9th International IEEE EMBS Conference on Neural Engineering
San Francisco, CA, USA, March 20 - 23 , 2019

Transfer-learning for differentiating epileptic patients who respond to treatment based on chronic ambulatory ECoG data

Sharanya Arcot Desai, Thomas Tcheng, and, Martha Morrell

***Abstract*— the aim of this study was to evaluate whether transfer-learning with pre-trained deep convolutional neural networks (deep CNNs) can be used for assessing patient outcomes in epilepsy. Transfer-learning with the GoogLeNet InceptionV3 CNN model pre-trained on the large ImageNet dataset (~1.2 million images) was able to differentiate upper (n=12) and lower (n=9) response quartile mesiotemporal lobe epilepsy patients in the NeuroPace® RNS® System clinical trials with ~76% classification accuracy based on chronic ambulatory baseline electrocorticographic (ECoG) data. These promising findings justify further research using deep CNNs for assessing patient outcomes in epilepsy.**

I. INTRODUCTION

A major challenge in epilepsy is assessing effectiveness of treatment [1, 2]. Using measures derived from interictal chronic ambulatory intracranial electrographic recordings (ECoGs) to assess patient outcomes may help with quickly iterating through therapy options and optimizing treatment for individual patients [3].

Patient-specific modeling studies have explored relationships between interictal electrographic features (e.g. interictal spikes, spectral power bands, etc.) and clinical outcomes with promising results [4, 5]. However, patient specific-models require prior collection of substantial amounts of data from a given patient before the trained model can be applied to the patient, thus reducing their applicability to new patients or patients with limited ECoG data. Electrographic features and models which can generalize well in new epileptic patients remain an unmet need. One limitation of previous studies is the use of hand-crafted features which rely on accurately identifying and extracting the features. Supervised feature engineering also misses the opportunity to discover novel patterns [6].

Deep learning algorithms can perform end-to-end learning from data and do not rely on feature-engineering [7]. Deep convolutional neural networks (deep CNNs) have made breakthroughs in the field of computer vision with CNN depths often positively correlated with model performance, since similar to the human visual processing system, deeper layers of CNNs can learn more complex data patterns compared to the initial shallow CNN layers [7]. Training deep CNNs from scratch for complex problems requires very large datasets usually on the order of millions of data points. However, through the process of transfer learning, the weights learned by a deep CNN on a large dataset may be applied to a small dataset with some customization to tune the network to the small dataset [8]. Transfer learning with pre-trained deep CNNs has demonstrated impressive performance in several domains [8, 9].

The goals of the analyses in this paper were to determine (1) whether pre-trained deep convolutional neural networks are capable of identifying ECoG similarities within groups of epileptic patients with similar clinical outcomes and (2) whether deep CNNs can be used to assess treatment outcome in a new patient whose data was not used for training.

Although transfer-learning with pre-trained deep CNNs has been widely reported in computer vision, speech and text processing [7, 8], their role in analyzing neural signals has been very limited. A few recent studies have trained CNNs for seizure detection and forecasting [10], but to the best of our knowledge, this is the first study to use pre-trained deep CNNs for assessing therapy outcomes in epilepsy using baseline chronic human ambulatory ECoG data.

II. METHODS

The dataset used here comes from patients implanted with the NeuroPace® RNS® System in clinical trials (n = 256 patients) [11] with a median follow-up period of 8.97 years. Since the objective of this analysis is to test the utility of transfer-learning for differentiating patients with different clinical outcomes, we selected two groups of patients with vastly different clinical outcomes. That is, patients in the upper and lower seizure reduction response quartiles after 7 years of treatment. Further, in this analysis we only included patients with mesiotemporal lobe (MTL) epilepsies to limit the analysis to a fairly homogeneous patient population.

A. The RNS System

The NeuroPace® RNS® System was approved by the FDA in 2013 for the adjunctive treatment of patients with partial onset epilepsy having 1-2 seizure foci. Details about the RNS System can be found in several previous publications [11, 12]. In brief, the RNS System consists of a closed-loop responsive neurostimulator device implanted in the skull, with 1 or 2 quadripolar depth or strip leads connected (Fig. 1). Leads are implanted at the seizure foci. MTL patients received mostly bilateral hippocampal depth leads. The neurostimulator continuously monitors differentially recorded neural activity using a patient-specific detection algorithm and delivers electrical stimulation when abnormal activity is detected. ECoGs with 1-4 channels are recorded from the implanted leads and stored in the neurostimulator, then wirelessly downloaded by the patient using an internet-connected Remote Monitor. ECoG records have a data sampling rate of 250 Hz per channel and are typically 90 seconds in duration (Fig. 2A); however the length of stored ECoGs can vary and is determined by the treating physician,

All authors: NeuroPace, Mountain View, California, 94043
Phone: 650-237-2700 E-mail:sdesai@neuropace.com

with a maximum length of 180 seconds. ECoG storage is most frequently triggered by detection of epileptiform activity but may also be scheduled based on time of day. Approximately 1/3 of stored ECoGs are scheduled and contain baseline interictal activity.

