



Published in final edited form as:

*Seizure*. 2021 April ; 87: 61–68. doi:10.1016/j.seizure.2021.03.001.

## Machine Learning Models to Predict Electroencephalographic Seizures in Critically Ill Children

Jian Hu, MS<sup>1</sup>, France W. Fung, MD, MSc<sup>3,4</sup>, Marin Jacobwitz, NP<sup>3</sup>, Darshana S. Parikh<sup>2,7</sup>, Lisa Vala, RN REEGT<sup>7</sup>, Maureen Donnelly, REEGT<sup>7</sup>, Alexis A. Topjian, MD, MSCE<sup>2,5</sup>, Nicholas S. Abend, MD, MSCE<sup>3,4,5,6,7</sup>, Rui Xiao, PhD<sup>1,6</sup>

<sup>1</sup>Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine

<sup>2</sup>Department of Anesthesia and Critical Care Medicine, Children's Hospital of Philadelphia

<sup>3</sup>Department of Pediatrics (Division of Neurology), Children's Hospital of Philadelphia

<sup>4</sup>Departments of Neurology and Pediatrics, University of Pennsylvania Perelman School of Medicine

<sup>5</sup>Department of Anesthesia & Critical Care, University of Pennsylvania Perelman School of Medicine

<sup>6</sup>Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine

<sup>7</sup>Department of Neurodiagnostics, Children's Hospital of Philadelphia

### Abstract

**Objective:** Determine whether machine learning techniques would enhance our ability to incorporate key variables into a parsimonious model with optimized prediction performance for electroencephalographic seizure (ES) prediction in critically ill children.

**Methods:** We analyzed data from a prospective observational cohort study of 719 consecutive critically ill children with encephalopathy who underwent clinically-indicated continuous EEG monitoring (CEEG). We implemented and compared three state-of-the-art machine learning methods for ES prediction: (1) random forest; (2) Least Absolute Shrinkage and Selection Operator (LASSO); and (3) Deep Learning Important Features (DeepLIFT). We developed a ranking algorithm based on the relative importance of each variable derived from the machine learning methods.

---

Corresponding Author: Rui Xiao, PhD, 423 Guardian Drive, Blockley Hall, Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104-6021, xiaor@email.chop.edu.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Statistical Analysis:** Jian Hu, MS and Rui Xiao, PhD.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Results:** Based on our ranking algorithm, the top five variables for ES prediction were: (1) epileptiform discharges in the initial 30 minutes, (2) clinical seizures prior to CEEG initiation, (3) sex, (4) age dichotomized at 1 year, and (5) epileptic encephalopathy. Compared to the stepwise selection-based approach in logistic regression, the top variables selected by our ranking algorithm were more informative as models utilizing the top variables achieved better prediction performance evaluated by prediction accuracy, AUROC and F1 score. Adding additional variables did not improve and sometimes worsened model performance.

**Conclusion:** The ranking algorithm was helpful in deriving a parsimonious model for ES prediction with optimal performance. However, application of state-of-the-art machine learning models did not substantially improve model performance compared to prior logistic regression models. Thus, to further improve the ES prediction, we may need to collect more samples and variables that provide additional information.

