



Published in final edited form as:

*J Neurosci Methods*. 2021 March 01; 351: 108966. doi:10.1016/j.jneumeth.2020.108966.

## Deep active learning for interictal ictal injury continuum EEG patterns

Wendong Ge<sup>a,b,1</sup>, Jin Jing<sup>a,b,1</sup>, Sungtae An<sup>t,1</sup>, Aline Herlopian<sup>d</sup>, Marcus Ng<sup>e</sup>, Aaron F. Struck<sup>f</sup>, Brian Appavu<sup>g</sup>, Emily L. Johnson<sup>h</sup>, Gamaleldin Osman<sup>i</sup>, Hiba A. Haider<sup>j</sup>, Ioannis Karakis<sup>j</sup>, Jennifer A. Kim<sup>d</sup>, Jonathan J. Halford<sup>k</sup>, Monica B. Dhakar<sup>j</sup>, Rani A. Sarkis<sup>l</sup>, Christa B. Swisher<sup>m</sup>, Sarah Schmitt<sup>k</sup>, Jong Woo Lee<sup>l</sup>, Mohammad Tabaeizadeh<sup>n</sup>, Andres Rodriguez<sup>j</sup>, Nicolas Gaspard<sup>o</sup>, Emily Gilmore<sup>p</sup>, Susan T. Herman<sup>q</sup>, Peter W. Kaplan<sup>r</sup>, Jay Pathmanathan<sup>s</sup>, Shenda Hong<sup>t</sup>, Eric S. Rosenthal<sup>a,b</sup>, Sahar Zafar<sup>a,b</sup>, Jimeng Sun<sup>c</sup>, M. Brandon Westover<sup>a,b,\*</sup>

<sup>a</sup>Massachusetts General Hospital, Department of Neurology, Boston, MA, United States

<sup>b</sup>Harvard Medical School, Boston, MA, United States

<sup>c</sup>University of Illinois at Urbana-Champaign, College of Computing, Champaign, IL, United States

<sup>d</sup>Yale University, United States

<sup>e</sup>University of Manitoba, Canada

<sup>f</sup>University of Wisconsin Madison Department of Neurology, United States

<sup>g</sup>University of Arizona College of Medicine, Phoenix, United States

<sup>h</sup>Johns Hopkins School of Medicine, United States

<sup>i</sup>Henry Ford Hospital, United States

<sup>j</sup>Emory University School of Medicine, Georgia

<sup>k</sup>Medical University of South Carolina, United States

<sup>l</sup>Brigham and Women's Hospital, United States

<sup>m</sup>Duke University Hospital, United States

<sup>n</sup>Baylor College of Medicine, United States

<sup>o</sup>Université Libre de Bruxelles, Hôpital Erasme and Yale University, Belgium

<sup>p</sup>Yale University, Yale New Haven Hospital, United States

<sup>q</sup>Barrow Neurological Institute, Phoenix, AZ, United States

<sup>r</sup>Johns Hopkins University, United States

\*Corresponding author at: Massachusetts General Hospital, Department of Neurology, Boston, MA, United States. mwestover@partners.org (M. Brandon Westover).

<sup>1</sup>Co-first authors.

Declaration of Competing Interest

The authors report no declarations of interest.

<sup>s</sup>University of Pennsylvania, PA, United States

<sup>t</sup>Georgia Institute of Technology, College of Computing, Atlanta, GA, Georgia

## Abstract

**Objectives:** Seizures and seizure-like electroencephalography (EEG) patterns, collectively referred to as “ictal interictal injury continuum” (IIIC) patterns, are commonly encountered in critically ill patients. Automated detection is important for patient care and to enable research. However, training accurate detectors requires a large labeled dataset. Active Learning (AL) may help select informative examples to label, but the optimal AL approach remains unclear.

**Methods:** We assembled >200,000 h of EEG from 1,454 hospitalized patients. From these, we collected 9,808 labeled and 120,000 unlabeled 10-second EEG segments. Labels included 6 IIIC patterns. In each AL iteration, a Dense-Net Convolutional Neural Network (CNN) learned vector representations for EEG segments using available labels, which were used to create a 2D embedding map. Nearest-neighbor label spreading within the embedding map was used to create additional pseudo-labeled data. A second Dense-Net was trained using real- and pseudo-labels. We evaluated several strategies for selecting candidate points for experts to label next. Finally, we compared two methods for class balancing within queries: standard balanced-based querying (SBBQ), and high confidence spread-based balanced querying (HCSBBQ).

**Results:** Our results show: 1) Label spreading increased convergence speed for AL. 2) All query criteria produced similar results to random sampling. 3) HCSBBQ query balancing performed best. Using label spreading and HCSBBQ query balancing, we were able to train models approaching expert-level performance across all pattern categories after obtaining ~7000 expert labels.

**Conclusion:** Our results provide guidance regarding the use of AL to efficiently label large EEG datasets in critically ill patients.

## Keywords

Electroencephalography(EEG); Seizure; Ictal interictal injury continuum; Machine learning; Active learning; Convolutional neural network; Embedding map

## 1. Introduction

EEG monitoring in critical care has grown dramatically over the past 20 years with the discovery that a large proportion of ICU patients suffer from potentially harmful subclinical seizures and seizure-like electrical events, often collectively called “ictal-interictal-injury continuum” (IIIC) patterns, detectable only by electroencephalography (EEG) (Zafar et al., 2020; Koren et al., 2016). Besides seizures, clinically-important IIIC patterns include seizure-like patterns designated Lateralized Periodic Discharges (LPDs), Generalized Periodic Discharges (GPDs), Lateralized Rhythmic Delta Activity (LRDA) and Generalized Rhythmic Delta Activity (GRDA) (Newey et al., 2017; Rubinos et al., 2018; Sivaraju and Gilmore, 2016). The growth in ICU EEG monitoring has created a clinical need for more rapid and accurate EEG interpretation. IIIC patterns have the potential to damage the brain and contribute to permanent neurologic disability, yet detection by visual review is typically

not possible to conduct in real time. Automated identification of these patterns would allow targeted treatment (and prevention of iatrogenically harmful treatments where they were not warranted). There is a critical unmet need for automated EEG monitoring for IIC patterns (Bajaj and Pachori, 2011; Pohlmann-Eden and Newton, 2008; Temko et al., 2011).

Prior work on seizure and IIC labeling and classification includes (Jing et al., 2018), in which our group applied unsupervised machine learning methods to achieve efficient pre-clustering of NCS and IIC patterns in prolonged EEG recordings into a small number of clusters; the present work is an outgrowth of this approach. Others have explored deep learning approaches for seizure detection or prediction either directly from the raw EEG signal (Thodoroff et al., 2016) or from images (Jana et al., 2019). The latter reference used a Dense Convolutional Network (Dense-Net), similar to the model architecture we use in the present work. However, these prior efforts focused on relatively small data sets.

To develop an automated monitoring system for IIC patterns, an adequately large set of labeled training data for supervised machine learning models must be developed first. For this purpose, we have collected over 200,000 h of EEG signals from > 1000 patients, comprising over 350 million EEG Segments. Our goal is ultimately to obtain a label every 2-second segment of EEG data from multiple experts. However, our dataset contains 350 million 2-second EEG segments corresponding to 200,000 h of EEG. Human experts cannot feasibly label all these EEG segments. Thus, we sought to design Active Learning (AL) schemes to select which EEG segments are most valuable to ask experts to label (Settles, 2009; Osugi et al., 2005; Dasgupta and Hsu, 2008).

Many different AL approaches have been proposed for various tasks involving EEG data, although none of them completely fits our present task. Among these Wu et al. (2016) used a technique called enhanced batch-mode active learning, in which the method improved reliability, representativeness, and diversity of the selected samples in driver drowsiness estimation based on EEG signals. Lawhern et al. (2015) used AL to improve the reliability of existing EEG artifact classifiers with minimal user interaction. Yang et al. (2012) applied AL to classify motor imagery from EEG signals in a brain computer interface. Gupta et al. (2017) developed methods that use sparse recovery and AL techniques for near real-time artifact identification and removal in EEG recordings. Macas et al. (2018) demonstrated that for semiautomatic annotation of sleep EEG signals, an AL strategy leads to statistically significantly smaller mean class error than a simple random sampling strategy. Joao et al. (2018) proposed a heuristic AL process that improves the performance of a previously developed epileptic seizure prediction algorithm. Yang et al. (2016) investigated the effectiveness of using an active learning algorithm to automatically annotate a large EEG for seizure and some of IIC patterns. Karuppiah Ramachandran et al. (2018) proposed an active learning-based prediction framework that aims to improve the accuracy of the seizure prediction with a minimum amount of labelled data. Balakrishnan and Syed (2012) described an active learning framework for patient-specific seizure onset detection in continuous EEG signals.