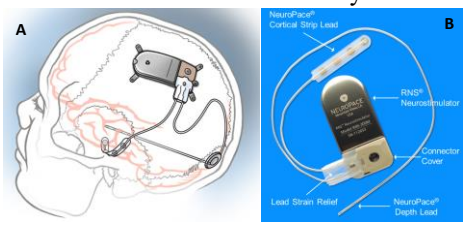

Figure 1. (A) Illustration of the RNS System with the neurostimulator implanted in the skull and connected to a strip lead and a depth lead. (B) Photograph of the RNS System.

B. Patient selection for this study

Out of 256 patients enrolled in the RNS System clinical trials, 111 had mesiotemporal lobe (MTL) epilepsy. Of these, 79 patients had at least 7 years of follow-up after the initial implant of the RNS System [11]. The 20 (out of 79) patients with the greatest reduction in patient-reported clinical seizures at year 7 post-implant compared to a 3-month pre-implant baseline period were classified as the upper response quartile (URQ; -96.5% median change in clinical seizure rate), and the 20 patients with the least seizure reduction were classified as the lower response quartile (LRQ; -17.4% median change). An inclusion criteria of at least 5 scheduled ECoGs captured in year 7 was set for a patient to be included in the analysis.

C. Computing spectrograms

Since the goal of this analysis was to classify patients into upper and lower response quartiles based on baseline non-ictal ECoG data, only scheduled ECoG records containing interictal baseline activity were used. To simply analysis, only data from ECoG channel 1 were used for training and testing. Python 3.5 was used for all analyses. Any stimulation artifacts in the ECoG data were deleted along with 40 ms before and 120 ms after the artifact, concatenating the portions of the ECoG before and after the artifacts. After removal of any stimulation artifacts, ECoG records <85 seconds long were excluded from analysis. Only the first 90 seconds of ECoG records longer than 90 seconds were used for analysis. This was done to ensure that the data used for computing spectrograms were of similar lengths. Spectrogram images were computed using matplotlib.pyplot.specgram. A window size of 256 data points (~1 second of ECoG data) and step size of 128 data points was used for computing the spectrograms between ~1 Hz and 125 Hz. The resulting spectrograms were saved as 299 (height) x 299 (width) pixel PNG images using a gray color scale.

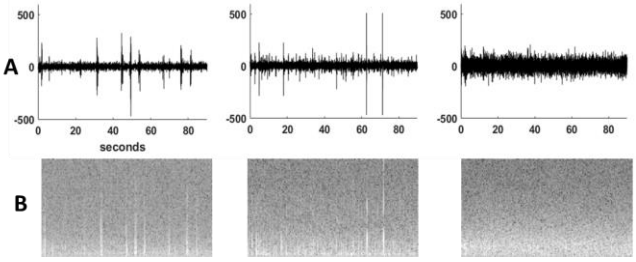

Figure 2. (A) Example time-series of channel 1 data from scheduled ECoGs. (B) Corresponding grayscale spectrogram images derived from the time-series ECoG.

D. Model training and testing

GoogLeNet Inception v3 CNN pre-trained on the ImageNet dataset (image-net.org) consisting of 1.2 million everyday images in 1,000 categories was used to analyze spectrogram data using methods inspired by Esteva et al. [8] where pre-trained deep CNNs were used for classifying skin cancer images.

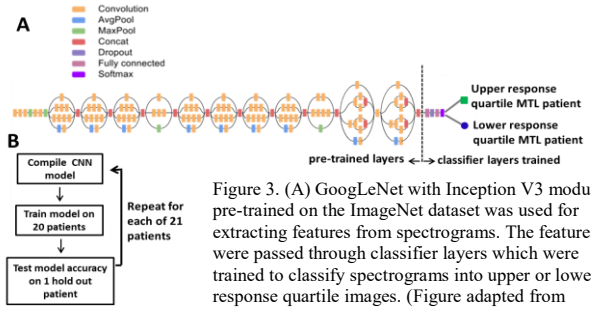

Figure 3. (A) GoogLeNet with Inception V3 modules pre-trained on the ImageNet dataset was used for extracting features from spectrograms. The features were passed through classifier layers which were trained to classify spectrograms into upper or lower response quartile images. (Figure adapted from https://cloud.google.com/tpu/docs/inception-v3-advanced.)

