C²SP-Net: Joint Compression and Classification Network for Epilepsy Seizure Prediction

Di Wu, Yi Shi, Ziyu Wang, Jie Yang, and Mohamad Sawan, Fellow, IEEE

Abstract—Recent developments in brain-machine interface technology have rendered seizure prediction possible. However, the transmission of a large volume of electrophysiological signals between sensors and processing apparatuses and the related computation become two major bottlenecks for seizure prediction systems due to the constrained bandwidth and limited computational resources, especially for power-critical wearable and implantable medical devices. Although many data compression methods can be adopted to compress the signals to reduce communication bandwidth requirement, they require complex compression and reconstruction procedures before the signal can be used for seizure prediction. In this paper, we propose C²SP-Net, a framework to jointly solve compression, prediction, and reconstruction without extra computation overhead. The framework consists of a plug-and-play in-sensor compression matrix to reduce transmission bandwidth requirements. The compressed signal can be utilized for seizure prediction without additional reconstruction steps. Reconstruction of the original signal can also be carried out in high fidelity. Compression and classification overhead from the energy consumption perspective, prediction accuracy, sensitivity, false prediction rate, and reconstruction quality of the proposed framework are evaluated using various compression ratios. The experimental results illustrate that our proposed framework is energy efficient and outperforms the competitive state-of-the-art baselines by a large margin in prediction accuracy. In particular, our proposed method produces an average loss of 0.6% in prediction accuracy with a compression ratio ranging from 1/2 to 1/16.

Index Terms—Seizure prediction, EEG, convolutional neural network, compressive sensing (CS), hardware-friendly.

I. INTRODUCTION

Epilepsy is a neurological disease that causes recurrent seizures in the brain, influencing the lives of more than 50 million people [1], [2]. Patients experience unconsciousness, movement disorders, and other control loss of body parts when a seizure occurs. Currently, many wearable and implantable circuits and systems [3] have been developed to detect [4] or predict [5] the occurrence of seizures to provide time for emergent preparation under risky scenarios without medical attendance, such as driving and operating heavy machines.

Fig. 1 (a) shows the system built-up of commonly seen seizure prediction systems. The system usually consists of electrodes to collect electrophysiological signals, which reflect the brain’s neuron activities. The signals are further transmitted to a processing apparatus such as a micro controller unit (MCU) and dedicated signal processors through weak and short distance RF communication [6].

Due to the recurrent nature of epilepsy, these systems are required to be wearable or implantable; hence both the communication bandwidth and computation ability are limited due to the miniaturization and power constraints. In-sensor signal compression is therefore essential to ensure the usability of wearable/implantable real-time prediction systems [3]. As illustrated in Fig. 1-(a), the conventional electrophysiological signals analyzing process requires devices to have high bandwidth in data transmission and massive computation in subsequent analysis.

In-sensor signal compression could vastly reduce the transmission bandwidth. Conventional compression methods such as average downsampling [7] and uniform downsampling [8] are substantially limited by the Nyquist sampling rate, and the compressed data can not be reconstructed afterward. Compressive sensing (CS) has been proposed to compress signals at a sub-Nyquist sampling rate in recent years. The compression process can be denoted as a matrix multiplication between the signal and a sensing matrix, which is computationally efficient and hardware friendly [9]. In addition, CS enables simultaneous sampling and compression, which significantly alleviates on-chip transmission bandwidth and storage space. Fig. 1-(b) illustrates how CS methodology is involved in seizure detection systems [10].

However, existing methods [11] take compression and reconstruction as two separate processes. We argue that signals compressed (randomized CS measurements) with current CS approaches are feasible for reconstruction under a sparsity assumption but might not be suitable for downstream tasks such as seizure prediction. Moreover, typical CS
algorithms that widely rely on convex optimization [12], greedy algorithms [13], and Bayesian learning [14] for reconstruction are computationally expensive, rendering them impractical under real-time scenarios.

To mitigate the previously discussed limitations and provide a feasible solution for a reliable and efficient seizure prediction system, we propose a novel end-to-end deep learning framework that jointly solves the signal compression, reconstruction, and seizure prediction tasks. Our proposed framework ensures that the sensing matrix learned is optimized both for reconstruction and prediction purposes. Compared with random sensing matrices, our learned matrix captures informative features based on various downstream tasks. The learned sensing matrix could be deployed onto electroencephalography (EEG) sensors as a general plug-and-play solution for low-cost data compression and transmission. Moreover, the compressed signal can be directly applied for seizure prediction or EEG reconstruction. Unlike traditional CS algorithms, Our proposed compression and reconstruction mechanism is parametric and could be directly applied to previously unseen data as opposed to solving an optimization task on new data. To the best of our knowledge, we are the first to explore the feasibility and reliability of CS-involved seizure prediction using EEG. Moreover, in section II, we introduce previous works in the areas of both seizure prediction and compressive sensing. In section III, we provide a detailed description of this proposed framework. The performance of this proposed framework is evaluated in Section IV. The paper is concluded in Section V.