However, several novel challenges arise for our IIC pattern labeling and pattern classification problem that prior AL studies have not considered:

- **Inter-Rater Agreement (IRA):** In our labeling efforts, each EEG segment is labeled by multiple human experts. Experts frequently disagree regarding the correct label for an EEG segment.
- **Imbalance:** EEG datasets from hospitalized patients commonly exhibit substantial class imbalance, with most segments belonging to an “other” category, consisting of neither seizures nor IIC patterns.
- **Learning from raw EEGs:** Optimal features for IIC pattern classification are not known, therefore a data-driven method that discovers suitable features is desirable.
- **Leveraging unlabeled data:** As only a small amount of EEG data can be manually labeled, it is advantageous to design label spreading schemes to generate high-quality pseudo-labels to augment the size and diversity of the training data. Label spreading means assigning labels samples that have not been labeled by experts, to augment the training data. Methods for creating pseudo-labels are described below.

In this paper we present several AL methods to address each of these challenges. We evaluate how quickly and how well the labels provided by each proposed AL labeling scheme enable a deep learning model to learn to classify IIC patterns in a large and representative EEG dataset.

## 2. Methods

### 2.1. Data

We selected 10-second EEG segments from 1435 patients from the Massachusetts General Hospital (MGH) EEG Archive from 2012 to 2017. All data was collected as part of routine clinical care. Retrospective analysis of the data was approved by the local Institutional Review Board (IRB), which waived the need for written informed consent. These data were labeled as part of an ongoing effort to develop automated seizure and IIC pattern detectors. So far, we have collected 88,341 labels for 9,808 EEG segments from 25 human experts from 12 hospitals. The median number of labels gathered from each expert was 4015, and the range was 1005–6813.

### 2.2. Seizure and Interictal Ictal Injury Continuum (IIC) pattern definitions

In this work, we study Active Learning for efficiently annotating seizures and IIC patterns in continuous EEG recordings from critically ill and acutely hospitalized patients. All EEGs were recorded using the 10–20 system, an internationally standardized method for the placement of scalp electrodes (Fig. 1a) (Homan et al., 1987). The events of interest, seizures and IIC patterns (LPD, GPD, LRDA, GRDA), are defined according to standardized American Clinical Neurophysiology Society (ACNS) nomenclature and the Salzburg criteria; (Hirsch et al., 2013; Beniczky et al., 2013; Leitinger et al., 2015) samples that do not fall into any of the preceding patterns were labeled as “other”. Definitions of IIC patterns are available at ANON (2021a). Example images of these patterns are shown in Fig. 1.

ACNS terminology also defines BIPDs (bilaterally independent periodic discharges), which are similar to GPDs except that the discharges on the left and right sides are not synchronous. BiPDs is a rare pattern, therefore we did not create a separate category for this pattern. Instead, experts were instructed to label any instances of BiPDs as GPDs.

### 2.3. Seizures and IIC dataset

The 9,808 labeled EEG segments were obtained in a balanced manner, such that approximately half belonged to IIC categories. Nevertheless, the proportion of the IIC pattern data per unit time is lower in most ICU EEGs. Thus, we add 29,551 samples from EEGs reported as containing no IIC patterns, and assigned these a label of “Other”, to yield 39359 samples for the AL experiments.

We divided data into a training dataset (70 %), validation dataset (10 %) and test dataset (20 %). Data was separated at the patient level, such that any given patient’s EEG data appeared in only one of the subsets, as summarized in Table 1.

The samples were imbalanced across the six pattern categories, as shown in Table 2.

### 2.4. Vector representation for EEG segments

EEG signals were recorded with 19 electrodes (FP1, F3, C3, P3, F7, T3, T5, O1, FZ, CZ, PZ, FP2, F4, C4, P4, F8, T4, T6, O2; see Fig. 1a). The sampling rate was 256 or 512 Hz; all signals were resampled to 200 Hz. EEG signals were arranged into bipolar montage, consisting of 16 channels formed by taking differences involving the following pairs of spatially adjacent channels: "FP1-F7", "F7-T3", "T3-T5", "T5-O1", "FP2-F8", "F8-T4", "T4-T6", "T6-O2", "FP1-F3", "F3-C3", "C3-P3", "P3-O1", "FP2-F4", "F4-C4", "C4-P4", "P4-O2". For purposes of annotation, EEG signals were divided into 2-second EEG segments. Each 2-second segment was provided with a 4-second context signal on both sides. Thus each 2-second segment was represented as a 16 channels x 2000-samples (10 s) signal. Here, ‘samples’ refers to digital data points. The sampling rate is 200 Hz, hence the number of samples in each 10 s signal segment is 2000.

Dense-Net, a type of Convolutional Neural Network, was used to learn a vector representation of each  $16 \times 2000$  EEG segment (Huang et al., 2017). The structure of the Dense-Net used is illustrated in Fig. 2. The Dense-Net includes 7 Dense Blocks. Each Dense Block includes 4 Dense-Layers. Each Dense Layer includes 2 Convolutional layers and 2 exponential linear unit (ELU) activations. Additionally, there are 6 transition blocks among the 7 Dense Blocks, where each transition block includes ELU activation, a convolutional layer, and an average pooling layer. We extracted the vectors (256 dimensions) from the second-to-last Fully Connected (FC) layer as the vector representation for EEG segments.

### 2.5. Embedding map for seizure and IIC patterns

After mapping EEG segments into a high-dimensional vector representation using the Dense-Net model, we used the Uniform Manifold Approximation and Projection (UMAP) algorithm to reduce data dimensionality and visualize similarity relationships between segments (McInnes et al., 2018). In subsequent sections, we describe how the 2D UMAP

visualization, hereafter referred to as a “UMAP”, is used to facilitate active learning. Fig. 3 shows an example of a UMAP for our data after 10 rounds of AL. The colored points have been labeled; the grey points are unlabeled points.

## 2.6. Active learning

Our AL framework is illustrated in Fig. 4. The steps are as follows:

- Step 0: Randomly select EEG segments for human experts to label as initial seed points.
- Step 1: Use labeled data to train a first CNN.
- Step 2: Use the first CNN to process labeled and unlabeled data to obtain embeddings.
- Step 3: Use the EEG segment embeddings to create a 2D UMAP.
- Step 4: Use label spreading strategies to assign pseudo-labels to unlabeled data.
- Step 5: Use labeled data and pseudo-labeled data to train a second CNN.
- Step 6: Use query and balance strategies to select the most informative unlabeled data in the current iteration as candidate segments for expert labeling.
- Step 7: Ask multiple human experts to label the new candidate segments. Add the labeling results to the labeled dataset. Repeat steps 1-7.

In addition to the steps above, the second CNN is used to evaluate performance on a test dataset in each iteration.

In the following, we introduce details of the label propagation, query, and balance strategies explored in our AL scheme. We denote the number of label categories as  $M$ . In each iteration, the number of labeled EEG segments is  $K$ , and the  $k$ -th labeled EEG segment is denoted  $b_k$ , where  $1 \leq k \leq K$ . The number of unlabeled EEG segments is  $N$ , and the  $n$ -th unlabeled point is denoted  $a_n$ , where  $1 \leq n \leq N$ .

**2.6.1. Label spread strategies**—The goal of label spreading is to propagate labels from labeled EEG segments to similar unlabeled EEG segments. Label spreading generates pseudo-labels for unlabeled data to increase the number of training samples. In this work we propagate labels using the embedding map (UMAP), as follows. In each iteration of AL, unlabeled samples receive pseudo-labels from neighboring labeled points, as follows.

