Highly Efficient and Accurate Seizure Prediction on Constrained IoT Devices

Farzad Samie, Sebastian Paul, Lars Bauer, and Jörg Henkel,
Chair for Embedded Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany
Emails: {samie;lars.bauer;henkel}@kit.edu

***Abstract*—In this paper we present an efficient and accurate algorithm for epileptic seizure prediction on low-power and portable IoT devices. State-of-the-art algorithms suffer from two issues: computation intensive features and large internal memory requirement, which make them inapplicable for constrained devices. We reduce the memory requirement of our algorithm by reducing the size of data segments (i.e. the window of input stream data on which the processing is performed), and the number of required EEG channels. To respect the limitations of the processing capability, we reduce the complexity of our exploited features by only considering the simple features, which also contributes to reducing the memory requirements. Then, we provide new relevant features to compensate the information loss due to the simplifications (i.e. less number of channels, simpler features, shorter segment, etc.). We measured the energy consumption (12.41 mJ) and execution time (565 ms) for processing each segment (i.e. 5.12 seconds of EEG data) on a low-power MSP432 device. Even though the state-of-art does not fit to IoT devices, we evaluate the classification performance and show that our algorithm achieves the highest AUC score (0.79) for the held-out data and outperforms the state-of-the-art.**

***Index Terms*—Epileptic Seizure Prediction, EEG, Healthcare, IoT, Embedded Systems, Constrained Devices.**

I. Introduction

Epilepsy is a chronic neurological disorder that, according to the Epilepsy Foundation, affects more than 65 million people worldwide [1]. It is associated with recurrent and sudden seizures which are due to temporary electrical disturbance and excessive neuronal discharge in the brain [2]. Epileptic seizures result in altered consciousness or a whole body convulsion with uncontrolled movements. Due to the random nature of seizure occurrences, they are life-threatening and may increase the risk of serious injuries, especially if they occur while the patient is driving, exercising, etc.

For many years, the medical community believed that epileptic seizures are abrupt transitions and could not be anticipated [3]. However, studies show that the Electroencephalographic (EEG) signals from patients reflect the changes prior to the epileptic seizures [3]. EEG is a record of the brain electrical activity, which can be measured using wearable or implantable sensors. EEG is accepted and utilized widely for monitoring and diagnosing the seizures. The evidences from studying EEG show that seizures develop over time, hours before the clinical symptoms [4].

The quantitative analyses of epileptic EEG recordings have shown that brain activity in subjects with epilepsy has three distinct states: *ictal*, *preictal*, and *interictal*. Ictal refers to the state of the brain during actual seizures while preictal is the state of the brain prior to seizure onset (i.e. one or several hours [4]). Interictal refers to the seizure-free state of the EEG recording.

The discrimination between ictal, interictal and preictal states has been proved and validated by visual analysis of the EEG recordings by the neurologists. From an engineering perspective, pattern recognition techniques are able to distinguish these three states using EEG signals. Several methods have been proposed to *detect* seizure "after" its occurrence [5–7]. These retrospective methods are important for automatic analysis of EEG signals and reducing its time and costs. They are developed to replace visual seizure detection, which has not been proven to be very efficient specially for long-term epileptic EEG recordings and Big Data analysis [8].

However, *prediction* or *forecast* of epileptic seizures before they occur has much greater usages and potential applications, as they will enable the patients to take appropriate precautions (e.g. taking medication, pulling aside, etc.). Especially with the emergence of Internet of Things (IoT), portable low-power devices with network connectivity can be used to assist patients (e.g. predict seizures and notify them) [9].

Machine learning techniques have been shown to be effective in developing seizure prediction solutions [10]. Classification methods including neural networks, support vector machines, decision trees, logistic regression, etc. are trained using large amount of labeled EEG signals to create a model to classify whether the state is *preictal* or *interictal* [11, 12]. Most of the seizure prediction methods suffer from the issues like low specificity (too many false alarms) and unreliable performance across patients [2, 7]. One of the reasons is the asymmetric nature of seizure data (very rare ictal and preictal status, while most of the time the patient is in the interictal state) which is also known as unbalanced classification.

