FETCH: A Fast and Efficient Technique for Channel Selection in EEG Wearable Systems

Alireza Amirshahi[1] ALIREZA.AMIRSHAHI@EPFL.CH
Jonathan Dan[1] JONATHAN.DAN@EPFL.CH
José Miranda[1] JOSE.MIRANDACALERO@EPFL.CH
Amir Aminifar[2] AMIR.AMINIFAR@EIT.LTH.SE
David Atienza[1] DAVID.ATIENZA@EPFL.CH
[1] *Embedded Systems Laboratory, EPFL, Switzerland*
[2] *Department of Electrical and Information Technology, Lund University, Sweden*

Abstract

The rapid development of wearable biomedical systems now enables real-time monitoring of electroencephalography (EEG) signals. Acquisition of these signals relies on electrodes. These systems must meet the design challenge of selecting an optimal set of electrodes that balances performance and usability constraints. The search for the optimal subset of electrodes from a larger set is a problem with combinatorial complexity. While existing research has primarily focused on search strategies that only explore limited combinations, our methodology proposes a computationally efficient way to explore all combinations. To avoid the computational burden associated with training the model for each combination, we leverage an innovative approach inspired by few-shot learning. Remarkably, this strategy covers all the wearable electrode combinations while significantly reducing training time compared to retraining the network on each possible combination. In the context of an epileptic seizure detection task, the proposed method achieves an AUC value of 0.917 with configurations using eight electrodes. This performance matches that of prior research but is achieved in significantly less time, transforming a process that would span months into a matter of hours on a single GPU device. Our work allows comprehensive exploration of electrode configurations in wearable biomedical device design, yielding insights that enhance performance and real-world feasibility.

Data and Code Availability This paper uses two publicly available datasets: the Temple University Hospital EEG Seizure Corpus -v2.0.0 (Shah et al., 2018) and the motor imagery paradigm in the OpenBMI dataset (Lee et al., 2019). The code for this work is available at https://github.com/esl-epfl/FETCH.

Institutional Review Board (IRB) This research does not require IRB approval.

1. Introduction

Electroencephalography (EEG) is the main tool to monitor and diagnose various health conditions such as epilepsy, sleep disorders, stroke or coma (Siuly et al., 2016). In clinical environments, a full-channel EEG system, which involves a cap equipped with numerous electrodes, is the gold standard for monitoring the brain's electrical activity. Despite the high-quality data this system provides, it is not well suited to home-based monitoring due to the social stigma of its daily use (Duun-Henriksen et al., 2020). Therefore, there is a growing trend towards wearable EEG devices that involve fewer electrodes and a selection of a subset from the full channel set (Kim et al., 2019; Ingolfsson et al., 2021).

However, the electrode subset selection raises the fundamental question of how to design a wearable device that balances optimal performance with design and comfort constraints. Addressing this challenge requires a thorough exploration of the wearable design space to find the optimal wearable EEG solution.

An approach to explore the wearable design space is **channel selection** (Moctezuma and Molinas, 2020b; Dan et al., 2023). These studies investigate algorithms to identify the optimal electrode subset based on parameters such as the number of electrodes and abnormality detection performance. However, the number of combinations explored in the design

space is very limited. Additionally, other parameters of the wearable system, such as energy consumption or ergonomic factors, are not directly integrated into the selection process of the electrode subset.

Another approach to tackle the design space exploration challenge is **exhaustive search** methods in which models are trained and evaluated exhaustively for all possible electrode combinations (Lee et al., 2022b). In these methods, models are trained for each electrode combination to perform a detection task. Although this approach enables a comprehensive examination of wearable device constraints, it encounters scalability issues due to the time-consuming nature of training models for all possible electrode subsets. The inherent computational demand of training models in exhaustive search methodologies raises concerns about their feasibility, especially when tasked with selecting combinations from a larger number of channels or when employing complex machine learning (ML) models.

In this paper, we propose a novel approach to address the aforementioned challenges by reducing the number of required trained models in wearable design space exploration. We introduce this strategy that we call *FETCH*: a Fast and Efficient Technique for CHannel selection. Rather than training a model for each electrode combination, FETCH performs an inference step on all combinations of electrodes. Considering that the time dedicated to model inference is orders of magnitude less than that required for training, the exploration of the design space using FETCH is remarkably rapid. With FETCH, the exploration of all electrode configurations unlocks opportunities to consider a wide range of factors and parameters in the design of wearables.

The basis of FETCH is inspired by few-shot learning principles, in which a model is trained to perform classification using only a few examples per class in inference (Wang et al., 2020; Finn et al., 2017). By customizing the few-shot learning definition to our scenario, we leverage the power of meta-learning to accelerate the inference process.

To demonstrate the efficacy of our approach, we focus on a case study task of epileptic seizure detection using 19-electrode EEG signals, applied to a complex transformer model (Vaswani et al., 2017). We demonstrate that the performance of FETCH is comparable to exhaustive search methods, but our approach significantly reduces the time required to obtain the optimized electrode configuration. The process, which previously took 3500 hours, can now be completed in as little as 26 hours on a single GPU device.

The contributions of this study are as follows:

- We have designed a novel method inspired by few-shot learning, referred to as FETCH, proposed to facilitate a rapid design space exploration for channel selection.
- Given that FETCH can deliver performance results for all electrode configurations, we have optimized a design cost for wearable devices. This optimization takes into account factors such as performance, energy consumption, and ergonomic metrics.
- We have performed an evaluation of FETCH compared to exhaustive search and channel selection methods. Our results demonstrate that the performance is not significantly different, and simultaneously, the time required to optimize the electrode configuration is drastically reduced by up to 135 times.

The rest of this paper is organized as follows. Section 2 discusses related work, offering important insights on channel selection and exhaustive search methods. Section 3 provides a background on few-shot learning. Section 4 introduces our proposed methodology, FETCH, and its combination with a transformer model. In Section 5, we detail the experimental setup employed in our study. Moving on to Section 6, we present and analyze the results obtained, comparing them with previous works. Furthermore, this section shows the design cost of optimizing various parameters. Finally, Section 7 summarizes our conclusions.