Let us denote segments that have been labeled as  $b_k$ , and their labels as  $L_k = [l_1^k, \dots, l_m^k, \dots, l_M^k]$ ; where  $M=6$ , and the 6 patterns are Seizure, LPD, GPD, LRDA, GRDA, Other. The numbers  $l_m^k$  represent the number of votes received from experts for sample  $k$  belonging to pattern  $m$ . We assign a pseudo-label vector for unlabeled EEG segment  $a_n$  by spreading labels to it from neighboring labeled samples,

$$\hat{L}_n = [\hat{l}_1^n, \dots, \hat{l}_m^n, \dots, \hat{l}_M^n] = \sum_{b_k \in D} (w_{n,k} \cdot L_k), \quad (1)$$

Where  $D$  is a neighborhood surrounding the unlabeled example  $a_n$ , and

$$w_{n,k} = \exp(-\eta \cdot \|a_n - b_k\|^2) \quad (2)$$

Where  $\|\cdot\|^2$  is the squared Euclidean distance, and  $\eta$  is a decay coefficient. The normalized pseudo-label for the unlabeled example  $a_n$  is  $\bar{L}_n = [\bar{l}_1^n, \dots, \bar{l}_m^n, \dots, \bar{l}_M^n]$ , where

$$\bar{l}_m^n = \frac{\hat{l}_m^n}{\sum_{m'=1}^M \hat{l}_{m'}^n} \quad (3)$$

If there is no labeled EEG segment in  $D$ , a uniform label distribution is used for the pseudo-label, representing no prior label knowledge. In this way, each EEG segment is assigned a “soft label”, where the percentage of votes assigned to each class serves as an estimate of the probability of that class.

**2.6.2. Query strategies**—The goal of the query strategy is to select the most informative unlabeled EEG segment for labeling. We considered six query strategies: three are model-based, including confidence sampling, margin sampling and entropy sampling; three are model-free, including random sampling, density sampling and diversity sampling. These strategies are described as follows:

**a) Random sampling:** In each iteration, unlabeled EEG segments are chosen at random with equal probability.

**b) Confidence sampling:** (Settles, 2011) The feature of the  $n$ -th unlabeled EEG segment  $a_n$  is denoted  $x_n$ , where  $1 \leq n \leq N$ . The label of  $a_n$  is denoted as  $y_{n,m}$ , where  $1 \leq m \leq M$  is the index of the IIC pattern. The classifier in the previous iteration is used to calculate the label probabilities for  $a_n$  denoted as  $p_{n,m} = Pr(y_{n,m}|x_n)$ , where  $\sum_{m=1}^M p_{n,m} = 1$ ,  $1 \leq n \leq N$ . The confidence score for  $a_n$  is then represented as the maximum difference between one and any of the class probabilities,

$$s_n = \max_m (1 - p_{n,m}) \quad (4)$$

In each iteration, the normalized probability of sampling the  $n$ -th unlabeled EEG segment  $a_n$  is

$$p_n^{\text{sample}} = \frac{\exp(s_n)}{\sum_{n=1}^N \exp(s_n)} \quad (5)$$

**c) Margin sampling:** (Settles, 2011) The margin score of the n-th unlabeled EEG segment  $a_n$  is represented as

$$s_n = \frac{1}{\max_m p_{n,m} - \max_{m, m \neq \hat{m}} p_{n,m}} \quad (6)$$

Where  $\hat{m} = \operatorname{argmax}_m p_{n,m}$ . The normalized probability of sampling the n-th unlabeled EEG segment  $a_n$  is calculated via Eq. (5). The denominator represents the difference (margin) between the most and second most probable labels, and margin sampling favors data points where this margin is small.

**d) Entropy sampling:** (Settles, 2011) The entropy of the n-th unlabeled EEG segment  $a_n$  is represented as

$$s_n = - \sum_{m=1}^M (p_{n,m} \cdot \log p_{n,m}) \quad (7)$$

The normalized probability of sampling the n-th unlabeled EEG segment  $a_n$  is calculated via Eq. (5).

**e) Density sampling:** In the embedding map, the similarity between unlabeled EEG segment  $a_n$  and labeled EEG segment  $b_k$  is represented as

$$\text{Similarity}(a_n, b_k) = \frac{1}{\|a_n - b_k\|^2} \quad (8)$$

The density for unlabeled EEG segment  $a_n$  is defined as (Hu et al., 2010)

$$s_n = \text{Density}(a_n) = \sum_{b_k \in \Omega} \text{Similarity}(a_n, b_k) \quad (9)$$

Where

$$\Omega = \{b_k \mid \text{Similarity}(a_n, b_k) \geq \gamma\} \quad (10)$$

Where  $\gamma$  is a threshold that defines the neighborhood of  $a_n$ . The normalized probability to sample the n-th unlabeled EEG segment  $a_n$  is calculated via Eq. (5).



**f) Diversity Sampling:** The diversity for the unlabeled EEG segment  $a_n$  is defined as (Hu et al., 2010)

$$s_n = \text{Diversity}(a_n) = \frac{1}{\max_{1 \leq k \leq K} \text{Similarity}(a_n, b_k)} = \min_{1 \leq k \leq K} \|a_n - b_k\|^2 \quad (11)$$

The normalized probability to sample the n-th unlabeled EEG segment  $a_n$  is calculated via Eq. (5).

**2.6.3. Balance strategies**—Class imbalance is a common problem in training machine learning models, and leads models to prefer classes with greater numbers of samples. The goal of balancing schemes is to balance the selected unlabeled EEG segments queried across IIC patterns. As seizures and IIC patterns in our data are imbalanced, querying without a balancing scheme is likely to select a class-imbalanced collection of unlabeled EEG segments, which may be detrimental to classifier training for underrepresented categories (e.g. seizures). We explored two balancing strategies: spread-based, and spread-based with high confidence.

**A) Standard spread-based balanced querying (SSBBQ):** Based on the label propagation schemes described above, we assign a pseudo label to each unlabeled EEG segment. These pseudo labels can be used to balance queries of new EEG segments. Namely, if the number of EEG segments to be queried is  $Q$ , we select (based on pseudo-labels)  $Q/M$  EEG segments from each of the  $M$  categories to query.

**B) High confidence spread-based balanced querying (HCSBBQ):** In the standard spread-based balancing scheme (A), pseudo-labels are derived by spreading based on votes and distances in the UMAP to balance query points across categories. However, this scheme is potentially problematic due to the extreme class imbalance of our dataset, as illustrated in Fig. 5. In this example, the blue area represents the majority category (“Other”), and the red area represents the minority category (“Seizure”). The yellow crosses are labeled points, namely three “Other” points and one “Seizure” point. Based on the label propagation scheme, the separation for pseudo-labels “Other” and “Seizure” might be the purple dashed line; that is, suppose points to the left of this line are pseudo-labeled “Other” and those to the right are pseudo-labeled “Seizure”. In balancing scheme (A), most sampled points in the “Seizure” pseudo-label region area are actually “Other” points rather than “Seizure” points. Thus, in balancing scheme (B), an area of high confidence is considered for the minority categories. If we assume a high confidence threshold of  $\phi$ , the high confidence area is defined as the points  $x_n$  such that  $Pr(y_n, \hat{m} | x_n) \geq \phi$ , where  $\hat{m}$  is the target minority category. The balancing scheme B samples  $Q/M$  points for each category as follows:  $Q/2M$  points are sampled using balancing scheme (A), and  $Q/2M$  are sampled from high confidence areas.

**2.6.3.1. Division of data into training, validation, and test sets.:** We divided the total data (39,359 samples) into a training dataset (26,752, 70 %), validation dataset (3,666, 10 %) and test dataset (8,941, 20 %), such that data from any one patient occurred in only one of these sets. We analyzed results over 10 iterations of AL. In each iteration we selected 200

new query EEG segments from the training dataset, obtained labels for these from the group of experts, and added their labels to the training dataset. In each iteration of AL we used the updated labeled dataset to train new models, used the validation dataset to tune hyperparameters, and tested model performance on the test dataset.