In 2014, the American Epilepsy Society, Epilepsy Foundation of America, and National Institutes of Health held a competition on *Kaggle.com* to advance the state-of-the-art in seizure prediction. They provided EEG data from seven subjects (see Section IV-A for details) labeled for interictal and preictal EEG training signals [13]. The contestants used supervised machine learning techniques to build classifiers using a wide range of features from the provided training data. The top participants published their models, algorithms and results in [13]. However, the state-of-the-art seizure prediction algorithms in [13] are not suitable or tailored for portable constrained IoT devices. The reasons are twofold: (i)

978-3-9819263-0-9/DATE18/©2018 EDAA

the complexity of the features mandates high computational effort, and (ii) the high number of EEG channels and the large segment size (i.e. the window of input data on which classification is performed) increase the amount of required memory. These shortcomings are more critical in IoT devices where the computation resources, memory, and power source are constrained and scarce.

The novel contributions of this work are as follows:

- We present a highly efficient seizure prediction technique by proposing new features for the classification model.
- The introduced features allow us to eliminate other complex features and reduce the number of required EEG channels while still achieving accurate prediction.
- The proposed model not only respects the scarce internal memory and constrained computation resources on the low-power devices, it also outperforms the state-of-the-art in classification performance.

Experimental results include evaluating the area under curve (AUC), the common metric for classification, which shows the tradeoff between False positive rate and True positive rate for different values. Our proposed algorithm outperforms state-of-the-art by achieving the highest AUC (0.79) for held-out data (unseen data during training and test phases) and it is implementable on IoT devices.

Paper organization: In Section II, we provide an overview on seizure prediction and related works. Then, we present the details of our proposed algorithm in Section III. Evaluations and experimental results are presented in Section IV, while Section V concludes the paper.

II. Preliminary Background & Related Work

A. Performance metrics

The main metrics used to evaluate a classifier are listed in Table I. Since most of the times the status of the EEG signal is interictal, the *True Negative* (TN) number shall be high. Hence, a classifier whose prediction is always ‘non-seizure’ would achieve high accuracy, high specificity and low False Positive Rate (FPR), but it misses the seizure events. On the other hand, a classifier whose prediction is always ‘seizure’ would achieve high True Positive Rate (TPR) and sensitivity. In the seizure prediction application, it is critical not to miss seizure events (preictal status), therefore a high TPR is desired. In addition, the FPR must be reasonably low. The AUC metric evaluates the tradeoff between TPR and FPR for different threshold values. AUC ranges from 0.0 to 1.0, where 1.0 corresponds to perfect classification (i.e. TPR=1 and FPR=0). The AUC is specifically a useful metric for dealing with unbalanced data [14] and is widely used for evaluating the performance of seizure prediction techniques [13, 15]. We, too, use this metric to evaluate our method and compare it against the state-of-the-art algorithms in [13].

B. EEG sub-bands

Traditional EEG analysis methods (even for other applications such as sleep monitoring and brain-machine-interface) split the frequency spectrum of an EEG signal into several frequency sub-bands. Even though there are various definitions of sub-bands in the community, they are only slightly different. We use one of the most common ranges similar to [13]: δ (0.1 – 4 Hz), θ (4 – 8 Hz), α (8 – 12 Hz), β (12 – 30 Hz), low-γ (30 – 70 Hz) and high-γ (70 – 180 Hz). Some studies split the low and high-γ bands further into 2 sub-bands. Different techniques including Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT) and Finite Impulse Response (FIR) can be used to obtain the sub-bands from an EEG signal.

TABLE I: Performance measures of binary classification (N is the number of events to be classified)

Name	Formula	desired
Accuracy:	$\frac{TP+TN}{N}$	high
False positive rate (FPR):	$\frac{FP}{FP+TN}$	low
True positive rate (TPR), Sensitivity (SEN) or Recall:	$\frac{TP}{TP+FN}$	high
Specificity (SPC):	$\frac{TN}{FP+TN}=1-FPR$	high

TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

C. State-of-the-art

Five state-of-the-art algorithms are presented and evaluated in [13], which are the top participants of the *Kaggle* seizure prediction competition. The QMSDP method is an ensemble of three distinct algorithms: the first algorithm uses 60 s segments. It uses many features including spectral entropy and Shannons entropy of sub-bands, Shannons entropy in dyadic sub-bands, spectral correlation between dyadic sub-bands, time series correlation matrix and its eigenvalues, and fractal dimensions. The second algorithm uses 8 s segments and calculates the following features: sums of FFT power, time series correlation matrix, and time series variance. The classifier is based on support vector machines (SVM). The third algorithm is similar to the second one but uses a different classifier (Random Forests [16]).