(B) GoogleNet with Inception V3 modules along with the top classifier layers was compiled and trained on spectrograms from 20 patients and tested on one holdout patient's data. This was repeated once with each patient as the holdout.

The only preprocessing performed was the standard procedure of normalizing pixel values to a range between 0 and 1 by dividing by 255. As shown in Fig. 3A, the initial CNN layers used weights from the GoogLeNet model pre-trained on the ImageNet dataset, and the final classification layers were trained to classify the input spectrograms into upper and lower response quartile classes. Informed by preliminary experiments, training was performed for a maximum of 10 epochs. Learning rates between 0.01 and 0.00001, with and without a learning rate decay factor of 10% over each subsequent training epoch, were evaluated using a train and test batch size of 32 and 64. Stochastic gradient descent with momentum of 0.9 for the backpropagation process was used for training optimization. All analyses were performed using Keras v2.2.2 with Tensorflow v1.10.1 backend on an Ubuntu 16.04 computer with 4 NVIDIA GeForce GTX 1080 Ti GPUs. All 4 GPUs were used in parallel for model training and testing.

To assess whether CNNs can be used to assess treatment outcome in a new patient, the CNN was compiled and trained (Fig. 3B) for each patient held out as the test patient. For each iteration, the training dataset consisted of spectrograms from all patients minus holdout, and the test dataset included only the holdout patient.

E. Dimensionality reduction and feature visualization

Inspired by previous studies in computer vision, a two-step dimensionality reduction process was used for visualizing the features extracted by the CNNs [8]. First, principal component analysis (sklearn.decomposition.PCA) was used to reduce the number of features extracted by the final pre-trained model layer from ~131,000 to 50. Second, t-distributed stochastic neighbor embedding (sklearn.manifold.TSNE) further reduced the number of dimensions from 50 to 2 which were used for the 2D visualizations (Fig. 4).

III. RESULTS

Nine patients from the lower quartile and 12 patients from the upper quartile met the analysis inclusion criteria of having at least 5 ECoGs captured during Year 7 after initial implant of the RNS System. The number of ECoGs from URQ patients ranged from 19 to 536, and totaled 1447. LRQ patients had between 6 to 404 ECoGs each, and totaled 990.

A. Visualization of features extracted by GoogLeNet

A two-dimensional representation of the features extracted by the pre-trained GoogleNet Inception v3 model from upper and lower response quartile patients is shown in Fig. 4 A-C. Visualization of spectrograms from two extreme end regions and the mid region of the two dimensional feature space (dotted boxes a, b and c respectively in Fig. 4A) shows stratification of spectrograms within the feature space. Spectrograms containing a large amount of spiking and other signal features characteristic of the epileptic brain (i.e., feature-rich spectrograms) are concentrated at the top of the plot, spectrograms containing moderate amount of spiking and other epileptic signal features are grouped in the center, relatively quiet spectrograms with little spiking are concentrated at the bottom of the plot. The URQ patients contribute relatively more spectrograms to the lower portion of the feature space with relatively quiet spectrograms, whereas the LRQ patients contribute relatively more spectrograms to the upper portion of the feature space. A large overlap in spectrograms from the LRQ and URQ patients can be seen towards the center of the feature space. Spectrograms from within the same LRQ and URQ patient can be either tightly (e.g. B3 and C3) or loosely clustered (e.g. B2 and C2) in the 2D feature space.

