [47]. Similarly, an ictal prototype vector is generated from a smaller window of 10–30 s depending on seizure duration.

The output of the HD classifier for an unseen 1 s window is the comparison of its H vectors to the related two prototypes of the AM using the Hamming distance metric. Intuitively, lower distances from the ictal prototypes imply the beginning of a seizure, whereas very short distances from the interictal prototypes imply the absence of ictal activity. Note that both the H vector and the two prototypes are different for each feature, thereby producing six distances (two per feature) to be fed into the linear layer for final labeling.

C. Linear Layer and Postprocessing

As the last part of the algorithm, we combine all the distances produced by the three HD classifiers using a linear layer that generates the final label. We use one hidden layer with 3 neurons and ReLu activation and a single neuron with sigmoid activation as output, to bound the classification label in the range [0,1] (Figure 3). Adam optimizer, lr=0.01, lr_{decay}=0.01, step_decay=200 epochs and a total of 800 epochs complete the list of parameters to build the layer. Note that the linear layer is trained after the HD classifiers, and takes as training inputs the 6 distances produced for every 1 s window. After training, this layer learns how to combine the six input distances of the unlabeled time window to generate the correct label.

Finally, to reduce the latency, we use a short postprocessing window: only a small four-samples checker is used in patients with ideal training sensitivity (100%). The checker entails a window of 2 s where a seizure flag is raised when all four consecutive labels are ictal. This postprocessing window is removed for patients where we can not correctly detect all the trained seizures (P3, P7, P9, P10, P12) to avoid low sensitivity. The window is removed prior to any further analysis and statistics, thus influencing specificity (lower w/o the postprocessing), sensitivity (higher), and latency (lower).

V. Experimental Results

A. The SWEC-ETHZ iEEG Database

We use the short-term SWEC-ETHZ iEEG database [12], [17] that comprises 16 patients (P1–P16) with a total of 99 seizure recordings[1]. The clinical data and a detailed description of the dataset can be found in [17]. For each patient, the number of seizures varies from 2 to 14. Each seizure is stored in a separate file with 3 minutes of interictal state preceding it and 3 minutes of postictal period immediately following. The number of electrodes and their implantation scheme was derived entirely based on clinical necessity. The iEEG signals provided in the dataset were recorded by either implanted strip, grid, and depth electrodes or by a combination of these different electrode types. The iEEG recordings were either sampled at 512 or 1024 Hz, depending on whether the implantation scheme comprises more or less than 64 electrodes. All recordings were then down-sampled to 512 Hz prior to further analysis. The iEEG signals were re-referenced against the median of all the electrodes free of permanent artifacts (e.g. the line noise at 50 Hz) as judged by visual inspection [48], [49]. The data were converted using a 16-bit ADC and band-pass filtered between 0.5 and 150 Hz using a fourth-order Butterworth filter. An EEG board-certified and experienced epileptologist (K.S.) visually inspected all the iEEG time series for the exclusion of electrodes, always corrupted by artifacts and seizure onset marking.

[1]This dataset is publicly available at http://ieeg-swez.ethz.ch.

TABLE II: Performance of our algorithm compared to SoA algorithms (LBP+HD) [12] and Mixture+SVM (Three SVM classifiers, each per feature type, fused by a linear layer, and followed by the postprocessing). The table shows the results in terms of latency, sensitivity, and specificity across all the patients.
Abbreviations: ℓ: functional latency of seizure onset detection, Spc.: specificity, Sen.: sensitivity, M.: mean, and Acc: macro averaging accuracy.

ID	Our algorithm			LBP+HD [12]			Mixture+SVM		
	ℓ [s]	Spc. [%]	Sen. [%]	ℓ [s]	Spc. [%]	Sen. [%]	ℓ [s]	Spc. [%]	Sen. [%]
P1	3.8	100	100	6.3	100	100	2.0	99.20	100
P2	10.4	100	100	15.1	100	100	14.0	100	83.30
P3	4.1	96.14	86.43	21.8	79.97	91.03	0.0	90.30	58.30
P4	24.9	100	100	34.5	100	100	26.0	100	100
P5	19.3	99.42	100	20.9	100	100	19.0	100	100
P6	8.0	100	100	6.3	100	100	9.0	100	100
P7	0.5	89.27	90.00	5.0	49.90	88.57	0.0	88.30	75.0
P8	12.9	100	100	13.2	100	100	10.0	100	100
P9	0.0	88.33	100	16.2	96.31	96.43	6.0	99.40	100
P10	0.8	97.44	94.52	3.9	98.41	94.41	0.0	97.53	96.40
P11	2.0	100	100	7.0	100	100	10.0	100	100
P12	0.0	93.22	95.00	15.9	96.88	80	9.0	96.20	80.0
P13	7.0	100	100	10.0	100	100	10.0	100	100
P14	7.9	99.53	76.17	10.5	95.94	85.71	9.0	99.54	90.50
P15	22.6	93.91	100	36.4	100	100	10.0	100	100
P16	17.3	100	100	32.3	100	100	10.0	100	100
M.	**8.81**	97.31	96.38	15.96	94.84	96.01	9.94	98.15	92.72
Acc	**96.85**			95.42			95.43		

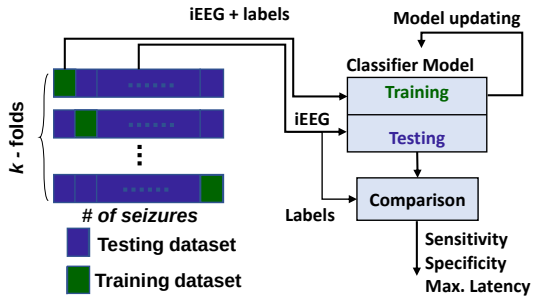

Fig. 4: k - fold validation procedure proposed in this work.