2. Related Works

2.1. Channel selection

Channel selection methods are designed to identify a subset of electrodes from full-channel EEG signals that maximize performance. A popular approach is the greedy search method (Moctezuma and Molinas, 2020a). In this method, the process starts with the full-channel EEG configuration. Subsequently, a performance metric is evaluated for configurations with one electrode removed. Among these, the configuration yielding the highest performance is selected. This iterative process of removing electrodes one by one, known as sequential backward selection (SBS),

continues until the desired electrode count is reached. Alternatively, greedy searches can begin with a single electrode, and incrementally add the most informative electrode at each step (Aydemir and Ergün, 2019; Choi et al., 2012). This process of progressively building the configuration is called Sequential Forward Selection (SFS).

Another approach involves non-dominant sorting genetic algorithms II (NSGA-II) (Deb et al., 2002). As Moctezuma and Molinas (2020b) suggests, these algorithms start with various random electrode configurations called initial population. Iteratively, electrodes are either removed or added in successive steps called generations. In Moctezuma and Molinas (2020b), the algorithm pursues a multi-objective goal balancing accuracy and electrode count. The process continues until this multi-objective function is met or a predetermined number of generations elapses.

While these channel selection methods balance performance accuracy and electrode count during training, they may not explore all possibilities, potentially converging on sub-optimal solutions (Narayanan et al., 2020). Furthermore, the goal to optimize is limited to the accuracy and number of electrodes, while the relation of the electrodes and the ergonomic factors of a wearable integrating the electrodes are not considered. On the other hand, FETCH ensures that no potential configuration is overlooked. Our approach is scalable and encompasses multiple objectives, including electrode count, energy consumption, and ergonomic factors of wearable systems.

2.2. Exhaustive search

Lee et al. (2022b) introduced an approach for selecting the ideal electrode configuration in EEG-based wearable devices featuring 32 scalp electrodes, focusing on detecting mild cognitive impairment. They employed a Support Vector Machine (SVM) to train and evaluate 4396 distinct models corresponding to two-, four-, six-, and eight-electrode symmetric configurations. Gelbard-Sagiv et al. (2023) explored a similar approach for seizure detection using Light Gradient Boosting Machine (LightGBM) models. They trained 17000 models on various random electrode configurations and identified eight electrodes as optimal. A time-consuming analysis of all eight-electrode combinations was then performed to identify the most effective subset.

Despite the promising results of the two methods mentioned (Lee et al., 2022b; Gelbard-Sagiv et al., 2023), their methodologies face scalability challenges. In particular, they do not report the time spent training all possible electrode combinations, which is exponentially related to the number of electrodes. This is an even more critical factor in more complex diagnostic tasks. For example, in a dense array EEG system with 256 electrodes (Holmes, 2008), the number of combinations will increase with combinatorial complexity. Furthermore, their reliance on relatively simple models like SVM and LightGBM limits applicability. However, advanced models, including convolutional neural networks, graph neural networks, and transformers, are commonly used in biomedical applications (Busia et al., 2024; Tang et al., 2023; Gómez et al., 2020; Hannun et al., 2019). These models require substantially longer training times, making them impractical for training in every electrode combination. On the contrary, FETCH is a scalable approach compatible with complex deep learning models and removes the need for training of every electrode combination, relying instead on training a few models and a rapid inference process.

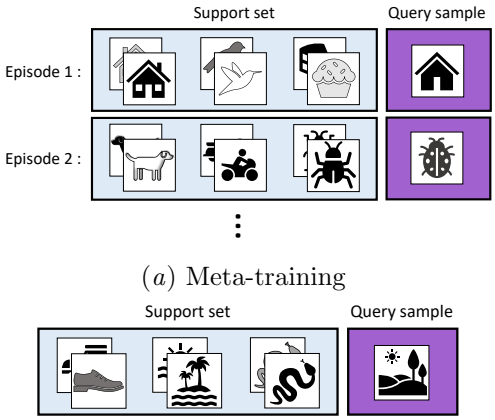
3. Background

FETCH is inspired by few-shot learning. Before describing our methodology, we first provide a concise overview of the background of few-shot learning in this section. Subsequently, in the following section, we explain the derivation of FETCH within the framework of the original few-shot learning context.

Few-shot learning methods aim at designing a model that classifies a test sample based on a very limited number of previously seen labeled examples (Finn et al., 2017). Figure 1(b) shows an example of a few-shot learning *task*. In this figure, the query sample needs to be classified using only a small group of labeled samples, referred to as *support set*.

Few-shot learning approaches address this challenge by simulating the scenario of limited labeled samples during training. The model is trained on a large dataset to learn the similarities and differences among a few objects. This model training step is called *meta-train*, shown in Figure 1(a).

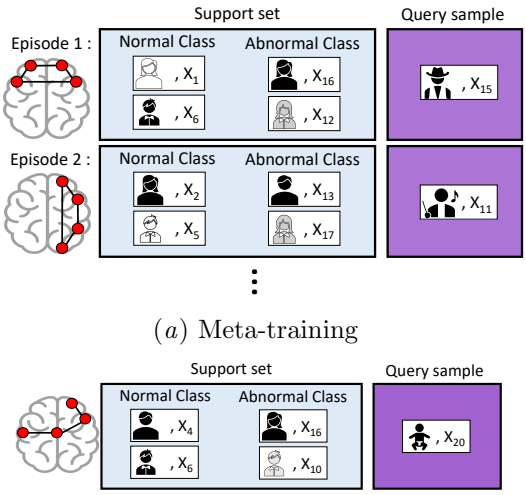
The training step, as shown in Figure 1(a), involves several *episodes*. In each episode, N different classes are randomly selected from the large training dataset, and k samples are chosen from each class. These samples collectively form a k-shot N-way support set for each episode (2-shot 3-way in Figure 1(a)).



(a) Meta-training

(b) Inference

Figure 1: Meta-training and inference phases in few-shot learning, highlighting the creation of episodes with three distinct classes and two samples randomly selected from a dataset. The inference process involves new classes in the support set and query sample, mirroring the tasks in the meta-training phase.