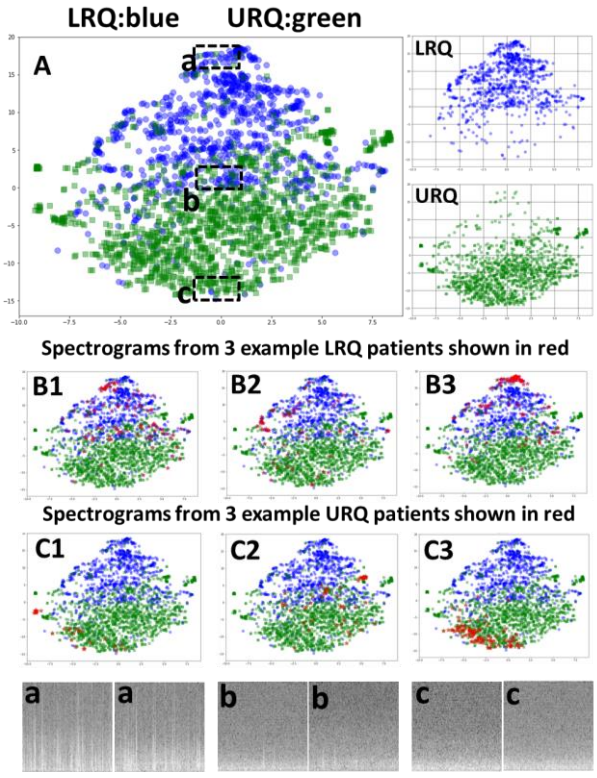

Figure 4: GoogLeNet Inceptionv3 ECoG features extracted from grayscale spectrograms from upper response quartile (URQ) patient ECoGs (n = 1447) and lower response quartile (LRQ) patient ECoGs (n = 990) are shown in green squares and blue circles, respectively. B1-3: ECoG features from 3 individual lower quartile patients are shown with red stars. C1-3 ECoG features from 3 individual upper quartile patients are shown with red stars.Spectrograms labeled a,b,c: Example spectrograms from the 2 extreme end regions (a,c) and the mid region (b) of the 2 dimensional feature space depicted with dotted boxes a,b,c in panel A.

B. Classification performance of the CNN

The majority of training conditions (12 out of 16) produced above chance-level classification accuracy on the test data. That is, the majority of spectrograms were correctly classified in at least 50% of all patients. Among the different initial learning rates (ILR) and learning rate decay factors (LRDF) tested, lower ILRs (≤0.0001) comparatively produced better classification performances. An ILR of 0.0001 with no learning rate decay and batch size of 32 produced the highest classification accuracy of 76.2% with a similar percentage of LRQ (77.8%) and URQ (75%) patients correctly classified. Example trends of training and test accuracies with incremental training in 3 test patients with ILR 0.0001 and no LRDF are shown in figure 5. Under these training conditions, a rapid increase in accuracy was generally observed across all test patients over the 1st 3-5 training epochs. This was followed by saturation, improvements or decline in test accuracy with subsequent training. Training accuracy improvements were nearly uniform across the 21 training iterations (one training iteration for each patient held out as test) and reached 80.2% accuracy by the 5th training epoch and 85.2% accuracy by the 10th training epoch. Among the 15 different ILRs and LRDFs experimented with, results from the top experiment are summarized in Table 1.

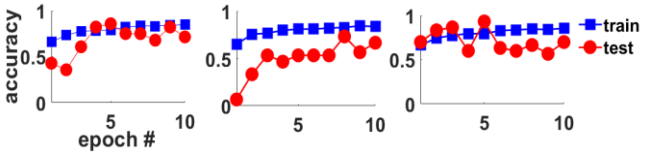

Figure 5: Training (blue) and test (red) accuracy for 3 example test/holdout patients over 10 training epochs for the model summarized in Table 1. A large increase in test accuracy over the 1st 3-5 training epochs was observed across all patients.

Table 1: Classification performance

Initial Learning rate (ILR), Learning decay rate (LDR), Batch Size (BS), Number of training epochs (TE)	Fraction of patients correctly classified†	%of URQ and LRQ patients correctly classified†	Average test accuracy in correctly classified patients
ILR = 0.0001 LDR = none BS = 32 TE= 10	16/21 (76.2%)	URQ: 75% LRQ: 77.8%	72.8%

†Note: Final test accuracies used for generating metrics in this table were computed as the average of test accuracies during final 3 training epochs to dampen noise.

IV. DISCUSSION AND CONCLUSION

The findings in this paper are novel and significant for several reasons: (1) ECoG and clinical data from the NeuroPace RNS System clinical trials provide a unique opportunity to study electrographic correlates of clinical outcomes in epilepsy since it is the only dataset available that contains long-term (mean follow-up period of ~9 years)

chronic ambulatory human ECoG data [11]. (2) To the best of our knowledge, transfer learning with pre-trained deep convolutional neural networks has not been explored before as a tool for assessing patient outcomes in epilepsy based on chronic interictal ECoG data. (3) Despite the presence of large patient-to-patient variations in ECoG features between and within the MTL patients (Figure 4 B1-3 and C1-3), deep CNNs are able to extract and learn ECoG patterns which can differentiate clinical responder and non-responder MTL patients (Figure 4 A). (4) We have shown that transfer learning with weights from a CNN pre-trained on everyday images can be applied to spectrograms of brain activity, a completely different type of dataset, to achieve well above the performance of uninformed random selection (Table 1). These promising results justify further expansion of this work.