B. Seizure Onset Detection

To evaluate the performance of our algorithm, we measure the following metrics given a limited number of training seizure examples:
(1) Sensitivity is computed as the percentage of correctly detected seizures in the test dataset:

$$Sens_{Pi} = \frac{\sum_{i=1}^{N}(Seiz_i == ictal)}{N}$$

TABLE III: Comparison with SoA algorithms with added postprocessing. Abbreviations: Latency: maximum latency of seizure onset detection, Sen.: sensitivity, Spc.: specificity, Accuracy: macro averaging accuracy, and OPS: number of operations.

Work	Algorithms	Sen. [%]	Spc. [%]	Accuracy [%]	Latency [s]	Memory [KB]	Arithmetic OPS [M]
[This work]	Mixture + Ensemble HD	96.38	97.31	**96.85**	**8.81**	**17.8 (1×)**	32.8 (1×)
[This work]	Mixture + SVM	92.72	**98.15**	95.43	9.94	89.6 (5.0×)	**4.2** (0.13×)
Burrello et al, 2019 [12]	LBP + HD	96.01	94.84	95.42	15.96	16.3 (0.9×)	31.9 (0.98×)
Jaiswal et al, 2017 [7]	LBP + SVM	96.00	92.74	94.37	14.4	1163.2 (65.3×)	7.4 (0.23×)
G. O'Leary et al, 2018 [15]	Spectral + SVM	**99.77**	89.31	94.54	11.57	34.6 (1.9×)	1.4 (0.05×)
Jaiswal et al, 2017 [7]	LGP+MLP	97.11	92.37	94.74	15.4	655.5 (36.8×)	4.1 (0.13×)
Baldassano et al, 2017 [5]	Time & Frequency + RF	95.93	90.81	93.37	11.6	56.2 (3.2×)	44.4 (1.4×)
Baldassano et al, 2017 [5]	Correlations + RF	95.89	89.06	92.48	11.9	178.7 (10.0×)	69.9 (2.1×)
Truong et al, 2018 [9]	STFT + CNN	97.73	83.61	90.67	17.9	6753.3 (379.4×)	232.8 (7.1×)
Hussein et al, 2018 [6]	LSTM	95.14	94.77	94.96	14.7	185.4 (10.4×)	32771 (999×)

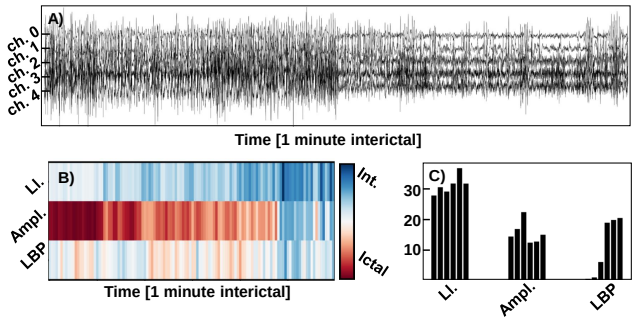

Fig. 5: Panel A. One minute of interictal signal from 5 channels of the seizure onset zone of seizure 4 of P3. **Panel B**. $\Delta_{Ict.} - \Delta_{Int.}$ for each feature. Higher values imply higher reliability of detecting interictal outcome solely by that feature. **Panel C**. Distribution of the weights in the linear layer. For each feature, we show the two relative weights of each of the three neurons.

where N is the number of seizures in the test dataset and $Seiz_i$ the label of the algorithm for a seizure.
(2) Specificity is defined as the percentage of correctly classified interictal 1 s windows:

$$Spec_{Pi} = \frac{\sum_{i=1}^{M}(Window_i == interictal)}{M}$$

where $window_i$, the label provided by the algorithm and M is the number of interictal windows in the test set.
(3) Macro averaging accuracy is defined as the mean of sensitivity and specificity:

$$Accuracy_{Pi} = \frac{Sens_{Pi} + Spe_{Pi}}{2}$$

As opposed to the micro averaging accuracy (number of windows correctly labeled divided by the total number of windows), it equally weights the two classes, i.e. ictal and interictal.
(4) Latency is measured as the time that an algorithm takes to classify an unseen seizure after the seizure onset time marked by the expert:

$$l = \frac{\sum_{i=1}^{N}(t_{detection} - t_{SOE})}{N}$$

where t_{SOE} is the time of seizure onset marked by the expert. Importantly, this is the maximum latency due to the non-ideal classification (i.e. the amount of time used to flag a seizure after its onset with the algorithm sensitivity). On the other hand, the minimum latency of our algorithm corresponds to the input window shift (0.5 s) plus its execution time ($<$ 350 ms on a simple optimized RISCY-based System on Chip (SoC) [18]), comparable to other seizure detectors that reported latency of $<$1 s, e.g. [15], [50], [51], [21]. In addition, we statistically test all our results using a t-test, with a level of significance, p-value, of 0.05.

