EBioMedicine 45 (2019) 422–431

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Identifying signal-dependent information about the preictal state: A comparison across ECoG, EEG and EKG using deep learning

Christian Meisel [a,b,*], Kimberlyn A. Bailey [a]

[a] Technical University of Dresden, 01069 Dresden, Germany
[b] Boston Children's Hospital, Boston, USA

ARTICLE INFO

Article history:
Received 30 April 2019
Received in revised form 8 June 2019
Accepted 1 July 2019
Available online 9 July 2019

Keywords:
Epilepsy
Seizure prediction
Electrocardiogram
Deep neural networks
Precision medicine

ABSTRACT

Background: The inability to reliably assess seizure risk is a major burden for epilepsy patients and prevents developing better treatments. Recent advances have paved the way for increasingly accurate seizure preictal state detection algorithms, primarily using electrocorticography (ECoG). To develop seizure forecasting for broad clinical and ambulatory use, however, less complex and invasive modalities are needed. Algorithms using scalp electroencephalography (EEG) and electrocardiography (EKG) have also achieved better than chance performance. But it remains unknown how much preictal information is in ECoG versus modalities amenable to everyday use – such as EKG and single channel EEG - and how to optimally extract that preictal information for seizure prediction.

Methods: We apply deep learning - a powerful method to extract information from complex data - on a large epilepsy data set containing multi-day, simultaneous recordings of EKG, ECoG, and EEG, using a variety of feature sets. We use the relative performance of our algorithms to compare the preictal information contained in each modality.

Results: We find that single-channel EKG contains a comparable amount of preictal information as scalp EEG with up to 21 channels and that preictal information is best extracted not with standard heart rate measures, but from the power spectral density. We report that preictal information is not preferentially contained in EEG or ECoG channels within the seizure onset zone.

Conclusion: Collectively, these insights may help to devise future prospective, minimally invasive long-term epilepsy monitoring trials with single-channel EKG as a particularly promising modality.

1. Introduction

The inability to reliably assess seizure risk is a major burden for epilepsy patients [1]. From a clinician's perspective, it prevents developing better treatment strategies to optimise antiepileptic drug regimen and timely intervention strategies to avert impending seizures, needed especially for the 30% of patients refractory to current treatments [2,3].

The search for reliable seizure risk assessment methods - or "seizure forecasting" - has a long history in epilepsy research. After initial attempts [4], a recent surge of increasingly accurate seizure forecasting algorithms has lent renewed support to such a project [5–8]. In a prospective application, electrocorticography (ECoG) has proven useful for such as task [9], retrospective studies have similarly shown that also scalp electroencephalography (EEG) and electrocardiography (EKG) are capable of differentiating preictal from interictal states with better-than-random chance [10–13]. However, to make seizure risk assessment available for broad clinical and ambulatory use methods less invasive than ECoG and less complex than scalp EEG are needed. Single channel EKG or EEG are particularly interesting in this regard as they can be measured with minimal complexity and no invasiveness allowing implementation for broad ambulatory use [3].

What has thus far remained critically under-explored is how much information about an impending seizure is specifically contained in modalities like EKG or EEG compared to ECoG and how this preictal information can best be extracted from these modalities for identifying preictal periods. Assessing the preictal information contained in these modalities with respect to preictal state detection capabilities is an essential step for optimally designing prospective, long-term seizure forecasting trials. Furthermore, reliable characterization of where and when information about a preictal state is prevalent in the body may indirectly provide insights into the pathophysiology of seizure generating mechanisms.

Deep learning has elevated information extraction from complex data to new levels and has been repeatedly shown to achieve top performance on classification tasks [14]. Deep neural networks are a powerful

* Corresponding author.
E-mail address: christian@meisel.de (C. Meisel).

https://doi.org/10.1016/j.ebiom.2019.07.001

Research in context

Evidence before this study

The unpredictability of seizures severely hinders the quality of life of epilepsy patients and prevents developing better treatments to control or avert seizures. Efforts to reliably forecast seizures have made considerable progress in recent years using recordings from electrocorticography, i.e. recordings from electrodes implanted beneath the skull (ECoG). However, to make seizure forecasting not just possible but amenable to widespread use, it is necessary to develop seizure forecasting that uses data from a less invasive, easy to use modality than electrocorticography.

Added value of this study

We used ECoG and recordings of modalities more amenable to everyday use - single sensor electrocardiography (heart activity monitor, EKG) and electroencephalography (electrodes temporarily adhered to the scalp, EEG). We applied modern data science technology, deep learning algorithms, to train algorithms to distinguish times shortly before a seizure from times far from a seizure, using each modality. We compared how well these alternate modalities performed relative to ECoG. We find that, EKG performs better than a single-channel EEG and comparably to all-channel EEG when the whole EKG signal is considered and not just heart rate, as used in more traditional approaches.

Implications of all the available evidence

In the search for a modality that is not just predictive of seizures but amenable to widespread clinical use, our work suggests EKG as a promising modality that warrants more research. Collectively, our results and previous seizure forecasting work suggests EKG should be included in future clinical trials.

tool fit for comparing information content in different data streams valuable for a given task. Provided with enough data, deep networks can extract information from data by approximating a function capturing the relationship between the input (in this case, segments of EKG, EEG, or ECoG data) and the output (in this case, whether a given segment is classified as preictal or interictal). Trained to detect preictal states from data of each modality, deep learning algorithms can, in other words, effectively extract the preictal information contained in each modality.

Here, we systematically study the information content of ECoG, EEG and EKG in terms of its capability to reliably detect preictal states. For our approach, we use a comprehensive epilepsy data set containing multi-day, simultaneous recordings of all three modalities [15]. To assess the preictal information contained in each, we train deep learning algorithms to classify data segments as either pre- or interictal and compare the performance of networks trained on each modality.

2. Materials and methods

2.1. Simultaneous electrocorticography, scalp encephalography and electrocardiography recordings

The ECoG, EEG, and EKG data sets consisted of multi-day, simultaneous, continuous recordings from 10 patients undergoing presurgical monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany [15]. The Ethics Committee of the University of Freiburg approved the use of the data for research. ECoG, EEG and EKG were recorded at 256, 512 or 1024 Hz and later downsampled to 256 Hz. The number of ECoG electrodes varied between patients (between $n = 30$ and $n = 114$ electrodes). Electrode placement was determined solely by clinical considerations and included both surface and depth electrodes. Scalp EEG was recorded with 21 channels. EEG was not always recorded for the whole duration of EKG and ECoG recordings, and one patient (patient 114) did not have scalp EEG recorded Table 1.