(a) Meta-training

(b) Inference

Figure 2: Meta-training and inference in FETCH. In meta-training, the electrode configuration varies through the episodes, enabling the model to quickly adapt to a new electrode combination during the inference phase.

Let $\mathcal{D}^{\text{train}}$ be the dataset for meta-training. The support set for an episode is randomly selected from $\mathcal{D}^{\text{train}}$ to create $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{k \cdot N}$. The goal in this episode is to classify new query samples $\mathbf{x}_q$ that belong to one of these N classes. The loss function during meta-training aims to minimize the log-likelihood of predicted class $\hat{y}_q$ of the query sample compared to the corresponding ground truth y_q.

In the inference process, we again have a support set with the same number of classes and samples, and the goal is to classify the unseen query sample. This task mirrors the tasks in the meta-train.

4. Method

This section is structured into the following subsections. Initially, in Section 4.1, we present our proposed FETCH method. In the subsequent section (Section 4.2), we explain the combination of FETCH with a transformer model for a seizure detection task. Finally, in Section 4.3, we propose a design cost to show the impact of FETCH on selecting the optimal electrode combination, considering seizure detection performance, energy efficiency, and ergonomic levels in a wearable device.

4.1. Channel adaptation with few-shot learning

Figure 2 illustrates the meta-training and inference phases defined in FETCH. As shown in Figure 2(*a*), similar to the original definition of few-shot learning, FETCH consists of multiple episodes in meta-training. Within each episode, a random electrode set is selected, and subsequently, the support set is formed, encompassing k random samples in each class based on the selected electrode combination. The samples can be drawn from various users in the dataset. The query sample in this episode undergoes the same electrode configuration but from a different user than the support set. This episode formation is repeated multiple times in meta-training. Subsequent episodes involve different electrode sets but share the same classes.

During the inference phase, illustrated in Figure 2(*b*), the system encounters an electrode configuration that may be entirely new to the system. Given its meta-training on various electrode configurations, the model is now capable of quickly adapting to the new electrode combination and classifying new samples in the inference. The adaptation process is accelerated because it necessitates only a few data points

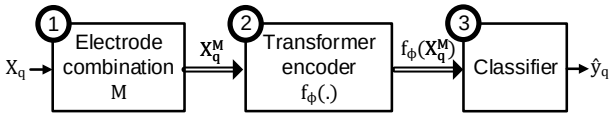


Figure 3: The processing sequence in FETCH and its integration into a transformer model.

or shots from the support set that share the same electrode configuration.

While our approach is derived from few-shot learning, FETCH diverges from conventional few-shot learning in three key aspects. First, it is tailored to biomedical abnormality detection tasks, which involve two classes, such as seizure and non-seizure. Consequently, each episode in FETCH uniformly encompasses these classes. The second difference is that we structure these few-shot learning episodes so that the support sets and query samples are derived from the same electrode configuration, which differs from the configurations in the other episodes. This arrangement enables the model to learn to adapt rapidly when a new electrode configuration is introduced into the query sample. Lastly, in the conventional version of few-shot learning (Finn et al., 2017), the main limitation is the number of annotated samples in the unseen classes. However, in FETCH, the constraint is not the lack of annotated data in the target electrode configuration, but the computational cost of adapting to it.

4.2. Integration into a transformer model

To establish the principles of FETCH and its integration into a transformer model, we represent an input EEG sample as $\mathbf{x}_q$, encompassing data from all electrodes (full-channel). We define an electrode combination, denoted by M, as a specific subset of electrodes chosen from the full set. Figure 3 illustrates this concept as the initial block in the FETCH pipeline. Consequently, $\mathbf{x}_q^M$ denotes the modified version of $\mathbf{x}_q$, where only the electrodes belonging to set M retain their original data. The other electrodes are masked with a predetermined value. In the context of FETCH, M varies with each episode during meta-training, indicating that a different set of EEG electrodes is selected randomly from feasible sets.

Our approach utilizes a transformer model inspired by Ma et al. (2023). Initially, the EEG signal $\mathbf{x}_q^M$ is converted into a Short-Time Fourier Transform (STFT) representation, which is then fed into the transformer encoder. This transformer model's architecture parallels that of a Vision Transformer (Dosovitskiy et al., 2020), with STFT as the input image. In the STFT representation, the y-axis corresponds to the frequency range, while the x-axis encompasses both the time and channel dimensions. The transformer encoder is shown in Figure 3 as the second block in the processing sequence.

The Vision Transformer model has multiple encoder layers. Following the established practice in transformers (Devlin et al., 2018), a special class token [CLS] in the form of a vector is prepended to the input fed into the transformer. The feature vector produced by the final transformer encoder layer for this class token is expressed as $f_\phi(\mathbf{x}_q^M)$. Here, $f_\phi(.)$ denotes the function of the transformer encoder.

The classifier block, in Figure 3, follows the principles of prototypical networks (Snell et al., 2017). In each episode, *prototypes* $\mathbf{p}_n^M$ are computed for each class n within the support set S_n^M. The prototypes are the average of the feature vectors $f_\phi(\mathbf{x}_i^M)$, corresponding to all instances $\mathbf{x}_i$ belonging to class n in the support set S_n^M. Mathematically, the prototypes are calculated as follows:

$$\mathbf{p}_n^M = \frac{1}{|S_n^M|} \sum_{(\mathbf{x}_i^M, y_i) \in S_n^M} f_\phi\left(\mathbf{x}_i^M\right), \qquad (1)$$

Finally, the classification of the query $\mathbf{x}_q$ is achieved by comparing its feature vector, $f_\phi(\mathbf{x}_q^M)$, to the class prototypes $\mathbf{p}_n^M$. The query is assigned a class label $\hat{y}_q$ corresponding to the nearest prototype in this embedding space.

To obtain the seizure detection performance for all electrode configurations and explore the entire wearable design space, we process each sample $\mathbf{x}_q$ in the validation and unseen test sets through the pipeline depicted in Figure 3 for all possible electrode combinations M. The rationale behind acquiring performance metrics for all combinations is discussed in the following section.