## 2.7. Evaluation

Labels for EEG segments were obtained as votes from human experts. We denote the number of human experts as  $I$ , the number of labeled EEG segments as  $K$ , and the number of label categories as  $M$ . We denote the vote of the  $i$ -th human expert for the  $k$ -th EEG segment as  $v_{i,k}$ , where  $v_{i,k} \in \{NA, 1, 2, \dots, M\}$ , and  $NA$  means that the  $i$ -th human expert didn't cast a vote for the  $k$ -th EEG segment. Thus, the  $k$ -th EEG segment is given a label vector, containing the number of votes received for each pattern class across all experts, denoted  $L_k = [J_{k,1}, J_{k,2}, \dots, J_{k,M}]$ , where

$$l_{k,m} = \sum_{i=1}^I \mathbb{1}(v_{i,k}, m), \mathbb{1}(v_{i,k}, m) = \begin{cases} 1 & \text{if } v_{i,k} = m \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

We also define the normalized label vector for the  $k$ -th EEG segment as  $\tilde{L}_k = [\tilde{l}_{k,1}, \tilde{l}_{k,2}, \dots, \tilde{l}_{k,M}]$ , where

$$\tilde{l}_{k,m} = \frac{l_{k,m}}{\sum_{i=1}^M l_{k,i}} \quad (13)$$

The goal in training the deep learning model is that the predicted labels should be close to the labels given by human experts. To formalize this goal, we used two evaluation methods: KL divergence and pairwise agreement between human experts and model.

**2.7.1. KL divergence**—KL divergence is used to measure the distance between the label distribution from human experts vs the trained CNN (DenseNet) model. We denote the predicted normalized score vector from the model as  $\hat{L}_k = [\hat{l}_{k,1}, \hat{l}_{k,2}, \dots, \hat{l}_{k,M}]$ . The KL divergence between the normalized expert label vector  $\tilde{L}_k$  and the normalized predicted label vector  $\hat{L}_k$  is defined as

$$KL_{average} = \frac{1}{K} \sum_{k=1}^K KL(\tilde{L}_k, \hat{L}_k) = \frac{1}{K} \sum_{k=1}^K \sum_{m=1}^M \left[ \tilde{l}_{k,m} \cdot \log \frac{\tilde{l}_{k,m}}{\hat{l}_{k,m}} \right] \quad (14)$$

**2.7.2. Pairwise agreement**—Pairwise agreement is used to measure agreement between human and human (model) for each category. The pairwise agreement between the  $i$ -th expert and the  $j$ -th expert for the  $m$ -th category is defined as

$$A_{i,j} = \frac{1}{2} \left\{ \frac{\sum_{i=1}^I \sum_{j=1}^I [\| (v_{i,k}, m) \cdot \| (v_{j,k}, m) ]}{\sum_{i=1}^I \| (v_{i,k}, m) } + \frac{\sum_{i=1}^I \sum_{j=1}^I [\| (v_{i,k}, m) \cdot \| (v_{j,k}, m) ]}{\sum_{j=1}^I \| (v_{j,k}, m) } \right\} \quad (15)$$

The average agreement between experts, called expert-expert agreement (EEA), is defined as

$$EEA = \frac{1}{I(I-1)} \sum_{i=1}^I \sum_{j=1, j \neq i}^I A_{i,j}$$

The average agreement between the algorithm and experts, called expert-algorithm agreement (EAA), is defined as

$$EAA = \frac{1}{I} \sum_{i=1}^I A_{i, \text{model}} \quad (16)$$

## 2.8. Hyper parameter tuning

Hyper-parameters for CNN training were as follows. Batch size was set to 256. Each EEG segment included 16 channels, where each channel consisted of a 10-second EEG signal. As the EEG sampling rate was set to 200 Hz, the dimensions of the input tensor for the CNN were (256, 16, 2000). The dimensions of the representation vector in the second to last layer was set to be (255,1). The objective function was the weighted Kullback-Leibler divergence. Adam (an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments) was used as the optimizer for the CNN (Kingma and Ba, 2014). The initial learning rate was 0.0001 and weight decay was 0.001. Dropout was used as regularization, with a dropout rate of 0.2. The python libraries we used included pytorch, sklearn, scipy, numpy.

Hyper-parameters for UMAP were set as follows (ANON, 2021b). The dimensions of the output map were set to (2,1), to allow plotting in 2D. The size of the local neighborhood was set 15. The minimum distance apart that EEG segments can be in the low dimensional representation was set to 0.1. The distance metric was set to be Euclidean Distance. The initial embedding positions assignment was set to be random.

## 3. Results

### 3.1. Patient characteristics

Demographics and the presenting medical and neurologic problems leading to hospital admission for the 1435 patients are summarized in Table 3. Note that some patients had more than one presenting problem, resulting in a total greater than 1435.

### 3.2. Comparison of AL schemes

We performed two types of experiments. In the first, we used only labeled data (9,808 labeled EEG segments + 29,551 “Other” EEG segments); this allowed us to compare the various AL schemes in detail. In the second, we added 120,000 unlabeled EEG segments to the 9,808 examples, to for AL schemes, to allow more intensive use of label spreading to improve model performance. These experiments were designed to examine the degree to which AL can help make it feasible for models to achieve expert-algorithm agreement that approaches expert-expert agreement.

**1) Comparison of Query and Label Propagation schemes**—Fig. 6 compares the KL divergence for the test dataset using different query and label propagation schemes. In these experiments, we select 200 new query EEG segments in each AL iteration, and test performance of the different AL schemes over ten iterations. Accordingly, the number of labeled data points available for training is 200 in iteration 1, 400 in iteration 2, 600 in iteration 3, etc, until reaching 2000 in iteration 10. In Fig. 6, the blue lines are for the AL scheme without label spread, and the red lines are the KL divergences in the AL scheme with label spread. Different marker shapes indicate the different query schemes.

From this figure, we see that schemes with label spread are substantially better than schemes without label spread over the first five iterations (up to 1000 labels); results for the schemes with vs without spread become close after obtaining 1000 labels. This suggests that the pseudo-labels obtained by label spreading do enhance the model training when the number of available labels is low.

We also see from the figure that there is no clear difference among the different query schemes. In particular, given more than 200 segments, none of the query methods appreciably outperforms the simplest method of selecting query points at random.

**2) Comparison of Query Balancing Schemes**—Fig. 7 compares the KL divergence for test data using different balancing schemes. In these experiments, we similarly used 10 iterations and selected 200 new query EEG segments to have their labels revealed in each iteration. Label spreading based on votes and distance were used in all schemes.

In Fig. 7, red lines indicate KL divergences for AL schemes without balancing, green lines are for schemes with standard spread-based balancing (SSBBQ), and purple lines are for schemes with high confidence spread-based balanced querying (HCSBBQ).

There is no clear difference between AL schemes without balance (red) and AL schemes with SSBBQ (green). This is attributable to the problem illustrated in Fig. 5. By contrast, AL schemes with HCSBBQ (purple) achieve the best (lowest) KL divergences. This is attributable to the highly class-imbalanced nature of the data, and that the confidence-based balance scheme better samples the minority class.

KL divergences for the different query schemes were all similar. In particular, no scheme was substantially better than selecting queries randomly.

**3) Agreement between Models and Human Experts**—We performed experiments to compare agreement between Models and Human Experts, as follows. We collected 9,808 labeled EEG segments and divided them into a training dataset (6974, 70 %), validation dataset (967, 10 %) and test dataset (1867, 20 %) at the patient level. 120 K unlabeled EEG segments were added to the training data to enhance the ability of label spreading to assist model training, as shown in Fig. 4. We performed 7 AL iterations. The numbers of EEG segments in these 7 iterations were (100, 200, 400, 800, 1600, 3200, 6974). In each iteration, if an unlabeled EEG segment was selected, the closest labeled EEG segment to it in the embedding map was utilized to replace it, to approximate being able to obtain a label for selected segment from an expert.

Fig. 8 compares pairwise agreements between the algorithm and human experts (EAA) with pairwise agreement among experts (EEA) for each of the six pattern categories. In the figure, the blue line is the average agreement between the algorithm and experts (EAA) using random query to obtain training labels, without label spreading or balancing; the orange line is the average agreement between the algorithm and experts (EAA) using the best-performing AL scheme (diversity query, label spread based on votes and distance, and high confidence spread-based balanced querying, HCSBBQ). The green line is the average agreement between pairs of experts (EEA). From this figure, we make the following observations.