2.2. Preictal and interictal data segments

Signal segments were divided into 30-second sub-segments and classified as either preictal or interictal using the following criteria. All interictal segments were at least two hours away from preceding or proceeding seizures. Similar to other epilepsy forecasting research, we focused on lead seizures and only considered preictal segments, if the corresponding seizure was at least two hours away from the preceding seizure. Preictal segments were composed of 20 min taken from the 70 to ten minutes period prior to a seizure. This preictal window was chosen to be commensurate with the preictal window length of other EKG-based seizure prediction studies [10–13]. The ten-minute buffer period between the defined preictal period and seizure onset time was chosen to ensure that preictal data was entirely uncontaminated by seizure data and that, in real-world conditions, there would be potential for an interventional treatment to prevent seizures [8]. Using these criteria - in addition to a few data quality criterion (See subsection "Extracting heart rate from EKG") - a total of 1275 interictal hours and 20 preictal hours corresponding to 60 seizures were used for ECoG and EKG. A detailed breakdown of interictal/preictal data for ECoG and EKG for each patient is provided in Table 2. EEG was not always recorded for the whole duration and one patient (patient 114) did not have scalp EEG recorded. For the nine patients who had EEG, ECoG and EKG data was matched to the same times EEG was available (on average 7650 $\pm$ 3146 min interictal data, 120 $\pm$ 53 min of preictal data (mean $\pm$ standard deviation)).

2.3. Feature sets

Deep neural networks are unique among machine learning algorithms for achieving remarkably good performance without feature engineering. Using interconnected layers, deep learning networks can extract information from data and learn abstract representations in the data useful for the given task [14]. Still, machine learning performance generally depends upon well-designed features, which may be even more important in relatively small data sets such as ours [16]. Our feature sets thus fall into two major types: (1) features that give the networks the greatest advantage of working with the raw data, in effect, giving networks free reign to extract information from the data optimal for preictal state detection, and (2) preprocessed features the literature endorses as important for capturing preictal effects. We denote our (1) raw data feature sets as "PSD" for power spectral density-based features and (2) the literature-informed feature sets as "SEL" for "select" features. For both feature set types, a 50 Hz notch filter was used prior to filter out any line noise.

Table 1
Details for each patient's seizure focus location and EEG channels considered to reside within the seizure focus.

Patient ID	Seizure focus	EEG channels considered over focus
565	Temporal: L	T1, T3, T5
375	Temporal: L/R, Frontal: R	T1, T2, T3, T4, T5, T6, FP2, F4, F8
264	Temporal: L/R	T1, T2, T3, T4, T5, T6
970	Temporal: R	T2, T4, T6
253	Temporal: L/R	T1, T2, T3, T4, T5, T6
108	Frontal: L	FP1, F3, F7
109	Temporal: L	T1, T3, T5
958	Temporal: L	T1, T3, T5
922	Frontal/Central: L	Fp1, F3, F7, C3

Table 2
Details of each patient's interictal and preictal data.

Patient ID	Seizures	Preictal minutes	Interictal minutes
565	5	100	10,920
375	7	140	6000
264	3	60	6540
970	9	180	7740
253	5	100	13,500
108	12	240	8760
114	3	60	3900
109	4	80	7620
958	7	140	9480
922	5	100	2040
Total	60	1200	76,500

2.4. Feature sets based on power spectral density (PSD)

The power spectral density (PSD) of each data segment was used as a compact representation of the raw data, a standard way to represent a time series in many fields, including epilepsy research and machine learning [7,17–19]. The PSD was computed for the raw ECoG, EEG, and EKG signals using Fast Fourier Transform with five-second, non-overlapping Hanning windows. This generally resulted in a feature vector of 641 length. For ECoG, three feature sets were produced from the PSD to compare predictive performance: (1) The average PSD across all subdural electrodes (ECOG-PSD), (2) the PSD from each individual subdural electrode (ECOG-PSDSingleChannel), where the PSD of each channel comprised its own feature set, and (3) the collected PSDs of each individual channel as one feature vector (ECOG-PSDAllChannels). Similarly, for EEG we created two feature sets: (1) the PSD from each individual subdural electrode (EEG-PSDSingleChannel) where the PSD of each channel comprised its own feature set and (2) the collected PSDs of each individual channel as one feature vector (EEG-PSDAllChannels). For EEG/ECOG-PSDAllChannels, the PSD was computed with a resolution of 2.5 Hz, producing, for each 30-second segment, feature vectors of 1344 length for scalp EEG (64 × 21 channels) and (64 × number of channels a given patient had) for ECoG.

2.5. Select EKG features based on heart rate (EKG-SEL)

We used standard measures based on heart rate in the time and frequency domains [20,21]. These features were similar to the feature sets used by other EKG-based seizure prediction research [10–13]. Measures of heart rate and heart rate variability (HRV) have been shown to be clinically useful markers of pathology, including epilepsy, and fruitful for autonomic nervous system research [20–23]. We used 13 features in total from the literature, denoted as EKG-SEL. For details, see the supplementary materials.

2.6. Select ECoG features (ECOG-SEL)

For the ECOG-SEL feature set, measures were in part chosen from the literature [4], and in part from our recent submission to a seizure prediction competition [6]. The latter were based on a dynamical systems theory that proposes a critical transition between low and high network excitability underlying seizure onset [24]. This feature set had 129 measures in total. The complete feature list is described in the supplementary materials.

2.7. Deep learning networks

Deep neural networks were trained for binary classification to classify segments as either preictal or interictal. We used a five-fold cross validation approach where data sets for each patient were randomised and divided into 70% training and 30% test sets in each fold. Algorithms were trained per-patient. We used balanced learning to handle our severely imbalanced training set, repeating preictal segments in the training set until a 50–50 preictal-interictal balance was achieved. A data augmentation approach, where Gaussian noise (zero mean, 1000 variance) was added to preictal PSD data, yielded very similar results (Supplementary Fig. 4). Deep neural networks were built and trained using Keras with Tensorflow backend (batch size = 256, learning rate 0·00001, ADAM optimizing for binary cross-entropy). Training was stopped after 5000 epochs (500 epochs for ECOGPSDAllChannels, due to larger network size taking longer to train) at which point learning curves had reached a plateau. A hyperparameter search for network size - changing the number of nodes in layers three and five - demonstrated robustness of results as a function of network size. Dropout was used to prevent overfitting. Full algorithm topology and hyperparameter details are outlined in Table 3.

2.8. Baseline machine learning methods

We used the following three algorithms to compare deep learning approaches to: k-Nearest Neighbors (k-NN), linear classifier using SGD (SGD), and AdaBoost. To determine the optimal value of k in k-NN we used GridSearchCV. SGD fit a linear support vector machine (SVM) with l2 penalty. AdaBoost used a decision tree classifier, maximum number of 50 estimators.