4.3. Design cost representation

Having efficiently obtained results for all feasible configurations through FETCH, we can now define the design cost for wearable devices. This enables selecting the most suitable configuration based on specific

target constraints. Equation (2) shows the suggested design cost for wearable devices.

$$\text{Cost}(M) = w_1 \cdot P(M) + w_2 \cdot E(M) + w_3 \cdot U(M) \quad (2)$$

The design cost in Equation (2) incorporates three primary constraints. The first constraint, P, is performance. This performance shows how accurately a target wearable device can perform seizure detection. This value is obtained using FETCH applied to the transformer model on the validation/test set.

The second constraint, E, is the energy consumption of the target wearable device. In our case, to estimate energy consumption for each electrode configuration, we implemented the transformer model and performed inference on X-HEEP, an ultra-low power system specifically designed for biomedical signal processing in wearable devices (Machetti et al., 2024). Further information regarding the estimated energy consumption can be found in Section 6.

The third constraint, represented by U, assesses whether the wearable device, configured with the target electrode arrangement, is user-friendly and ergonomically suitable for daily use. To the best of our knowledge, no existing research has introduced a metric specifically designed to evaluate ergonomic or user-friendly factors. In this context, without loss of generality, we propose a metric as follows. We represent the electrode configuration as a graph, where the nodes represent the electrodes and the edges represent the connections. In this graph, the edges are weighted according to their pairwise Euclidean distance, as calculated and used in Tang et al. (2021). In this work, we suggest using the diameter of the graph as an ergonomic metric. The diameter is the longest path among all shortest paths between any two nodes. This metric implies that, with the same number of electrodes, a configuration in which the electrodes are closer to each other is considered more user-friendly. This is a basic proposition for the ergonomic metric, and any similar approach would be applicable in our study.

In Equation (2), w_1, w_2, and w_3 are the weights assigned to each constraint, reflecting their relative importance in the overall design cost. These weights can be determined based on the specific requirements and priorities of the target wearable design. Adjusting these weights allows wearable designers to emphasize certain aspects based on the design priorities.

5. Experimental Setup

5.1. Dataset

The dataset used in our study comprises the publicly available Temple University Hospital EEG Seizure Corpus (TUSZ)-v2.0.0 (Shah et al., 2018), encompassing data from 675 subjects with a cumulative duration of 1476 hours. The dataset features heterogeneity in the sampling frequency and number of EEG electrodes. Therefore, to ensure uniformity of sampling frequency, all files are re-sampled to 250 Hz. Also, as suggested in Dan et al. (2024), N=19 EEG electrodes are used in this study.

The spatial arrangement of EEG channels in TUSZ-v2.0.0 follows the 10-20 system, and we adopt the longitudinal bipolar montage (also called a double banana montage). In a bipolar montage, each electrode is paired with another electrode in the graph. Hence, when identifying all feasible subsets, we adhere to the rule that an electrode configuration is considered valid only if it does not result in any isolated nodes in the graph, as per the longitudinal bipolar montage setup. To maintain the generality of the FETCH method, in this application no additional restrictions are imposed on the validity of the subset.

In this dataset, the training set, validation set, and test set are provided in separate directories. We follow the same split in our experiments. The TUSZ dataset has been widely used in previous works for the seizure detection task (Lee et al., 2022a; Rahmani et al., 2023).

5.2. TSD model configuration

The baseline model follows the architecture proposed in Ma et al. (2023), referred to by the authors as a Transformer for Seizure Detection (TSD). Designed for a seizure detection task, this model processes the input EEG signal by transforming it into STFT. In our framework (refer to Figure 3), each TSD model is specifically trained for a unique electrode combination by excluding (rather than masking) the non-selected electrodes from the input signal (First block). Notably, while both FETCH and TSD share the same transformer encoder (second block), TSD incorporates a different classifier (Third block). The classifier is a fully connected layer designed to classify the input signal as seizure or non-seizure.

For the training process, the authors use Adam optimizer (Kingma and Ba, 2014), with a learning rate of 3e-5. We set the maximum epochs of 100 with

early stopping criteria on the validation set performance metric and patience of 10 epochs.

5.3. FETCH model configuration

In the context of our experimental setup, our models are trained on EEG windows of duration 12 seconds, a parameter that is chosen by the previous works (Tang et al., 2021; Ma et al., 2023) to balance the need for temporal context and computational efficiency. In our study, the training process is guided by the cross-entropy loss function.

For the optimization process, we employ the same optimizer as TSD. The FETCH model is trained for a maximum of 500 epochs, with early stopping criteria implemented to prevent overfitting. This early stopping is predicated on the accuracy of the validation set, halting the training if the accuracy does not increase for 10 consecutive epochs. The few-shot learning parameters are chosen based on the results of the validation set as 200 episodes in each epoch and $k = 25$ shots for each class.

The computational resources utilized for this experiment are equipped with 512 GB of RAM and a single Tesla V100 GPU with 32 GB of memory.

In our study, we used the AUC (Area Under the Receiver Operating Characteristics) metric, a widely recognized evaluation measure for binary classification tasks such as seizure detection (Tang et al., 2021). It offers a comprehensive view of the model's discriminative power across various classification thresholds between the true positive rate and the false positive rate.

6. Results

6.1. Comparison with previous works

In this section, we compare FETCH with channel selection methodologies, including Sequential Forward Selection (SFS), inspired by Aydemir and Ergün (2019); Choi et al. (2012), Sequential Backward Selection (SBS), based on Moctezuma and Molinas (2020a), and NSGA-II as discussed in Moctezuma and Molinas (2020b). The NSGA-II method is trained on a population size of 20 over a maximum of 20 generations. Although the mentioned references use diverse models, datasets, and tasks, for a fair comparison, we applied these methodologies to a consistent framework comprising the TSD model and the TUSZ dataset, specifically for the seizure detection task. Therefore, we reimplemented and performed model training and evaluation for all the methods.

6.1.1. Performance comparison

Initially, the performance of channel selection methods, exhaustive search methods, and FETCH as the proposed method are compared. Each method identifies the top five electrode configurations that yield the best AUC performance on the validation set. Subsequently, these top five configurations for each method are evaluated on the unseen test set, and the average performance is reported in Table 1. For a fair comparison, we train separate TSD models for the top five configurations identified by FETCH and evaluate them on the test set. Selecting the five best channel configurations aims to ensure robust results, accounting for potential performance variations between the validation and test sets.

