A Transfer Learning-based Model for Individualized Clustered Seizure Prediction Using Intracranial EEG

Yurui Cao
University of Illinois
Urbana-Champaign, IL, USA
yuruic2@illinois.edu

Krishnakant V. Saboo
University of Illinois
Urbana-Champaign, IL, USA
ksaboo2@illinois.edu

Vaclav Kremen
Mayo Clinic
Rochester, MN, USA
kremen.vaclav@mayo.edu

Vladimir Sladky
Mayo Clinic
Rochester, MN, USA
sladky.vladimir@mayo.edu

Nicholas M. Gregg
Mayo Clinic
Rochester, MN, USA

Paul M. Arnold
Carle Illinois College of Medicine
Urbana, IL, USA

Suguna Pappu
Carle Illinois College of Medicine
Urbana, IL, USA

Philippa J. Karoly
University of Melbourne
Melbourne, VIC, Australia

Dean R. Freestone
Seer Medical Pty Ltd.
Melbourne, VIC, Australia

Mark J. Cook
University of Melbourne
Melbourne, VIC, Australia

Gregory A. Worrell
Mayo Clinic
Rochester, MN, USA
Worrell.Gregory@mayo.edu

Ravishankar K. Iyer
University of Illinois
Urbana-Champaign, IL, USA
rkiyer@illinois.edu

2023 11th International IEEE/EMBS Conference on Neural Engineering (NER) | 978-1-6654-6292-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/NER52421.2023.10123862

***Abstract*—Clustered seizures are prevalent among people with epilepsy and can increase mortality risk. While past research has mainly focused on seizure cluster detection, a few recent studies predict seizure clustering by determining whether there will be more seizures in the next 24 hours after the termination of a seizure. Moreover, personalized prediction of clustered seizures in the presence of limited and imbalanced data remains an outstanding problem. We address this problem using a novel transfer learning model to predict seizure clustering within a 24-hour window. To compensate for the limited and imbalanced available data, for each target patient, the model combines trained individual-level predictive models of the target patient and two other patients whose seizure patterns are similar to those of the target patient. Approximate Kullback-Leibler divergence is used to measure the similarity between patients in high-dimensional data. The proposed model is evaluated on a long-term ambulatory intracranial EEG dataset. Compared with individualized predictive models, the proposed model improves F1 scores for patients with limited or highly imbalanced data by up to 51.0%. In addition, the proposed model achieves an average F1 score of 0.702 and an area under the precision-recall curve of 0.809. Our model can be clinically helpful in guiding the treatment of clustered seizures.**

***Index Terms*—Individualized prediction, Intracranial EEG, Clustered seizures, Deep learning, Transfer learning**

I. Introduction

Clustered seizures, defined as repetitive occurrences of seizures [1] within 24 to 48 hours, are prevalent among people with epilepsy [2]. Untreated clustered seizures can progress to status epilepticus and cause higher mortality risk [3]. Although treatments exist for terminating clustered seizures [4], unnecessary overtreatment increases the risk of respiratory complications [5]. Predicting whether more seizures will occur in the next 24 hours after the termination of a seizure (i.e., seizure clustering prediction) can guide the treatment for seizure clustering termination [4].

While patient-specific differences in characteristics of seizures [6]–[8] motivate personalized prediction, challenges exist for developing appropriate personalized prediction models. First, the performance of personalized models is negatively impacted when a patient's number of seizures (including clustered and isolated ones) is small (e.g., some patients have hundreds of seizures within a year, while others may have fewer than ten seizures [9]). Second, imbalanced data can hurt the performance of personalized clustered seizure prediction models (e.g., there is an imbalance if a patient experiences clustered seizures more frequently than isolated seizures [10]).

We have developed a personalized seizure clustering prediction model to address these challenges with transfer learning. To assist with the personalized prediction for a target patient, the model combines trained individual-level models from the target patient and two patients whose seizure data are most similar to that of the target patient. We measure the similarity between patients by computing approximate Kullback-Leibler divergence (KL divergence) [11] of the pre-ictal and ictal intracranial EEG (iEEG) features of seizures. Our deep-learning-based model demonstrated an $11.1\% - 51.0\%$ improvement in F1 scores on patients with limited or highly imbalanced data compared to individual-level models.