2.9. Layer-wise relevance propagation

Although deep neural networks are a powerful tool, a major downside of deep neural networks is that they are uninterpretable "black boxes". Several methods have sought to make deep neural networks interpretable [25]. Here we used one such method, layer-wise relevance propagation (LRP) [26], to assess the value of our chosen inputs.

LRP measures how much each feature influences a trained network's predictions by computing the "relevance" for each feature in a feature set fed into a trained network. To understand what this relevance measure means within the context of our work, consider a network trained to perform binary classification. For such a network, predictions may fall between zero and one, with predictions ≥0·5 corresponding to a given class A, and predictions <0·5 corresponding to a given class B. Such a trained network might produce a prediction of 0·9 when fed a particular example, indicating the network classified that example as class A. LRP distributes that prediction back onto the inputs, measuring how much each feature influenced the network to make its prediction of 0·9. LRP might, for example, compute 0·1 relevance for feature X, 0·005 relevance for feature Y, and so on, such that the relevance across all features sums up to the final 0·9 prediction. We can then assess how much each feature contributed to that prediction. Consequently, what is key for understanding relevance values is not their magnitude, but their scale relative to the relevance values of other features in the feature set. Within this example, feature X with relevance 0·1 - making up a large fraction of 0·9 - can be interpreted to have far more strongly pushed the network to classify the example as class A (prediction ≥0·5) than feature Y, which had a much smaller relevance value of 0·005.

Within our work, we are chiefly concerned about reliably identifying preictal segments (≥0·5 predictions). Thus, we wanted to explore what features reliably identified preictal segments. In other words, we wanted to identify features that had reliably high relevance values for preictal segments, consistently pushing the network to make high (≥0·5) predictions for preictal segments across all preictal segments. Thus, we computed the relevance values for each preictal segment and identified the features that, across all preictal segments, had consistently high relevance relative to the other features. Those features with consistently high relevance - or in the case of our feature sets comprised of the power spectral density, those frequency ranges with consistently high relevance - may be interpreted as reliable preictal biomarkers.

Table 3
Details of deep neural network architecture. Results were robust under different network sizes quantified by the factor (0.5, 1, 2, 3, 4) by which the number of nodes in layer 3 and 5 was multiplied. Results in the manuscript are based on networks with factor = 0.5. Dense, Dropout and ReLU denote fully-connected layer, dropout, and rectified linear unit non-linearity respectively. Learning rate was set to 0·00001.

Layer	Layer type	Number of nodes	Layer parameters
1	Dense	Number of features	Activation = ReLU, weight constraint = max norm(3)
2	Dropout	NA	Dropout rate = 0·2
3	Dense	Number of features x factor	Activation = ReLU, weight constraint = max norm(3)
4	Dropout	NA	Dropout rate = 0·2
5	Dense	Number of features x factor	Activation = ReLU, weight constraint = max norm(3), weight initialization = normal
6	Dense	20	Activation = ReLU, weight constraint = max norm(3), weight initialization = normal
7	Dense	1	Activation = sigmoid, weight initialization = normal

There are several variants for how LRP computes feature "relevance", with some variants more suited to some network architectures and applications than others [26]. We used the α-β decomposition formula to compute relevance, with α set as one. This variant is appropriate for network architectures that use the rectified linear activation unit (ReLu) activation function, as is used within our networks (Table 3). This LRP variant only distributes relevance onto input features that successfully propagated all the way through the network - that is, features that never contributed to a negative input value at a node with ReLu activation function - so that relevance is only computed for features that contributed to the final prediction. Our networks use ReLu activation functions for all but the output layer. Thus, we used this particular LRP variant to only compute relevance for the input features for each example that actually contributed to the prediction.

2.10. Evaluating algorithm performance

Performance on our binary classification task was assessed using two performance metrics: the area under the receiver operating curve (AUC) and improvement over chance based on F1 scores (IoC-F1). AUC is a commonly used performance metric to assess classification performance. We used IoC-F1 as a metric particularly well-suited to the characteristics of our classification task: a highly imbalanced data set for which accurately classifying the minority class is a high priority. The F1 score is the harmonic mean between precision and recall which is particularly suitable in highly imbalanced data sets such as ours [27]. Specifically, precision and recall are metrics that do not take into account how well the classifier correctly identified majority class examples (true negatives). A performance metric based on true negatives may indicate an excellent performance even if the classifier performs poorly on the minority class, by virtue of the fact that the majority class vastly outnumbers minority class examples. This is important for preictal state detection, for which interictal periods vastly outnumbers preictal periods. As a metric particularly well suited to our classification task, we use F1 as our primary performance metric.

To compare the performance of our algorithms to that of a chance predictor, we used improvement over chance based on F1. The F1 score was computed for random predictions (i.e., randomly shuffled test set labels, which maintains the numbers of pre- and interictal classifications but destroys any correlation to the data) averaged across 1000 randomizations. We compared the F1 scores of our trained networks (F1(Data)) to the F1 score of a random predictor (F1(Chance)) by calculating the improvement over chance (IoC), defined as IoC = (F1(Data) - F1(Chance))/(1-F1(Chance)). IoC is often used in epilepsy forecasting research [9,17].

2.11. Statistical tests

To compare performance metrics across algorithms, the Mann-Whitney U test was used. In case the comparison involved more than two groups, the Mann-Whitney U test was performed post-hoc following a one-way ANOVA.

2.12. Data availability

The data analyzed in this study is publicly available at the European Epilepsy Database [15]. The code used to analyze the data is available from the authors upon reasonable request.

3. Results

As a first step, deep learning networks were trained on ECoG or EKG, where features were either the power spectral density (PSD) or select measures based on the literature (SEL; Fig. 1A) as detailed in methods. Multi-layer neural networks (Table 3) were trained separately for each feature set and patient (Supplementary Fig. 1). Performance - i.e. how accurately the networks classified preictal and interictal segments - was measured by calculating the IoC-F1 and AUC scores averaged over out-of-sample data using a five-fold cross-validation approach.