In Table 1, the performance results on the unseen test set, 95% confidence intervals (CI), and GPU time spent for identifying the top configurations are presented. The table shows the results for channel selection, exhaustive search, and the proposed FETCH methods for 2-, 4-, and 8-electrode configurations. Given that the transformer model used in our experiments is larger than SVM and LightGBM used in the exhaustive search references (Lee et al., 2022b; Gelbard-Sagiv et al., 2023), the extensive time required to train all subsets with a transformer model makes a direct performance comparison impractical for cases in which more than hundreds of models should be trained.

As shown in the table, considering both the CI and the average AUC, the proposed FETCH method significantly outperforms SFS and SBS in most cases. Furthermore, there is no significant difference in performance between FETCH, NSGA-II, and exhaustive search methods. However, the time required to achieve the top electrode configurations with FETCH is up to 135x less than the other methods.

While the performance of FETCH is comparable to established methods, it achieves this level of performance in significantly less time. This highlights the efficiency of our proposed method for identifying suitable electrode configurations. Additionally, FETCH enables exploration of the entire design space, unlike channel selection methods that are limited to identifying only a few configurations. Obtaining the performance results for all the electrode combinations becomes even more important when incorporating other

Table 1: Results comparison with Aydemir and Ergün (2019); Choi et al. (2012); Moctezuma and Molinas (2020a,b); Lee et al. (2022b); Gelbard-Sagiv et al. (2023)

	2-electrode AUC (Mean ± CI)	Time	4-electrode AUC (Mean ± CI)	Time	8-electrode AUC (Mean ± CI)	Time	Training time complexity
			Channel selection methods				
SFS	0.853 ± 0.006	27H	0.858 ± 0.017*	31H	0.878 ± 0.016*	39H	$\mathcal{O}(N^2)$
SBS	0.813 ± 0.005*	82H	0.834 ± 0.013*	78H	0.907 ± 0.010	68H	$\mathcal{O}(N^2)$
NSGA-II	0.837 ± 0.023	258H	0.903 ± 0.009	258H	0.923 ± 0.003	258H	$\mathcal{O}(G \cdot P)$
			Exhaustive search methods				
All comb.	0.853 ± 0.006	27H	0.897 ± 0.015	175H	Timeout	≈ 3500H	$\mathcal{O}(2^N)$
Random K	0.853 ± 0.006	27H	0.897 ± 0.015	175H	Timeout	≈ 1000H	$\mathcal{O}(K \cdot N)$
			Proposed method				
FETCH	0.862 ± 0.014	4H	0.893 ± 0.013	5H	0.917 ± 0.003	26H	$\mathcal{O}(N)$

* The proposed method FETCH significantly outperforms these results.

factors, such as user-friendliness into wearable system design. We will discuss and present the design space exploration capabilities of FETCH in Section 6.3.

6.1.2. Training time complexity

In Table 1, we have also reported the training time complexity of each experiment, quantified by the number of models required to be trained in each methodology. In this table, N denotes the total count of electrodes in a full-channel setup. For example, within the 10-20 EEG configuration, N equals 19, whereas, in a dense array EEG system, which features a more concentrated arrangement of EEG electrodes, N is set at 256 (Holmes, 2008).

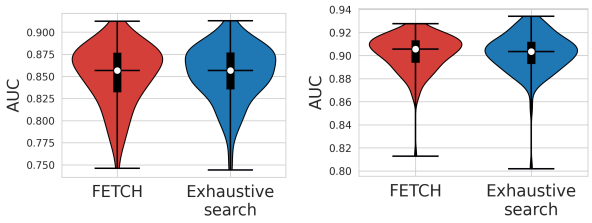
For SFS and SBS methods as the greedy search ones, in total $\frac{N(N+1)}{2}$ models must be trained. However, in the NSGA-II method which uses a genetic algorithm, the complexity of the training time is $\mathcal{O}(G \cdot P)$, where G represents the number of generations and P the population size. These parameters are chosen based on the problem's complexity. For more complex problems, larger population and generation sizes are required to adequately explore the solution space (Roeva et al., 2015). Specifically, in the context of selecting EEG electrodes for seizure detection, the number of electrodes N in the full-channel system can be seen as indicative of the complexity within the optimization's solution space. For instance, while Moctezuma and Molinas (2020b) uses 22 EEG channels, P is selected experimentally as 20, and G averaged over the cross-validation folds is 39.

Exhaustive search methods are more time-consuming. For example, in Lee et al. (2022b), training individual models for each configuration leads to exponential complexity, scaling as $\mathcal{O}(2^N)$. Similarly, in Gelbard-Sagiv et al. (2023), random K configurations are selected for each electrode count, with $K = 1000$, which is significantly larger than N. All these configurations are then trained, resulting in the complexity of $\mathcal{O}(K.N)$.

In contrast to previous methodologies, our proposed FETCH method requires training only a single model per desired number of electrodes. While training a single model can take up to 4 hours in our case, inference on the validation and test sets is significantly faster, taking only 17 seconds and 6 seconds, respectively. This highlights the key advantage of FETCH – its training complexity scales linearly with the number of electrodes, $\mathcal{O}(N)$, significantly lower than channel selection and exhaustive search methods. Although the inference step technically has a complexity of $O(2^N)$, its practical impact is negligible due to the three orders of magnitude different timescales of training and inference. It is important to note that Table 1 includes the inference time in the reported time values.

6.2. FETCH as an alternative to exhaustive search

Due to the exponential complexity of training models for all possible electrode combinations using exhaustive search, a direct performance comparison becomes impractical. To address this limitation, we randomly selected a total of 200 combinations of 4- and 8-electrode configurations from the feasible set. TSD models were trained on each selected configuration, generating a sample group of AUC performance representative of the exhaustive search. The correspond-



(*a*) 4 electrodes (*b*) 8 electrodes

Figure 4: The comparison between FETCH and exhaustive search models on 200 different random electrode configurations with 4 and 8 electrodes.

ing results for the same configurations were obtained using FETCH. We present the following analyses on these results:

6.2.1. Result distribution comparison

The first analysis compares the distribution of results obtained by FETCH and the exhaustive search method. Figure 4 depicts the distribution of AUC scores on the validation set, highlighting the ability of our method to achieve comparable results to the exhaustive search approach.