Based on these criteria, we compare our new algorithm to its previous version (SoA on the SWEC-ETHZ iEEG database) in Table II. We choose the number of trained seizures (TrS) equal to our previous studies [12], [17] to have a fair comparison. [2]

We also use the same k-fold cross-validation scheme: k has been chosen as the total number of seizures minus the number of seizures in the training set (TrS) plus one [17]. We then rotate the trained seizures among all available seizures. The training data is used first to train HD classifiers, and then the linear layer and postprocessing (see Fig. 2). Fig. 4 showcases the k-fold cross-validation scheme.

As shown in Table II, our algorithm achieves 97.31% mean specificity and 96.38% sensitivity, surpassing LBP+HD algorithm by a more than 1% higher accuracy. Importantly, our algorithm significantly lowers the average latency to 8.8 s (15.96 s in LBP+HD, $p < 0.00001$). By including more features and their proper combination in HD computing, we simultaneously increase both sensitivity and specificity, while significantly reducing the latency of detection.

Figure 5 highlights the advantages of using more features fused by our patient-specific trained linear layer. Panel A shows one minute of the interictal state from P3, which exhibits a very low specificity of 79.97–81.33%, using only the LBP feature with both HD computing and SVM classifier [17]. As shown in Panel B, the LBP feature often fails in detecting the correct interictal label. On the other hand, the line length, which captures the frequency trace of the iEEG signal, correctly identifies the interictal state, demonstrating for this patient that the frequency features are the most informative. Panel C shows how the linear layer can automatically draw this conclusion, by strongly weighting the line length feature for the final prediction for P3, while giving relatively lower weights to both amplitude and LBPs. Therefore, our algorithm weights the most important feature for each patient by allowing (i) to maintain the same high precision for patients for which the LBP is reliable and (ii) to increase the precision for the patients for which the other features result to be more informative.

Thanks to the high reliability of the classifier, we can reduce the postprocessing window without impairing the specificity, but reducing the maximum latency. A maximum latency $\leq$ 10 s is sufficient for several applications because the electrical onset often precedes the clinical onset by more than 20 s [28]. For instance, it may improve diagnostic procedures [52] such as ictal

[2]Compared to the training scheme proposed in [12], [17], training all the patients with a single seizure leads to a significant reduction of the specificity by 5.72 % (p-value < 0.01) while maintaining a constant latency.

single-photon-emission-computer-tomography (SPECT) [53], and control of seizure activity by electrical stimulation [54]. Moreover, considering that the median duration of mesial temporal lobe seizures and neocortical extratemporal seizures are 106 s, and 78 s, respectively [55], it may help to prevent further spreading and secondary generalization [56], [57], [28], [58]. Furthermore, by removing the postprocessing window, we can reduce the delay of our algorithm to 3.74 s, enabling earlier interventions, at the cost of 2% lower specificity and not having any patient with 100% accuracy.

1) State-of-the-art (SoA) Comparisons: We also compare our algorithm with other SoA methods [7], [15], [5], [12], including deep learning [9], [6]:

Mixture + SVM. To assess the net effect of our HD method, we replicate the algorithm in Fig. 2, with the same features and postprocessing but with three SVM classifiers (same settings of [7]) instead of three HD classifiers. In this setting, the linear layer takes as inputs the three scores and the three classification outcomes of SVM models, each trained with different features.

LBP + SVM. The histogram of LBPs with l=6 is used as input to feed a linear support vector machine; more details are given in [7].

Spectral + SVM. The energies from five frequency bands are first passed to an exponential decaying memory (EDM) with five coefficients (from 2 to 16) and finally to the SVM classifier; for more information on the setup refer to [15].

LGB + MLP. Akin to LBP, local gradient patterns (LGP) are proposed together with a MLP neural network that outperforms LBP+SVM in [7].

Time & Frequency / Correlations + Random Forest. Both algorithms feed a wide range of features in the time/frequency domain (derivatives, correlations, etc.) to a 1000-tree and a 3000-tree forest, respectively. The codes are taken from the first, and the third-ranked Kaggle competition on seizure detection available online [5].

LSTM. We feed raw iEEG samples to a long-short term memory (LSTM) network. The LSTM network and its variations are now the SoA algorithm in many temporal-series based tasks. For our comparison, we consider the model in [6]. This method is the leader in seizure detection on the Bonn Dataset [59], outperforming in terms of accuracy other algorithms such as [60]. The code used is provided by the authors.

STFT + CNN. We also consider a convolutional neural network (CNN) coupled with a short-time Fourier transform (STFT) used for seizure prediction [9] and detection [27]. The code used is provided by the authors.

Table III summarises all the metrics as well as memory and number of operations (OPS). To have an identical setup, we also consider our postprocessing steps for all other methods. On average, our algorithm outperforms the methods mentioned above in accuracy and latency while using less memory. More precisely, our algorithm surpasses all methods in terms of accuracy (96.85% vs. 94.96%, $p < 0.05$, highest LSTM) and latency (8.81 s vs. 11.57 s, $p < 0.05$, lowest Spectral + SVM). Note that we have already compared the performance of the LBP+HD algorithm for long-time seizure detection on a subset of these methods in [61].