The ability to distinguish between pre- and interictal states from individual ECoG channel data varied greatly between channels (Fig. 1B). Across patients, channels considered to fall within the seizure onset zone (SOZ; Fig. 1B, turquoise bars) did not on average exhibit better performance than channels considered outside of the SOZ (NonSOZ; Fig. 1B, C). Performance metrics for all feature sets and modalities varied among patients (Fig. 1D). In every case, algorithms exceeded the performance of a random predictor. Across patients, preictal state identification was best using the PSD of all ECoG channels (ECOG-PSDAllChannels), significantly outperforming all other approaches (IoC-F1: EKG-PSD p = 0·003; ECOG-SEL p = 0·001; average ECoG power, ECOG-PSD p = 0·0007; ECOG-PSDSingleChannel p = 0·00001; Fig. 1E left). Analyses using AUC as performance metric closely mimicked these results (AUC: EKG-PSD p = 0·032; ECOG-SEL p = 0·011; average ECoG power, ECOG-PSD p = 0·001; ECOG-PSDSingleChannel p = 0·0003; Fig. 1E right). The average single channel performance for ECOG-PSDSingleChannel was computed as the average performance across all single channel networks. Despite only containing one channel of data, EKG-PSD exhibited the second-best discriminative performance, with significantly better performance than the average performance of a single ECoG channel (IoC-F1: ECOG-PSDSingleChannel p = 0·001; AUC: p = 0·0005). When using the AUC as the performance metric as opposed to IoC-F1, EKG-PSD also outperformed the average ECoG PSD (ECOG-PSD: p = 0·013). In contrast to these EKG PSD results, EKG features based on heart rate generally performed significantly worse than all other approaches (Fig. 1E). Results were robust over a wide range of network sizes, as tested by varying the factor by which the number of nodes in layers three and five was multiplied (Table 3).

Deep learning approaches generally outperformed more conventional machine learning methods (k-NN, SGD, AdaBoost; Supplementary Fig. 3) in our data which highlights the overall superior capabilities of deep learning to extract information content from data. The relative performance between data modalities, however, was maintained also in the more conventional machine learning methods. For

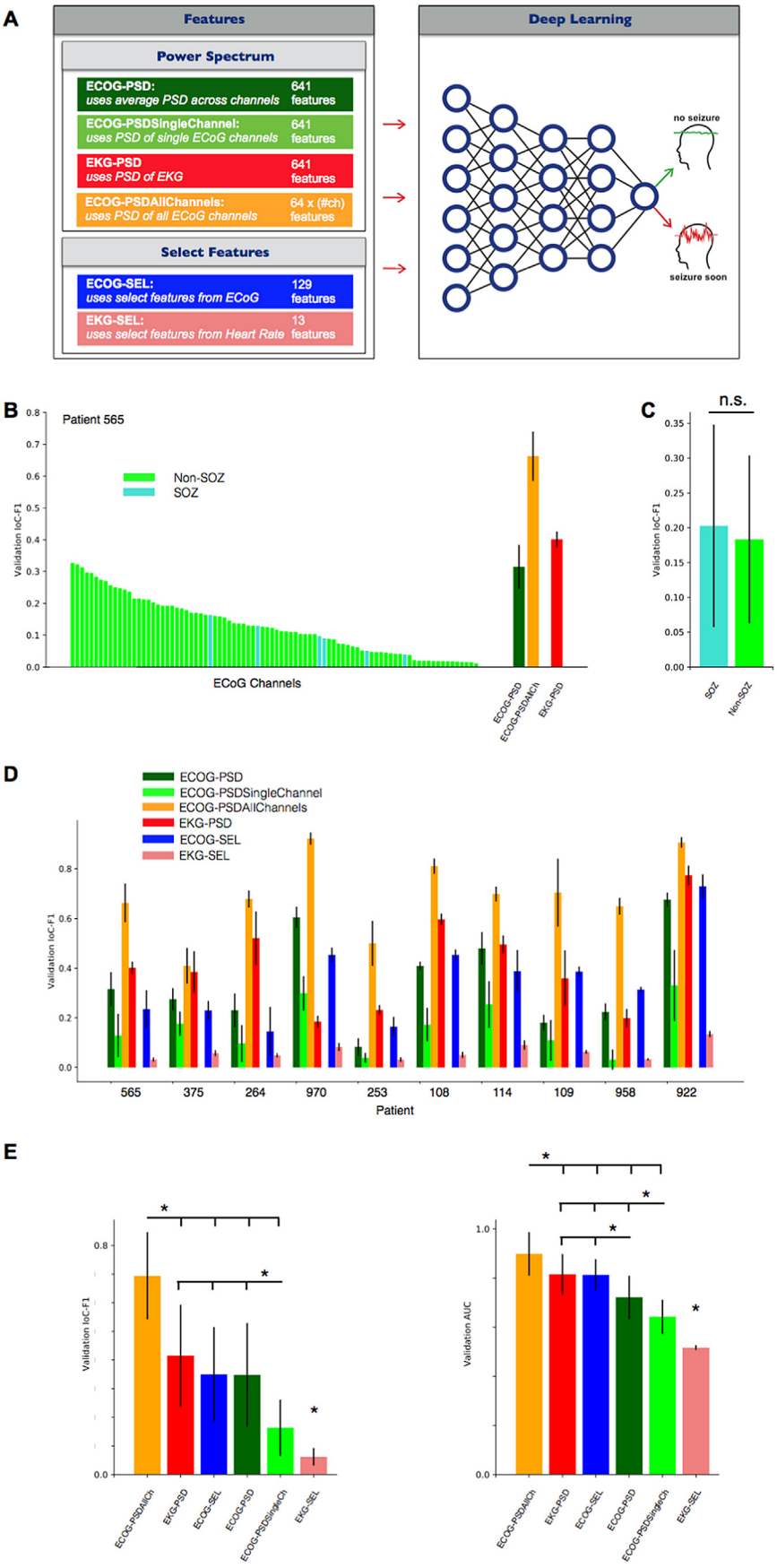

Fig. 1. Comparing information from ECoG and EKG in terms of their abilities to distinguish pre- from interictal states. (A) Illustration of the different feature sets paired with deep learning to assess the information contained in sensor data. (B) Classification performance for individual ECoG channels in one patient. (C) Average performance of all channels within the seizure onset zone (SOZ) and channels not within the SOZ (Non-SOZ). (D) Average performances across multiple network runs for each patient and all feature sets. (E) Mean performance metrics across all patients. Except for ECOG-PSDSingleChannel (averages across channels), all plots reflect mean values across five network runs. Whiskers denote standard deviation. Performance was assessed using improvement over chance (IoC-F1) and area under the ROC curve (AUC). * $p \leq 0{\cdot}05$.

example, EKG-PSD generally performed better than EKG-SEL in all approaches (Supplementary Fig. 3). This invariance to machine learning methods further supports the distinct differences in information content across data modalities.

Taken together, these findings demonstrate that the power spectrum of the EKG signal contains information for preictal state identification that is not contained in heart rate and which is only surpassed when information collectively from all ECoG channels are used. The amount of information in EKG is greater than in an average ECoG channel alone, irrespective of choosing this channel to be close to the seizure focus.