Previous research has focused on clustered seizure detection, with a few studies predicting seizure clustering using long-term iEEG data. Several studies have used statistical methods such as the Hurst exponent [12] to retrospectively detect clustered seizures and study their characteristics [13], [14]. However, these studies did not predict seizure clustering. A recent study [15] developed a support vector machine model to predict lead seizures of clusters using iEEG data from a canine population. This population-level model might fail to capture patient-specific seizure characteristics [6]–[8]. Another

study [10] developed individualized models using the random forest to predict clustered seizures. However, the performance of individualized models can be adversely affected if patients have very few seizures or imbalanced data [9], [10].

We evaluated the prediction performance of the proposed model on a large ambulatory iEEG dataset [9], which contains up to 2 years of iEEG recordings each for 15 people with drug-resistant epilepsy. We make the following contributions:

- We develop a novel transfer learning model for personalized predictions of clustered seizures. It combines trained models from 3 patients with similar seizure data.
- We propose a novel method for comparing pre-ictal and ictal iEEG features of patients with approximate KL divergence. The method effectively selects patients with similar seizure data for transfer learning.
- The model demonstrates an $11.1\% - 51.0\%$ improvement in F1 scores for patients with limited or highly imbalanced data compared to individual-level models. On average, the model has a 0.702 F1 score and 0.809 area under the precision-recall curve (AUPRC).

Our model can easily be adapted to new patients and can be clinically helpful in guiding the treatment of clustered seizures.

II. Methods

Our analysis pipeline is described below (Figure 1).

A. iEEG Data

We evaluated our model on the NeuroVista dataset [9], which contains up to 2 years of iEEG recordings for 15 patients with medically refractory epilepsy. For each patient, 16 electrodes were placed on the presurgically assessed seizure onset zone. Because some patients had significant data drops or very few seizures of one category, we included only eight patients in our analysis, each with 458.1 ± 320.0 seizures.

B. Definition of Seizure Clusters

We labeled a seizure as cluster-non-last (C-NL) if it was the first/intermediate seizure within a cluster, and cluster-last (C-L) if it was the last. C-NL seizures are followed by subsequent seizures within 24 hours after their termination, while isolated (I) or C-L seizures are not. Therefore, we can predict clustered seizure occurrence in the next 24 hours by classifying the just-observed seizure as C-NL rather than I/C-L.

C. Data Preprocessing

We detected the ictal period of seizures from the iEEG data with a published method [16] and defined 10 minutes before seizure onset as the pre-ictal period. Next, we filtered the iEEG data into five frequency bands — delta ($0.5-4$ Hz), theta ($4-8$ Hz), alpha ($8-12$ Hz), beta ($12-25$ Hz), and gamma ($25-45$ Hz) — according to a published method [10] and divided the signals in each band into 2.5-second non-overlapping segments. We computed relative entropy (REN), a bivariate feature measuring the dissimilarity between signals from pairs of electrodes, for all pairs of electrodes in each segment for the five bands. Finally, we computed *mean RENs* by averaging the REN across all pairs of electrodes and segments for each band in the pre-ictal and ictal periods separately so that each seizure was represented by ten features.

D. Source Patients Selection

Transfer learning trains a model on *source* data and applies it to *target* data to transfer knowledge from the source data. For each *target* patient, we identified the two most similar patients as the corresponding *source* patients. The similarity was measured with approximate KL divergence [11] on the mean REN distributions of patients' seizures. With ten-dimensional mean REN data, we did not have enough samples for some patients to compute the approximate KL divergence. Therefore, we reduced the data dimensionality to 2D using the Uniform Manifold Approximation and Projection (UMAP) method to compute the approximate KL divergence.

E. Transfer Learning-based Model

The proposed model uses the mean REN features of a seizure (X) to predict the probability that the seizure will be C-NL (Y). It consists of three parts: (1) a *transfer block*, (2) a *combination block*, and an *output block*. We built a personalized model for each target patient separately.

1) Transfer Block: The transfer block consists of trained neural network layers from the two source patients and the target patient. To obtain the trained layers (referred to as "transfer layers" in the rest of the paper), we trained a multilayer perceptron (MLP) with three layers to predict seizure labels using mean REN data for each patient separately, and we picked the first two layers as the transfer layers. The number of neurons in the first layer varied across patients, while the second layer had ten neurons to ensure consistent output size for combination block computation. Transfer layers learn patient-specific representations that are useful for distinguishing seizure types. Let $Z_{i,j}$ represent the output of the transfer layer obtained from patient j for target patient i's data, $W_{t,j,2}, W_{t,j,1}$ be the trained weights, X_i be the data on a seizure from patient i, and ϕ be the rectified linear unit (ReLU) activation function. Then we can write the output of the transfer layer as $Z_{i,j} = W_{t,j,2}^{\top}[\phi(W_{t,j,1}^{\top} X_i)]$.

