Deep Learning provides exceptional accuracy to ECoG-based Functional Language Mapping for epilepsy surgery

Harish RaviPrakash[a], Milena Korostenskaja[b,c,d], Eduardo M. Castillo[c,d], Ki H. Lee[d], Christine M. Salinas[d], James Baumgartner[d], Concetto Spampinato[e,a], Ulas Bagci[a,*]

[a]*Center for Research in Computer Vision, University of Central Florida, FL, USA*
[b]*Functional Brain Mapping and Brain Computer Interface Lab, Florida Hospital, FL, USA*
[c]*MEG Lab, Florida Hospital, FL, USA*
[d]*Florida Epilepsy Center, Florida Hospital, FL, USA*
[e]*Department of Electrical, Electronics and Computer Engineering, University of Catania, Catania, Italy*

Abstract

Successful surgical resection in epilepsy patients depends on preserving functionally critical brain regions while removing pathological tissues. Being the gold standard, Electro-cortical Stimulation Mapping (ESM) helps surgeons localize the function of eloquent cortex through electrical stimulation of the electrodes placed directly on the cortical brain surface. Due to the potential hazards of ESM, including increased risk of provoked seizures, Electrocorticography based Functional Mapping (ECOG-FM) was introduced as a safe alternative approach. However, ECoG-FM has a low success rate in localization of eloquent language cortex compared to the ESM. In this study, we address this limitation by developing a new deep learning algorithm for ECoG-FM with an accuracy comparable to the ESM when identifying eloquent language cortex. In our experiments with 11 epilepsy patients who underwent presurgical evaluation (through deep learning-based signal analysis on 637 electrodes), the proposed algorithm made an exceptional 34% improvement over the conventional ECoG-FM analysis, reaching the state-of-the-art accuracy of ~89%. Our findings have demonstrated, for the first time, that deep learning powered ECoG-FM can serve as a stand-alone modality and avoid potential hazards of the ESM in epilepsy surgery.

Keywords: Deep Learning, Electro-cortical Stimulation Mapping (ESM), Electrocorticography (ECoG), Real Time Functional Mapping (RTFM)

1. Introduction

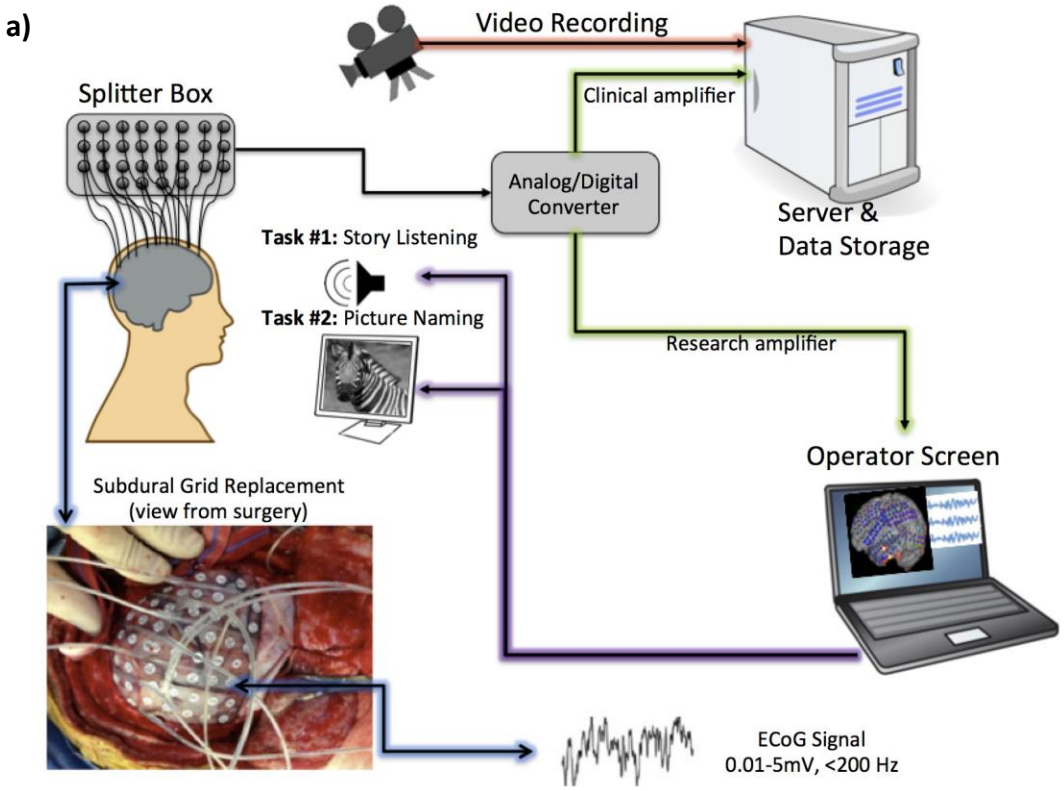
Epilepsy is a chronic neurological disorder characterized by recurrent, unpredictable seizures[1]. There are over 65 million reported cases of epilepsy around the world[1]. Approximately 20% of these patients are diagnosed with *drug-resistant epilepsy* and the only possible treatment in a majority of these cases is surgical intervention. During epilepsy surgery, a pathological brain tissue, associated with seizures, might be surgically removed. While epilepsy surgery is a curative option for drug-resistant epilepsy, neurosurgeons need to avoid removing tissues associated with language, sensory, and motor functions. This calls for accurate identification and localization of these functionally significant brain regions. The more precise their localization, the more accurate the surgical procedure can be performed to prevent corresponding post-surgical neurological/functional deficits and ensure improved patients' quality of life post-surgery.