Next, we compared the predictive performance of these EKG and ECoG-based algorithms to scalp EEG-based algorithms. One patient, patient 114, did not have scalp EEG recorded and was thus excluded from these analyses. EEG was typically not recorded for the whole duration of the EKG and ECoG recordings. Thus, ECoG and EKG segments were matched to the segments for which EEG was available (Fig. 2A). As for ECoG, EEG channels considered close to the seizure focus (Table 3) did not exhibit better performance compared to other channels (Fig. 2C). Performance metrics for the different feature sets varied among patients (Fig. 2D), with algorithms based on the PSD of all ECoG channels (ECOG-PSDAllChannels) again performing best on average (IoC-F1: EEG-PSDAllChannels $p = 0{\cdot}026$; EKG-PSD $p = 0{\cdot}002$; EEG-PSDSingleChannel $p = 0{\cdot}0007$; Fig. 2E left). Analyses using AUC as performance metric closely mimicked these results (AUC: EEG-PSDAllChannels $p = 0{\cdot}005$; EKG-PSD $p = 0{\cdot}001$; EEG-PSDSingleChannel $p = 0{\cdot}0004$; Fig. 2E right). As for ECoG, EEG-PSDSingleChannel performance was computed as the average taken over all channels. EKG-PSD, despite having only one data channel, exhibited comparable performance to EEG using all 21 channels (IoC-F1: EEG-PSDAllChannels $p = 0{\cdot}239$; AUC: $p = 0{\cdot}239$), and outperformed single-channel EEG (F1-IoC: EEG-PSDSingleChannel $p = 0{\cdot}026$; AUC: $p = 0{\cdot}003$; Fig. 2E).

Next, we assessed which features - or, in the case of PSD-based algorithms, which frequency ranges - were most relevant for identifying preictal segments. For this purpose, we used layer-wise relevance propagation (LRP; for details, see methods) [26]. The features which may be considered the best preictal biomarkers are those that reliably push the network to classify preictal segments as preictal - in other words, those features that produce reliably high levels of relevance across all preictal segments. To identify those high relevance features, we produced relevance heatmaps, where each row corresponds to a preictal segment, each column corresponds to an input feature and more yellow colours indicate higher relevance. Fig. 3A shows heatmaps for ECOG-PSD of two patients (heatmaps for all other patients in Supplementary Fig. 2). The ECoG heatmaps indicated that relevance for each feature/frequency range was similar across preictal segments within a given patient but varied considerably between patients (Fig. 3A). For EKG, relevance was more consistent across patients, with generally high relevance levels in the lower frequency ranges (approx. ≤40 Hz; Fig. 3B). To quantify the similarity of relevance values across patients, we calculated the average cross-correlation between pairs of median feature relevance across patients (Fig. 3D). In line with visual assessments of the heatmaps, similarity of relevance levels between patients was much higher for EKG-PSD than for ECOG-PSD (Fig. 3D; $p = 1e\text{-}7$). We also observed highly similar relevance levels across patients for ECOG-SEL compared to ECOG-PSD (Fig. 3D; $p = 1e\text{-}10$). To investigate which ECoG-SEL features were most relevant to identify preictal states, we computed the average relevance across patients for each feature. ECoG-SEL features were composed of many measures computed for several frequency ranges for each channel. We grouped those features computed for each measure and calculated the average for each group (Fig. 3C). LRP indicated the highest average relevance for those features based on phase synchrony R, a measure that was recently implied in correlating well with cortical excitability levels [24,28].

4. Discussion

Efforts to make seizure forecasting available for broad clinical and ambulatory use have been underway for decades. But only recent work has shown reliable detection of preictal states, a crucial step toward seizure forecasting, is indeed feasible [8]. Seizure forecasting builds on the idea that the preictal state, the time right before seizure onset, can be reliably distinguished from the interictal state, i.e. times far from a seizure, using biophysical sensor data [4]. An increasing amount of long-term ECoG and EEG data of epilepsy patients and significant advances in machine learning have paved the way for a wealth of studies achieving better-than-chance performance on preictal state detection [5,7–9]. EEG, however, is not amenable to everyday, easy use and ECoG is too complex and invasive to be a good option for most patients. To make seizure risk assessment available for broad clinical use less invasive and more easily accessible methods are needed. Here, we compared how specifically non-invasive, more easily implementable modalities, like EKG or single channel EEG, compare to state-of-the-art ECoG approaches and what sort of features work best for these modalities.

Overall, in the search for a modality that is not just predictive, but amenable to wide clinical and ambulatory use, our results suggest EKG is a viable option. Research has demonstrated a variety of preictal effects on heart activity [29]. Indeed, we found that the single-sensor EKG contains considerable preictal information, comparable to scalp EEG (using all 21 channels). Furthermore, we found EKG is also more informative than a random single EEG or ECoG channel and more traditional algorithms based on heart rate [21–23]. For single EEG or ECoG channels, positioning the channel closer to the seizure onset zone does not necessarily improve seizure predictability. Our results suggest that ECoG, when using all channels, performs best out of all three modalities, and EKG and all-channel EEG achieve comparable performance below that of ECoG.

Our work suggests several recommendations for EKG-based preictal state detection research. First, the PSD or related features should be considered for EKG-based feature sets. Only our PSD-based EKG algorithms achieved performance competitive with that of ECoG and EEG. Second, heart rate and heart rate variability measures should not be the only measures considered for an EKG-based feature set. Our results suggest that the most valuable preictal information within EKG is not wholly contained in standard heart rate measures. Previous studies that developed EKG-based preictal state detection algorithms have used feature sets only composed of standard heart rate measures [10–13]. Our EKG-SEL algorithms - also composed of measures based on heart rate - performed considerably worse than EKG PSD-based algorithms. Deep neural networks have the unique capability among machine learning algorithms to learn abstract representations of the data valuable for the given task, in effect doing their own sort of feature engineering [14]. We provided the PSD of each EKG segment to our deep neural networks as a compact representation of the raw signal, giving networks free reign to extract relevant preictal information not necessarily yet identified by researchers. LRP results show that the lower (<40 Hz) frequency range - which contains heart rate information – is most informative in this regard [30]. But algorithm performance clearly benefitted from the freedom to extract both heart rate measures and other information.