### Keywords

Seizure; Pediatric; Electroencephalogram; EEG Monitoring; machine learning

---

### Introduction

Electroencephalographic seizures (ES) occur in 10-40% of children with acute encephalopathy who undergo continuous EEG monitoring (CEEG),<sup>1-32</sup> increasing evidence indicates that high ES exposure is associated with unfavorable neurobehavioral outcomes,<sup>8,11,15,23,24</sup> and ES are often treatable with anti-seizure medications.<sup>33-35</sup> Thus, CEEG use is increasing for ES identification and management,<sup>36</sup> and guidelines recommend that children with acute encephalopathy undergo 24-48 hours of CEEG to identify ES.<sup>37-39</sup> However, given CEEG is resource-intensive,<sup>40,41</sup> a clinical prediction tool that enabled evidence-based targeting of CEEG resources to patients most likely to experience ES would be of great clinical value. Unfortunately, developing an ES prediction model with sufficiently high specificity to meaningfully reduce CEEG utilization and sufficiently high sensitivity to avoid failing to identify patients experiencing ES has been difficult. Prior studies have identified risk factors for ES in critically ill children,<sup>2,5-7,9-12,15,16,18,19,21,22</sup> and generated multi-variable prediction models using standard regression approaches.<sup>31,32,42,43</sup> In a prospective observational study of 719 consecutive critically ill children with acute encephalopathy who underwent CEEG, we used a logistic regression approach to develop an ES prediction model. Variables associated with increased ES risk included age (1 year or <1 year), acute encephalopathy category (epilepsy-related, acute structural, acute non-structural), clinical seizures prior to CEEG initiation (present or absent), EEG background category (normal/sleep, slow-disorganized, discontinuous, burst-suppression, or attenuated-featureless), and epileptiform discharges (present or absent). The model yielded an area under the receiver operating characteristic curve (AUROC) of 0.80. If model cutoff was selected to emphasize sensitivity to avoid failing to identify a child experiencing ES, then the model would limit 29% of patients from undergoing CEEG while failing to identify 8% of patients with ES.<sup>31</sup> A subsequent study indicated that the model was well calibrated and well discriminating in a separate validation cohort.<sup>43</sup>

Machine learning techniques could improve prediction performance. In this study, we aimed to determine whether several machine learning techniques, ranging from widely used traditional machine learning models to novel neural network approaches, would enhance our ability to incorporate key variables into a parsimonious model with optimized prediction performance. In addition, given the power of the machine learning techniques, we aimed to develop a ranking algorithm for the features to better understand which features are most important for seizure prediction. This ranking information could provide guidance to prioritize data collection for future seizure risk prediction studies.

## Methods

### Study Subjects

The Institutional Review Board approved the study with a waiver of consent. We registered the study with [clinicaltrials.gov](https://clinicaltrials.gov) (NCT03419260) and applied observational study reporting standards.<sup>44</sup> Details regarding the cohort characteristics and a standard logistic regression model for ES prediction have been published.<sup>31-33,43</sup>

We performed a prospective observational cohort study of consecutive critically ill children with encephalopathy treated in the pediatric intensive care unit (PICU) of a quaternary care hospital who underwent clinically-indicated CEEG to screen for ES based on a guideline-adherent<sup>37-39</sup> institutional pathway<sup>45</sup> between April 2017 and February 2019. We excluded: (1) neonates (<30 days old); (2) patients who received brief post-operative epilepsy surgery care in the PICU; and (3) patients admitted after >2 days of care for refractory status epilepticus at a different institution since the available initial clinical and EEG data were often insufficient. We performed continuous video-EEG monitoring using Natus Neuroworks (Middleton, WI) using the international 10-20 system for electrode placement. The inpatient Electroencephalography Service EEG interpreted the EEG studies, and the Critical Care Medicine and Neurology Consultation Services managed the patients.

We prospectively collected clinical and EEG data using a Research Electronic Data Capture (REDCap) database.<sup>46</sup> Clinical data included age, sex, prior neurodevelopmental disorders, medications, CEEG indication, hospital and PICU admission and discharge dates, presence of clinically evident seizures prior to CEEG, acute encephalopathy category (epilepsy-related, acute structural, or acute non-structural) based on the primary presenting problems/diagnoses available at the time of admission, and mental status (comatose or not; baseline or not). A pediatric electroencephalographer (FWF) scored the EEG through review of both tracings and clinical reports using standardized<sup>47</sup> and critical care EEG terminology<sup>48</sup> which has good inter-rater reliability<sup>49</sup> and has been used in prior studies of CEEG in critically ill children.<sup>11,23,24,31-33,42,43,50-52</sup> EEG data included initiation and discontinuation date and time, background features (normal/sleep, slow-disorganized, discontinuous, burst-suppression, or attenuated-featureless) epileptiform discharges, and ES presence and characteristics. We defined ES as abnormal paroxysmal events that were different from the background, lasted longer than 10 seconds, had a plausible electrographic field, and had a temporal-spatial evolution in morphology, frequency, and amplitude, as consistent with proposed definitions<sup>53</sup> and prior studies.<sup>16,23,31-33,52</sup> We defined epileptiform discharges based on a glossary of EEG terms as transients distinguishable from the background with