Electro-cortical stimulation mapping (ESM) has been considered the gold standard for functional cortex localization in epilepsy surgery. ESM is an invasive procedure that uses electrodes placed on the surface of the brain (grid electrodes) or within the brain (depth electrodes). It is considered vital for reducing the risk of language deficits post-surgery and, therefore, expanding surgical options. ESM has a long history of serving as the main modality for functional mapping in surgical patients. Acute electrical cortical stimulation was successfully performed in 1950 during epilepsy surgery by Penfield and colleagues[2,3]. During ESM, pairs of electrodes covering the region of interest (in our case - eloquent cortex) are stimulated by delivering a brief electric pulse. The stimulation temporarily disables/inhibits the cortical area of interest (creates a temporary functional lesion). Behavioral changes such as unusual sensation, involuntary movements or language impairments (i.e. speech paucity), observed during stimulation, indicate that the tested area is essential to that task performance and its resection might lead to functional deficits. Ojemann et al.[4] studied language localization, using ESM, on a large dataset of 117 patients. The study found that there was sufficiently large individual variability in the exact location of language function. The study concluded that there was a need for an improved language localization model. Much later, more standard and effective tasks for language localization, such as verb generation[5] and picture naming[6] , were tested with the increased use of ESM.

One of the major drawbacks of ESM, however, is its potential to induce after-discharges[7], which could result in seizures. Since stimulation provoked seizures can occur rather frequently during ESM procedures[8], ESM tests often need to be repeated, leading to extended time and effort from medical professionals (neuropsychologists and/or neurologists). In some cases, ESM cannot be completed due to repeated seizure activity and/or its consequences.

Due to limitations of ESM (e.g., its association with the increased risk of seizures), there is a need for establishing other independent functional mapping modalities to identify eloquent cortex. Till now, none of the existing neuro-imaging modalities are

flexible enough to provide functional mapping results in real time in the operating room. Therefore, the search for a stand-alone methodology for functional eloquent cortex localization has been continuing and resulted in attempts to use electrocorticography (ECoG). ECoG is the invasive version of electroencephalography (EEG) and is sometimes also referred to as intracranial encephalography (iEEG). Importantly, ECoG equipment is portable and can be utilized both at the patients' bed-side and intra-operatively. ECoG demonstrates excellent temporal resolution. It also overcomes the poor spatial resolution problem, as it records the activity of interest directly from the cortical brain surface and avoids the problem of electrical signal change (caused by electric signal's propagation throughout the tissues surrounding the brain). To record ECoG signals, a craniotomy (removal of the skull section: *bone flap*) is performed and the dura is opened to access the brain. The arrays of grid of electrodes (**Figure 1a, left**) are then placed on the exposed cerebral cortex. Following this, ECoG-based functional mapping (ECoG-FM) is performed, when task-based responses from grid electrodes are recorded. Since there is no external electrical stimulation during this process, ECoG-FM is a safer alternative to ESM. When performed in real time[9], ECoG-FM procedure can be referred to as real-time functional mapping (RTFM)[10–13]. **Figure 1a** demonstrates the general setup for ECoG-FM recordings.



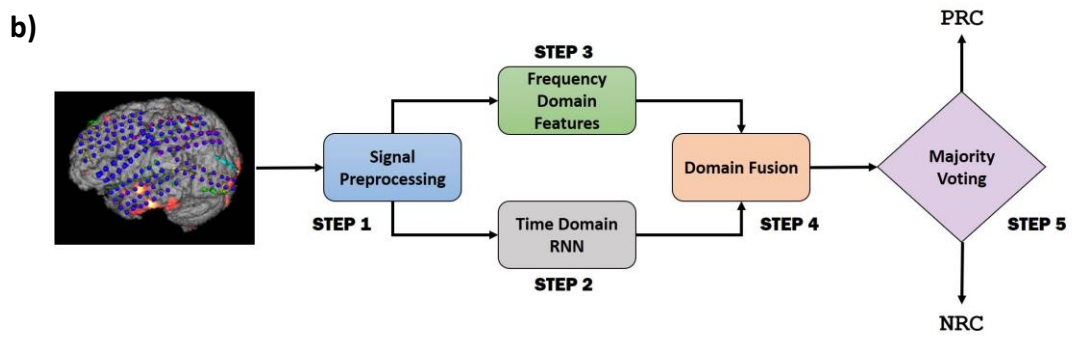


Figure 1: Overview of a) the language localization framework with ECoG-based functional mapping (ECoG-FM) approach. ECoG signal recording, data transfer, storage, research and clinical paths, and tasks are illustrated. ECoG signals are obtained in response to task-related changes (e.g., picture naming) from grid electrodes implanted on the cortical surface in the subdural space, b) the proposed ECoG signal classification approach for each channel on the cortical surface in the subdural space. PRC - Positive Response Channel & NRC - Negative Response Channel. Each individual step is considered as a "module" in the overall system design.

ECoG-based approaches have been used successfully for motor cortex localization[10,11,14–16]. When compared to motor cortex localization, the functional language cortex appears far more complex and challenging[17]. Current localization approaches are based on detecting positive response channels (active channels). A baseline recording of each channel (electrode) at resting-state is used to determine signal characteristics at specific frequency ranges, and most often, the power of the ECoG signal is focused on in the α, β and primarily, the high-γ (70Hz-170Hz) frequency bands[9,18]. These values are compared with the signal power measured during the language task execution. The results of this approach for language mapping do not achieve the desirable accuracy. For example, Arya et al.[19] studied the high-γ response during spontaneous conversation of 7 patients. The results showed low specificity and accuracy. In a follow up study, Arya et al.[20] demonstrated high-gamma modulation for the story-listening task and achieved a high specificity but sensitivity remained low. Korostenskaja et al.[21] showed that, similar to the results in motor cortex, the ECoG-FM can be used for eloquent language cortex localization complimentary to ES, but not as a stand-alone modality. Finally, it has been demonstrated that ECoG-FM can be used as a guiding tool for ESM, reducing the time of ESM procedure and, therefore, decreasing the risk of provoked seizures[13,22]. Despite its promise, current ECoG-FM approach is not capable of being used as a stand-alone methodology for accurate language mapping. To address these challenges and provide ECoG-FM more independence in eloquent language cortex localization, we fill in the following currently existing methodological gaps:

i Available approaches compare a channel's signal with its resting-state (baseline) recording. They do not compare the channels' characteristics to other recorded channels.

ii Further, the signal in the frequency range beyond high-*γ* has not been explored yet.

iii There has been limited work on validation of ECoG-FM for language cortex localization.