II. RELATED WORK

A. Compressive Sensing

CS was initially introduced for the acquisition of low-rate images, and was gradually developed into various fields including video processing [15], face recognition [16], [17], and magnetic resonance imaging (MRI) acquisition [18]. CS performs signal compression by sampling a few measurements, i.e. \( z = Ax \), where \( x \in \mathbb{R}^N \) denotes the original signal, \( z \in \mathbb{R}^M \) is the sampled measurements (compressed signal) and \( A \in \mathbb{R}^{N \times M} \) represents the sampling/sensing matrix with \( M < N \).

Most studies involving CS use random matrices as sensing matrices. Zeng et al. [19] selected the Bernoulli matrix for compressive sensing. The authors regard EEG compressibility as a kind of feature in terms of seizure detection and use four classifiers, including decision tree, K-nearest neighbor (K = 5), discriminant analysis, and support vector machine (SVM) to classify the features. Using this method, the highest prediction accuracy achieves 76.7%. Abdulghani et al. [20] used Gaussian matrix to compress the EEG data in their work, and investigated the performance of different implementations of the CS theory involved in EEG signals. These random matrices fail to embed data-specific or downstream task-related features into the compressed signals.

The signal reconstruction process from compressed measurements is to solve an optimization task which can be formulated as follows:

\[
\min_x \mathcal{R}(x), \ s.t. \ z = Ax, \tag{1}
\]

where \( \mathcal{R}(\cdot) \) is a regularization term.

To solve the under-determined optimization function, a variety of algorithms have been developed to reconstruct the signals. The Greedy algorithm [21] is popular due to its low computational complexity. Mallat et al. [22] developed a Greedy algorithm referred to as matching pursuit That decomposes signals into a linear expansion of waveforms selected from a redundant dictionary of functions. However, it requires prior knowledge of the sparsity of the underlying signal. Bayesian learning is another algorithm that is common in CS studies. The block sparse Bayesian learning (BSBL) algorithm is an example of Bayesian learning. Initially, the BSBL framework was proposed for signals with a block structure. Zhang et al. [14] adopted BSBL algorithm in their research, introducing the technique to the telemonitoring of EEG. Their experiment shows good reconstruction quality with an average normalized mean square error (NMSE) of 0.116 and an average structural similarity index measure (SSIM) of 0.81. The Bayesian method has a high speed,
but it depends on preliminary knowledge and causes massive computation costs. To overcome the limitation of conventional CS methods, we propose to construct a sensing matrix during the optimization process of reconstruction and seizure prediction tasks in our framework.

**B. Seizure Prediction**

Epilepsy influences 1% of the world’s population, of which up to 35% cannot be cured by pharmaceutical or medical treatment [23]. Inevitably, these people suffer from unexpected seizure onset. Therefore, efforts have been made towards accurate alarm before seizure onset to provide better lives for epileptic sufferers [24], [25], [26]. As an important source for monitoring brain activities in the entire process of epileptic seizure, EEG became the primary focus of the seizure prediction studies. Early machine learning based approaches utilize support vector machines (SVMs) [27], [28] and multi-layer perceptrons (MLP) [29] for seizure prediction. However, these methods rely on hand-engineered features that require considerable prior knowledge [30]. Park et al. [31] introduced SVM to seizure prediction. They calculate spectral power in nine bands from the EEG of the Freiburg EEG database using four pre-processing methods, namely raw, bipolar, time-differential, and bipolar/time-differential. SVM with double cross-validation is applied for classification. However, manual feature extraction is a time-consuming process and the features extracted lack generalization ability. With the successful application of deep learning (DL) methods in many fields, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are extensively used in recent studies [32], [33], [34], [35], [36]. Convolution operations can be considered as filters and thus act as a learnable automatic feature extractor. RNNs are adopted to better model the relationship between time sequences. Wang et al. [33] proposed to merge CNN with the directed transfer function to model information exchange among different EEG channels. Mapping raw signals to the frequency domain via short-time Fourier transform (STFT) is also a popular technique adopted for seizure prediction. Truong et al. [35] proposed a generalized retrospective and patient-specific seizure prediction method based on CNN. The authors utilize STFT to enhance time-frequency information and then apply a CNN-based model for feature extraction and binary classification between pre and interictal states. Similarly, Yang et al. [37] proposed a dual self-attention residual network to combine local features and global features via spectrum attention with STFT as input. In addition, graph neural networks [38] can also be adopted to model the relationship between different seizure activities and brain regions.