2) Combination Block: This block obtains the combined representation, $Z_{c,i}$, by combining transfer block outputs Z_i, where $Z_i = \begin{bmatrix} Z_{i,i} & Z_{i,j} & Z_{i,k} \end{bmatrix}^{\top}$, and $Z_{i,i}, Z_{i,j}, Z_{i,k}$ are representations of target patient i generated by transfer layers from target patient i and source patients j, k. The output of the combination block can be expressed as $Z_{c,i} = \phi(W_{c,i}^{\top} Z_i)$, with $W_{c,i} = \begin{bmatrix} w_{i,i} & w_{i,j} & w_{i,k} \end{bmatrix}^{\top}$. The combination block learns to weigh contributions from individual patients' transfer layers, with source patients more similar to the target patient potentially being weighted more.

3) Output Block: The output block is a fully connected layer with a sigmoid activation function (σ). The probability, Y_i, that a seizure X_i will be cluster-non-last (C-NL) can be written as $Y_i = \sigma(W_{o,i}^{\top} Z_{c,i})$, where $W_{o,i}$ are the weights.

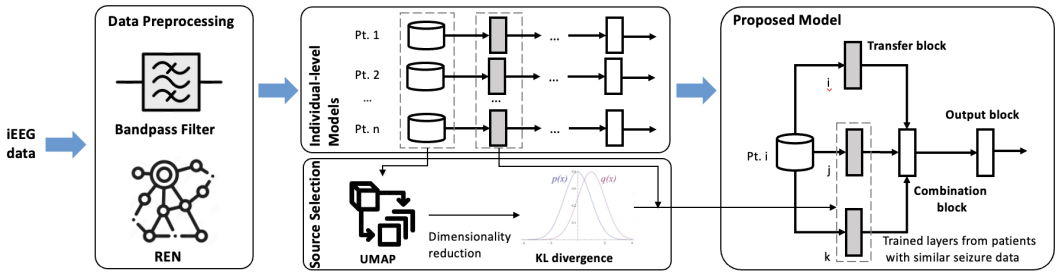

Fig. 1. Analysis pipeline. Assuming the target patient is Pt. i, and Pts. j & k are the two patients whose data are the most similar to those of Pt. i, then the trained layers of Pts. i, j, & k will be used to construct the transfer block for the model of Pt. i. Abbreviations: iEEG, intracranial EEG; REN, relative entropy of network; UMAP, Uniform Manifold Approximation and Projection; KL divergence, Kullback–Leibler divergence; Pt., patient.

F. Experimental Setup

We randomly split the data of each patient into training and testing sets with an 80:20 ratio and developed the proposed model using the training set. The development of the proposed model has two steps: (1) training individual-level models for each patient and (2) training transfer learning models using transfer layers from the corresponding individual-level models. For each step, we selected the optimal sets of hyperparameters (number of neurons, learning rates, batch size, and epochs) using Optuna [17] with 5-fold inner cross-validation, and we fine-tuned the model using the Adam optimizer.

We evaluated the performance of our proposed model using 5-fold cross-validation and compared it against two baselines: (1) individual-level models and (2) a cohort-level model. The cohort-level and individual-level models were both three-layer MLPs to share similar model complexity as the proposed model. The cohort-level model was trained with the training sets on all patients. The individual-level models and our proposed model were trained using training sets on the respective patients. We also compared our model performance with another transfer learning method, the Leave One Patient Out (LOPO) method. The LOPO method trains a cohort-level model on all patients except the target patient and then fine-tunes the trained model with data from the target patient. In a recent study [18], Dissanayake et al. applied this method to seizure prediction using scalp EEG signals. The LOPO method improved the prediction performance significantly compared to the cohort-level model. We used precision, recall, F1 score, and AUPRC as the evaluation metrics.

III. Results

A. Seizure Data Similarity

Seizures of some patients were very different from those of others, while some patients had similar seizure data (Figure 2). That motivated the selection of source patients for transfer learning. Approximate KL divergence was computed for each pair of patients using the distribution of seizures in the UMAP space (Figure 3). As in Figure 2, some patients were more similar to each other than other patients.

B. Prediction Performance

The proposed model was better than or comparable to the baseline models for patient-specific predictions (Table I).