Birchwood’s method uses 50 s non-overlapping segments. It forms 18 equal frequency sub-bands from 1 to 50 Hz, then calculates the logarithm of FFT magnitude in these bands for the features. The inter-channel covariance and eigenvalues are calculated as the features for both sub-bands and time domain. The classifier is based on SVM.

The method of ESAI CEU-UCH uses 60 s segments with 30 s of overlap. The feature set includes inter-channel correlation matrix of frequency sub-bands and its eigenvalues. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are applied to the frequency sub-bands. The classifier is based on k-nearest neighbors (k = 40).

Michael Hills’ method uses non-overlapping 75 s segments. The feature set includes time domain and frequency domain correlation matrix and their sorted eigenvalues, the logarithm of FFT magnitude over sub-bands, and fractal dimension. The classifier is based on SVM.

Wei Wu’s method uses non-overlapping 60 s segments. The features include statistics (i.e. mean, maximum, minimum and standard deviation) in time and frequency domains, inter-channel covariance matrices in time and frequency domains, etc. The average spectral power for frequency sub-bands is added to the feature set. They use SVM classifiers to train for individual subject as well as a global one for all training data.

Finally, Golondrina's method uses non-overlapping 60 s segments. It also calculates the standard deviation and average spectral power in frequency sub-bands for each channel. The classifier is based on the convolutional neural network (CNN).

Ref. [17] uses spectral power of frequency sub-bands as well as cross-channel correlation between each two channels as features. The classification algorithm is based on SVM. However, the goal is to predict seizure for 1 hour segments based on 20 s windows.

These methods are not suitable for low-power IoT devices due to the following reasons: (i) the complexity of their features is high and their computation requires more processing capabilities than what is usually offered by low power devices, and (ii) the internal memory that is required to buffer the segments (at least 60 seconds of 16 EEG channels) and perform the feature extraction is much larger than what is available on low power IoT devices (more details in Section III-A). Alternatively, transmitting unprocessed raw data to another powerful device (e.g. gateway or cloud) would consume huge amount of power and therefore is not efficient for low-power portable devices (see Section IV-B).

III. Highly Efficient Seizure Prediction

A. General overview of seizure prediction methods

Figure 1 shows the high-level overview and stages for developing the seizure prediction method using machine learning classification based on the general method in [18]. The EEG signal can be acquired from implantable devices or on-skin electrodes. The segmentation includes buffering the data for a specific window size (e.g. 60 seconds of EEG signal in [13]). The subsequent operations are performed on this chunk of data. Band-pass filters are used either to remove higher frequencies and noise or to obtain the frequency sub-bands. Many features including time-domain and frequency-domain information as well as the correlation between channels can be extracted from EEG channels, but not all have the same importance. The next step is to select those features with maximum relevance and minimum redundancy. Afterwards, the classifiers (e.g. support vector machine, artificial neural network, logistic regression, etc.) are trained. The output of the classifiers are smoothed using post-processing techniques [18]. If the performance of the model is not satisfactory, then the parameters are changed and a new model will be trained.