Furthermore, a Wilcoxon matched-pairs test (Conover, 1999) was conducted to statistically analyze the performance difference between the two methods. The null hypothesis states that the median difference between paired observations in the AUC distributions is zero. The results for both 4- and 8-electrode configurations indicate that we cannot reject the null hypothesis (p-value > 0.05). Given that there is insufficient evidence to reject the null hypothesis and considering the visual representation in Figure 4, this experiment suggests that the performance of our FETCH method is not significantly different from the exhaustive search method.

6.2.2. Rank preservation

In our study, we investigate the relationship or rank among configurations in the two methods of FETCH and exhaustive search. Preserving the rank order is particularly important in this context, as it ensures that if the configurations are sorted by their performance in one method (e.g., FETCH), the corresponding order in the other method (e.g., exhaustive search) remains consistent.

We employed Spearman's rank correlation coefficient, a non-parametric statistic that evaluates the degree and direction of association between two ranked variables (Dodge, 2008). This analysis revealed a Spearman's rank correlation coefficient of 0.86 and 0.81 for 4- and 8-electrode configurations, respectively, with p-value < 0.05. This finding is important in understanding the alignment between the two sets of rankings, as it allows us to use FETCH as an alternative to computationally expensive exhaustive search methods.

6.3. Design cost representation

We measure the power consumption for each electrode configuration by performing the transformer model inference on X-HEEP equipped with 12 on-chip memory blocks (each 32 KB). On the X-HEEP system operating at 75 MHz and 0.8 volts, the system processes STFT, the transformer encoder, and prototype comparison in 2.7 seconds, 9.2 seconds, and 6.4 microseconds, respectively. Given a 12-second EEG signal window, this translates to real-time seizure detection with a total latency of 11.9 seconds. The computation consumes 3.7 mW of power.

As discussed in Section 4.2, for each electrode configuration, we apply a mask to the input STFT representations for any absent electrodes in the given configuration. This approach leaves the number of active electrodes as the primary factor influencing total device power consumption. For signal acquisition, we utilize ADS1298 (TexasInstruments, 2015) as an analog-digital converter (ADC) device. This 8-channel ADC expends 0.8 mW of power per enabled channel during operation and 2 mW of static power. Since the number of electrodes is limited to 8 in this device, depending on the number of electrodes in our wearable device, we need to use one, two or three ADC devices.

Utilizing the defined design cost in Equation (2), we illustrate the results in Figure 5. In this figure, each point corresponds to a specific electrode configuration. The coordinates of these points represent two dimensions: the performance achieved by FETCH, and the diameter of the corresponding graph, which reflects the ergonomic function. Also, power consumption is depicted by the point size in the plot. The coefficients w_1, w_2, and w_3 in Equation (2), which represent the weights for performance, power consumption, and ergonomic factor respectively, are

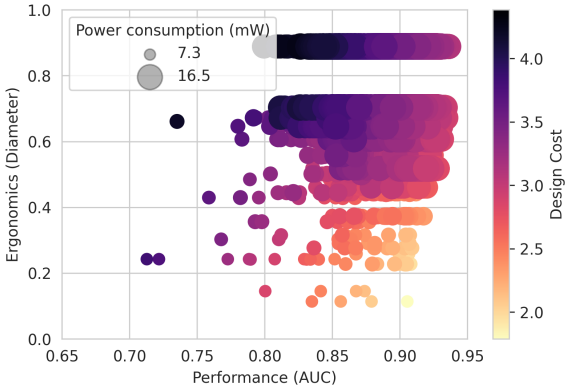


Figure 5: Wearable design space exploration based on a custom design cost.

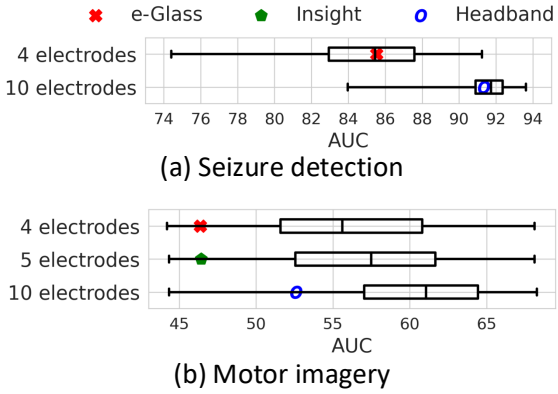


Figure 6: Comparing the existing wearable devices to their corresponding n-electrode distribution for two different tasks. The TUSZ dataset electrode positions are not compatible with the Insight device.

set to 10, 0.1, and 1. This selection is made to enhance the clarity of the visualization in Figure 5.

This type of figure aids wearable designers in making more informed decisions regarding the target electrode configuration, taking into account their application and design constraints. The efficiency of FETCH is particularly noteworthy, as it allows the generation of this figure in a significantly reduced time frame, even when dealing with substantial machine learning models or a large number of electrodes.

6.4. FETCH with existing wearable devices

This section leverages FETCH as a framework to systematically analyze the performance of existing wearable devices within the context of all feasible electrode combinations. The existing devices have shown promising results in their own dataset using their own models. This comparison is conducted solely to demonstrate the utility of FETCH in wearable device design, not to highlight any shortcomings of these existing devices.

Traditionally, wearable device performance is solely evaluated against full-channel EEG caps, which provides limited insights into their true potential across diverse electrode configurations. FETCH enables a more comprehensive evaluation by benchmarking wearable devices against the performance distribution of all feasible electrode combinations. This systematic approach offers valuable insights for wearable designers beyond a simple comparison to a full-channel cap.