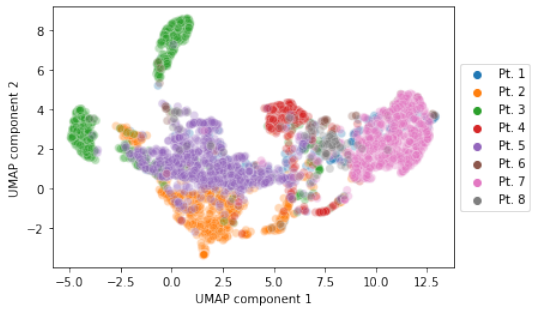

Fig. 2. Seizure data reduced to 2D by UMAP. Each point represents a seizure sample. Samples from the same patient are labeled in the same color.

Compared to individual-level models, the proposed model had better performance on patients with limited data (patients 1, 6 & 8) and patients with highly imbalanced data (patients 2 & 3), with $11.1\% - 51.0\%$ higher F1 scores. For patients with larger amounts of data (patients 5 & 7), the proposed model achieved similar performance as individual-level models.

Compared to the cohort-level model, the proposed model achieved $12.8\% - 97.3\%$ higher F1 scores on patients 6, 7 & 8 by appropriately capturing patient-specific features. In contrast, the cohort-level model was negatively impacted by training with data from dissimilar patients. The proposed model had a lower F1 score than the cohort-level model for patient 4, because patient 4 was similar to multiple patients (see Figure 3), and our selection method used only patients 6 & 8 as source patients. The cohort-level model also learned from the other patients, which provided more helpful information for the prediction for patient 4.

We compared the averaged performance across patients of the proposed model, the LOPO method [18], and the baselines (Table II). The LOPO method outperformed the baselines by capturing patient-specific seizure patterns and overcoming data limitations for some patients. The proposed model achieved the highest scores on all metrics (with $5\% - 8\%$ improvement across the metrics compared to the LOPO method). This performance improvement resulted from our selection method that properly transfers learning from patients with similar data. On the contrary, the LOPO method also transferred learning from patients whose data distributions differed far from that

TABLE I

PREDICTION PERFORMANCE OF THE PROPOSED MODEL FOR EACH PATIENT AGAINST THE INDIVIDUAL- AND COHORT-LEVEL MODELS. THE MEAN OF THE METRICS OVER 5-FOLD CROSS-VALIDATION IS PROVIDED. THE HIGHEST METRICS ACROSS MODELS ARE IN BOLD.

Pt. #	Sample Size		Proposed Model				Individual-level Models				Cohort-level Model			
	I/C-L	C-NL	Precision	Recall	F1	AUPRC	Precision	Recall	F1	AUPRC	Precision	Recall	F1	AUPRC
1	73	107	**0.820**	**0.419**	**0.462**	**0.719**	0.585	0.314	0.389	0.714	0.527	**0.419**	0.460	0.626
2	28	648	**0.969**	0.893	0.921	0.972	0.948	0.324	0.610	0.968	0.961	**0.919**	**0.937**	**0.976**
3	83	646	0.899	**0.729**	**0.794**	0.892	**0.909**	0.564	0.694	**0.910**	0.907	0.633	0.744	0.906
4	132	148	**0.601**	0.352	0.440	**0.624**	0.537	0.200	**0.699**	0.601	0.528	**0.572**	0.606	0.592
5	100	793	0.888	**0.933**	0.905	0.894	0.898	0.861	0.877	0.919	**0.899**	0.923	**0.910**	**0.923**
6	28	10	**0.700**	**0.800**	**0.733**	**0.907**	0.558	0.500	0.535	0.797	0.583	0.300	0.650	0.830
7	228	540	**0.817**	0.650	0.722	**0.830**	0.804	**0.685**	**0.738**	0.825	0.692	0.278	0.366	0.685
8	83	18	**0.625**	**0.533**	**0.643**	**0.639**	0.542	0.333	0.579	0.625	0.412	0.400	0.439	0.537

TABLE II

AVERAGED PREDICTION PERFORMANCE ACROSS PATIENTS FOR EACH MODEL. THE HIGHEST METRICS ACROSS MODELS ARE IN BOLD.

Model	Precision	Recall	F1	AUPRC
Proposed Model	**0.789**	**0.664**	**0.702**	**0.809**
LOPO Method [18]	0.738	0.612	0.667	0.770
Individual-level Models	0.723	0.473	0.640	0.795
Cohort-level Model	0.689	0.511	0.639	0.759

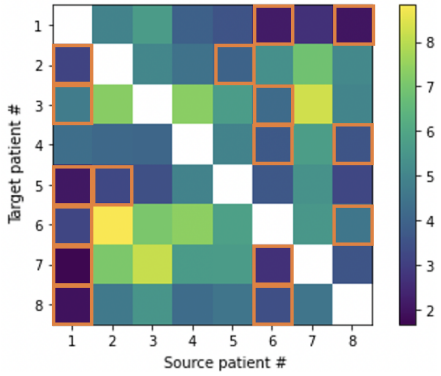

Fig. 3. Approximate KL divergence between patients' seizure data. The lower the KL divergence, the more similar the data. Two source patients with the smallest KL divergence from the target patient are boxed in orange.

of the target patient, leading to a performance drop.