In addition, Table II highlights that our algorithm achieves higher accuracy (96.85% vs. 95.44%, $p < 0.05$) and lower latency (8.8 s vs. 9.9 s, $p < 0.05$) than the SVM method fed with the same features (Mixture+SVM). Mixture+SVM achieves the second-best latency among all methods. This indicates the suitability of our features, linear layer, and postprocessing that improve the performance of both HD and SVM classifiers compared to their counterparts in Table III. Finally, we also compare our algorithm to the other algorithms by removing the postprocessing window to evaluate whether some algorithms benefit more from its usage. We notice that all the algorithms achieve a negligible gain in sensitivity ($<$ 0.5%), while the specificity loss ranges from 2% to 8%. On the other hand, the latency reduction varies from 2.5 s to 6.3 s. With this configuration our algorithm outperforms all the algorithms in specificity (best 95.73%) and delay (3.74 s), while reaching a slightly lower sensitivity (97.13% vs 99.89% of [15]).

TABLE IV: Impact of different type of noise and/or artifacts on the specificity and sensitivity of our algorithm. Abbreviations: Sen.: sensitivity, Spc.: specificity, and SNR: signal-to-noise ratio

Spc.,Sen.	SNR [dB]				
	40	30	20	10	0
Gaussian	95.6, 95.2	91.3, 95.2	70.0, 91.4	40.1, 87.9	37.2, 75.6
Flicker	96.6, 95.9	96.2, 95.4	80.5, 93.7	62.6, 90.7	36.4, 84.1
Muscle	96.0, 95.2	91.3, 95.3	70.4, 91.9	39.7, 88.0	37.1, 76.4
Eye-blink	95.6, 96.1	90.8, 95.0	68.7, 91.1	40.1, 88.6	39.3, 81.1
All	95.9, 95.6	94.2, 94.5	75.2/92.5	46.2, 89.2	32.4, 78.6

VI. Discussion

To better assess whether a seizure detection algorithm is suitable for a real-life scenario, we analyze the performance degradation of our algorithm under noisy conditions. We also profile the execution of our algorithm with regard to the SoA implementations.

A. Impact of Noise and Artifacts

Several noise sources and artifacts can affect the iEEG signals, impairing the performance of an algorithm. We simulate this behavior by considering many noise types and sweeping their amplitudes. In particular, we consider the noise sources shown in [6], [62], as well as muscle, and eye-blinking artifacts. Moreover, we consider flicker noise, as suggested in [63], [64], to take into account the noise of the electronic system. The two artifact types and the white noise are simulated as suggested in [6], while flicker noise is generated using the *colorednoise* library of python based on the algorithms in [65]. Table IV shows the impact of various noisy conditions on specificity and sensitivity values. With the overlapping of all the noise types at SNR=40 dB, the algorithm yet achieves a mean accuracy of 95.76%, outperforming all the other methods. Further, adding all the artifacts and noises individually or overlapped at SNR=30 dB, the algorithm maintains an accuracy > 92.88 %. This high accuracy confirms that also during a real-life execution, while different sources of errors could impair the signal recording chain, the algorithm is yet reliable.

We also verify the performance of our algorithm while changing the effective number of bits (ENOB). Since we do not have access to the raw ADC data, we simulate the ENOB by considering the ADC range equal to the highest iEEG signal in the available recordings (i.e. during the seizure) and adding noise following the formula $ENOB = {}^{(SNR-1.76)}/_{6.02}$, where the SNR is computed using a sinusoidal wave with an amplitude equal to the range of the ADC.

Figure 6 shows the performance at different ENOB. Using ENOB greater than 11-bit does not further improve the accuracy, thus allowing the recording system to relax the quality of the signal constraints needed to have reliable results.

TABLE V: Comparison with embedded implementations.

Metrics	Manzouri et al [20]	Heller et al [19]	This work
Alg.	Rand. Forest	CNN	HD
MCU	MSP430FR5994	MSP430FR serie	Quentin [18]
Max. fr.	25 MHz	8 Mhz	670 MHz
# of Elec.	Variable	4	4 / 64
Memory	256 KB	18.8 KB	10.3 / 17.8 KB
Ex. cycles	N.R.	4.4 M	2.1/33.1 M
Energy	210 μJ x Elec.	802 μJ	18.5 / 287.9 μJ [1]

[1] Best power density @ 0.52V, 187 MHz.

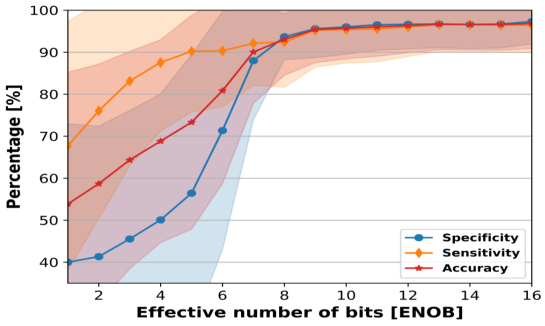

Fig. 6: sensitivity, specificity, and accuracy with varying effective number of bits. Light-colored bands indicate mean±standard deviation.

B. Hardware Requirements Analysis

We compare our algorithm with the SoA methods, both in terms of theoretical memory/computation trade-off and of energy/cycles on a general-purpose micro-controller implementation. We use hypervectors of $d = 1,000$ that have maintained the same performance for most iEEG recordings.