- a) In “Seizure”, “LPD”, “GPD”, “LRDA” and “GRDA” categories, agreement for the proposed AL scheme is initially higher than the baseline AL scheme, suggesting that label spreading and data balancing are beneficial.
- b) For the “Other” category, agreement for the proposed AL scheme is initially lower than the baseline AL scheme, which can be ascribed to the fact that the baseline scheme without balancing selects more “other” data (the dominant class).
- c) The baseline AL scheme eventually approaches the proposed AL scheme, because they select the same 6974 labeled EEG segments in our experimental setup. After reaching this number of labeled examples, the model is able to label “GPD” patterns as well as human experts. A small performance gap remains for “LRDA”, “GRDA” and “Other”. Relatively large gaps remain for “Seizure” and “LPD” patterns, suggesting that acquisition of additional expert labels is needed to achieve expert-algorithm agreement (EAA) equal to expert-expert agreement (EEA).

## 4. Discussion

In this work, we compared a wide variety of AL schemes for developing algorithms that automatically label seizure and IIC patterns in electroencephalography data from critically ill patients. Based on the experimental results, we reached the following conclusions. First, AL label spreading is beneficial. Specifically, label spreading based on expert votes and distance in an embedding map converged faster, with fewer human labels needed, than AL without label spread. Second, in obtaining new labels from experts, balancing the queries based on label confidence is beneficial. Specifically, HCSBBQ (high confidence spread

based balanced querying) performs substantially better than SSBBQ (standard spread based balanced querying). Third, all six query criteria produced similar results. Specifically, after choosing a label spreading and query balancing scheme, it did not matter which metric was used to select new samples for labeling: random sampling, confidence sampling, margin sampling, entropy sampling, density sampling, and diversity sampling all produced similar results. We conclude that random sampling, because it is the simplest, is the preferred strategy. Thus, in future work, in which we aim to use AL to obtain labels for 200,000 h EEG data from 25 human experts, we plan to combine 1) label spreading based on expert votes and embedding map distances, 2) high confidence spread based balanced querying (HCSBBQ), and 3) random sampling.

We were surprised that random query works approximately as well as any other scheme. In our experiments, we compared random query with five other query strategies including three model-based strategies (confidence, margin, entropy), and two model-free strategies (density and diversity). The model-based strategies prefer to select uncertain examples based on the model output which are usually located at category boundaries. The model-free strategies prefer to select examples based on data distribution information such as density or diversity. We believe two reasons may help explain this finding. The first is the problem of imperfect inter-rater agreement among experts regarding seizure and IIC patterns. The uncertain EEG samples at boundaries between these categories may not be the best points to query from the standpoint of improving model performance, since these are likely to be labeled in a noisier manner by experts. The second reason is dataset imbalance. As most examples are “Other” and only a small amount of labeled data is available to estimate category boundaries to balance AL sampling, boundary estimation is noisy. Consequently, most queries find samples in the “other” category, which largely cancels the advantages of both model based and model-free query strategies.

Three prior studies used AL for developing automated methods to detect seizure or IIC EEG patterns, such as (a) (Yang et al., 2016), (b) (Karuppiiah Ramachandran et al., 2018), (c) (Balakrishnan and Syed, 2012). For the classification category, (b) and (c) only considered seizures, (a) considered two IIC patterns namely LPDs and GPDs, whereas our work considers seizure and four IIC patterns namely LPD, GPD, LRDA and GRDA. Regarding features, (a) (b) (c) use handcrafted features to train classifiers, while our work extracts features from raw EEG via a CNN. For the AL schemes, (a) used what is essentially a semi-supervised learning scheme, and (b) used the Bernoulli-Gaussian Mixture model which is essentially a model-based query scheme, (c) used a Distance and Diversity combined query scheme based on an SVM model. Our work considers a broader range of query strategies including three model-based strategies (confidence, margin, entropy), and three model-free strategies (random, density and diversity), and combined these with label spreading schemes.

Our study has important limitations. First, our data was from a single site. Second, our data set is highly imbalanced, which likely influences the experimental results for the AL strategies evaluated. Our results may not generalize to other datasets with less extreme data imbalance. Second, we assumed a standard CNN architecture, DenseNet, and held this fixed throughout all experiments. We believe that this was justified by high degree of flexibility

afforded by the DenseNet architecture, and by our desire to focus the work on evaluating competing design choices for AL schemes proposed in the literature. Nevertheless, although we consider it unlikely, it is possible that other CNN architectures could lead to qualitatively different results. Thus optimization of the core deep learning model architecture for use in AL remains an important unexplored topic.

Third, not all EEG segments were labeled by all 25 human experts. Further, we only consider 10-second EEG segments to train the classifier, whereas human experts were allowed to review larger contextual EEG windows when labeling. In the future work, we plan to explore varying levels of context, and to explore using Recurrent Neural Networks to learn temporal dependencies in the EEG to improve classification, particularly for seizures. Finally, given the imperfect levels of interrater agreement for seizure and IIC patterns among experts, it is possible that these categories could benefit from further refinement. In fact, using data-driven approaches such as deep learning to delineate new EEG patterns is an exciting area for future research. Possible approaches might include data-driven identification of EEG features that facilitate higher levels of inter-rater agreement among experts. Alternatively, data-driven pattern discovery seek to identify EEG patterns that are highly predictive of neurologic outcomes (e.g. level of disability or death). These approaches contrast with the approach we have followed here, which takes expert-defined EEG pattern categories as given.

The experimental results and label spreading and query balancing methods presented herein provide valuable guidance for active learning involving multiple experts and large EEG datasets. By using these methods to obtain labels for over 200,000 h of EEG from critically ill patients from over 25 human experts, we expect to soon be able to create deep learning algorithms that are able to identify seizures and IIC patterns with a level of reliability equivalent to human experts.

## Acknowledgments

During this research, Dr. Westover was supported by the Glenn Foundation for Medical Research and the American Federation for Aging Research through a Breakthroughs in Gerontology Grant; through the American Academy of Sleep Medicine through an AASM Foundation Strategic Research Award; by the Football Players Health Study (FPHS) at Harvard University; from the Department of Defense through a subcontract from Moberg ICU Solutions, Inc, and by grants from the NIH (1R01NS102190, 1R01NS102574, 1R01NS107291, 1RF1AG064312).

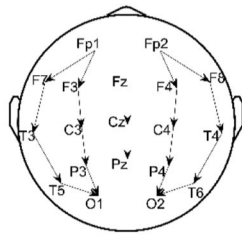
## References

- <https://www.acns.org/pdf/guidelines/Guideline-14-pocket-version.pdf>.  
<https://umap-learn.readthedocs.io/en/latest/>.
- Bajaj V, Pachori RB, 2011. Classification of seizure and nonseizure EEG signals using empirical mode decomposition. *IEEE Trans. Inf. Technol. Biomed* 16 (12 (6)), 1135–1142. [PubMed: 22203720]
- Balakrishnan G, Syed Z, 2012. Scalable personalization of long-term physiological monitoring: active learning methodologies for epileptic seizure onset detection. *Artificial Intelligence and Statistics*, pp. 73–81. 3 21.
- Beniczky S, Hirsch LJ, Kaplan PW, Pressler R, Bauer G, Aurlien H, Brøgger JC, Trinka E, 2013. Unified EEG terminology and criteria for nonconvulsive status epilepticus. *Epilepsia* 54 (9), 28–29.
- Dasgupta S, Hsu D, 2008. Hierarchical sampling for active learning. *Proceedings of the 25th International Conference on Machine Learning* 208–215. 7 5.