sharp or spiky morphology, different wave duration than the background, waveform asymmetry, and sometimes followed by an after-going slow-wave.<sup>47</sup>

### Machine Learning Methods

We implemented three machine learning methods to predict ES occurrence. First, random forest, is a popular tree-based machine learning algorithm for classification.<sup>54</sup> It was preferred for variable ranking for this dataset because all the variables in this study were categorical. In addition, random forest is more robust and interpretable than standard logistic regression. In our implementation, we built a random forest model consisting of 30 unpruned decision trees, each of which was built over a random extraction of the observations and a random extraction of the variables from the data. Thus, each tree only used partial variables and partial observations, which guaranteed that the trees were de-correlated and therefore, less prone to over-fitting. We built random forest models with different numbers of tree estimators, among which the model with best performance according to the five-fold cross validation was selected (Supplemental Table 1). After building the random forest, the entropy decreases from each variable were averaged across the decision trees to determine the importance of each variable. Second, Least Absolute Shrinkage and Selection Operator (LASSO),<sup>55</sup> is a regression analysis method performing simultaneous variable selection and regularization to enhance prediction accuracy and interpretability.<sup>55</sup> This method uses a penalty factor to control the number of variables kept in the model through an alpha parameter to regulate the weight of the penalty factor. A higher value of the alpha indicates that fewer variables are considered important and remain in the model. Starting from zero (all variables were included in the model), we gradually increased the value of alpha to 0.05 by 0.0001 at a time, and the model dropped variables which were less important. We tracked the set of remaining variables and ranked the importance of variables by the reverse order of each variable being dropped from the model. Third, Deep Learning Important FeaTures (DeepLIFT), is a state-of-the-art method which derives feature importance scores based on neural networks.<sup>56</sup> DeepLIFT is a novel algorithm for decomposing the output prediction of a neural network on a specific input by backpropagating the contributions of all neurons in the network to every variable of the input. It can assign importance score to each input variable for a given output. With the power of DeepLIFT, we were able to prioritize variables critical for ES prediction and rank them directly using the derived importance score. We implemented fully connected neural networks with different structures, i.e., different number of hidden layers (Supplemental Table 2), and we selected the one with best performance in ES prediction. Then, we used DeepLIFT to rank the variables based on the importance scores. Neural network approaches have shown superior performance beyond traditional statistical methods in a variety of fields.<sup>57</sup>

### Model Performance Evaluation and Variable Ranking

We evaluated the model performance in predicting ES using five-fold cross validation with three metrics: (1) prediction accuracy including training and validation accuracy, which measures the percentage of observations, both positive and negative, that were correctly classified or predicted by the model in the training set and validation set, respectively; (2) AUROC; and (3) F1 score which combines precision and recall into one metric by

calculating the harmonic mean between the two, and therefore considers the false positive and false negative predictions simultaneously. The standard logistic regression model previously reported from this cohort<sup>31</sup> was the baseline for model comparisons.

We performed variable ranking using the random forest, LASSO, and DeepLIFT models. Subsequently, we calculated an *overall composite ranking* for each variable that combined the results across the three models. The overall composite ranking was calculated as a weighted sum of the three rankings using their validation accuracies as the weights. Variables with multiple categories were coded as dummy variables for each level against the baseline. The highest rank among the corresponding dummy variables was used as the rank for the multilevel variable. As comparison, we also obtained the order of the variables selected into the logistic regression model by the stepwise selection algorithm.