Similar to other classification techniques, the state-of-the-art algorithms in [13] buffer the EEG data into smaller segments (with or without overlap) and extract the features from these segments. The internal memory which is required for each seizure prediction algorithm depends on the following:

- *Segment size:* It not only determines the initial buffer size, but also affects the size of temporary variables and arrays during processing and classification. The size of buffer to store one data segment depends on the length of segment (e.g. 60 seconds in the state-of-the-art), the sampling frequency of EEG data (e.g. 400 Hz in our dataset), the number of EEG channels, and the size of each sample (e.g. 2 bytes) as shown in Eq. (1).

$$B = W_{[\mathrm{sec}]} \times f_{[\mathrm{Hz}]} \times \#\mathrm{ch} \times \mathrm{size\_of(sample)} \quad (1)$$
$$= 60 \times 400 \times 16 \times 2\,\mathrm{Bytes} = 750\,\mathrm{kB} \quad (2)$$

  Equation (2) shows the initial segment size in state-of-the-art algorithms. This already exceeds the total available memory of many low-power IoT platforms[1].
- *Feature complexity:* Some of the features that are used in the state-of-the-art, including correlation-based features, eigenvalues or fractal dimensions, are not only time consuming but require large memory to keep the required data.

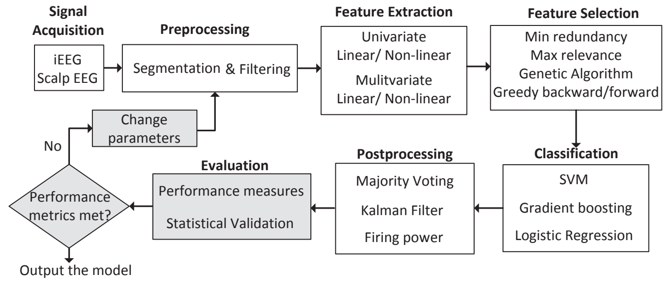

Fig. 1: General overview of the stages for seizure prediction using EEG signals adopted from [18].

B. Our Proposed Optimizations and Simplifications

Our approach to develop an efficient and accurate classification model for seizure prediction is based on *greedy backward elimination* [19]. We start with a complex and accurate model (e.g. similar to state-of-the-art), and repeatedly simplify it by eliminating the least useful information until reaching an efficient and fairly accurate model. We start with the most straightforward candidates. We simplify and optimize our model by reducing the EEG sampling rate (down-sample the existing data), reducing the segment size, reducing the number of EEG channels, and using features with lower complexity.

One possibility to reduce the size of the segmented data is to reduce the sampling frequency of EEG data (e.g. by down-sampling). However, this would lead to loss of information in higher frequencies due to Nyquist theorem. Since the seizure is associated with high-frequency activities in EEG, reducing the sampling frequency undermines the discriminative features, and consequently decreases the performance (i.e. sensitivity) of the classifier. Following our observations, we keep the EEG data at the original sampling frequency (i.e. 400 Hz).

We reduced the segment size down to 5.12 seconds which corresponds to $5.12\,\mathrm{s} \times 400\,\mathrm{Hz} = 2048$ samples. Using a 2048-point FFT will result in $\frac{400}{2048} \simeq 0.20$ resolution in frequency bins. Reducing the window size further will cause degradation in classifier performance (because of coarser resolution of frequency bins after applying FFT).

Another improvement in the size of segmented data is achieved by reducing the number of EEG channels. The state-of-the-art techniques usually use 16 channels of EEG data or

[1]For instance, ARM Cortex-M cores and Intel Quark SE microcontrollers that are used in wearable devices have less than 200 kB internal memory.

more [2, 6, 11, 17]. The provided dataset contains 16 EEG channels [13]. We employ a greedy algorithm to identify and use the most discriminative channels which is illustrated in Figure 2. The least number of channels with which we could achieve acceptable accuracy (compared to state-of-the-art) is 3 or 2 channels. Our proposed model has two variants: the one with 3 channels outperforms the state-of-the-art. The second variant (2 channels) is slightly less accurate but requires less memory, computational power and energy.