The number of operations and memory occupation of our algorithm and SoA methods is reported in Table III. Our algorithm shows the lowest memory consumption thanks to the employment of 1-bits vectors, saving 1.9× memory compared to the SVM classification of [15]. The memory footprint includes both the IMs and the AMs. From the computational point of view, without any optimizations, our algorithm has comparable complexity to conventional machine learning methods (upper part of Table III), ranging from 20× more to 2× fewer operations, while it outperforms deep learning methods by at least 7.1×. Most importantly, it respects the real-time constraint of 0.5 s when running at a minimum frequency of 70 MHz.

Noteworthy, while the expensive float32 MAC operations are predominant in the workload of other methods, our algorithm spends 98% of its workload doing 1-bit operations (XOR, and popcount). This workload distribution allows for extensive hardware optimizations to potentially accelerate our algorithm up to 1,000 times, due to the inherent parallelism of the HD and the 1-bit operations. In [61], we demonstrated that a GPU implementation renders our algorithm faster than an SVM based approach, by exploiting the *__ballot_sync* operation, used to make a 32x32 1-bit transpose operation in a single instruction and thereby saving up to 1,000 cycles for the HD bundling operation. Similarly, 1-bit low-power hardware accelerators have already been proposed for quantized convolutional neural networks demonstrating speed increases proportional to the bandwidth of the accelerator [66].

Figure 7 shows the computation–memory trade-off between

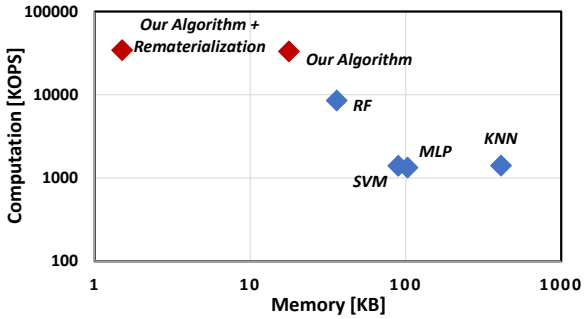

Fig. 7: Computation versus memory for different classifiers. We consider the same set-up as our proposed algorithm, i.e. ∼80 s training time and with the three extracted features.

the most popular machine learning classifier with identical features (Line length, amplitude, and LBPs). While SVM achieves the lowest number of operations to execute, our algorithm demonstrates the smallest memory footprint, allowing us to save energy and area. In the graph, we also show our algorithm enhanced with the *rematerialization* of the IMs (see [47] for more details). The memory-hungry IMs can be efficiently replaced by a cellular automaton working with a random seed with a small cycle overhead; hence, there is no need to store the IM [47]. As a consequence, our algorithm requires storing only the contents of the AMs, $2 \times d$ bits for the two prototypes of each of the three AMs, achieving a 24×/60× memory reduction compared to the RF/SVM.

Finally, we also benchmark our algorithm on a general-purpose RISC-V SoC (Quentin) [18]. The algorithm is implemented using the RISC-V instruction set, exploiting the 1-bit manipulating operations (*bitinsert* and *bitextractu*), the *bitwise XOR*, and the *popcount* operation to minimize the execution cycles. In Table V, we compare our algorithm to a random forest [20] and a convolutional neural network embedded on an MSP430FR MCU [19]. Our algorithm shows a memory occupation of 10.3 KB, 2.1× lower than the best competitor, and 2.1 M execution cycles using 4 electrodes, 1.8× lower when compared to [19]. Besides, due to the ultra-low energy Quentin SoC, our algorithm saves 43.3× energy compared to both methods [19], [20]. Note that our hardware platform is a general-purpose programmable processor that cannot be compared with specialized ASIC/FPGA implementations [21], [15], [50], [51].

However, we highlight that the energy consumption of our method, while running on a programmable core, is already comparable to ASIC implementations which range from from 2.73 μJ [21] to 168.6 μJ [50] per classification, suggesting that a dedicated hardware implementation has the potential to outperform all other methods.

C. Identification of Discriminative Electrodes

In this section, we describe the analyzability of the HD algorithm and automatic electrode ranking through HD scoring. Our proposed algorithm is inspired by the seizure onset detection algorithm (Section IV) followed by statistical ranking using individual electrodes to measure how important they are for discriminating between ictal and interictal states.

Contrary to the detection algorithm, we generate the prototypes for only one electrode. The flow is identical to the detection algorithm, but with the spatial record containing a

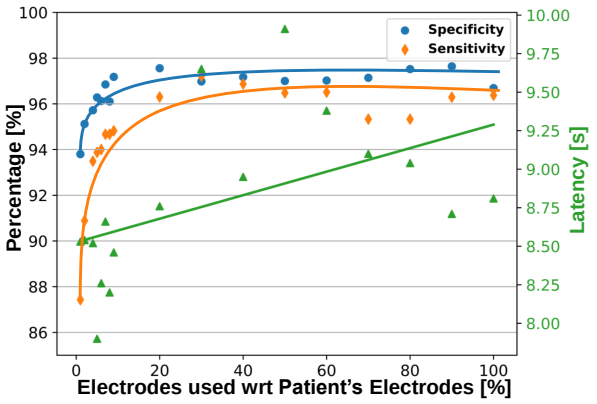

Fig. 8: Evaluation of specificity, sensitivity, and latency of detection as a function of number of electrodes employed in classification. On the x-axis, we report the percentage of patient's electrodes employed for classification.