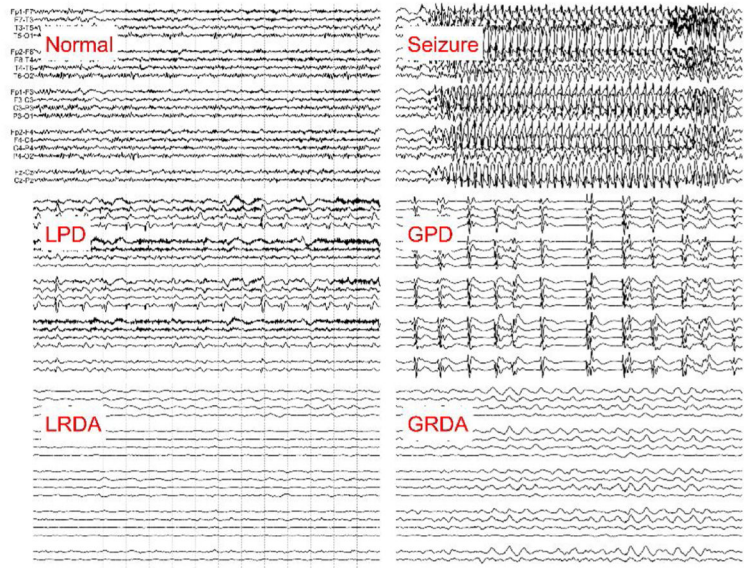
- Gupta M, Beckett SA, Klerman EB, 2017. On-line EEG denoising and cleaning using correlated sparse signal recovery and active learning. *Int. J. Wirel. Inf. Netw* 24 (6(2)), 109–123. [PubMed: 29983539]
- Hirsch LJ, LaRoche SM, Gaspard N, Gerard E, Svoronos A, Herman ST, Mani R, Arif H, Jette N, Minazad Y, Kerrigan JF, 2013. American clinical neurophysiology society's standardized critical care EEG terminology: 2012 version. *J. Clin. Neurophysiol* 30 (2 (1)), 1–27. [PubMed: 23377439]
- Homan RW, Herman J, Purdy P, 1987. Cerebral location of international 10–20 system electrode placement. *Electroencephalogr. Clin. Neurophysiol* 66 (4 (4)), 376–382. [PubMed: 2435517]
- Hu R, Delany SJ, Mac Namee B, 2010. EGAL: exploration guided active learning for TCBR. In: *International Conference on Case-Based Reasoning*. Springer, Berlin, Heidelberg, pp. 156–170. 7 19.
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, 2017. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4700–4708.
- Jana R, Bhattacharyya S, Das S, 2019. Epileptic seizure prediction from EEG signals using DenseNet. In: *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, pp. 604–609. 12 6.
- Jing J, d'Angremont E, Zafar S, Rosenthal ES, Tabaeizadeh M, Ebrahim S, Dauwels J, Westover MB, 2018. Rapid annotation of seizures and interictal-ictal continuum eeg patterns. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 3394–3397. 7 18.
- Joao M, Cerdeira HA, Tanaka E, de Vitor C, Gomez P, 2018. Heuristic active learning for the prediction of epileptic seizures using single EEG channel. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, pp. 2628–2634. 12 3.
- Karuppiah Ramachandran VR, Alblas HJ, Le DV, Meratnia N, 2018. Towards an online seizure advisory system—an adaptive seizure prediction framework using active learning heuristics. *Sensors* 18 (6(6)), 1698.
- Kingma DP, Ba J, 2014. Adam: a Method for Stochastic Optimization arXiv preprint arXiv:1412.6980. 12 22.
- Koren JP, Herta J, Pirker S, Fürbass F, Hartmann M, Kluge T, Baumgartner C, 2016. Rhythmic and periodic EEG patterns of 'ictal–interictal uncertainty' in critically ill neurological patients. *Clin. Neurophysiol* 127 (2 (2)), 1176–1181. [PubMed: 26679421]
- Lawhern V, Slayback D, Wu D, Lance BJ, 2015. Efficient labeling of EEG signal artifacts using active learning. In: *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, pp. 3217–3222. 10 9.
- Leitinger M, Beniczky S, Rohrachner A, Gardella E, Kalss G, Qerama E, Höfler J, Lindberg-Larsen AH, Kuchukhidze G, Dobesberger J, Langthaler PB, 2015. Salzburg consensus criteria for non-convulsive status epilepticus—approach to clinical application. *Epilepsy Behav.* 49 (8), 158–163. [PubMed: 26092326]
- Macas M, Grimova N, Gerla V, Lhotska L, Saifutdinova E, 2018. Active learning for semiautomatic sleep staging and transitional EEG segments. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, pp. 2621–2627. 12 3.
- McInnes L, Healy J, Melville J, 2018. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction arXiv preprint arXiv:1802.03426. 2 9.
- Newey CR, Sahota P, Hantus S, 2017. Electrographic features of lateralized periodic discharges stratify risk in the interictal–ictal continuum. *J. Clin. Neurophysiol* 34 (7 (4)), 365–369. [PubMed: 28166083]
- Osugi T, Kim D, Scott S, 2005. Balancing exploration and exploitation: a new algorithm for active machine learning. In: *Fifth IEEE International Conference on Data Mining (ICDM'05)*. IEEE, pp. 8–pp. 11 27.
- Pohlmann-Eden B, Newton M, 2008. First seizure: EEG and neuroimaging following an epileptic seizure. *Epilepsia* 49 (1), 19–25.
- Rubinos C, Reynolds AS, Claassen J, 2018. The ictal–interictal continuum: to treat or not to treat (and how)? *Neurocrit. Care* 29 (8 (1)), 3–8. [PubMed: 29139014]



- Settles B, 2009. Active Learning Literature Survey. University of Wisconsin-Madison Department of Computer Sciences.
- Settles B, 2011. From theories to queries: active learning in practice. Active Learning and Experimental Design Workshop In Conjunction With AISTATS 2010 1–18. 4 21.
- Sivaraju A, Gilmore EJ, 2016. Understanding and managing the ictal-interictal continuum in neurocritical care. *Curr. Treat. Options Neurol* 18 (2 (2)), 8. [PubMed: 26874841]
- Temko A, Thomas E, Marnane W, Lightbody G, Boylan G, 2011. EEG-based neonatal seizure detection with support vector machines. *Clin. Neurophysiol* 122 (3(3)), 464–473. [PubMed: 20713314]
- Thodoroff P, Pineau J, Lim A, 2016. Learning robust features using deep learning for automatic seizure detection. *Machine Learning for Healthcare Conference* 178–190. 12 10.
- Wu D, Lawhern VJ, Gordon S, Lance BJ, Lin CT, 2016. Offline EEG-based driver drowsiness estimation using enhanced batch-mode active learning (EBMAL) for regression. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp. 000730–000736. 10 9.
- Yang H, Guan C, Ang KK, Pan Y, Zhang H, 2012. Cluster impurity and forward-backward error maximization-based active learning for EEG signals classification. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 569–572. 3 25.
- Yang S, Lopez S, Golmohammadi M, Obeid I, Picone J, 2016. Semi-automated annotation of signal events in clinical EEG data. In: 2016 IEEE Signal Processing Medicine and Biology Symposium (SPMB). IEEE, pp. 1–5. 12 3.
- Zafar SF, Subramaniam T, Osman G, Herlopian A, Struck AF, 2020. Electrographic seizures and ictal-interictal continuum (IIC) patterns in critically ill patients. *Epilepsy Behav.* 1 (5 (106)), 107037.