Although CNN models are commonly employed in many research activities, some authors claim that RNNs might be better at isolating temporal characteristics. Tsiouris et al. [32] first applied a two-layer long short-term memory (LSTM) network to seizure prediction. Different lengths of preictal windows (ranging from 15 minutes to 2 hours) are used in the seizure prediction task. Prior to classification between preictal and interictal classes, the LSTM model extracts time and frequency domain features between EEG channels cross-correlation and graph-theoretic features. Despite the success of deep learning based methods on seizure prediction in terms of high prediction accuracy, minimal to no effort has been dedicated to efficient seizure prediction using compressed EEG data.

**III. METHODOLOGY**

In this section, we formalize our proposed C^2-SP-Net framework, as illustrated in Fig. 2. The primary purpose of our proposed framework is to compress the original EEG signal for efficient real-time seizure prediction with minimum degradation in prediction performance. Unlike previous CS-based compression methods, which fail to capture signal statistics, our compression strategy embeds prior knowledge of reconstruction and prediction into the compressed signal. Furthermore, our proposed framework is able to reconstruct the original signal for human visualization. The proposed framework contains a compression function \( C(\cdot) \), a reconstruction function \( R(\cdot) \) and a prediction function \( P(\cdot) \). \( x \in \mathbb{R}^{N \times C} \) is an EEG signal sequence of length \( N \) and \( y \) is its corresponding one-hot label, which indicates interictal or preictal state, where \( C \) denotes the number of channels.
The input signal $x$ is compressed as follows:

$$z = C(x),$$

where $z \in \mathbb{R}^{M \times C}$ denotes the compressed signal. Naturally, the signal compression ratio $r$ is defined as $\frac{M}{N}$ with $M < N$. Then, the reconstructed signal $\hat{x}$ is given by $R(z)$ while a prediction result $\hat{y}$ is given by $P(z)$. We design three deep learning based networks to implement these three functions. Next, we describe them in detail.

### A. Compression Network

Since the purpose of compression is to reduce the computation and transmission cost on wearable/implantable devices, we consider using compression matrix with different bit precision as the compression network. In particular, we choose the floating-point compression matrix $W \in \mathbb{R}^{N \times M}$ and binary compression matrix $W \in \{0, 1\}^{N \times M}$. Then the signal from each channel is compressed by multiplying the compression matrix with the original signal, $y^c = Wx^c$, where $y^c$ and $x^c$ denotes the compressed signal and the original signal of channel $c$ respectively. For simplicity, we use the same compression matrix for all signal channels.

### B. Prediction Network

To directly capture the patterns of different granularity levels from compressed raw EEG signals, we designed a CNN based network inspired by the popular ResNet [39], as shown in Fig. 3. Specifically, the input signal is first fed to a stem layer that consists of a convolution and a max-pooling operation. Then, the output signal from the stem layer goes through a cascade of basic convolution blocks. Each basic block consists of convolution layers with each followed by a batch normalization layer and an activation layer. In our work, we adopt the rectified linear unit (ReLU) as the activation function. Each basic convolution block contains a residual connection defined as follows:

$$f_{l,k} = \delta(BN(conv_{l,k}(f_{l,k-1}))),$$

$$f_l = f_{l,0} + BottleNeck(f_{l,2}),$$

where $f_{l,k}$ denotes the intermediate feature map of after passing $k^{th}$ convolution layer of block $l$, namely $conv_{l,k}$. $BN$ and $\delta$ stands for batch normalization and activation function respectively. In this paper, we set $k = 2$ and $l = 4$. The BottleNeck operation reduces the number of channels of $f_{l,2}$ to the same number of $f_{l,0}$. The final output $f_l$ of block $l$ is then given by addition (residual connection).

Note that most existing CNN-based approaches treat EEG signal as greyscale images, meaning that the channel dimension and the time dimension of EEG signal are considered the height and width dimension analogous to an image. Thus, 2D convolution is commonly adopted. The convolution operation is considered to be able to extract local spatial information with the assumption that nearby pixels are of high correspondence, which is valid for valid images. However, for the same correspondence to hold for multichannel EEG signals, nearby EEG signals from the channel dimension should be functionally correlated. Unfortunately, due to the different choice of placement of electrodes during EEG signal acquisition, we hypothesize that most existing CNN-based methods fail to capture the discriminative correlation patterns across channels. We seek to capitalize on the interactions among different channels of EEG signal by setting the kernel size of the channel dimension the same as the number of channels of input signal/intermediate feature map, which is also known as 1D convolution.

The prediction result $\hat{y}$ is then given by:

$$\hat{y} = \sigma(FC_2(\delta(FC_1(GAP(\hat{f}_{4}))))),$$

where $GAP$ denotes global average pooling and $FC_i$ denotes the $i^{th}$ fully connected layer. $\delta$ and $\sigma$ represent the LeakyReLU and Softmax activation functions respectively.

We define the prediction loss function as:

$$\mathcal{L}_{pred} = \mathcal{L}(\hat{y}, y),$$

where $\mathcal{L}$ represents any loss function to evaluate the prediction performance.