IV. CONCLUSION

We propose a novel transfer-learning-based model that supports personalized predictions of clustered seizures using pre-ictal and ictal iEEG data. On average, the proposed model outperforms the individual-level and cohort-level models. Furthermore, the generality of our source selection method and the modular structure of the model allows it to be easily applied to new patients. Thus, our approach can guide treatment selection for terminating clustered seizures.

Limitations. Our work has several limitations. First, we did not consider the temporal dynamics of iEEG, which can help with prediction. Second, we used two patients as sources even if more subjects had similar seizure data. Using data from more source patients could further improve performance. Lastly, we evaluated our model on data from only eight patients because of the challenges of collecting ambulatory iEEG data. Evaluation on larger amounts of more easily obtained data, such as scalp EEG data, will be important to assess the generalizability of the results. We will address these limitations in future work.

ACKNOWLEDGMENTS

This work was partly supported by NSF award CNS-1624790 (CCBGM), the Jump ARCHES Foundation, Carle Foundation Hospital, and a Mayo Clinic-Illinois Technology-Based Healthcare Research Fellowship. We thank Jenny Applequist, Yogatheesan Varatharajah, Chang Hu, Anirudh Choudhary, and Mosbah Aouad for their valuable feedback.

REFERENCES

[1] S. Haut, "Seizure clustering," *Epilepsy & Behavior*, vol. 8, 2006.
[2] S. Jafarpour et al., "Seizure cluster: Definition, prevalence, consequences, and management," *Seizure*, vol. 68, pp. 9–15, 2019.
[3] S. Haut, S. Shinnar, and S. Moshé, "Seizure clustering: Risks and outcomes," *Epilepsia*, vol. 46, no. 1, pp. 146–149, 2005.
[4] B. Gidal and K. Detyniecki, "Rescue therapies for seizure clusters: Pharmacology and target of treatments," *Epilepsia*, vol. 63, 2022.
[5] M. Spatola, V. Alvarez, and A. Rossetti, "Benzodiazepine overtreatment in status epilepticus is related to higher need of intubation and longer hospitalization," *Epilepsia*, vol. 54, no. 8, pp. e99–e102, 2013.
[6] M. Cook et al., "Human focal seizures are characterized by populations of fixed duration and interval," *Epilepsia*, vol. 57, no. 3, 2016.
[7] P. Karoly et al., "The circadian profile of epilepsy improves seizure forecasting," *Brain*, vol. 140, no. 8, pp. 2169–2182, 2017.
[8] D. Freestone, P. Karoly, and M. Cook, "A forward-looking review of seizure prediction," *Current Opinion in Neurology*, vol. 30, 2017.
[9] M. Cook 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," *The Lancet Neurology*, vol. 12, no. 6, 2013.
[10] K. Saboo et al., "Individualized seizure cluster prediction using machine learning and ambulatory intracranial EEG," in *2022 IEEE International Conference on Bioinformatics and Biomedicine*, 2022, pp. 1157–1163.
[11] F. Perez-Cruz, "Kullback-Leibler divergence estimation of continuous distributions," in *2008 IEEE Int. Symp. Inf. Theory*, 2008.
[12] S. Chiang et al., "Individualizing the definition of seizure clusters based on temporal clustering analysis," *Epilepsy Research*, 2020.
[13] A. Asadi-Pooya et al., "Seizure clusters in drug-resistant focal epilepsy," *Epilepsia*, vol. 57, no. 9, pp. e187–e190, 2016.
[14] V. Ferastraoaru et al., "Termination of seizure clusters is related to the duration of focal seizures," *Epilepsia*, vol. 57, no. 6, pp. 889–895, 2016.
[15] H. Chen, H. Shiao, and V. Cherkassky, "Online prediction of lead seizures from iEEG data," *Brain Sciences*, vol. 11, no. 12, 2021.
[16] V. Sladky et al., "Distributed brain co-processor for tracking spikes, seizures and behavior during electrical brain stimulation," *Brain Communications*, vol. 4, no. 3, 2022.
[17] T. Akiba et al., "Optuna: A next-generation hyperparameter optimization framework," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.
[18] T. Dissanayake et al., "Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9377–9388, 2021.