After dimensionality reduction for feature visualization, stratification of spectrograms was observed in the 2 dimensional features space with segregation of those containing apparent spiking and other visible signal features characteristic of an epileptic brain (i.e., feature-rich) from those with little or no visible spiking on the other side (Fig. 4). Although loose or no obvious clustering of spectrogram features from within the same patient was observed in several cases (Fig. 4, B2 and C2), spectrograms from the LRQ MTL patients were generally concentrated in the feature-rich region of the 2D feature space and those from the URQ MTL patients were generally concentrated in the feature-poor region, suggesting that LRQ MTL patients have more interictal epileptiform discharges compared to the URQ patients. This type of data visualization could inform hand engineering of a relatively small feature set that may be directly incorporated into computationally-constrained implanted neurostimulators for online assessment of patient brain states and continuous modulation of neurostimulation.

Neural network classification layers trained with an initial learning rate of 0.0001 on features extracted with pre-trained GoogLeNet InceptionV3 layers produced a classification accuracy of ~76% i.e., 16/21 patients were correctly classified . However, only a small range of three hyperparameters i.e., the initial learning rate, the learning rate decay factor and batch size were explored. The initial learning rate and learning rate decay factor are among hyperparameters often having the greatest impact on model performance [13] and hence these were characterized first. In addition to further exploring these hyperparameters, future studies will be devoted to testing other model architectures and training hyperparameters by leveraging the automated neural network architecture search and hyperparameter optimization tools offered by cloud service providers such as the Google Cloud Platform. Other spectrogram color scales, window lengths and step sizes, along with the size and resolution of the saved PNG images will also be explored.

In this analysis, only the final classification layers of the deep CNN were trained, with the rest of the model employing pre-trained weights. Future studies should be devoted to examining the effects of fine-tuning the number of pre-trained networks layers on the classification performance. Additionally, features extracted at several different depths of the CNN should be visualized and used for training neural network and other types of machine learning classifiers.

Only one channel of data per ECoG record was used for training and classification. This was done to simplify the training and test process since the number of channels per ECoG can vary across patients and within a patient over time. Although the selection of the 1st channel of ECoG data from all ECoG records should not bias the analysis in any way, future analyses should be designed to include data from all available channels in a patient for training and testing. In fact, a multi ECoG-channel approach may improve model performance and produce greater model output confidence compared to a single-channel approach.

Finally, the aim of this paper was not to optimize deep CNNs for assessing clinical outcomes in epilepsy, but instead to test the feasibility of transfer learning for finding similar ECoG patterns within epilepsy clinical responders and non-responders. Through preliminary data visualization and classification training experiments, transfer learning with the pre-trained GoogLeNet Inception-V3 model has shown promise in differentiating MTL epilepsy patients who respond to treatment from those who do not.

.

References

[1] C. Hoppe, A. Poepel, and C. E. Elger, "Epilepsy: accuracy of patient seizure counts," *Arch Neurol,* vol. 64, no. 11, pp. 1595-9, Nov 2007.

[2] H. Blumenfeld, "Impaired consciousness in epilepsy," *Lancet Neurol,* vol. 11, no. 9, pp. 814-26, Sep 2012.

[3] J. Engel, Jr. *et al.*, "Epilepsy biomarkers," *Epilepsia,* vol. 54 Suppl 4, pp. 61-9, Aug 2013.

[4] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia,* vol. 52, no. 10, pp. 1761-70, Oct 2011.

[5] L. Kuhlmann *et al.*, "Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG," *Brain,* vol. 141, no. 9, pp. 2619-2630, Sep 1 2018.

[6] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Brief Bioinform,* May 6 2017.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-44, May 28 2015.

[8] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 7639, pp. 115-118, Feb 2 2017.

[9] R. K. Samala, H. P. Chan, L. M. Hadjiiski, M. A. Helvie, K. H. Cha, and C. D. Richter, "Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms," *Phys Med Biol,* vol. 62, no. 23, pp. 8894-8908, Nov 10 2017.

[10] N. D. Truong *et al.*, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Netw,* vol. 105, pp. 104-111, Sep 2018.

[11] G. K. Bergey *et al.*, "Long-term treatment with responsive brain stimulation in adults with refractory partial seizures," *Neurology,* vol. 84, no. 8, pp. 810-7, Feb 24 2015.

[12] M. J. Morrell and R. N. S. S. i. E. S. Group, "Responsive cortical stimulation for the treatment of medically intractable partial epilepsy," *Neurology,* vol. 77, no. 13, pp. 1295-304, Sep 27 2011.

[13] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks,* vol. 1, no. 4, pp. 295-307, 1988.