### C. Reconstruction Network

We attempt to assess the feasibility and reliability of a seizure prediction system using EEG signals with various compression ratios. In addition to prediction, we also take reconstruction into consideration for other possible applications which require the original signal such as human expert visualization. Moreover, the reconstruction
Reconstruction Network with Compression Ratio $r$

![Architecture overview of the reconstruction network with compression ratio $r$. It is composed of $\left\lfloor \log_2 r \right\rfloor + 1$ up-sampling blocks with the BottleNeck operation mapping the signal to the original number of channels. The first $\left\lfloor \log_2 r \right\rfloor$ blocks upsample the signal to twice the length, while the last block upsamples the signal to original signal length.](image)

**TABLE I**

| Patient ID | Total seizure | Lead seizure | Sample Count |
|------------|---------------|--------------|--------------|
| chb01      | 7             | 3            | 354          |
| chb05      | 5             | 2            | 353          |
| chb06      | 10            | 6            | 685          |
| chb07      | 3             | 2            | 238          |
| chb08      | 5             | 3            | 355          |
| chb09      | 4             | 3            | 357          |
| chb10      | 7             | 6            | 547          |
| chb14      | 8             | 4            | 456          |
| chb18      | 6             | 3            | 247          |
| chb19      | 3             | 3            | 238          |
| chb22      | 3             | 3            | 301          |

**D. Joint-Objective Training**

We simultaneously optimize the compression network, prediction network, and reconstruction network using a joint objective function:

$$
\mathcal{L}_{\text{joint}} = \mathcal{L}_{\text{pred}} + \lambda \ast \mathcal{L}_{\text{recon}},
$$

where $\lambda$ is the weighting parameter to balance the two losses. Without loss of generality, we choose cross entropy and mean square error as the loss function for $\mathcal{L}_{\text{pred}}$ and $\mathcal{L}_{\text{recon}}$, respectively. The compression matrix is obtained by minimizing the joint object and is thus capable of compressing EEG signals in a way for better reconstruction and prediction purposes.

**IV. RESULTS**

**A. Dataset and Preprocessing**

We evaluate the effectiveness of our proposed framework on the CHB-MIT sEEG database [40], which is available through open access. The database contains sEEG signals from 23 epileptic patients (17 females, five males, and one person missing gender information). The sEEG signals were collected at a rate of 256 samples per second with 16-bit resolution. Most cases had their sEEG data recorded from 23 channels, and the electrodes were placed according to the International 10-20 system. Each case has a descriptive document, which illustrates relevant information, including case ID, channel information, seizure start time, and end time. The seizure onset time and end time were annotated by clinical experts through visual inspection.

According to the annotation documents, there are channel changes (channels added or removed) in some cases. Therefore we choose the cases that have fixed channels during the acquisition. Furthermore, we are more interested in patients with at least two lead seizures and 1-hour-long preictal time in total. Here, the preictal states are collected only before lead seizures since lead seizures have higher value clinically according to previous work [41]. In our study, the lead seizure is defined as a seizure preceded by four-hour of seizure-free period. We list all patient IDs that qualify our requirements in Table I.

The seizure prediction horizon (SPH) and preictal interval length (PIL) are two critical parameters in determining preictal
segments. SPH is a short interval between the end of preictal states and seizure onset. The PIL refers to the duration of preictal states. As is shown in Fig. 5, if an alarm occurs at any point within PIL + SPH before seizure onset, it is considered a successful prediction.

The SPH and PIL are still controversial and are usually chosen based on assumptions. SPH offers time for patients to prepare themselves. If the SPH is too large, patients might suffer from anxiety for too long, and the preictal data length might not be sufficient for training; if the SPH is too small, there might not be enough time for the patients to adjust themselves to safe positions.

According to the analysis above, the SPH and PIL are chosen as 5 minutes and 30 minutes respectively in this study, which means the preictal is defined as 5 to 35 minutes ahead of the seizure. In addition, the interictal period is defined as 30 minutes after the seizure offset and before the prediction period of subsequent seizure. To reduce the imbalance of preictal and interictal samples, we extracted interictal samples from EEG recordings with a sliding window of 20 seconds without overlapping. We apply a 20-second-long sliding window for the preictal samples with 25% overlapping between two consecutive window sets. We normalize the data by subtracting its mean and dividing its standard deviation.