In our previous work[23,24], we showed the feasibility of machine learning methods to perform classification of channel response by using the whole signal spectrum (not limited to *α, β* & *high γ*) and without using the baseline recording. This was the first machine learning based approach in this field and demonstrated the potential of ECoG-FM signal to be analyzed more accurately compared to the conventional methods.

Summary of our contributions:

We propose an innovative deep learning algorithm for accurately classifying the channel response of eloquent cortex based on a deep learning family of algorithms, alleviating the current challenges of ECoG-FM. Overview of the proposed system is illustrated in **Figure 1b**.

1. We pre-process the ECoG signal to eliminate undefined signal components and make the ECoG signals to be of equal length sequences. We then divide the ECoG signals into sub-blocks of data, aiming to learn more *discriminative signal patterns* and eventually reduce the computational load (Step 1).
2. We learn different sets of signal features independently: frequency domain (i.e. auto-regression) and time domain (deep learning-based features) in Step 2 and Step 3, respectively.
3. After we combine the learned features, we train a recurrent neural network (RNN), a class of deep learning algorithm tailored to analyze sequential data, to classify the sub-blocks of signals (Step 4).
4. Finally, we use a statistical majority voting algorithm to combine the sub-block labels and determine the channel label as *Positive Response Channel (PRC)* or *Negative Response Channel (NRC)* (Step 5).

In the following section, we describe each module of our proposed system in detail.

2. Materials and Methods

2.1. Dataset

The study was approved by the Institutional Review Board at Florida Hospital, Orlando, USA. We recruited eleven patients with drug-resistant epilepsy, who underwent pre-surgical evaluation with ECoG grid implantation. All patients provided their written informed consent/assent to participate in the study. The patients were teenagers and adults with an average age of 23.18 ± 11.61 yrs (See Supplementary **Table 1** for summary of the patient demographics).

2.2. Experimental Setup and Task Paradigm

The basic setup for ECoG-FM is shown in **Figure 1a**. ECoG signals from the implanted subdural grids are split into two streams: one for continuous clinical seizure monitoring and the other for ECoG-FM. The tool used to record the incoming ECoG signal was BCI2000[9]. A baseline recording of the cortical activity was first acquired to capture the "resting-state" neuronal activity. Following the *baseline* recording step, paradigms similar to those employed in ESM or fMRI were used to record the task-related ECoG signal for functional mapping[25]. **Figure 2** shows one such paradigm, mimicking details of the experimental setup for the language comprehension task that we used. Alternate 30 second blocks of ECoG data during "control" and "active" conditions were recorded continuously at a fixed sampling rate of 1200 Hz.

For the language comprehension task, the *active condition* implies listening to a story, while the control task involves listening to broadband noise[26]. The active condition (listening to a story) was different for each block in order to keep the patient attentive and responsive. The system recorded information from 128 ECoG channels (128 electrodes (channels) in **Figure 2**) by using *g.USBamp biosignal amplifiers* (g.tec Medical Engineering GmbH, Austria) Subdural ground and reference electrodes were used.

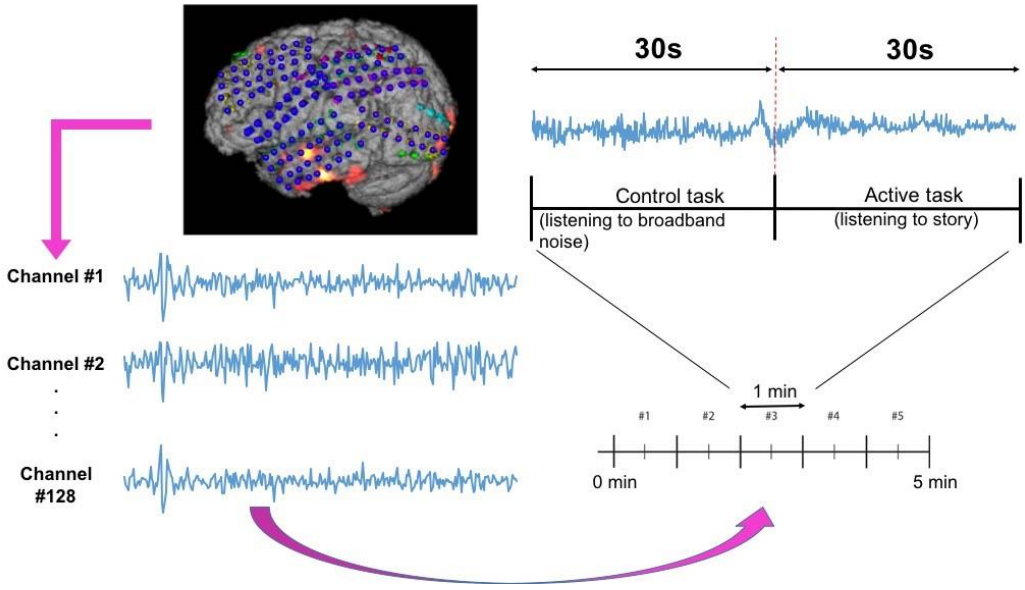


Figure 2: Subdural grid localization and position of ECoG electrodes (128 of them) on the brain surface of individual patient are illustrated (left). For a sample of 1 min language test, signals from both control and active tasks are illustrated (right).

2.3. Step 1: Pre-processing ECoG Signal

As a first step of preparing the recorded data for deep learning-based analysis, non-task/control time points in the signal were eliminated: these correspond to the

spontaneous activity recording before the 0-min illustrated in **Figure 2** and any trailing signals at the end of the experiment (See Supplementary for more details about this step).

2.4. Step 2: Time Domain RNN

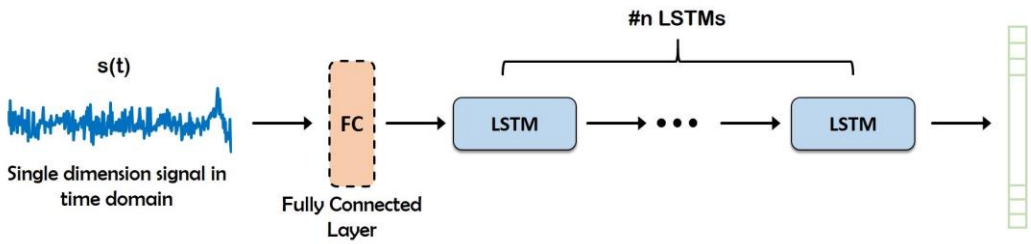


Figure 3: Deep network structure for extraction of time domain features from the input signal. Note that these features are combined with frequency domain features (Auto-Regressive) later for final prediction of the channels.