To show the generalizability of FETCH, we applied FETCH to another dataset. We evaluate FETCH on the motor imagery (MI) task of the OpenBMI dataset (Lee et al., 2019). The dataset consists of 54 subjects, each subject did two sessions, and during each session, subjects performed 100 trials. Subjects were shown visual cues (right/left arrows) and grasped a handle with the corresponding hand for 4 seconds while EEG data was recorded. We used the 62 EEG channels available in the dataset and the same transformer architecture described in Section 4. However, the input STFT was modified to account for the 62 electrodes, and the target frequency range was set to [8-30 Hz], aligning with existing literature (Lee et al., 2019). Since the task is left/right discrimination, we assumed that the combinations should be symmetric.

In particular, we compare FETCH to three existing wearable devices for: Ceribell rapid response EEG (Vespa et al., 2020), where 10 electrodes are positioned in a headband, Emotiv Insight (Emotiv, 2022) equipped with 5 electrodes, and the e-Glass (Sopic et al., 2018) device, wherein electrodes are integrated into eyeglasses. The electrode configurations used in these wearable devices have been widely used in the literature (Amirshahi et al., 2022; Pascual et al., 2020; Baghersalimi et al., 2022).

To evaluate performance across different electrode configurations, we analyzed the AUC distribution obtained by FETCH on the TUSZ dataset. Figure 6(a) depicts the positioning of e-Glass and headband devices within the distribution for both 4-electrode (170 configurations) and 10-electrode (7258 config-

urations) scenarios. Similarly, Figure 6(b) showcases the placement of three wearable devices in the AUC distribution for the motor imagery task, involving 813, 2834, and 557 845 different subsets of 4-, 5-, and 10-electrode configurations, respectively. Notably, FETCH identified the distribution of 10-electrode configurations within 27 hours, whereas an exhaustive search retraining the model for each configuration would take one year.

This analysis enables us to understand how the performance of these existing wearable devices compares to the spectrum of possibilities offered by all feasible electrode combinations, as identified by FETCH. This information can be instrumental in guiding future wearable design choices and optimizations.

7. Conclusion

In conclusion, our study demonstrates the efficacy of FETCH in optimizing electrode configurations for EEG-based wearable devices. By employing a transformer model integrated with few-shot learning, we successfully achieved high-performance metrics while significantly reducing training times. Our approach, which is adaptable to various electrode arrangements, proved to be comparable with traditional exhaustive search and channel selection methods, as we achieved an AUC value of 0.917 with configurations using eight electrodes. Also, Spearman's rank correlation of 0.86 showed that FETCH can be considered as an alternative to the exhaustive search methods. Furthermore, through the application of our design cost function, we showcased the potential to create ergonomic and energy-efficient wearable devices. Comparisons with existing devices were further validated FETCH as a robust tool for designers of wearable technology, enabling them to significantly improve device performance and user comfort.

8. Acknowledgement

This work has been partially supported by the PEDESITE Swiss NSF Sinergia project (GA No. SCRSII5 193813/1), the RESoRT grant from Fondation Botnar (REG-19-019), the WASP Program of the Knut and Alice Wallenberg Foundation, and the Wyss Center for Bio and Neuro Engineering through funding for ESL-EPFL in the Non-invasive Neuromodulation of Subcortical Structures project of the Lighthouse Partnership Agreement with EPFL.

References

Alireza Amirshahi, Anthony Thomas, Amir Aminifar, Tajana Rosing, and David Atienza. M2d2: Maximum-mean-discrepancy decoder for temporal localization of epileptic brain activities. *IEEE Journal of Biomedical and Health Informatics*, 27 (1):202–214, 2022.

Onder Aydemir and Ebru Ergün. A robust and subject-specific sequential forward search method for effective channel selection in brain computer interfaces. *Journal of neuroscience methods*, 313: 60–67, 2019.

Saleh Baghersalimi, Alireza Amirshahi, Farnaz Forooghifar, Tomas Teijeiro, Amir Aminifar, and David Atienza. Many-to-one knowledge distillation of real-time epileptic seizure detection for low-power wearable internet of things systems. *arXiv preprint arXiv:2208.00885*, 2022.

Paola Busia, Andrea Cossettini, Thorir M Ingolfsson, Simone Benatti, Alessio Burrello, Victor JB Jung, Moritz Scherer, Matteo A Scrugli, Adriano Bernini, Pauline Ducouret, et al. Reducing false alarms in wearable seizure detection with eegformer: A compact transformer model for mcus. *IEEE Transactions on Biomedical Circuits and Systems*, 2024.

Kup-Sze Choi, Yugu Zeng, and Jing Qin. Using sequential floating forward selection algorithm to detect epileptic seizure in eeg signals. In *2012 IEEE 11th International Conference on Signal Processing*, volume 3, pages 1637–1640. IEEE, 2012.

William Jay Conover. *Practical nonparametric statistics*, volume 350. john wiley & sons, 1999.

Jonathan Dan, Mette Thrane Foged, Benjamin Vandendriessche, Wim Van Paesschen, and Alexander Bertrand. Sensor selection and miniaturization limits for detection of interictal epileptiform discharges with wearable eeg. *Journal of Neural Engineering*, 20(1):016045, 2023.

Jonathan Dan, Una Pale, Alireza Amirshahi, William Cappelletti, Thorir Mar Ingolfsson, Xiaying Wang, Andrea Cossettini, Adriano Bernini, Luca Benini, Sándor Beniczky, David Atienza, and Philippe Ryvlin. Szcore: A seizure community open-source research evaluation framework for the validation of eeg-based automated seizure detection algorithms. *arXiv preprint arXiv:2402.13005*, 2024.

Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Yadolah Dodge. *The concise encyclopedia of statistics*. Springer Science & Business Media, 2008.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Jonas Duun-Henriksen, Maxime Baud, Mark P Richardson, Mark Cook, George Kouvas, John M Heasman, Daniel Friedman, Jukka Peltola, Ivan C Zibrandtsen, and Troels W Kjaer. A new era in electroencephalographic monitoring? subscalp devices for ultra–long-term recordings. *Epilepsia*, 61 (9):1805–1817, 2020.

Emotiv. Insight - 5 channel wireless eeg headset. https://www.emotiv.com/products/insight, 2022. Accessed: 2024-03-19.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017.