**TABLE II**

| Compression ratio | Metrics   | Lawhern [26] | Zhang [25] | Xu [24] | Truong [35] | This work (λ = 0) | This work (λ = 1) |
|------------------|-----------|--------------|------------|---------|-------------|------------------|------------------|
| Original signal  | Accuracy (%) | 87.2±1.3     | 89.9±0.8   | 84.8±3.7 | 83.4±1.4   | 92.5±1.2         | N/A              |
|                  | Sensitivity (%) | 87.9±2.4   | 93.3±1.4   | 85.6±3.4 | 89.7±2.9   | 94.2±1.7         | N/A              |
|                  | FPR (1/h)  | 0.24±0.03    | 0.14±0.03  | 0.18±0.08 | 0.24±0.05  | 0.09±0.02        | N/A              |
| r = 1/2          | Accuracy (%) | 85.2±1.4     | 88.7±0.9   | 84.3±3.9 | 83.1±1.6   | 89.4±1.4         | 90.3±1.3         |
|                  | Sensitivity (%) | 88.1±2.6   | 92.4±1.3   | 84.7±3.3 | 88.4±3.8   | 92.8±1.6         | 93.9±1.6         |
|                  | FPR (1/h)  | 0.25±0.04    | 0.15±0.04  | 0.19±0.10 | 0.25±0.06  | 0.13±0.06        | 0.13±0.05        |
| r = 1/4          | Accuracy (%) | 85.1±1.6     | 88.1±1.2   | 83.7±3.7 | 81.6±1.8   | 88.4±1.6         | 89.7±1.2         |
|                  | Sensitivity (%) | 87.9±2.2   | 91.9±1.4   | 82.6±3.4 | 86.4±3.7   | 92.6±1.7         | 93.5±1.5         |
|                  | FPR (1/h)  | 0.27±0.05    | 0.17±0.04  | 0.24±0.12 | 0.26±0.07  | 0.15±0.08        | 0.12±0.04        |
| r = 1/8          | Accuracy (%) | 84.2±1.7     | 88.3±1.3   | 82.9±4.2 | 80.7±1.6   | 86.7±1.6         | 89.8±1.2         |
|                  | Sensitivity (%) | 87.6±2.5   | 91.7±1.6   | 81.4±3.6 | 84.5±3.8   | 92.8±1.7         | 93.4±1.2         |
|                  | FPR (1/h)  | 0.26±0.11    | 0.18±0.05  | 0.27±0.16 | 0.26±0.09  | 0.19±0.05        | 0.13±0.03        |
| r = 1/16         | Accuracy (%) | 82.5±1.6     | 87.6±1.2   | 80.6±4.6 | 80.2±1.7   | 88.2±1.9         | 89.7±1.3         |
|                  | Sensitivity (%) | 79.8±2.7   | 90.2±1.4   | 82.2±3.5 | 83.3±4.1   | 90.4±1.8         | 93.3±1.3         |
|                  | FPR (1/h)  | 0.29±0.09    | 0.21±0.05  | 0.30±0.14 | 0.27±0.08  | 0.17±0.05        | 0.13±0.03        |

**TABLE III**

| Patient ID | Metrics | r = 1/2 | r = 1/4 | r = 1/8 | r = 1/16 |
|------------|---------|---------|---------|---------|----------|
| chb01      | PSNR    | 40.29   | 42.08   | 40.46   | 35.42    |
|            | PCC     | 0.97    | 0.97    | 0.96    | 0.91     |
| chb05      | PSNR    | 37.40   | 33.21   | 32.06   | 32.68    |
|            | PCC     | 0.96    | 0.94    | 0.91    | 0.90     |
| chb06      | PSNR    | 52.98   | 36.31   | 39.48   | 35.14    |
|            | PCC     | 0.96    | 0.95    | 0.93    | 0.89     |
| chb07      | PSNR    | 36.83   | 32.45   | 31.91   | 31.09    |
|            | PCC     | 0.92    | 0.81    | 0.82    | 0.79     |
| chb08      | PSNR    | 37.95   | 33.46   | 32.89   | 31.64    |
|            | PCC     | 0.96    | 0.86    | 0.86    | 0.85     |
| chb09      | PSNR    | 38.53   | 32.76   | 33.74   | 31.28    |
|            | PCC     | 0.91    | 0.76    | 0.83    | 0.73     |
| chb10      | PSNR    | 44.14   | 40.17   | 39.01   | 37.63    |
|            | PCC     | 0.98    | 0.96    | 0.94    | 0.93     |
| chb14      | PSNR    | 43.55   | 45.92   | 41.87   | 37.38    |
|            | PCC     | 0.99    | 0.99    | 0.99    | 0.97     |
| chb18      | PSNR    | 34.09   | 34.77   | 34.28   | 32.42    |
|            | PCC     | 0.77    | 0.85    | 0.77    | 0.79     |
| chb19      | PSNR    | 39.12   | 39.34   | 39.69   | 35.59    |
|            | PCC     | 0.93    | 0.93    | 0.91    | 0.87     |
| chb22      | PSNR    | 38.83   | 35.50   | 33.94   | 31.62    |
|            | PCC     | 0.93    | 0.85    | 0.84    | 0.85     |
| Average    | PSNR    | 40.34   | 36.91   | 36.30   | 33.81    |
|            | PCC     | 0.93    | 0.90    | 0.89    | 0.86     |

**Fig. 5.** Different states of an epileptic seizure. Seizure prediction horizon (SPH) is a short period between preictal state and seizure onset. The preictal interval length (PIL) is equal to the length of a preictal state.