Our goal was to find discriminative signal patterns from ECoG signals, which are time varying and non-stationary 1D sequences. They are non-stationary because task based ECoG recordings can have the signal statistics depend on the time relative to the events. Inspired by the effectiveness of Recurrent Neural Networks (RNN) in sequence classification tasks in different domains[27–29], we developed an RNN based deep neural network algorithm to extract discriminative features from time-domain ECoG data. We hypothesize that limitations of the conventional spectral (frequency-based) or time-based signal analysis methods can be overcome with RNN based methods because they keep higher contextual information by integrating a loop allowing information to flow from one step to the next.

This time domain feature learning module is illustrated in **Figure 3**. The complete ECoG signal contained both control and active task signals, thus; the input to this time domain RNN module was a mixture of sub-blocks of control and active task signals. To learn signal characteristics from such a noisy data, we first used a fully connected layer (*dense layer* in other words denoted as FC in **Figure 3**). We designed the next few layers as Long-Short Term Memory (LSTM) blocks. LSTM, introduced in 1997[30], is a type of RNN that has the ability to learn long-term dependencies of data. In the literature, LSTM and its variants have primarily been used in 1D sequence classification tasks[31,32] and prediction tasks [33,34]. In our experiments, we used multiple LSTM layers in different exploratory configurations to learn a more complex feature representation of the input signal.

2.5. Step 3: Frequency Domain Features

One of the objectives in our experiments was, to explore not only time-domain, but also frequency and hybrid (i.e., a combined time and frequency domain) domain-based analysis of signal characteristics, and do a thorough comparison, which was never

done before. In Step 2, we explained how to extract deep features from the time-domain, and in this step (Step 3), we focused on the spectral estimation of the signals. Conventional ECoG signal classification approaches are based on frequency-domain, where spectral analysis of the signal is performed to identify the channel response. Traditionally, spectral estimation of the signals is performed by fitting a parametric time domain model to the ECoG signals. One of the most commonly employed models/approaches in this category is the autoregressive (AR) model. An AR model for a discrete signal $x[n]$ can be represented as

$$x[n] = -\sum_{k=1}^{k=p} a_p[k]x[n-k] + w[n] \qquad (1)$$

where $a_p[k]$ are the AR coefficients, p is the order of the AR model, $w[n]$ is a zero mean white noise process with a variance ρ. AR parameters, once the model in Eq. (1)~~(1)~~ is solved, are used for characterization of the signal.

Methods to solve for the AR parameters are diverse, we used the reflection coefficient estimation-based methods (see Supplementary for details of the parameter selection procedure). Estimated AR parameters are used to characterize the ECoG signals from frequency-domain perspective.

2.6. Step 4: Domain Fusion (Hybrid Domain)

Using LSTMs in Step 2, we learned a different set of features (i.e., time domain) than the AR features that were generated in Step 3. In domain fusion step (Step 4), these two (largely) complimentary features were combined to obtain a hybrid signal representation model with a new deep network setting, Domain Fusion Network (DFN) (See **Figure 4**). Although the merging of the two feature vectors can be done in multiple ways, we used a concatenation approach to get the full benefit of each domain (time vs. frequency).

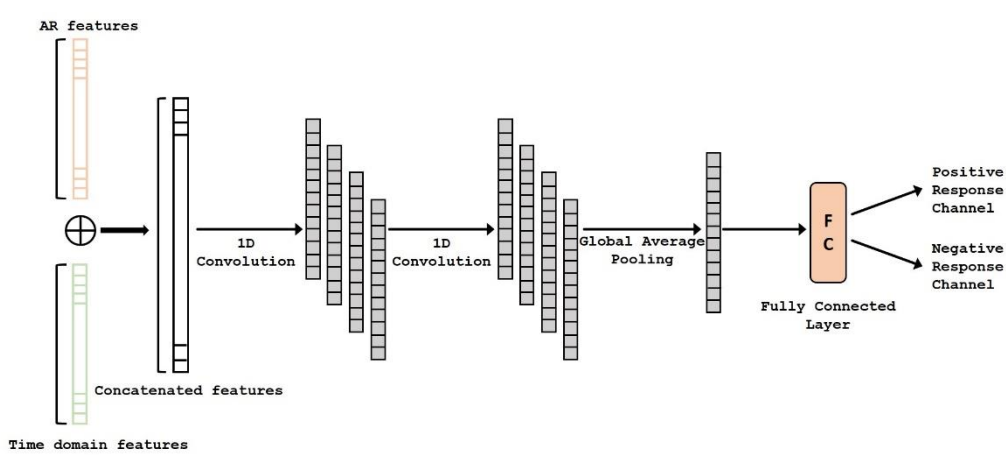


Figure 4: Deep network structure of the fusion module. AR (i.e., frequency) features (orange) and time-domain features (green) are concatenated and classified.

In concatenation, we assumed independence of features; hence, we did not use element-wise multiplication or other approaches for data merging. Since convolution helps identify local patterns and reduce redundant information in the data, the complete feature vector (after concatenation) was then passed through multiple layers of 1D convolution with an activation function, to weight each feature based on its contribution to the classification problem (positive response channels vs negative response channels). Following the 1D convolution layers, the output feature maps were spatially averaged using Global Average Pooling[35], making the DFN more robust to spatial translations of the input and introducing structural regularization to the feature maps. Finally, we inserted a single fully connected layer into the DFN and used a sigmoid activation to perform the final classification.

2.7. Step 5: Majority Voting

The output of the domain fusion model was a label for the input signal, which was a sub-block. Signal from each channel/electrode was made up of hundreds of sub-blocks of the signals. Therefore, a statistical majority voting of the sub-block labels pertaining to a channel is performed to generate the final channel response classification as a Positive Response Channel (PRC) or Negative Response Channel (NRC). For instance, if a channel included 354 sub-blocks and more than 177 sub-blocks indicated a positive response, that channel was labelled as a positive response channel.