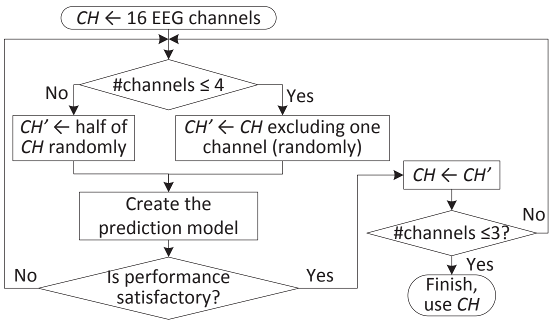

Fig. 2: Greedy approach to identify discriminative EEG channels

As mentioned, another reason that makes state-of-the-art unsuitable for low-power IoT devices is the complexity of some features. It results in computation-intensive operations and consequently increases the execution time, the energy consumption, and the probability to miss the deadline (i.e. not finishing the operation of one segment before the next one is ready to process). Some of these complex features are spectral correlation between channels in frequency sub-bands, time domain correlation matrix, frequency domain correlation matrix, eigenvalues and fractal dimensions. At design time in the 'feature selection' phase, we avoid these complex features.

C. Proposed Feature Engineering & New Features

Due to the necessary optimizations and simplifications (i.e. reducing the window size, reducing the number of channels and excluding complex features), we expect and observe a reduction in the performance of our seizure prediction classification. Hence, we introduce new features that only require low memory and are less computation intensive but still are discriminative and make up for the loss in accuracy.

EEG signals have shown higher activity during the preictal state, which is seen as high frequency fluctuations in time-domain signals. Therefore, the energy of high-frequency sub-bands increases compared to the low-frequency bands. We perform the inverse Fourier transform on the sub-bands to get their time-domain representation separately. Based on our observations, the time domain data associated with θ ($4-8$ Hz) and partial high-γ ($75-97$ Hz) contains the most information. We transform these two bands separately back into the time domain using inverse FFT (iFFT). We extract eight statistical features from these two time-domain sub-bands: minimum, maximum, variance, standard deviation, root mean square, Hjorth mobility, Hjorth complexity, and Kurtosis. These new features improve the performance of our classifiers to outperform state-of-the-art (see Section IV-B and Figure 3).

D. Classifiers & Post processing

We use logistic regression as one of our classification techniques. It is widely used in seizure prediction [18] and has a low memory requirement and light-weight computation, which makes it suitable for low-power IoT devices.

Along with logistic regression, we use another classifier that is called eXtreme Gradient Boosting (XGBoost) [20], which is widely used in machine learning applications. Gradient boosting is a method to combine multiple classifiers –called base-learners– that can complement each other. It uses multiple trees that –contrary to traditional decision trees– contain actual scores on each leaf. These multiple trees are combined in an ensemble model to form one strong classifier.

The classification stage is followed by a post-processing stage that aims at increasing the specificity of the seizure prediction system by decreasing the number of false alarms. We adopt a simple yet efficient post-processing scheme based on majority voting [17]. We use a sliding window of size four for voting. If three (or more) out of four consecutive segments are classified as preictal, the output of this sequence is preictal.

E. Computation Offloading

Even though XGBoost offers a strong classifier, it comes at the cost of larger storage requirements for the model. In our algorithm, the size of the XGBoost classification model is more than 300 kB, which exceeds the available Flash memory on many IoT devices. Using a powerful microcontroller or processor with large memory comes at the cost of higher power consumption and shorter battery lifetime. Our solution is to offload it to the gateway (e.g. smartphone), which is also in charge of the user interface, emergency calls, etc. Therefore, the postprocessing can also be done on the gateway as shown in Figure 4. The advantage of this design is that we only need to transmit the small amount of data (i.e. a few necessary features) to the gateway for the final step. Therefore, we can still benefit from the high accuracy of XGBoost model.

F. Summary of our algorithm

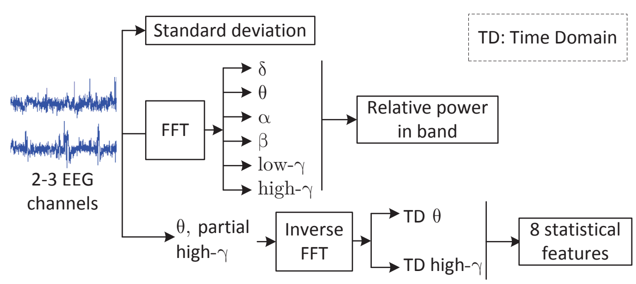

Fig. 3: Features extracted from each EEG channel

Figure 3 illustrates the features that we use in our algorithm. Relative spectral Power-In-Band (PIB) for six frequency sub-bands and standard deviation in the time domain (7 features in overall) are used in both classifiers. However, the XGBoost classifiers use additional features as follows. For the time-domain representation of the θ and partial high-γ sub-bands (see Section III-C), we calculate eight statistical features. Therefore, the number of features to be used by XGBoost for each EEG channel is $6 + 1 + 8 \times 2 = 23$.