**B. Experimental Setup**

We set the number of filters of the convolution layer of the first upsampling block in the reconstruction network to
be \( \# \text{filters}_{\text{recon}} \) with each subsequent block doubling the number of filters of the previous block. Due to the high variability of EEG signals across different individuals, we train the prediction network in a patient-specific manner. We refer to the number of filters of the stem layer as \( \# \text{filters}_{\text{stem}} \). For each convolution layer of basic convolution block \( l \), we set its number of filters to \( 2 \times l \times \# \text{filters}_{\text{stem}} \). We denote the hidden linear layer size of the prediction network as \( \text{size}_{\text{fc}} \). For optimization, we adopt the Adam optimizer with learning rate \( lr \in \{1e-5, 5e-5, 1e-4, 5e-4, 1e-3, 5e^{-3}, 1e-2\} \) and train each model with a fixed epoch number of 150. Then, for each patient, we sweep over \( \# \text{filters}_{\text{stem}} \in \{4, 8, 16, 32\} \), \( \text{size}_{\text{fc}} \in \{25, 50, 100\} \) and batch size in \( \# \text{filters}_{\text{stem}} \in \{4, 8, 16, 32\} \) to choose the hyper-parameter setup that gives the highest prediction accuracy on the validation set. Our model is implemented using the Pytorch framework and trained end-to-end on NVIDIA 2080Ti GPUs for acceleration.

C. Comparison With State-of-the-Art

We evaluate our proposed framework with the following metrics: accuracy, sensitivity, and false prediction rate (FPR) for seizure prediction; Pearson’s correlation coefficient (PCC) and peak signal-to-noise ratio (PSNR). Given signal \( x_i^c \) of channel \( c \) and time stamp \( i \) and its reconstructed version \( \hat{x}_i^c \), PCC and PSNR are defined as follows:

\[
PCC = \frac{1}{C} \sum_{c=0}^{C} \frac{\sum_{i=0}^{N}(x_i^c - \mu^c)(\hat{x}_i^c - \hat{\mu}^c)}{\sqrt{\sum_{i=0}^{N}(x_i^c - \mu^c)^2 \sqrt{\sum_{i=0}^{N}(\hat{x}_i^c - \hat{\mu}^c)^2}}},
\]

where \( \mu^c \) and \( \hat{\mu}^c \) denotes the average value at channel \( c \) of original signal and reconstructed signal, respectively.

\[
PSNR = 10 \cdot \log_{10}\left(\frac{\text{max}(x_i^c)}{MSE}\right),
\]

where \( MSE \) is defined as follows:

\[
MSE = \frac{1}{C \cdot N} \sum_{c=0}^{C} \sum_{i=0}^{N} \| x_i^c - \hat{x}_i^c \|^2,
\]

where \( C \) is the total number of channels and \( N \) is the total length of the signal, \( \text{max} \) stands for the maximum value. Higher PSNR and PCC values indicate better reconstruction performance. We compare the performance of our proposed method with the following deep learning based baselines:

- Lightweight solution [25]: The lightweight solution is based on CNN; it uses synchronization features calculated by Pearson correlation coefficient [42] on all EEG channels as input.
- End-to-End approach [24]: The End-to-End patient-specific approach is also based on CNN; it adopts 1-dimensional (1D) kernels in the early-stage convolution and 2D kernels in the late-stage.
- EEGNet [26]: EEGNet uses compact CNN architecture which contains temporal convolution, depthwise convolution, separable convolution, and pointwise convolution.
- STFT CNN [35]: The STFT CNN performs STFT to the EEG signals and performs the classification with a 3-layer CNN.

![Fig. 6. A demonstration of reconstructed signals w.r.t. the original signals of both training and testing set with compression ratio 1/8.](image)

We roughly divide the baselines into two typical types, i.e., methods using original EEG signal as input and methods using statistics of EEG signal as input. Truong et al. [35], and Zhang et al. [25] use the result of STFT and PCC of the original signal as the input to the neural network, respectively. We follow the same protocol on the original and compressed signal during comparison.