3. Experiments and Results

Our overall goal was to successfully (and automatically) identify positive response channels (PRC) and negative response channels (NRC) in ECoG-FM data using new machine learning models. The ground truth was inferred from the gold standard ESM results. Owing to the large imbalance in the number of PRCs and NRCs, NRCs outnumbering PRCs 3:1, we randomly selected equal number of NRCs to balance the data.

Each channel's signal comprised of blocks of active task data and control data and the discriminative power of both these blocks was unknown. To clearly understand the best way to use these different data inputs, we divided our experimental evaluation approaches into three main categories: Singular group (E_1), Separate (E_2) and Non-Separate (E_3) data classification approaches.

(E_1) Train the deep model using only one of active-task/control data.

(E_2) Train separate models for the active task data and control data and merge these models to learn more discriminative features.

(E_3) Train a single deep model using both the active task and the control data.

This structured experimental procedure helped us determine how useful each component of the signal was and provided insights into the response of brain regions (through channel responses) to different signals.

3.1. E_1 - Singular group data classification approach

In this approach, a distinction is made with regards to the response of channels to control signal and active task signal, i.e., the signal characteristics in the active task signal is assumed to be different from that in the control data and that processing only the active task data or control data is sufficiently discriminative. Thus, we learned individual models to test the discriminative power of these signals (i.e., classification was done using only active task data-based model or control data-based model).

Active task data model

In task-based experiments, a response is generally expected in the active task period and not in the control or rest period. Following this, we built our network by first testing the effect of using only the time domain features - *AT (ActiveTime)*, then by adding the AR features - $AT{-}AR^1$ *(ActiveTimeAR1)* and adding a fully connected layer as the first layer of the time domain RNN and Frequency Domain modules - $AT{-}AR^2$ *(ActiveTimeAR2)*. By convention, we used early stopping criteria to avoid overfitting. Supplementary Fig. 1 shows the training plots that helped determine the stopping criterion.

Control data model

To compare the information present in the active task data and the control data, we replicated the $AT{-}AR^2$ model and fed it control data - $CT{-}AR^2$ *(ControlTimeAR2)*. As can be seen from **Table 1**, the specificity of the control data model was lower than the active data model indicating that it had lower discriminative power.

3.2. E_2 - Separate data classification approach

This approach was based on the previous one in which active task and control data channel response is different. However, in the current E_2 approach, the channel's response to control signal was assumed to be different to the task signal. To test whether the two signals provide redundant or complimentary information towards classification, separate models were trained and then merged in an attempt to learn more discriminative features. Moreover, different approaches to domain fusion were tested with the use of fully connected layers in the place of 1D convolution layers. The layers used to classify upon merging the two separate models were also varied by interchanging 1D convolution and Global Average Pooling with stacked fully connected layers. These different *AC (ActiveControl)* models' performance is shown in **Table 1**.

Table 1: Channel classification accuracy for the different approaches.

Model	Sub-block Accuracy %	Sensitivity %	Specificity %	Accuracy %
Singular Group Data Approach (E_1)				
AT	59.26	100	60.09	70.99
***AT–AR*^{1}**	**72.20**	100	77.01	86.37
***AT–AR*^{2}**	66.53	100	**89.51**	**92.15**
***CT–AR*^{2}**	68.36	100	78.52	86.01
Separate Group Data Approach (E_2)				
***AC*^{1}**	68.00	100	58.33	79.10
***AC*^{2}**	73.11	100	**90.92**	**94.98**
***AC*^{3}**	**75.07**	100	68.72	82.62
***AC*^{4}**	74.80	100	68.76	81.96
***AC*^{5}**	50.00	100	49.99	66.67
Non-Separate Group Data Approach (E_3)				
***TimeAR*^{1}**	61.40	54.16	70.83	62.49
***TimeAR*^{2}**	**61.92**	100	45.86	69.78
***TimeAR*^{3}**	58.66	100	74.21	83.27
***TimeAR*^{4}**	60.84	100	**77.54**	**85.74**

3.3. E_3 - Non-Separate data classification approach

In this approach, it was assumed that channels that respond positively to a stimulus (story) also respond differently to the negative response channels during the control signal period of the experiment. That is, unlike E_2, where we separated blocks of the signal (active task signal and control data), in the E_3 approach, we did not discriminate between blocks of the channel's signal. We prepared the data as such and fed both active task and control data blocks to the model with the same label for a particular channel.

The models were sequence classification models and thus the first task was to determine the depth of stacked LSTM models. Towards this, we tested models with increasing depth.

Figure 5a shows the structure of *TimeAR*^{1} and *TimeAR*^{2} models with the difference being the depth of the fully connected layers. **Figure 5b** shows the network structure of *TimeAR*^{3} and *TimeAR*^{4} where fully connected layers are used as the first stage of the Step 2 and Step 3 modules. Further, they differ from the first two models with the Domain Fusion layers consisting of stacked 1D Convolution layers instead of stacked fully connected layers. The performance of these four models is shown in **Table 1**.