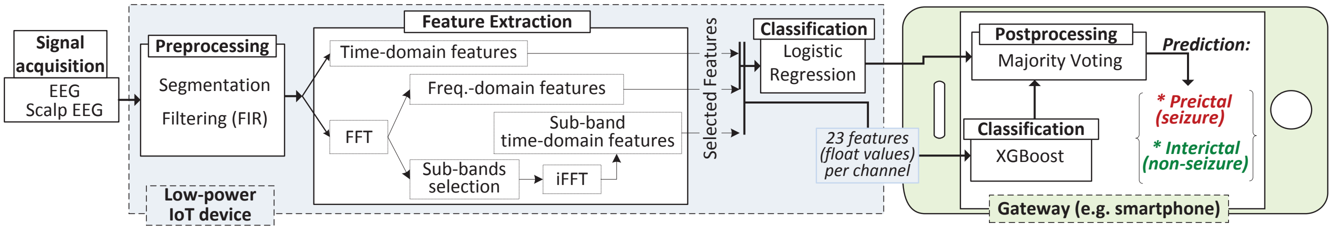

Fig. 4: General overview of the stages for our proposed seizure prediction method

Figure 4 summarizes our proposed algorithm. The EEG data is segmented and filtered on the low-power device. Then, the time-domain feature are calculated and the frequency domain and other time-domain features are obtained. The logistic regression model is implemented on the low-power device and calculates the probability of preictal state based on a subset of extracted features. The rest of the features along with the output of the logistic regression model are transmitted to the gateway to be used for XGBoost classification and post-processing. Finally, a majority voting approach does the post-processing stage. Table II lists the parameters and properties of our classifiers (logistic regression and XGBoost) for the model with 2 and 3 EEG channels.

TABLE II: Properties of our exploited classifiers in the proposed system with 2 and 3 EEG channels

	Logistic Regression		XGBoost (Gradient Boosting)	
#Channels	2	3	2	3
Size[kB]	0.19	0.26	300	330
Input features per channel	PIB (6 spectral bands), Standard deviation (time-domain)		PIB (6 spectral bands), Standard deviation (time-domain), 8 statistical features per band (2 band signals)	
#Features	7×2	7×3	23×2	23×3

IV. Experimental Results & Evaluation

A. Experimental Setup

We use the EEG dataset that is provided by *Kaggle* in a contest in 2014. The dataset was recorded from implanted electrodes in seven subjects with naturally occurring seizure. *Kaggle* hosted a similar contest in 2016, however the EEG data is subject to copyright and therefore cannot be used for any purpose including research.

EEG was recorded using an ambulatory monitoring system with 16 electrodes (i.e. channels) with 400 Hz sampling rate. The data is provided in 10 min clips with the label (either preictal or interictal). Around 13% of the training data is labeled as preictal, while the rest is interictal. The preictal data covers one hour prior to seizure with 5 minutes guard-band before the onset of seizure. Once the preictal stage is distinguished from interictal stage, there would be more than 5 minutes time to react before seizure occurs.

B. Experimental Results

To have a fair comparison against the state-of-the-art algorithms in [13], we used the very same metrics, input data and evaluation setup. The performance metric is AUC (see Section II) and it is calculated and reported by *Kaggle*, which is still available online (we upload our classification model, they evaluate it and report the AUC metric). The following AUC scores are calculated for the state-of-the-art and our algorithm by *Kaggle*:

Public score: Computed on a randomly sampled 40% subset of the test dataset.

Private score: Computed on the remaining 60% of the testing dataset, which makes it more representative.

Held-out score: Computed on a set of completely unseen and unlabeled data. This score shows the robustness and generality of the developed algorithms.