single electrode of interest. For instance, for the first electrode and amplitude feature, the related H is computed as: $H^1_A = [\sum A_{1(i)} \oplus E_1] \mid i \in [1, 64]$. Similarly, H^1_B and H^1_C are computed. After training the prototypes, we query the algorithm with a number of $1\,s$ segments, and record the output of the linear layer. This experiment is repeated n times, each time with a different electrode. This produces a set of tuple (x^i_t, l_t), where $x^i_t \in [0, 1]$ as the output of linear layer represents the probability of seizure for segment t from electrode i, and $l_t \in \{0, 1\}$ is the ground-truth label of segment t. For instance, $x^i_t > 0.99$ implies a high confidence of detecting a seizure by electrode i, while $x^i_t < 0.01$ reflects a high confidence of detecting an interictal state.

In the second step, we apply a Fisher score based ranking algorithm to rank the electrodes in descending order. The algorithm takes as input the sets of tuples generated for each electrode in the previous step and associates a ranking to each electrode:

$$x^i_{rank} = Fisher\_Score([(x^i_1, l_1) \dots (x^i_T, l_T)])$$

with $1 \dots T$ indexes of the classified segments and i index of the electrode. The larger x^i_{rank}, the better is the electrode i for discriminating between interictal/ictal states. Figure 8 shows the average behavior of specificity (blue), sensitivity (orange), and latency of detection (green) by reducing the percentage of electrodes employed for classification for all patients. Note that decreasing the number of electrodes to the extreme of one electrode, slightly decreases the latency of detection by $\approx$1 s. On the contrary, both sensitivity and specificity are profoundly affected when lowering the percentage of considered electrodes to less than 10%, decreasing respectively of $\approx$8% and $\approx$3% compared to the classification with all the electrodes. In particular, we can observe a knee around 10% of employed electrodes. Indeed, yet reducing the number of classification electrodes under this threshold causes the algorithm not to consider any electrodes coming from the seizure onset zone (SOZ) for three out of sixteen patients, impairing the mean sensitivity/specificity.

Overall, using only the top 10% of the electrodes ranked by our algorithm for all the patients, we achieve 95.98% mean accuracy, yet outperforming the other methods [7] using all the electrodes. Moreover, the electrodes in the SOZ are listed in the top 44$\pm$24% electrodes ranked by our algorithm.

VII. Conclusion

We present an efficient algorithm that shortens the latency of seizure detection, which is essential to improve both diagnostic and therapeutic procedures. The algorithm exploits simple iEEG features and binary operations during training and inference using brain-inspired HD computing. The algorithm's learning procedure is transparent and allows us to rank the electrodes according to their importance for seizure detection. Therefore, the most informative electrodes may be automatically selected to reduce the complexity of execution while maintaining almost the same accuracy. The algorithm further provides a universal and scalable interface that allows assessing iEEG recordings from patients with different numbers of implanted electrodes (e.g., 36–100). Using the dataset with 16 patients and 99 seizures, our algorithm outperforms LBP+SVM, Mixture+SVM, LGP+MLP, Mixture+RF, LBP+HD, and some deep learning methods with higher specificity and sensitivity and lower latency. Finally, we consider real-life scenarios by impairing the iEEG recordings with many noise sources and artifacts, and by implementation on a RISC-V microcontroller.

Future work will address an area of improvement, namely the development of an efficient specialized hardware implementation of our algorithm, including dedicated accelerators for bitwise operations (98% of the workload), and dynamic selection techniques to track the most informative electrodes over time.

References

[1] K. M. Fiest *et al.*, "Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies," *Neurology*, vol. 88, no. 3, pp. 296–303, Jan 2017.

[2] L. Kalilani *et al.*, "The epidemiology of drug-resistant epilepsy: A systematic review and meta-analysis," *Epilepsia*, vol. 59, no. 12, pp. 2179–2193, 12 2018.

[3] C. Rummel *et al.*, "Resected brain tissue, seizure onset zone and quantitative EEG measures: Towards prediction of post-surgical seizure control," *PLOS ONE*, vol. 10, no. 10, pp. 1–26, 10 2015.

[4] J. Parvizi and S. Kastner, "Promises and limitations of human intracranial electroencephalography," *Nat. Neurosci.*, vol. 21, no. 4, pp. 474–483, 04 2018.

[5] S. N. Baldassano *et al.*, "Crowdsourcing seizure detection: algorithm development and validation on human implanted device recordings," *Brain*, vol. 140, no. 6, pp. 1680–1691, 2017.

[6] R. Hussein *et al.*, "Robust detection of epileptic seizures using deep neural networks," in *Proc. IEEE ICASSP*, 2018.

[7] A. K. Jaiswal and H. Banka, "Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals," *Biomedical Signal Processing and Control*, vol. 34, pp. 81–92, 2017.

[8] Y. Zhang *et al.*, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 465–474, Feb 2020.

[9] N. D. Truong *et al.*, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, pp. 104–111, 2018.