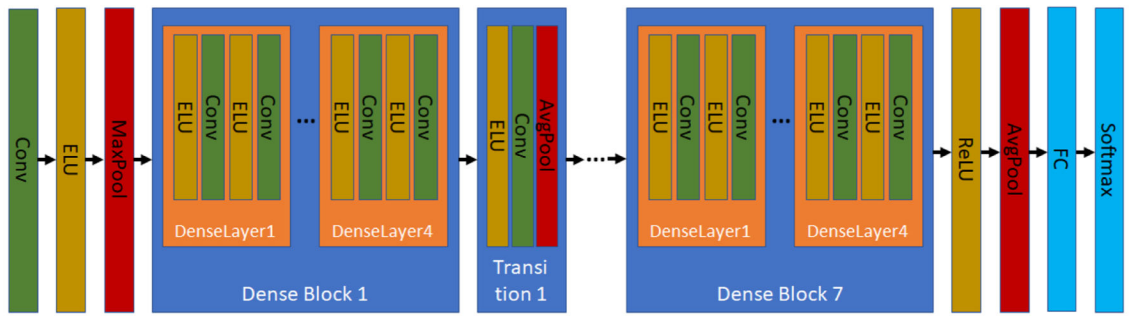


(1a) Bipolar Montage

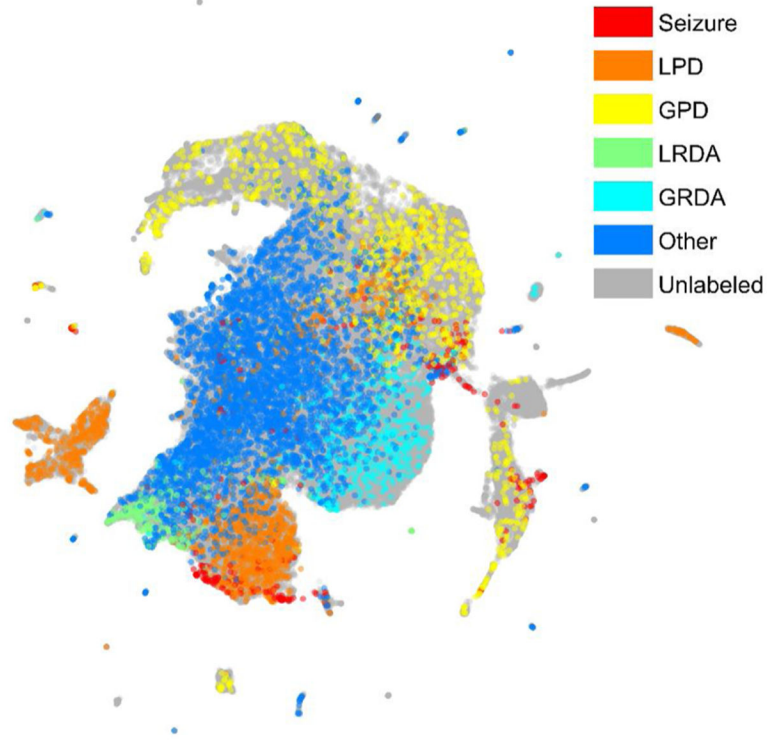


(1b) IIC EEG patterns

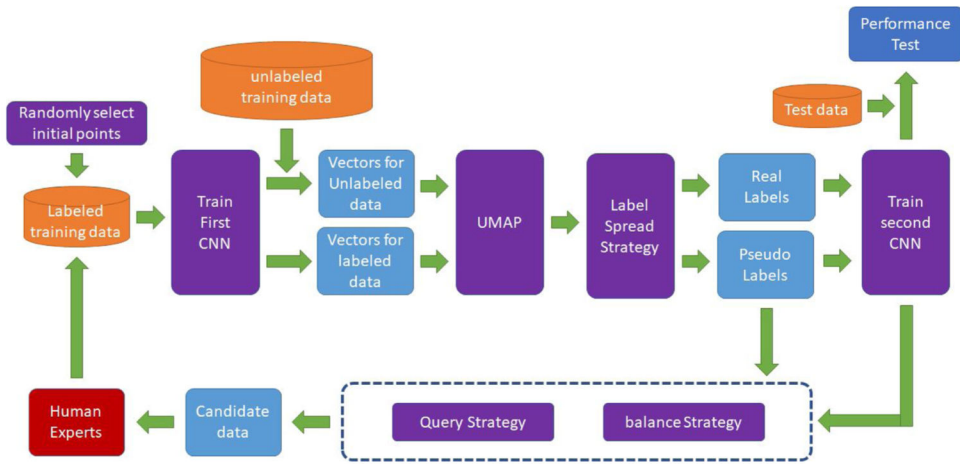
**Fig. 1.**  
Bipolar Montage and IIC EEG patterns.



**Fig. 2.**  
Structure of the Dense-Net CNN.



**Fig. 3.**  
The UMAP for Seizure and IIC.



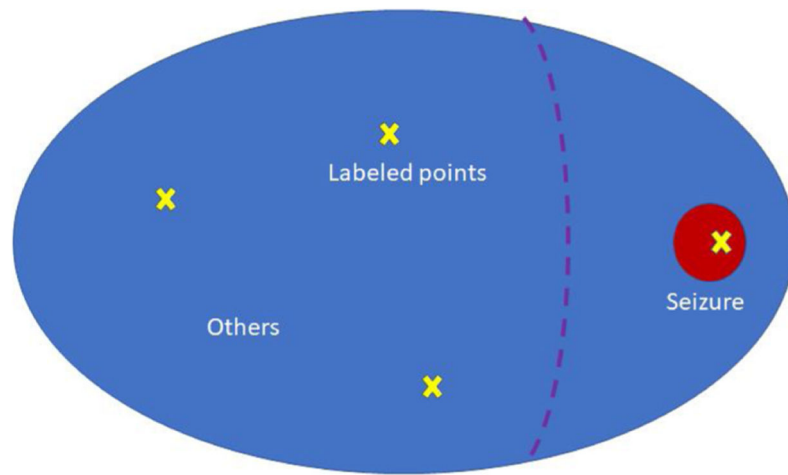
**Fig. 4.** Our framework for Active Learning.