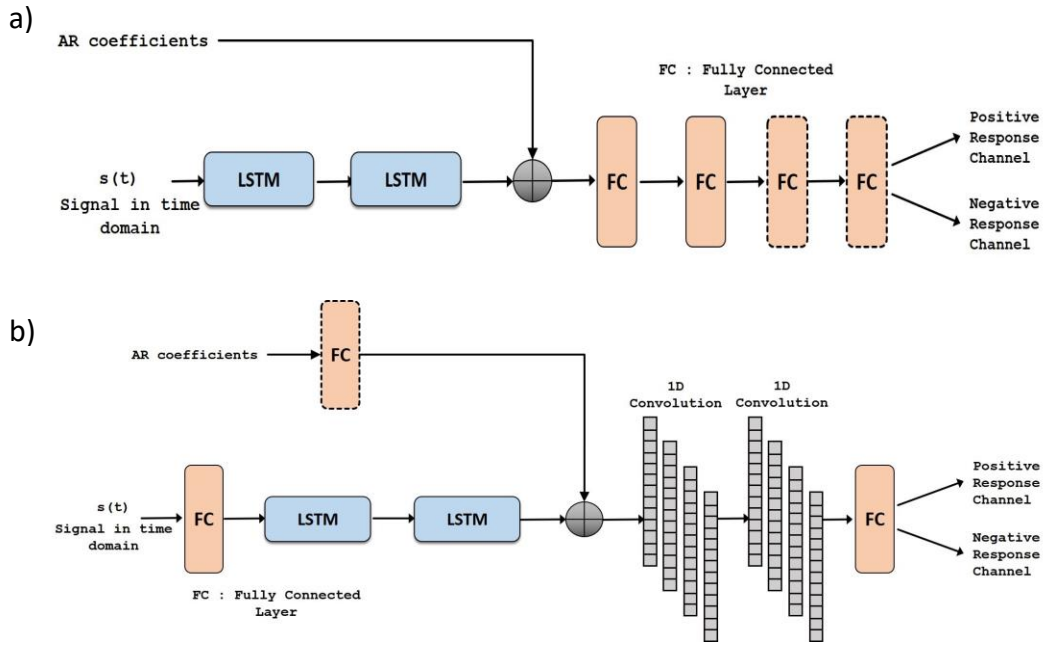


Figure 5: Deep network structure of a) *TimeAR*[1] and *TimeAR*[2] (with added FC in dotted boxes) and b) *TimeAR*[3] and *TimeAR*[4] (with added FC in dotted box).

3.4. Model Validation

In supervised learning, where a model is trained with ground-truth labels, the goal is to maximize predictive accuracy. However, therein lies the risk of memorizing the data rather than learning the optimal features. This problem of memorizing the data or learning the structure of the data to be the noise in the data is often referred to as overfitting[36]. It is important for a classification model to be able to generalize to unseen data and avoid the problem of overfitting. The method of testing how the analysis/model will generalize to an independent test dataset is known as cross-validation. When a completely independent dataset is not available, as is generally the case, the available data is split into training data and validation/test data. There are different types of cross-validation approaches like leave-one-out cross-validation, hold-out method, *k*-fold cross-validation, to name a few[37]. Previously, due to the time-consuming nature of the deep learning model training, we applied the hold-out method to validate the model. In this method, the model is trained on a part of the available data while the remaining data is held for testing/validating the model. Therefore, to test the generalization and robustness of the models, we further validate them using the *k*-fold cross-validation approach. In the *k*-fold cross-validation method, the data is split into *k* folds, and the model is then trained on $k - 1$ folds with the remaining fold used as validation data. This is repeated *k* times, each time using a different fold for validation. The results are averaged across the folds. This can be seen as performing the hold-out method *k* times with the data held for validation being

a different fold each time. We ensured that no data from an electrode (channel) is represented in both training and testing in order to provide a fair evaluation with a better generalization accuracy and to avoid overfitting. That is, if a sub-block of data was assigned to the training set, then all sub-blocks belonging to the same channel were assigned to the training set and likewise for the test set. For 5-fold cross validation, we repeated the experiments 5 times and used each of these distinct and non-overlapping training-testing sets to evaluate the model accuracies. Prediction accuracy was then calculated by averaging the results of these 5 experiments.

Following the structured experimental design, the top models from each category are chosen - $AT{-}AR^2$ from E_1 (owing to its higher specificity and accuracy), AC^2 from E_2 (as it is clearly superior to the other models in that group) and $TimeAR^3$ and $TimeAR^4$ from E_3 (since they had similar performance). Five-fold cross-validation (k = 5) is performed on these models where we repeat the experiments 5 times and used each of the distinct and non-over lapping training-testing sets to evaluate the model accuracies. Prediction accuracy was then calculated by averaging the results of these 5 experiments and the results are shown in **Table 2**.

Table 2: Channel classification accuracy - 5-fold cross-validation.

Model	Sub-block Accuracy %	Sensitivity %	Specificity %	Accuracy %
***TimeAR*^{3}**	59.99	100	66.56	76.84
***TimeAR*^{4}**	**71.42**	**100**	**84.16**	**89.73**
***AC*^{2}**	70.21	100	78.95	85.76
***AT−AR*^{2}**	65.40	100	82.01	86.57

The difference in results between the $AT{-}AR^2$ and AC^2 models hinted that the information learned from using control signal data might not provide additional information and could be redundant. This observation was supported by the fact that the experimental paradigm was defined such that the response of all channels to the control signal was different from the response to the active signals and somewhat similar to each other. The best model was $TimeAR^{\mathbf{4}}$, the model using non-separation of data approach, which achieved the highest accuracy as well as specificity.

All models were trained using linear activations in the 1D convolution layers. The reasoning behind usage of linear activations was that we did not want to make more sparse, the low dimensional feature vector. Not only did we test the linear activations but also different non-linear activations - *ReLU* and *Tanh*, as discussed in Supplementary section. Supplementary Fig. 2 compares the performance of $TimeAR^{\mathbf{4}}$ under the different activations.

For the overall computational setup, see Supplementary. Supplementary Table 2 details the network parameters.

4. Discussions and Concluding Remarks

In this paper, we proposed novel deep learning architectures to classify the channel response of ECoG signals. The results showed state-of-the-art classification accuracy of **89.7**% with high specificity and sensitivity of **84.1**% & **100**%, respectively, in determining whether the channel is positive (has a response) or negative (has no response) in relation to the task stimulus.

Traditionally, the accuracy of ECoG-FM is high for sensory and motor, but relatively low for language modality. On an average, ECoG-FLM has a lower sensitivity (62%) and higher specificity (75%) to detect language-specific regions [for a comprehensive review, see Korostenskaja et al.[26]], which is the opposite of what is observed for hand motor (100% sensitivity and 79.7% specificity) and hand sensory (100% sensitivity and 73.87% specificity) ECoG-based mapping[38]. The results of our current study demonstrate that the accuracy of language ECoG-FM can now be comparable to both sensory and motor ECoG-FM accuracies. Indeed, our achieved language ECoG-FM accuracy values are the highest among reported so far[39]. Although a number of studies have demonstrated a successful utilization of the ECoG-FM as a complimentary tool for ESM[13,22], there was not enough evidence to support the use of ECoG-FM as a stand-alone methodology for functional language mapping due to its relatively low accuracy compared with ESM[26,38]. The outcome of our research indicated the potential of ECoG-FM, to be considered as a stand-alone modality for eloquent language cortex localization.