Table III illustrates the AUC scores for the state-of-the-art algorithms in [13] and our proposed algorithm. The best of each score is highlighted. The QMSDP algorithm achieved the highest public and private scores, but it faces a fairly significant drop in the held-out score. It shows that this algorithm is very specific to the provided training data and is not general and robust. Our algorithm with 3 channels reached the highest held-out score (i.e. 0.79) compared to the state-of-the-art algorithms. Another metric that is reported by [13] is sensitivity at 75% specificity which is shown in the last column of Table III. Our algorithm achieves the best performance for this metric (jointly with Golondrina's model).

TABLE III: AUC scores for our proposed model compared to the state-of-the-art algorithms.

Model	Public score	Private score	Held-out score	SEN at 75% SPC
QMSDP [13]	**0.860**	**0.820**	0.75	0.71
Michael Hills [13]	**0.860**	0.793	**0.79**	0.73
Golondrina [13]	0.823	0.785	0.77	**0.75**
Proposed Model (3 channels)	0.799	0.810	**0.79**	**0.75**
Proposed Model (2 channels)	0.794	0.767	0.78	0.63

Since the held-out and private score are the most important evaluation metrics, our algorithm outperforms state-of-the-art. The closest competitor is Michael Hills' algorithm that achieves the same held-out score but has a lower private score. Note that none of the state-of-the-art algorithms in Table III are applicable for low-power IoT devices due to their large internal memory requirements. Therefore, our proposed algorithm not only outperforms them in classification metrics, it can also be implemented on IoT low-power devices as described in the hardware implementation (Section IV-C).

Our post-processing technique improves the average accuracy from 73.6% to 93.3%. Since it uses a majority voting over four sliding clips, it smooths out the false positive rate. Table IV shows the effect of our post-processing technique on the classifier performance for one of the subjects.

TABLE IV: Comparison of original and postprocessed predictions using performance measures (subject #2 in dataset)

	Original prediction	Postprocessed prediction
Accuracy	75.9%	95.8%
Sensitivity	85.6%	92.0%
Specificity	74.9%	96.1%
False Positive Rate	25.1%	3.9%

C. Hardware Implementation Results

We use *Texas Instrument's MSP432*, an ultra-low-power microcontroller that is optimized for IoT devices to implement our algorithm. It exploits a 32-bit *ARM Cortex-M4F RISC* processor providing an operating frequency up to 48 MHz, a floating-point unit, 64 kB or SRAM and 256 kB Flash memory.

Table V shows the measurements of the proposed algorithm on the low-power MSP432 device. The current and power consumption values are obtained using the built-in power profiler of TI. We measure the power and energy consumption during the active phase (processing the EEG data to predict the seizure) and the passive phase (low-power sleep mode and data acquisition to buffer the data). Even though the power consumption in passive phase is less than active phase, its energy consumption is more due to the longer time that device spends in passive phase. Moreover, the peak memory usage is about 54 kB which occurs during the FFT operation.

TABLE V: Measured execution time, power and energy consumption of our algorithm on an ultra-low power *MSP432* device.

Phase	t [ms]	I_{avg} [mA]	P_{avg} [mW]	E [mJ]
Active	565	1.99	6.57	3.71
Passive	4555	0.58	1.91	8.70
Total	5120	0.73	2.42	12.41

We consider a Bluetooth low energy device as transceiver. We use the energy model that is presented in [21] to estimate the energy consumption for data transmission. For each segment (5.12 s interval), the energy consumption for sending the features to the gateway is 0.72 mJ. Therefore, the total energy consumption of our algorithm for one segment is about $12.41 + 0.72 = 13.13$ mJ. Note that raw data transmission (with no processing) for the state-of-the-art algorithms would require sending of 57,344 bytes, which would consume more than 213.41 mJ. Hence, the total energy consumption would have approximately been $5.12\,\text{s} \times 1.91\,\text{mW} + 213.41\,\text{mJ} = 223.19\,\text{mJ}$, which is about $16\times$ more than our algorithm.

V. Conclusion

In this paper, we presented an efficient and accurate algorithm for seizure prediction on constrained IoT devices. To respect the limitation of processing capability as well as internal memory, we reduced the size of data segments, the number of required EEG channels, and the complexity of the exploited features compared to the state-of-the-art. We provided new relevant features to improve the prediction accuracy of our simplified model.