[10] W. Stacey *et al.*, "What is the present-day EEG evidence for a preictal state?" *Epilepsy Research*, vol. 97, no. 3, pp. 243 – 251, 2011, special Issue on Epilepsy Research UK Workshop 2010 on "Preictal Phenomena".

[11] Y. Nagahama *et al.*, "Intracranial EEG for seizure focus localization: evolving techniques, outcomes, complications, and utility of combining surface and depth electrodes," *J. Neurosurg.*, pp. 1–13, May 2018.

[12] A. Burrello *et al.*, "One-shot learning for iEEG seizure detection using end-to-end binary operations: Local binary patterns with hyperdimensional computing," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct 2018, pp. 1–4.

[13] P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cognitive Computation*, vol. 1, no. 2, pp. 139–159, 2009.

[14] Y. Kaya *et al.*, "1d-local binary pattern based feature extraction for classification of epileptic EEG signals," *Applied Mathematics and Computation*, vol. 243, pp. 209 – 219, 2014.

[15] G. O'Leary *et al.*, "Nurip: Neural interface processor for brain-state classification and programmable-waveform neurostimulation," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3150–3162, Nov 2018.

[16] Y. Yuan *et al.*, "A multi-view deep learning framework for eeg seizure detection," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 83–94, 2018.

[17] A. Burrello *et al.*, "Hyperdimensional computing with local binary patterns: One-shot learning for seizure onset detection and identification of ictogenic brain regions from short-time iEEG recordings," *IEEE Transactions on Biomedical Engineering*, pp. 1–1, 2019.

[18] P. D. Schiavone *et al.*, "Quentin: an ultra-low-power pulpissimo soc in 22nm fdx," in *2018 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*. IEEE, 2018, pp. 1–3.

[19] S. Heller *et al.*, "Hardware implementation of a performance and energy-optimized convolutional neural network for seizure detection," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2018, pp. 2268–2271.

[20] F. Manzouri *et al.*, "A comparison of machine learning classifiers for energy-efficient implementation of seizure detection," *Frontiers in systems neuroscience*, vol. 12, 2018.

[21] M. A. B. Altaf *et al.*, "A 16-channel patient-specific seizure onset and termination detection soc with impedance-adaptive transcranial electrical stimulator," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 11, pp. 2728–2740, Nov 2015.

[22] S. Khanmohammadi and C.-A. Chou, "Adaptive seizure onset detection framework using a hybrid pca–csp approach," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 154–160, 2017.

[23] Y. Liu *et al.*, "Automatic seizure detection using wavelet transform and svm in long-term intracranial EEG," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 20, no. 6, pp. 749–755, 2012.

[24] A. K. Tiwari *et al.*, "Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals," *IEEE journal of biomedical and health informatics*, vol. 21, no. 4, pp. 888–896, 2016.

[25] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2003–2015, 2017.

[26] M. Zabihi *et al.*, "Patient-specific seizure detection using nonlinear dynamics and nullclines," *IEEE journal of biomedical and health informatics*, 2019.

[27] N. D. Truong *et al.*, "Integer convolutional neural network for seizure detection," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 4, pp. 849–857, 2018.

[28] H. Martin *et al.*, "Latencies from intracranial seizure onset to ictal tachycardia: A comparison to surface EEG patterns and other clinical signs," *Epilepsia*, vol. 56, no. 10, pp. 1639–1647, 2015.

[29] B. C. Munsell *et al.*, "Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data," *Neuroimage*, vol. 118, pp. 219–230, 2015.

[30] W. Zhou *et al.*, "Epileptic seizure detection using lacunarity and bayesian linear discriminant analysis in intracranial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3375–3381, 2013.

[31] M. Shoaran *et al.*, "Hardware-friendly seizure detection with a boosted ensemble of shallow decision trees," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 1826–1829.

[32] M. Taghavi *et al.*, "A 41.2 nj/class, 32-channel on-chip classifier for epileptic seizure detection," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 3693–3696.

[33] K. Schindler *et al.*, "Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG," *Brain*, vol. 130, no. Pt 1, pp. 65–77, Jan 2007.

[34] R. Meier *et al.*, "Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns," *Journal of Clinical Neurophysiology*, vol. 25, no. 3, pp. 119–131, 2008.

[35] F. T. Sun and M. J. Morrell, "The RNS System: responsive cortical stimulation for the treatment of refractory partial epilepsy," *Expert Rev Med Devices*, vol. 11, no. 6, pp. 563–572, Nov 2014.

[36] T. S. Kumar *et al.*, "Classification of seizure and seizure-free EEG signals using local binary patterns," *Biomedical Signal Processing and Control*, vol. 15, pp. 33 – 40, 2015.

[37] K. Schindler *et al.*, "Forbidden ordinal patterns of periictal intracranial EEG indicate deterministic dynamics in human epileptic seizures," *Epilepsia*, vol. 52, no. 10, pp. 1771–1780, Oct 2011.

[38] K. Schindler *et al.*, "Ictal time-irreversible intracranial EEG signals as markers of the epileptogenic zone," *Clin Neurophysiol*, vol. 127, no. 9, pp. 3051–3058, 09 2016.

[39] A. Rahimi *et al.*, "A robust and energy efficient classifier using brain-inspired hyperdimensional computing," in *Low Power Electronics and Design (ISLPED), 2016 IEEE/ACM International Symposium on*, August 2016.