Apart from the heart rate signal, however, the PSD also contains information about the QRS complexes that typically fall below 30 Hz[30]. This frequency range also contains potential other preictal EKG abnormalities (T wave inversion, ST elevation/depression, etc.) that are not captured by heart rate, but that can be picked up by the PSD. These changes may thus have enabled the better performance of our EKG-PSD algorithms in comparison to algorithms based solely on the heart rate (EKG-SEL). The EKG PSD may have also contained electromyography (EMG) signal from the trunk muscles which might increase when patients move more prior to a seizure. EMG typically tends to

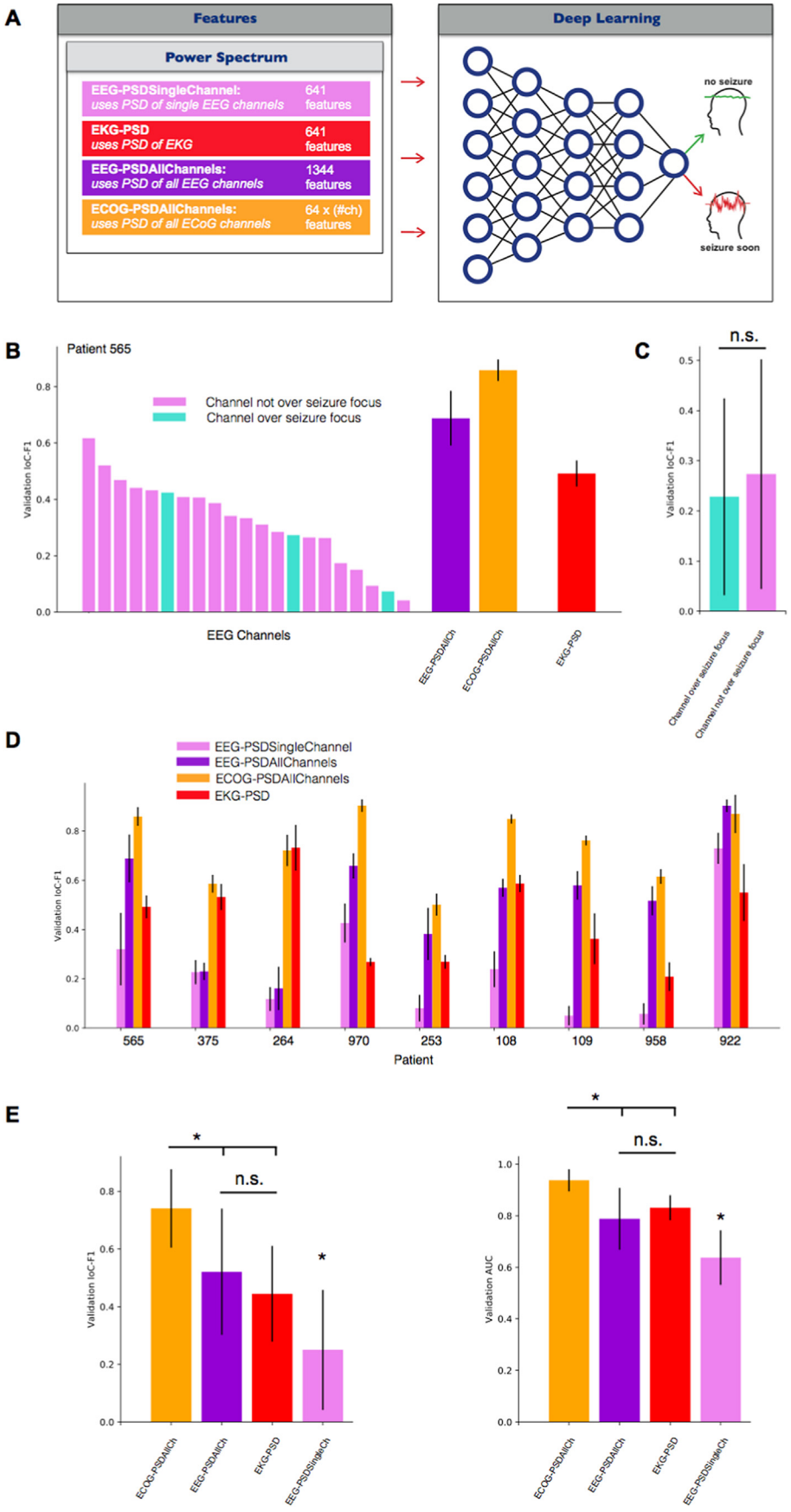

Fig. 2. Comparing information from scalp EEG, EKG and ECoG in terms of their abilities to distinguish pre- from interictal states. (A) Illustration of the different feature sets paired with deep learning to assess the information contained in sensor data. (B) Classification performance for individual EEG channels in one patient. (C) Average performance of all EEG channels over the seizure focus versus other EEG channels. (D) Average performances across multiple network runs for each patient and all feature sets. (E) Mean performance metrics across all patients. Except for EEG-PSDSingleChannel (averages across channels), all plots reflect mean values across five network runs. Whiskers denote standard deviation. Performance was assessed using improvement over chance (IoC-F1) and area under the ROC curve (AUC). * $p \leq 0{\cdot}05$.