To ensure impartial comparison, we closely follow the setup in the original work of the baselines to reproduce their approaches to the best of our effort using Pytorch. All models are trained and evaluated using the same dataset split and preprocessing method as described in Section IV-A. We also train each baseline method for 150 epochs and grid search training related hyperparameters as described in Section IV-B. We adopt a random compression matrix with each element subject to the Gaussian distribution as the sensing matrix for EEG compression of baseline methods. To demonstrate the stability of our approach, we report the averaged prediction accuracy, sensitivity, and FPR over all patients listed in Table I of all methods with compression ratios 1/2, 1/4, 1/8 and 1/16 in Table II. We also report the result on the original signal for reference. The best performance is marked in bold. N/A stands for results under reconstruction is not applicable since reconstruction is not required to the original signal with no compression applied. Notably, our method outperforms all other baseline algorithms in all metrics with compression ratios of 1/2, 1/4, and 1/16, which shows the effectiveness of the proposed framework. We yield slightly lower, but comparable sensitivity than Zhang et al. [25] with a compression ratio of 1/8. In addition, our model yields the minimum performance variation of 0.6% in accuracy with the compression ratio ranging from 1/2 to 1/16, indicating a robust seizure prediction performance under various compression ratios. The two methods [24], [26] using original EEG signal as input suffers more performance drop than methods using statistics [25], [35] as input under various signal compression scenarios. This finding shows that using signal statistics as input gives a more stable performance when using the compressed signals as input. Using random matrices such as Gaussian or Bernoulli as the sensing matrix cannot capture informative statistics/features for downstream tasks. Consequently, performing seizure prediction on compressed signals with information loss yields poor performance, which explains why the two original signal
based methods perform worse. However, using manually extracted statistics as input, on the other hand, discards information embedded in the original signal other than certain specific features and thus weakens the deep neural network’s generalized feature extraction ability. Our proposed framework kills two birds with one stone by optimizing the sensing matrix together with downstream tasks, which captures informative statistics/features during compression automatically by the network. Thus, compressing the original signal with the optimized sensing matrix embeds seizure prediction related information into the compressed signals. Additionally, Zhang et al.’s method outperforms other baseline methods. This performance gain may derive from the practice of extracting the cross-channel correlation coefficient, which shares a similar idea to adopting the 1D convolution in our proposed prediction network. From Table II, we also observe that jointly training reconstruction task together with the prediction task ($\lambda = 1$) yields higher accuracy and sensitivity compared to training prediction task alone ($\lambda = 0$). In addition to extracting features for seizure prediction, the reconstruction module also guides the model to learn hidden representations that are autoregressive so as to improve the model’s generalization ability. This finding corresponds to our intuition that the reconstruction task can serve as a regularizer to the prediction task.

To demonstrate the reconstruction performance of our proposed framework, we report PCC and PSNR w.r.t. different compression ratios in Table III. As shown in this table, with a compression ratio of 1/2, our approach yields an average PSNR of 40.63 and PCC of 0.94. With compression ratio 1/16, we observe a reconstruction performance degradation of 6.53 and 0.07 in PSNR and PCC, respectively. These results demonstrate the effectiveness and robustness of the reconstruction ability of our proposed framework regardless of different compression ratios. Besides numerical metrics, we also provide a visualization example of the reconstructed signal of our method in Fig. 6. It is observed from the figure that the reconstructed signal of both the training set and test set visually resembles the original signal to a large extent. Next, we consider the case where reconstruction for visualization is negligible, and the main focus is seizure prediction. In this case, we could further reduce the bit precision of the compression matrix to the binary scenario. We show seizure prediction performance without reconstruction under different compression ratios using a binary compression matrix in Table IV. For baseline methods, we adopt a random compression matrix with each element subject to the Rademacher distribution. As shown from Table IV, the four baseline methods suffer from an accuracy performance degradation of 8.9% on average with a signal compression ratio of 1/16. Our proposed method only shows a degradation of 3.1% in accuracy with a signal compression ratio of 1/16. Compared to using a random Gaussian matrix as compression matrix (floating-point), baseline methods demonstrate a drastic performance degradation as the compression ratio ranges from 1/2 to 1/16. In contrast, our proposed method shows similar seizure prediction performance when using floating-point and binary compression matrices. This performance further proves that learning compression matrices along with the downstream task can embed informative features into the compressed signal and thus yield an ideal trade-off between seizure prediction performance and signal transmission power consumption. Finally, we demonstrate a visualization of both the floating-point and binary compression matrices learned with a compression ratio of 1/16 for different patients in Fig. 7. The first row shows the binary compression matrices for subjects one, seven, and 14, while the second row shows the corresponding floating-point compression matrices. It is observed from Fig. 7 that the compression matrices vary significantly across patients regardless of the bit precision suggesting that the compression matrices extract patient-specific features since seizure-generating mechanisms.
TABLE IV
PREDICTION PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS USING BINARY COMPRESSION MATRIX