It is possible that some features of the ECoG signal, reflecting a complex nature of language processing, are omitted from consideration when restricting the language ECoG-FM analysis to the gamma frequency band only. Expanding analysis to the whole spectrum of frequencies in our study, therefore, has exceeded the results gained from prior analysis approaches. It contributed to the improved classification accuracy and confirmed the results of previous studies, pointing towards the complex nature of language processing that needs to be considered during the analysis of neurophysiological data. The results of our current study have a wide applicability. Not only our proposed approach can be utilized to prevent functional morbidity post-surgery in patients with pharmacoresistant epilepsy, but can also increase the accuracy of the eloquent cortex mapping in patients undergoing resections of brain tumours[40] and arteriovenous malformations[41]. Moreover, deep learning-based ECoG-FM approaches can be successfully applied in various fields of adaptive neurotechnologies (e.g., neuromodulation), where ECoG-based mapping is performed to determine the best area for responsive stimulation. For example, for defining the neural correlates of tics to be used for responsive stimulation in patients with Tourette's syndrome[42,43] and for bidirectional neurostimulation via fully implantable neural interfaces in Parkinson's disease[44]. The applicability of our proposed approach extends as well towards the fields of developmental disorders (e.g., autism)[45,46], psychiatry (e.g., major depression[47] and obsessive-compulsive disorder (OCD)[48], addiction[49], eating disorders and obesity[50,51], where neurostimulation can be potentially utilized to provide treatment and improve patients' quality of life.

Limitations and future perspectives

Despite the state-of-the-art results in ECoG based predictions, there are also some limitations of our work to be noted. First, our experimental paradigm (**Figure 2**) involves five different stories being played to the subject. The responses to these stories, have some inherent similarities, but overall are different. Therefore, training the deep learning model with a single label for the whole channel could add noise to the model. This could be the reason for the lower sub-block accuracy, i.e., the classification of the 500ms signal. We believe that these results can be improved by including additional data and then training the system individually for each story in the paradigm.

Second, the subjects used in this initial validation study were a mix of teenagers and adults. Smith[52] found that the effect of epilepsy and seizures on children and adults was different i.e. the rules learned about the behavior of the brain in adults is different for children. Hence, a more comprehensive study with focus on children/teenagers is needed. This is one of our targets to test the proposed machine learning based approach in different patient populations; however, patient recruitment is difficult due to the surgery nature of epilepsy, and disease prevalence.

Finally, in the proposed approach, due to the exploratory and research nature of our study, the classification was not performed in real-time. Having validated the innovations presented here, we are currently working on the real-time implementation of the algorithm as an extension of this study. We believe that implementing such a reliable technology will increase current presurgical and intra-operative functional mapping accuracy, expand surgical treatment opportunities, prevent post-surgical language morbidity and improve patient outcomes.

Acknowledgements

The authors acknowledge The Central Florida Health Research (CFHR) grant (PIs: Drs. M. Korostenskaja and U. Bagci) for supporting this study.

References

1. Patricia, S. About Epilepsy: The Basics. (2018).

2. Penfield, W. & Rasmussen, T. The cerebral cortex of man; a clinical study of localization of function. (1950).

3. Penfield, W. & Roberts, L. Mapping the speech area. *Speech brain Mech.* 103–118 (1959).

4. Ojemann, G., Ojemann, J., Lettich, E. & Berger, M. Cortical language localization in left, dominant hemisphere: an electrical stimulation mapping investigation in 117 patients. *J. Neurosurg.* **71,** 316–326 (1989).

5. Ojemann, J. G., Ojemann, G. A. & Lettich, E. Cortical stimulation mapping of language cortex by using a verb generation task: effects of learning and comparison to mapping based on object naming. *J. Neurosurg.* **97,** 33–38 (2002).

6. Edwards, E. *et al.* Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* **50,** 291–301 (2010).

7. Pouratian, N., Cannestra, A. F., Bookheimer, S. Y., Martin, N. A. & Toga, A. W. Variability of intraoperative electrocortical stimulation mapping parameters across and within individuals. (2004).

8. Bank, A. M., Schevon, C. A. & Hamberger, M. J. Characteristics and clinical impact of stimulation-evoked seizures during extraoperative cortical mapping. *Epilepsy Behav.* **34,** 6–8 (2014).

9. Schalk, G. *et al.* Real-time detection of event-related brain activity. *Neuroimage* **43,** 245–249 (2008).

10. Leuthardt, E. C. *et al.* Electrocorticographic frequency alteration mapping: a clinical technique for mapping the motor cortex. *Oper. Neurosurg.* **60,** ONS--260 (2007).

11. Miller, K. J. *et al.* Real-time functional brain mapping using electrocorticography. *Neuroimage* **37,** 504–507 (2007).

12. Kapeller, C. *et al.* CortiQ-based real-time functional mapping for epilepsy surgery. *J. Clin. Neurophysiol.* **32,** e12--e22 (2015).

13. Prueckl, R. *et al.* Passive functional mapping guides electrical cortical stimulation for efficient determination of eloquent cortex in epilepsy patients. in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 4163–4166 (2017).

14. Towle, V. L. *et al.* Identification of the sensory/motor area and pathologic regions using ECoG coherence. *Electroencephalogr. Clin. Neurophysiol.* **106,** 30–39 (1998).

15. Qian, T. *et al.* Fast presurgical functional mapping using task-related intracranial high gamma activity. *J. Neurosurg.* **119,** 26–36 (2013).

16. Vansteensel, M. J. *et al.* Task-free electrocorticography frequency mapping of the motor cortex. *Clin. Neurophysiol.* **124,** 1169–1174 (2013).

17. Alkawadri, R., Gaspard, N., Hirsch, L. J. & Spencer, D. D. T105. A novel method for ECOG-based localization of function. *Clin. Neurophysiol.* **129,** e43 (2018).

18. Prueckl, R. *et al.* cortiQ-Clinical software for electrocorticographic real-time functional mapping of the eloquent cortex. in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE* 6365–6368 (2013).