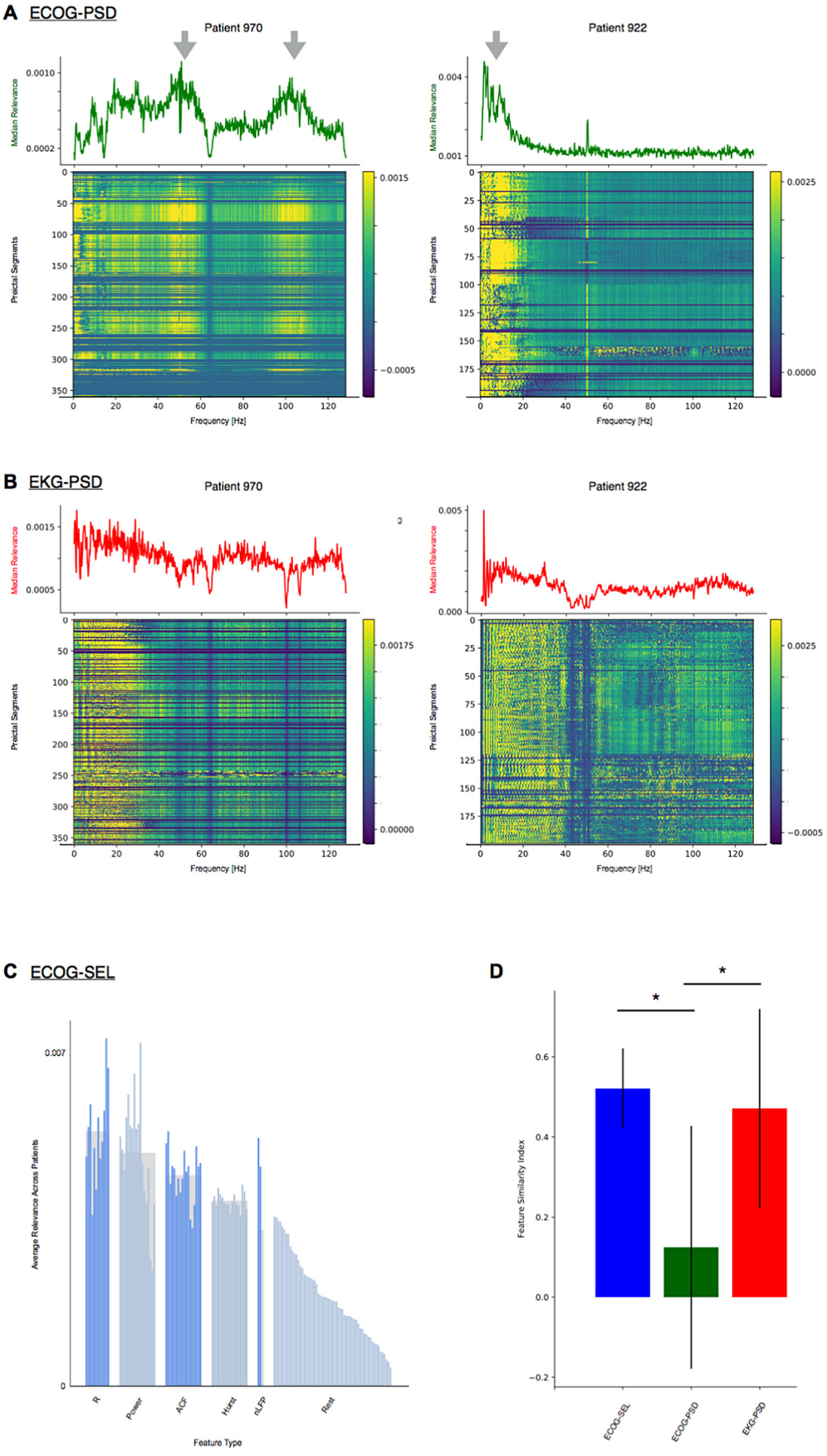

Fig. 3. Layer-wise relevance propagation (LRP) identifies informative preictal biomarkers/frequency bands (high relevance features) for each modality. (A) LRP reveals high relevance frequency bands in ECoG. Bottom: heatmap of feature relevance in two patients. Top: median feature relevance (green) across all preictal segments. ECoG signal frequency ranges with high relevance differs between patients exhibiting a more bimodal (left, grey arrows) or unimodal (right, grey arrow) distribution. (B) LRP reveals high relevance frequency bands (approx <40 Hz) in EKG. (C) Relevance of features calculated from the ECoG signal (ECOG-SEL). Average relevance across all preictal segments, divided into grouped columns by feature subtype and ordered by average relevance, greatest to least from the left. Averages for each grouped subtype are shown in grey. Features derived from synchrony R were the most informative on average. (D) Feature similarity index indicates that high relevance features are less similar within the ECOG-PSD group compared to ECOG-SEL and EKG-PSD. * $p < 1e\text{-}7$.

accumulate around 100 Hz, but also in lower frequencies and insofar overlaps with the EKG frequency spectrum. While an EMG contribution cannot be ruled out, our LRP results, which identified the most useful signal features to fall below 40 Hz — the frequency range that typically reflects EKG power — suggest that the EKG information within the PSD informed our predictions the most.

Making use of the richer stores of preictal information in the PSD of EKG may also help to improve performance of responsive vagus nerve stimulation (VNS). VNS is a method to treat drug-resistant epilepsy patients that triggers stimulation to control or avert seizures during periods of high seizure risk. Current VNS approaches mostly use changes in heart rate to detect high seizure risk periods. Our results, however, suggest, stimulation timing can be more accurate using the PSD instead.

Identifying those features that reliably identify preictal states can provide indirect insight into the physiology of seizure onset. A growing line of research suggests that changes in cortical excitability levels are conducive to seizure initiation and spread [31,32]. Our ECoG results support this line of research. Specifically, LRP results show measures previously identified to be closely related to excitability levels as features that reliably identify preictal states [28]. This suggests changes in excitability as one defining characteristic of the preictal state. Many of the measures used for ECOG-SEL, furthermore, were derived from a dynamical systems theory that proposes a critical transition underlie seizure onset [33]. The reliably good performance of these theory-driven features provides indirect support to this hypothesis and demonstrates how features informed by theory can be used to the benefit of machine learning algorithms.

Our findings have to be interpreted in the context of data collection and analysis. Our recordings covered only a few days of continuous data. We thus had comparably little data, including a limited number of seizures with which to properly train and evaluate our approach. Data sets that contain all three elements, EKG, EEG and ECoG, recorded over longer time spans, however, are currently not available. To optimally design clinical trials for long-term monitoring [3], it is important to incorporate the most informative sensors possible. A central goal of our work was to evaluate which modalities are worth incorporating for such a purpose. Future work should evaluate EKG-based seizure forecasting with longer data sets, ideally, in a prospective manner. Further validation of EKG- and EEG-based approaches will also be important to assess their performance under more real-life conditions. Clearly, conditions in the epilepsy monitoring unit are very different than in real life which might significantly impact performance. But again, to design such future long-term trials under more realistic conditions it seems necessary to first make decisions about sensor choice and methods informed by the data currently available. Another potential limitation relates to our validation approach. Given the relative short duration of our recordings, we validated our approach on an out-of-sample data partition after randomly shuffling data segments. Future work with more long-term data should validate the information content in each signal in a truly prospective, or at least pseudo-prospective manner. Deep learning is often desirable because it can produce good results without extensive feature selection. Generally, though, high dimensionality can harm performance, especially with smaller data sets like ours [34] and when features are correlated. In our data, however, there is no indication to suggest that the average performance declined with feature set size. Our best performing feature set based on the ECoG PSD, for example, had feature vectors with hundreds of features, while one of our worst performing feature sets, EKG SEL, had 14 features. Another potential limitation relates to subclinical seizures, whose inclusion could alter predictive performance, but were not considered here as seizure events. Clinical seizures often constitute only a small part of all abnormal electrical activity in the brain involved in epilepsy. But subclinical seizures have previously been shown to be predictable on par with clinical seizures and, for some patients, to precede clinical seizures, such that false positives for those seizures might be attributable to their preceding subclinical seizure activity [35,36]. In our comparative approach, however, all approaches had this same limitation, relative performance levels between them should thus not be affected too much.

In summary, we here attempted a comprehensive comparison to assess how much preictal information is contained in different data modalities (EEG, ECoG, EKG). Our results suggest ECoG contains the most preictal information (when all channels are used collectively), with EEG and EKG containing comparable amounts behind ECoG. In the search for a seizure forecasting device that it not just predictive but amenable to easy, everyday use, however, our work suggests single sensor EKG (utilizing the whole signal) as promising to improve preictal state detection algorithms. Our work demonstrates that EKG contains significantly more information about the preictal state than is contained in heart rate and its related measures. The findings reported here may help to improve other EKG-based seizure forecasting and detection methods, such a responsive VNS. More broadly, we expect these insights to help to devise future prospective long-term epilepsy monitoring trials and to substantiate the understanding of seizure pathophysiology.