[40] T. Wu *et al.*, "Brain-inspired computing exploiting carbon nanotube FETs and resistive RAM: Hyperdimensional computing case study," in *IEEE International Solid-State Circuits Conference, ISSCC*, 2018.

[41] A. Rahimi *et al.*, "Hyperdimensional biosignal processing: A case study for EMG-based hand gesture recognition," in *IEEE International Conference on Rebooting Computing*, October 2016.

[42] A. Moin *et al.*, "An EMG gesture recognition system with flexible high-density sensors and brain-inspired high-dimensional classifier," in *IEEE International Symposium on Circuits and Systems, ISCAS*, 2018.

[43] F. Montagna *et al.*, "PULP-HD: accelerating brain-inspired high-dimensional computing on a parallel ultra-low power platform," in *Proc. ACM/IEEE DAC*, 2018.

[44] A. Rahimi *et al.*, "Efficient biosignal processing using hyperdimensional computing: Network templates for combined learning and classification of ExG signals," *Proceedings of the IEEE*, vol. 107, no. 1, pp. 123–143, Jan 2019.

[45] P. Kanerva, "Binary spatter-coding of ordered k-tuples," in *ICANN'96, Proceedings of the International Conference on Artificial Neural Networks*, ser. Lecture Notes in Computer Science, , Ed., vol. 1112. Springer, 1996, pp. 869–873.

[46] P. Kanerva, "Sparse distributed memory." Cambridge, MA, USA: The MIT Press, 1988, p. 155.

[47] M. Schmuck *et al.*, "Hardware optimizations of dense binary hyperdimensional computing: Rematerialization of hypervectors, binarized bundling, and combinational associative memory," *ACM Journal on Emerging and Selected Topics in Circuits and Systems*, 2019.

[48] A. Nirkko *et al.*, "Die gemeinsame Referenz bei der Elektroencephalografie: Medianwerte im Vergleich zu Mittelwerten," *Klinische Neurophysiologie - KLIN NEUROPHYSIOL*, vol. 40, 03 2009.

[49] W. A. Ríos-Herrera *et al.*, "The influence of EEG references on the analysis of spatio-temporal interrelation patterns," *Frontiers in Neuroscience*, vol. 13, p. 941, 2019.

[50] T. Zhan *et al.*, "A resource-optimized vlsi architecture for patient-specific seizure detection using frontal-lobe EEG," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–5.

[51] W. Chen *et al.*, "A fully integrated 8-channel closed-loop neural-prosthetic cmos soc for real-time epileptic seizure control," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 232–247, Jan 2014.

[52] T. J. von Oertzen, "PET and ictal SPECT can be helpful for localizing epileptic foci," *Curr. Opin. Neurol.*, vol. 31, no. 2, pp. 184–191, 04 2018.

[53] S. Kim *et al.*, "Clinical value of the first dedicated, commercially available automatic injector for ictal brain spect in presurgical evaluation of pediatric epilepsy: comparison with manual injection," *Journal of Nuclear Medicine*, vol. 54, no. 5, pp. 732–738, 2013.

[54] Z. W. Ren *et al.*, "High-frequency and brief-pulse stimulation pulses terminate cortical electrical stimulation-induced afterdischarges," *Neural Regen Res*, vol. 12, no. 6, pp. 938–944, Jun 2017.

[55] P. Afra *et al.*, "Duration of complex partial seizures: an intracranial EEG study," *Epilepsia*, vol. 49, no. 4, pp. 677–684, Apr 2008.

[56] G. K. Bergey, "Neurostimulation in the treatment of epilepsy," *Exp. Neurol.*, vol. 244, pp. 87–95, Jun 2013.

[57] F. Bonini *et al.*, "Frontal lobe seizures: From clinical semiology to localization," *Epilepsia*, vol. 55, no. 2, pp. 264–277, 2014.

[58] M. Mirzaei *et al.*, "A fully-asynchronous low-power implantable seizure detector for self-triggering treatment," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, no. 5, pp. 563–572, Oct 2013.

[59] R. G. Andrzejak *et al.*, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.

[60] Y. Li *et al.*, "Epileptic seizure classification of EEGs using time–frequency analysis based multiscale radial basis functions," *IEEE journal of biomedical and health informatics*, vol. 22, no. 2, pp. 386–397, 2017.

[61] A. Burrello *et al.*, "Laelaps: An energy-efficient seizure detection algorithm from long-term human iEEG recordings without false alarms," 2019-03, design, Automation and Test in Europe Conference (DATE 2019); Conference Location: Florence, Italy; Conference Date: March 25-29, 2019.

[62] A. Delorme *et al.*, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.

[63] H. Kassiri *et al.*, "Rail-to-rail-input dual-radio 64-channel closed-loop neurostimulator," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 11, pp. 2793–2810, 2017.

[64] H. Chandrakumar and D. Marković, "A high dynamic-range neural recording chopper amplifier for simultaneous neural recording and stimulation," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 3, pp. 645–656, 2017.

[65] J. Timmer and M. König, "On generating power law noise." *Astronomy and Astrophysics*, vol. 300, p. 707, 1995.

[66] F. Conti *et al.*, "Xnor neural engine: A hardware accelerator ip for 21.6-fj/op binary neural network inference," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 11, pp. 2940–2951, Nov 2018.