Author Manuscript

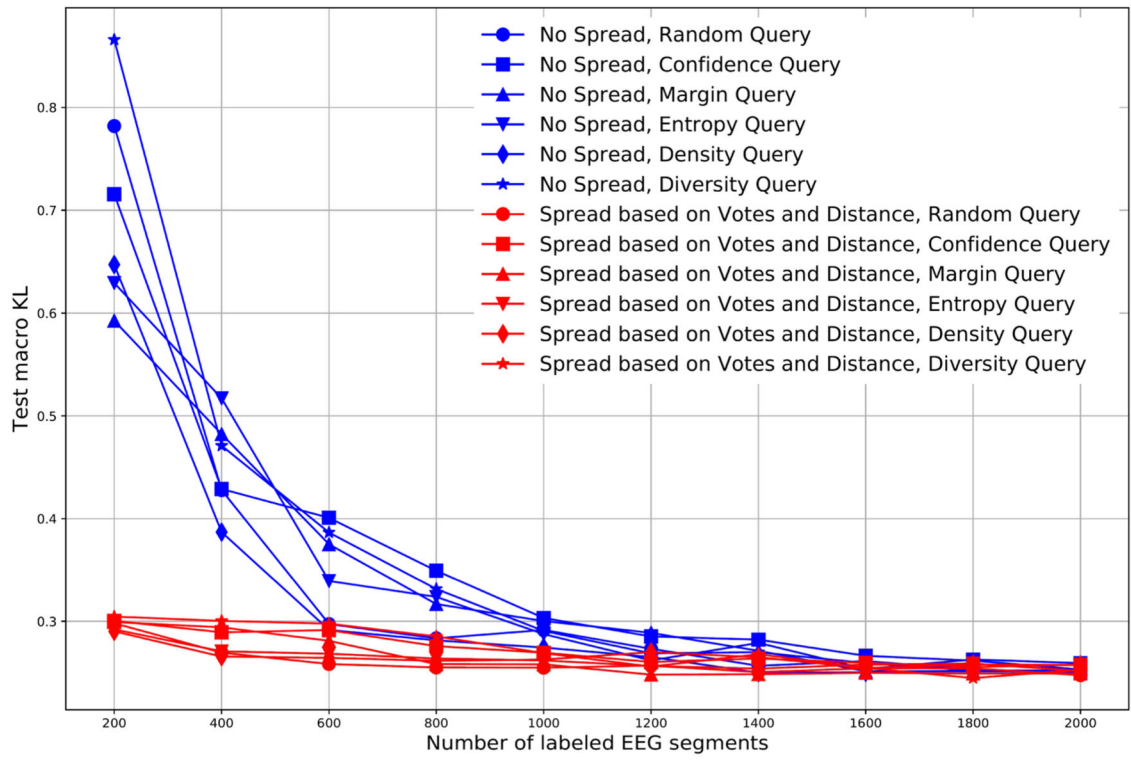
Author Manuscript

Author Manuscript

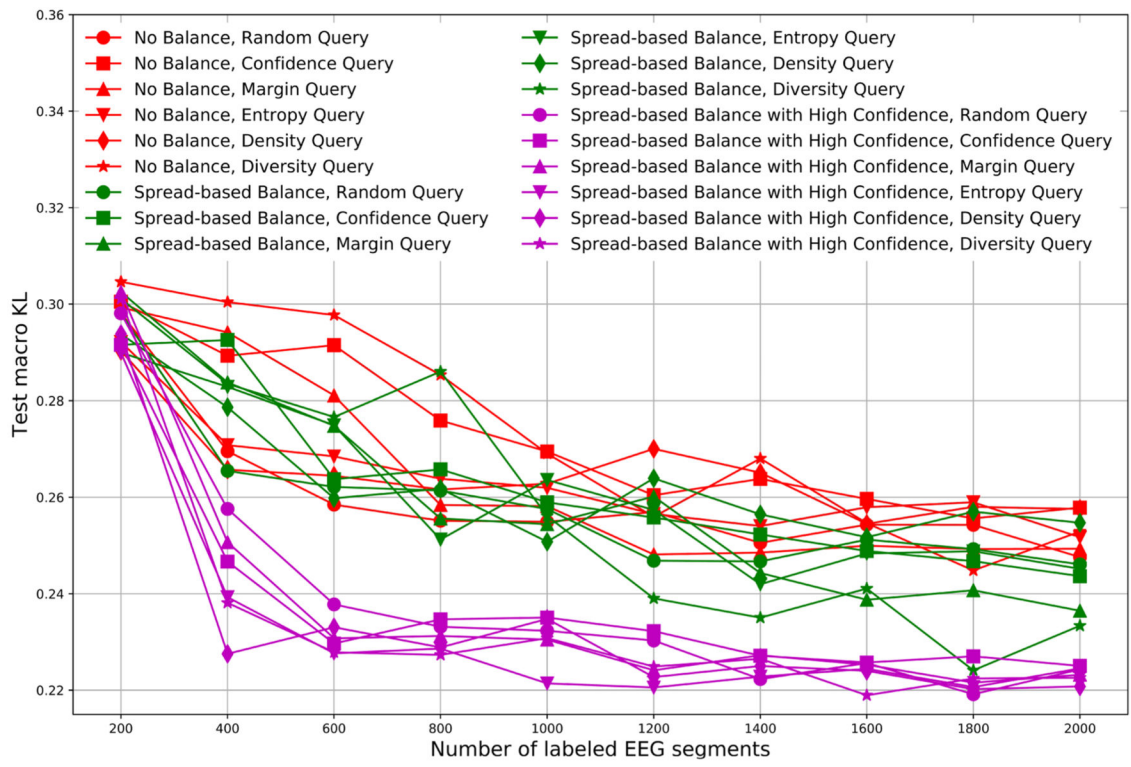
Author Manuscript



**Fig. 5.** Illustration of the balance problem. The blue region represents the majority class (“Other”). The red region represents the minority class (“Seizure”). Yellow x’s represent labeled points within each region. Labels spreading would correctly assign pseudo-labels as “Other” to points left of the purple dashed line, but would incorrectly pseudo-label as “Seizure” most points to the right of the line. To the right of the dashed line, only those in the red circle are actually seizures.

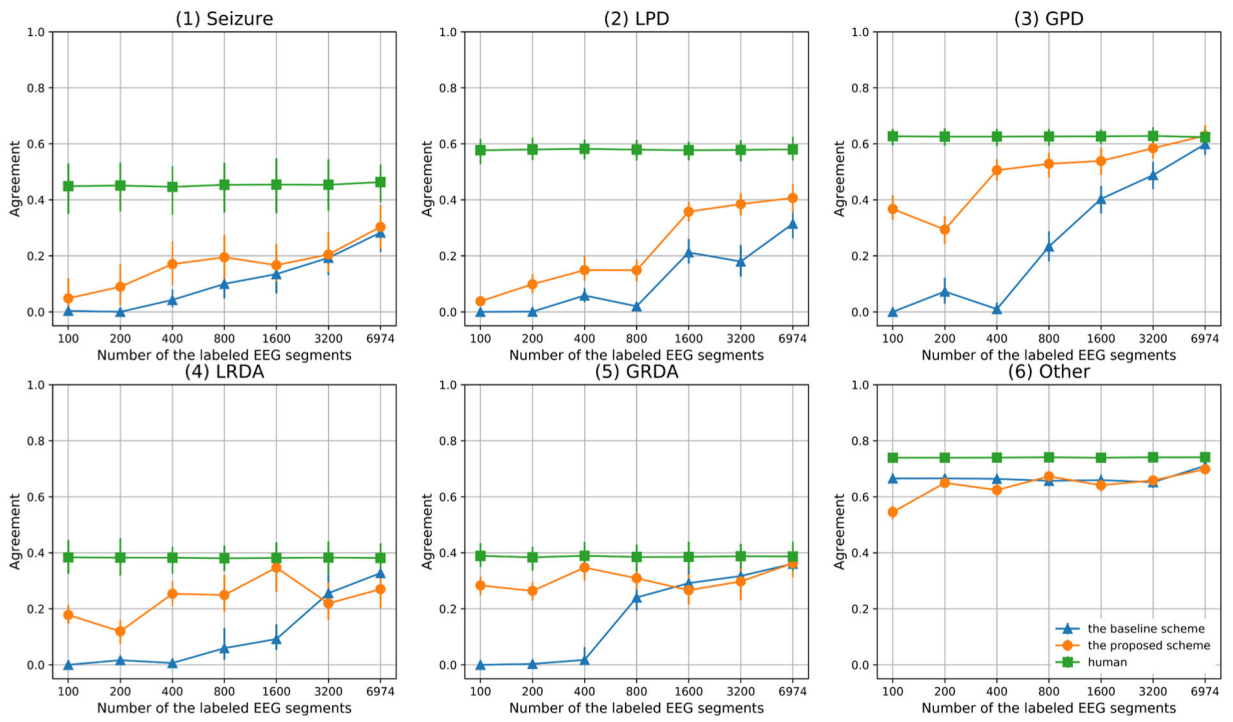


**Fig. 6.** Comparison between AL with label spread and AL without label spread.



**Fig. 7.**  
Comparison of different balancing schemes.





**Fig. 8.** Agreement Comparison.

**Table 1**

Dataset for Seizures and IIC.

	<b>Number of patients</b>	<b>Number of segments</b>	<b>Percentage</b>
Training	1003	26752	~70 %
Validation	144	3666	~10 %
Test	288	8941	~20 %
Total	1435	39359	100%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Number of EEG segments in different pattern categories.

	<b>Other</b>	<b>Seizure</b>	<b>LPD</b>	<b>GPD</b>	<b>LRDA</b>	<b>GRDA</b>
Training	23691(89 %)	247 (1 %)	1171 (4 %)	931 (3 %)	273 (1 %)	439 (2 %)
Validation	3051 (83 %)	53 (1 %)	212 (6 %)	229 (6 %)	73 (2 %)	48 (1 %)
Test	7961 (89 %)	57 (1 %)	353 (4 %)	331 (4 %)	66 (1 %)	173 (2 %)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Primary neurologic and medical problems leading to hospital admission.

<b>Diagnosis</b>			
<b>Diagnosis Neuro</b>	<b>Counts</b>	<b>Diagnosis Systemic</b>	<b>Counts</b>
Brain tumor/primary or mets	123	Dermatological/Musculoskeletal	9
CNS infection/inflammation	111	Endocrine emergency	23
Hypoxic ischemic encephalopathy/Anoxic brain injury	142	Gastrointestinal non-hemorrhagic	24
Other TBI (including SAH)	18	Genitourinary	43
Other neurosurgical (non tumor/trauma) (eg EVD/VPS etc)	10	Immunological	5
Other structural-degenerative diseases	35	Liver disorders	49
Primary psychiatric disorder	39	Malignancy (solid tumors/hematologic)	84
SAH	160	Non head trauma (including spine)	13
SDH	103	Other post operative (eg CT surgery, transplant)	7
SDH + other TBI (including SAH)	52	Primary hematological disorder	22
Seizures/status epilepticus	387	Renal failure	96
Spells (behavioral events suspicious for seizures)	6	Respiratory disorders	393
Toxic metabolic encephalopathy	161	Systemic hemorrhage	9
hemorrhagic stroke	137	cardiovascular disorders	251
ischemic stroke	91	<b>Total</b>	<b>1028</b>
Epilepsy/suspected Epilepsy (EEG)	252		
<b>Total</b>	<b>1827</b>		

Abbreviations: mets = metastases; CNS = central nervous system; TBI = traumatic brain injury; EVD = external ventricular drain; VPS = ventriculoperitoneal shunt; SAH = subarachnoid hemorrhage; SDH = subdural hematoma; EEG = electroencephalography; EMU = epilepsy monitoring unit; CT = cardiothoracic surgery.