19. Arya, R. *et al.* Electrocorticographic language mapping in children by high-gamma synchronization during spontaneous conversation: Comparison with conventional electrical cortical stimulation. *Epilepsy Res.* **110,** 78–87 (2015).

20. Arya, R. *et al.* Electrocorticographic high-gamma modulation with passive listening paradigm for pediatric extraoperative language mapping. *Epilepsia* **59,** 792–801 (2018).

21. Korostenskaja, M. *et al.* Electrocorticography-Based Real-Time Functional Mapping for Pediatric Epilepsy Surgery. *J. Pediatr. Epilepsy* **4,** 184–206 (2015).

22. Prueckl, R. *et al.* O202 Combining the strengths of passive functional mapping and electrical cortical stimulation. *Clin. Neurophysiol.* **128,** e243 (2017).

23. RaviPrakash, H. *et al.* Automatic response assessment in regions of language cortex in epilepsy patients using ECoG-based functional mapping and machine learning. in *Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on* 519–524 (2017).

24. Korostenskaja, M. *et al.* in *Brain-Computer Interface Research: A State-of-the-Art Summary 6* (eds. Guger, C., Allison, B. & Ushiba, J.) (Springer International Publishing: Cham, 2017).

25. Korostenskaja, M. *et al.* Real-time functional mapping: potential tool for improving language outcome in pediatric epilepsy surgery: Case report. *J. Neurosurg. Pediatr.* **14,** 287–295 (2014).

26. Korostenskaja, M. *et al.* Real-time functional mapping with electrocorticography in pediatric epilepsy: comparison with fMRI and ESM findings. *Clin. EEG Neurosci.* **45,** 205–211 (2014).

27. Graves, A. & Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* **18,** 602–610 (2005).

28. Baccouche, M., Mamalet, F., Wolf, C., Garcia, C. & Baskurt, A. Spatio-Temporal Convolutional Sparse Auto-Encoder for Sequence Classification. in *BMVC* 1–12

(2012).

29. Wang, J. *et al.* Cnn-rnn: A unified framework for multi-label image classification. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 2285–2294 (2016).

30. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9,** 1735–1780 (1997).

31. Liu, Q., Zhou, F., Hang, R. & Yuan, X. Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification. *arXiv Prepr. arXiv1703.07910* (2017).

32. Nowak, J., Taspinar, A. & Scherer, R. LSTM Recurrent Neural Networks for Short Text and Sentiment Classification. in *International Conference on Artificial Intelligence and Soft Computing* 553–562 (2017).

33. Flunkert, V., Salinas, D. & Gasthaus, J. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *arXiv Prepr. arXiv1704.04110* (2017).

34. Gammulle, H., Denman, S., Sridharan, S. & Fookes, C. Two Stream LSTM: A Deep Fusion Framework for Human Action Recognition. in *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on* 177–186 (2017).

35. Lin, M., Chen, Q. & Yan, S. Network in network. *arXiv Prepr. arXiv1312.4400* (2013).

36. Dietterich, T. Overfitting and undercomputing in machine learning. *ACM Comput. Surv.* **27,** 326–327 (1995).

37. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An introduction to statistical learning*. **112,** (Springer, 2013).

38. Kapeller, C. *et al.* CortiQ-based Real-Time Functional Mapping for Epilepsy Surgery. *J. Clin. Neurophysiol.* **32,** e12--e22 (2015).

39. Arya, R., Horn, P. S. & Crone, N. E. ECoG high-gamma modulation versus electrical stimulation for presurgical language mapping. *Epilepsy Behav.* **79,** 26–33 (2018).

40. Ogawa, H. *et al.* Rapid and minimum invasive functional brain mapping by real-time visualization of high gamma activity during awake craniotomy. *World Neurosurg.* **82,** 912--e1 (2014).

41. Kamada, K. *et al.* Novel techniques of real-time blood flow and functional mapping. *Neurol. Med. Chir. (Tokyo).* **54,** 775–785 (2014).

42. Molina, R. *et al.* Report of a patient undergoing chronic responsive deep brain stimulation for Tourette syndrome: proof of concept. *J. Neurosurg.* **129,** 308–314 (2018).

43. Shute, J. B. *et al.* Thalamocortical network activity enables chronic tic detection in

humans with Tourette syndrome. *NeuroImage Clin.* **12,** 165–172 (2016).

44. Swann, N. C. *et al.* Chronic multisite brain recordings from a totally implantable bidirectional neural interface: experience in 5 patients with Parkinson's disease. *J. Neurosurg.* **128,** 605–616 (2018).

45. Sinha, S., McGovern, R. A. & Sheth, S. A. Deep brain stimulation for severe autism: from pathophysiology to procedure. *Neurosurg. Focus* **38,** E3 (2015).

46. Sturm, V. *et al.* DBS in the basolateral amygdala improves symptoms of autism and related self-injurious behavior: a case report and hypothesis on the pathogenesis of the disorder. *Front. Hum. Neurosci.* **6,** 341 (2013).

47. Coenen, V. A. *et al.* Tractography-assisted deep brain stimulation of the superolateral branch of the medial forebrain bundle (slMFB DBS) in major depression. *NeuroImage Clin.* **20,** 580–593 (2018).

48. Vázquez-Bourgon, J. *et al.* Deep brain stimulation and treatment-resistant obsessive-compulsive disorder: A systematic review. *Rev. Psiquiatr. Salud Ment.* (2017).

49. Wang, T. R., Moosa, S., Dallapiazza, R. F., Elias, W. J. & Lynch, W. J. Deep brain stimulation for the treatment of drug addiction. *Neurosurg. Focus* **45,** E11 (2018).

50. Val-Laillet, D. *et al.* Neuroimaging and neuromodulation approaches to study eating behavior and prevent and treat eating disorders and obesity. *NeuroImage Clin.* **8,** 1–31 (2015).

51. Oterdoom, D. L. M. *et al.* Therapeutic potential of deep brain stimulation of the nucleus accumbens in morbid obesity. *Neurosurg. Focus* **45,** E10 (2018).

52. Smith, M. Lou. Neuropsychology in epilepsy: children are not small adults. *Epilepsia* **51,** 68–69 (2010).