Funding sources

CM acknowledges support by a NARSAD Young Investigator Grant by the Brain & Behavior Research Foundation and KB from a US Student Fulbright grant. These funding sources played no role in the collection, analysis or interpretation of data, the writing of the manuscript or in the decision publish the work.

Author contributions

CM, KB analyzed the data, interpreted results, wrote the paper, and produced figures and tables.

Declaration of Competing Interest

Dr. Meisel is part of patent applications to detect and predict clinical outcomes, and to manage, diagnose, and treat neurological conditions. The authors declare no other competing interests

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.07.001.

References

[1] 2016 Community survey [report on the Internet]. Landover, MD: Epilepsy Foundation; 2016 [cited 2019 June 5]. Available from: https://www.epilepsy.com/sites/core/files/atoms/files/community-survey-report-2016%20V2.pdf.
[2] Engel J. Approaches to refractory epilepsy. Ann Indian Acad Neurol 2014;17:S12–7.
[3] Dumanis SB, French JA, Bernard C, Worrell GA, Fureman BE. Seizure forecasting from idea to reality: outcomes of the my seizure gauge epilepsy innovation institute workshop. eNeuro 2017;4. https://doi.org/10.1523/ENEURO.0349-17.2017.
[4] Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. Brain 2007;130:314–33.
[5] Freestone DR, Karoly PJ, Cook MJ. A forward-looking review of seizure prediction. Curr Opin Neurol 2017;30:167–73.
[6] Kaggle [Internet]. San Francisco, CA: Kaggle Inc.; Melbourne University AES/MathWorks/NIH seizure prediction; 2016 [cited 2019 June 6]. Available from https://www.kaggle.com/c/melbourne-university-seizure-prediction.
[7] Mormann F, Andrzejak RG. Seizure prediction: making mileage on the long and winding road. Brain 2016;139:1625–7.
[8] Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure prediction-ready for a new era. Nat Rev Neurol 2018(10):618–30.
[9] Cook MJ, O'Brien TJ, Berkovic SF, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. Lancet Neurol 2013;12:563–71.
[10] Kerem DH, Geva AB. Forecasting epilepsy from the heart rate signal. Med Biol Eng Comput 2005;43:230–9.
[11] Valderrama M, Alvarado C, Nikolopoulos S, et al. Identifying an increased risk of epileptic seizures using a multi feature EEG–ECG classification. Biomed Signal Proc Control 2012;7:237–44.

[12] Fujiwara K, Miyajima M, Yamakawa T, et al. Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features. IEEE Trans Biomed Eng 2016;63:1321–32.
[13] Pavei J, Heinzen RG, Novakova B, et al. Early seizure detection based on cardiac autonomic regulation dynamics. Front Physiol 2017;8. https://doi.org/10.3389/fphys.2017.00765.
[14] Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw 2015;61:85–117.
[15] Ihle M, Feldwisch-Drentrup H, Teixeira CA, Witon A, Schelter B, Timmer J, et al. EPILEPSIAE - a European epilepsy database. Comput Methods Programs Biomed 2012;106:127–38.
[16] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35:1798–828.
[17] Kiral-Kornek I, Roy S, Nurse E, et al. Epileptic seizure prediction using big data and deep learning: toward a mobile system. EBio Medicine 2018;27:103–11.
[18] Truong ND, Nguyen AD, Kuhlmann L, et al. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. Neural Netw 2018;105:104–11.
[19] Eberlein M, Hildebrand R, Tetzlaff R, et al. Convolutional neural networks for epileptic seizure prediction. 2018 IEEE international conference on bioinformatics and biomedicine [conference proceedings on the Internet]. Madrid, Spain. New York: IEEE; 2018. p. 2577–82 cited 2019 June 6. Dec 3—6. (Available from: IEEE Xplore).
[20] Camm J, Malik M, Bigger T, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Eur Heart J 1996;17:354–81.
[21] Sassi R, Cerutti S, Lombardi F, et al. Advances in heart rate variability signal analysis: joint position statement by the e-cardiology ESC working group and the European heart rhythm association co-endorsed by the Asia Pacific Heart Rhythm Society. Europace 2015;17:1341–53.
[22] Ponnusamy A, Marques JL, Reuber M. Heart rate variability measures as biomarkers in patients with psychogenic nonepileptic seizures: potential and limitations. Epilepsy Behav 2011;22:685–91.
[23] Ponnusamy A, Marques JL, Reuber M. Comparison of heart rate variability parameters during complex partial seizures and psychogenic nonepileptic seizures. Epilepsia 2012;53. https://doi.org/10.1111/j.1528-1167.2012.03518.x.
[24] Meisel C. Linking cortical network synchrony and excitability. Commun Integr Biol 2016;9. https://doi.org/10.1080/19420889.2015.1128598.
[25] Montavon G, Samek W, Muller K. Methods for interpreting and understanding deep neural networks. Digit Signal Proc 2018;73:1–15.
[26] Bach S, Binder A, Montavon G, Klauschen F, Muller KR, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 2015;10. https://doi.org/10.1371/journal.pone.0130140.
[27] Maimon O, Rokach L. Data mining and knowledge discovery handbook. New York: Springer; 2009.
[28] Meisel C, Schulze-Bonhage A, Freestone D, Cook MJ, Achermann P, Plenz D. Intrinsic excitability measures track antiepileptic drug action and uncover increasing/decreasing excitability over the wake/sleep cycle. Proc Natl Acad Sci 2015;112:14694–9.
[29] Nei M. Cardiac effects of seizures. Epilepsy Curr 2009;9:91–5.
[30] Murthy K, Haywood J, Richardson J, et al. Analysis of power spectral densities electrocardiograms. Math Biosci 1971;12:41–51.
[31] Badawy R, Macdonell R, Jackson G, Berkovic S. The peri-ictal state: cortical excitability changes within 24 h of a seizure. Brain 2009;132:1013–21.
[32] Kimiskidis VK, Koutlis C, Tsimpiris A, Kalviainen R, Ryvlin P, Kugiumtzis D. Transcranial magnetic stimulation combined with EEG reveals covert states of elevated excitability in the human epileptic brain. Int J Neural Syst 2015;25. https://doi.org/10.1142/S0129065715500185.
[33] Meisel C, Kuehn C. Scaling effects and spatio-temporal multilevel dynamics in epileptic seizures. PLoS One 2012;7. https://doi.org/10.1371/journal.pone.0030371.
[34] Witten I, Frank E, Hall M, Pal C. Data mining: practical machine learning tools and techniques. Cambridge, MA: Morgan Kaufman; 2017.
[35] Feldwisch-Drentrup H, Ihle M, Quyen MV, et al. Anticipating the unobserved: prediction of subclinical seizures. Epilepsy Behav 2011;22:S119–26.
[36] Karoly PJ, Freestone DR, Boston R, et al. Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. Brain 2016;139:1066–78.