Hagar Gelbard-Sagiv, Snir Pardo, Nir Getter, Miriam Guendelman, Felix Benninger, Dror Kraus, Oren Shriki, and Shay Ben-Sasson. Optimizing electrode configurations for wearable eeg seizure detection using machine learning. *Sensors*, 23(13): 5805, 2023.

Catalina Gómez, Pablo Arbeláez, Miguel Navarrete, Catalina Alvarado-Rojas, Michel Le Van Quyen, and Mario Valderrama. Automatic seizure detection based on imaged-eeg signals through fully convolutional networks. *Scientific reports*, 10(1): 21833, 2020.

Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25 (1):65–69, 2019.

Mark D Holmes. Dense array eeg: methodology and new hypothesis on epilepsy syndromes. *Epilepsia*, 49:3–14, 2008.

Thorir Mar Ingolfsson, Andrea Cossettini, Xiaying Wang, Enrico Tabanelli, Giuseppe Tagliavini, Philippe Ryvlin, Luca Benini, and Simone Benatti. Towards long-term non-invasive monitoring for epilepsy via wearable eeg devices. In *2021 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 01–04. IEEE, 2021.

Jayoung Kim, Alan S Campbell, Berta Esteban-Fernández de Ávila, and Joseph Wang. Wearable biosensors for healthcare monitoring. *Nature biotechnology*, 37(4):389–406, 2019.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kwanhyung Lee, Hyewon Jeong, Seyun Kim, Donghwa Yang, Hoon-Chul Kang, and Edward Choi. Real-time seizure detection using eeg: A comprehensive comparison of recent approaches under a realistic setting. In *Proceedings of the Conference on Health, Inference, and Learning*, volume 174 of *Proceedings of Machine Learning Research*, pages 311–337. PMLR, 2022a.

Kyeonggu Lee, Kang-Min Choi, Seonghun Park, Seung-Hwan Lee, and Chang-Hwan Im. Selection of the optimal channel configuration for implementing wearable eeg devices for the diagnosis of mild cognitive impairment. *Alzheimer's Research & Therapy*, 14(1):170, 2022b.

Min-Ho Lee, O-Yeon Kwon, Yong-Jeong Kim, Hong-Kyung Kim, Young-Eun Lee, John Williamson, Siamac Fazli, and Seong-Whan Lee. Eeg dataset and openbmi toolbox for three bci paradigms: An investigation into bci illiteracy. *GigaScience*, 8(5): giz002, 2019.

Yongpei Ma, Chunyu Liu, Maria Sabrina Ma, Yikai Yang, Nhan Duy Truong, Kavitha Kothur, Armin

Nikpour, and Omid Kavehei. Tsd: Transformers for seizure detection. *bioRxiv*, pages 2023–01, 2023.

Simone Machetti, Pasquale Davide Schiavone, Thomas Christoph Müller, Miguel Peón-Quirós, and David Atienza. X-heep: An open-source, configurable and extendible risc-v microcontroller for the exploration of ultra-low-power edge accelerators. *arXiv preprint arXiv:2401.05548*, 2024.

Luis Alfredo Moctezuma and Marta Molinas. Classification of low-density eeg for epileptic seizures by energy and fractal features based on emd. *Journal of biomedical research*, 34(3):180, 2020a.

Luis Alfredo Moctezuma and Marta Molinas. Eeg channel-selection method for epileptic-seizure classification based on multi-objective optimization. *Frontiers in neuroscience*, 14:593, 2020b.

Abhijith Mundanad Narayanan, Panagiotis Patrinos, and Alexander Bertrand. Optimal versus approximate channel selection methods for eeg decoding with application to topology-constrained neurosensor networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:92–102, 2020.

Damian Pascual, Alireza Amirshahi, Amir Aminifar, David Atienza, Philippe Ryvlin, and Roger Wattenhofer. Epilepsygan: Synthetic epileptic brain activities with privacy preservation. *IEEE Transactions on Biomedical Engineering*, 68(8):2435–2446, 2020.

Abdellah Rahmani, Arun Venkitaraman, and Pascal Frossard. A meta-gnn approach to personalized seizure detection and classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

Olympia Roeva, Stefka Fidanova, and Marcin Paprzycki. Population size influence on the genetic and ant algorithms performance in case of cultivation process modeling. In *Recent Advances in Computational Optimization: Results of the Workshop on Computational Optimization WCO 2013*, pages 107–120. Springer, 2015.

Vinit Shah, Eva von Weltin, Silvia Lopez, James Riley McHugh, Lillian Veloso, Meysam Golmohammadi, Iyad Obeid, and Joseph Picone. The temple university hospital seizure detection corpus. *Frontiers in Neuroinformatics*, 12:357250, 11 2018. doi: 10.3389/FNINF.2018.00083.

Siuly Siuly, Yan Li, and Yanchun Zhang. Eeg signal analysis and classification. *IEEE Trans Neural Syst Rehabilit Eng*, 11:141–144, 2016.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

Dionisije Sopic, Amir Aminifar, and David Atienza. e-glass: A wearable system for real-time detection of epileptic seizures. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018.

Siyi Tang, Jared A. Dunnmon, Khaled Saab, Xuan Zhang, Qianying Huang, Florian Dubost, D. Rubin, and Christopher Lee-Messer. Self-supervised graph neural networks for improved electroencephalographic seizure analysis. In *International Conference on Learning Representations*, 2021.

Siyi Tang, Jared A Dunnmon, Qu Liangqiong, Khaled K Saab, Tina Baykaner, Christopher Lee-Messer, and Daniel L Rubin. Modeling multivariate biosignals with graph neural networks and structured state space models. In *Conference on Health, Inference, and Learning*, pages 50–71. PMLR, 2023.

TexasInstruments. Ads129x low-power, 8-channel, 24-bit analog front-end for biopotential measurements. https://www.ti.com/product/ADS1298, 2015. Accessed: 2024-01-30.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Paul M Vespa, DaiWai M Olson, Sayona John, Kyle S Hobbs, Kapil Gururangan, Kun Nie, Masoom J Desai, Matthew Markert, Josef Parvizi, Thomas P Bleck, et al. Evaluating the clinical impact of rapid response electroencephalography: the decide multicenter prospective observational clinical study. *Critical care medicine*, 48(9):1249, 2020.

Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020.