The experimental results show that our proposed algorithm outperforms the state-of-the-art algorithms in terms of classification performance. We achieved the highest AUC score for held-out data. While state-of-the-art methods are inapplicable for low-power portable devices (due to their large memory requirements and complex features), our algorithm only consumes 12.41 mJ for processing one segment (5.12 s) of EEG data, and takes 565 ms on an ultra-low power device.

References

[1] P. O. Shafer, "About epilepsy: The basics," Online: http://www.epilepsy.com/learn/about-epilepsy-basics, 2014.

[2] L. D. Iasemidis, "Epileptic seizure prediction and control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 549–558, 2003.

[3] B. Litt and J. Echauz, "Prediction of epileptic seizures," *The Lancet Neurology*, vol. 1, no. 1, pp. 22–30, 2002.

[4] H. Witte, L. D. Iasemidis, and B. Litt, "Special issue on epileptic seizure prediction," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 537–539, 2003.

[5] S. A. Imtiaz, L. Logesparan, and E. Rodriguez-Villegas, "Performance-power consumption tradeoff in wearable epilepsy monitoring systems," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1019–1028, 2015.

[6] A. S. Zandi, M. Javidan, G. A. Dumont, and R. Tafreshi, "Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1639–1651, 2010.

[7] D. Gajic, Z. Djurovic, S. Di Gennaro, and F. Gustafsson, "Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition," *Biomedical Engineering: Applications, Basis and Communications*, vol. 26, no. 02, 2014.

[8] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time–frequency analysis," *IEEE Trans. on information techn. in biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.

[9] F. Samie, L. Bauer, and J. Henkel, "IoT technologies for embedded computing: A survey," in *CODES+ISSS*, 2016, pp. 1–10.

[10] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[11] E. Dadgar-Kiani, C. Alkan, and A. Shameli, "Applying machine learning for human seizure prediction," Stanford University, Tech. Rep., 2016.

[12] N. Ahammad, T. Fathima, and P. Joseph, "Detection of epileptic seizure event and onset using EEG," *BioMed Research International*, 2014.

[13] B. H. Brinkmann, J. Wagenaar, D. Abbot, P. Adkins, S. C. Bosshard, M. Chen, Q. M. Tieng, J. He, F. J. Muñoz-Almaraz, P. Botella-Rocamora, J. Pardo, F. Zamora-Martinez, M. Hills, W. Wu, I. Korshunova, W. Cukierski, C. Vite, E. E. Patterson, B. Litt, and G. A. Worrell, "Crowdsourcing reproducible seizure forecasting in human and canine epilepsy," *Brain*, vol. 139, 2016.

[14] L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data–recommendations for the use of performance metrics," in *Affective Computing and Intelligent Interaction (ACII)*, 2013, pp. 245–251.

[15] A. Temko, A. Sarkar, and G. Lightbody, "Detection of seizures in intracranial EEG: UPenn and mayo clinic's seizure detection challenge," in *Int. Conf. of Eng. in Med. and Biology Society*, 2015, pp. 6582–6585.

[16] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] H.-T. Shiao, V. Cherkassky, J. Lee, B. Veber, E. E. Patterson, B. H. Brinkmann, and G. A. Worrell, "SVM-based system for prediction of epileptic seizures from iEEG signal," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 1011–1022, 2017.

[18] E. B. Assi, D. K. Nguyen, S. Rihana, and M. Sawan, "Towards accurate prediction of epileptic seizures: A review," *Biomedical Signal Processing and Control*, vol. 34, pp. 144–157, 2017.

[19] Y. Zhang and J. C. Rajapakse, *Machine learning in bioinformatics*. John Wiley & Sons, 2009, vol. 4.

[20] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Int. Conf. on Knowl. Discovery and Data Mining*, 2016, pp. 785–794.

[21] P. Kindt, D. Yunge, R. Diemer, and S. Chakraborty, "Precise energy modeling for the Bluetooth low energy protocol," *CoRR*, 2014.