| Compression ratio | Metrics       | Lawhern [26] | Zhang [25] | Xu [24] | Truong [35] | This work (λ = 0) |
|------------------|---------------|--------------|-----------|---------|-------------|------------------|
| Original signal  | Accuracy (%)  | 87.2±1.3     | 89.9±0.8  | 84.8±1.7 | 83.4±1.4    | 92.5±1.2         |
|                  | Sensitivity (%) | 87.9±2.4    | 93.3±1.4  | 85.6±3.4 | 89.7±2.9    | 94.2±1.7         |
|                  | FPR (fh)      | 0.24±0.03    | 0.14±0.03 | 0.18±0.08 | 0.24±0.05   | 0.09±0.02        |
| r = 1/2          | Accuracy (%)  | 84.8±1.4     | 88.2±1.0  | 80.2±1.4 | 83.2±1.4    | 90.8±1.4         |
|                  | Sensitivity (%) | 87.9±2.6    | 89.9±2.3  | 86.3±3.3 | 88.4±2.8    | 92.8±1.6         |
|                  | FPR (fh)      | 0.20±0.04    | 0.21±0.02 | 0.26±0.03 | 0.25±0.07   | 0.15±0.06        |
| r = 1/4          | Accuracy (%)  | 80.2±1.6     | 83.1±1.2  | 78.7±1.3 | 83.0±1.4    | 90.6±1.3         |
|                  | Sensitivity (%) | 83.4±3.2    | 86.4±2.8  | 86.4±2.8 | 87.3±3.8    | 92.6±2.3         |
|                  | FPR (fh)      | 0.30±0.05    | 0.29±0.04 | 0.29±0.04 | 0.20±0.05   | 0.18±0.04        |
| r = 1/8          | Accuracy (%)  | 77.7±2.2     | 82.0±1.4  | 76.7±1.2 | 80.4±1.9    | 89.9±1.4         |
|                  | Sensitivity (%) | 80.1±4.7    | 85.5±2.4  | 79.7±4.4 | 84.2±4.9    | 87.8±3.1         |
|                  | FPR (fh)      | 0.25±0.06    | 0.27±0.06 | 0.24±0.07 | 0.21±0.05   | 0.21±0.05        |
| r = 1/16         | Accuracy (%)  | 76.8±2.7     | 79.9±1.5  | 74.6±1.7 | 78.2±1.8    | 89.4±1.5         |
|                  | Sensitivity (%) | 79.3±4.0    | 83.4±2.4  | 76.8±5.0 | 83.8±4.6    | 85.8±2.8         |
|                  | FPR (fh)      | 0.27±0.06    | 0.30±0.07 | 0.30±0.05 | 0.29±0.06   | 0.25±0.06        |

TABLE V
MAC AND ENERGY CONSUMPTION FOR ONE INFERENCE UNDER DIFFERENT COMPRESSION RATIOS

| Compression Ratio | MACs (K) | Energy Consumption (μJ) |
|------------------|----------|-------------------------|
| Original Signal  | 1050     | 5.15                    |
| r = 1/2          | 528.84   | 2.59                    |
| r = 1/4          | 267.72   | 1.31                    |
| r = 1/8          | 137.16   | 0.67                    |
| r = 1/16         | 71.88    | 0.35                    |

vary across patients. Our proposed framework captures subject-specific informative features during the compression process, which in turn improves seizure prediction performance.

D. Framework Efficiency Analysis

In this study, we next show that the proposed network not only reduces the transmission bandwidth requirement and corresponding power consumption but also improves the performance of the seizure prediction module in terms of power and time consumption performance. Recall that our proposed framework reduces the input signal length via the compression module and that the number of neural network operations (addition and multiplication) associated with seizure prediction also decreases, leading to a real-time energy-efficient seizure prediction system. We use our proposed seizure prediction network as an example to show operation reduction and energy consumption resulting from our approach [43]. The energy consumption for various operations is calculated based on a 45 nm technology. In particular, we report Multiply–Accumulate Operations (MACs) and corresponding energy consumption for one inference in Table V using various compression ratios. The energy consumption and MACs of using the original signal for prediction are also provided as references. Notably, the reduction in the MACs associated with signal compression is approximately proportional to the signal compression ratios indicating a much faster inference speed through the proposed compression module. Consequently, we also observe a significant energy consumption reduction due to fewer MACs.

V. Conclusion

Signal transmission related power consumption is the main bottleneck for power-critical wearable and implantable systems. In this paper, we proposed a novel joint compression and classification framework that significantly reduces the power consumption of both the signal compression stage and classification stage. The learned compression matrix can be deployed on the wearable or implantable sensor end in a hardware-friendly fashion for signal compression to reduce transmission bandwidth requirement and corresponding power consumption. The compressed signals can be directly employed for prediction and reconstruction. Our proposed framework is generic and can be applied to any deep learning-based seizure prediction algorithm. Extensive experiments over a benchmark dataset show that the proposed approach not only significantly improves the system power efficiency but also helps improve prediction accuracy. Our framework also supports signal reconstruction for clinical expert diagnosis via visualization. Our proposed framework could be easily extended to scenarios such as brain-machine-interfaces and EMG measurements, showing great potential in real-world wearable and implantable biomedical applications. Two potential limitations of this work are the requirement of patient-specific modeling and poor reconstruction using a binary compression matrix. Patient-independent compression can alleviate the heavy burden of data annotation, while a low-precision compression matrix further reduces hardware memory consumption. We will also address the mentioned limitations in future works.