We considered three ‘forward’ scenarios using logistic regression to test the validity of the *overall composite ranking*. First, in the rank-based scenario, variables were added sequentially to the model one at a time in the order of the overall ranking starting from the highest ranking (i.e., the most important variable). Second, in the stepwise selection-based scenario, variables were added according to the stepwise model selection approach in a logistic regression. Third, in the random scenario, variables were added in a random order. Five-fold cross validation accuracies, AUROCs, and F1 scores were monitored during the process to evaluate the performance of these three scenarios. We hypothesized that to achieve similar performance, the prediction model would need fewer variables when it included the variables with the highest overall composite rankings as compared to including variables under the other two scenarios. This ranking algorithm would determine a relatively parsimonious model with the most important variables while achieving reasonably good ES prediction.

## Results

### Cohort

We enrolled 719 consecutive critically ill children. ES occurred in 184 subjects (26%). Table 1 provides the clinical and EEG characteristics of the full cohort and the subgroups with and without ES. We described the cohort previously as part of the standard logistic regression analysis to predict ES.<sup>31</sup>

### Model Performance

Table 2 provides performance of the models using five-fold cross validation. Training accuracy was higher for the random forest model (0.963) and DeepLIFT model (0.962) compared to the standard logistic regression (0.762) or LASSO regression model (0.791) with  $\alpha=0.01$ . All models were similar in validation accuracy, although the LASSO regression model achieved slightly higher validation (0.777) than the other models. Based on the F1 score, all models performed similarly with F1 scores ranging from 0.732-0.758.

## Variable Ranking

Table 3 provides rankings for all variables obtained from each model, as well as the *overall composite ranking*. According to the *overall composite ranking*, the top five variables for ES prediction were: (1) epileptiform discharges in the initial 30 minutes, (2) clinical seizures prior to CEEG, (3) sex, (4) age dichotomized at 1 year, and (5) epileptic encephalopathy. The last column of Table 3 lists the order of the variables selected into the logistic regression model by the stepwise selection algorithm. While most of the high-ranking variables were similar across the approaches, there were several differences. The variable with the top overall composite ranking, epileptiform discharges in the initial 30 minutes, was the fourth variable selected into the model by stepwise selection algorithm, while the EEG background initial 30 minutes was the eighth overall composite ranked variable yet was the third ranked variable by stepwise selection.

We evaluated the prediction performance of models with the top ranked variables. Figure 1 shows the changes of validation accuracies of the logistic regression models as we sequentially added variables one at a time to the model under the rank-based, stepwise selection-based, and random scenarios. When the variables were added to the model in the ranked order, the model started with a fairly high accuracy even with only the top variable included (0.777), peaked with the top three variables included (0.782), and then decreased slightly with addition of more variables. Comparatively, when variables were added in the stepwise selection-based order, the accuracy started relatively low (0.743), dropped as the top three variables were included (0.735), peaked with 13 variables added (0.779), and then decreased with additional variables added. Finally, when variables were added in a random order, the validation accuracy started low (0.742), dropped as other 8 variables were added (0.736), then increased and peaked with 19 variables included (0.783).

Figure 2 provides the comparison of AUROC of the logistic regression models for the rank-based, stepwise selection-based, and random scenarios. The AUROC for the rank-based scenario started lower with only one variable included (0.682) and peaked with seven variables (0.793). The AUROC for the stepwise selection-based scenario started from a lower (0.616) and peaked with seven variables (0.795). The AUROC for the random scenario started low (0.469) and peaked (0.782) with 15 variables.

Figure 3 provides the F1 score of the logistic regression models for the rank-based, stepwise selection-based, and random scenarios. For the rank-based scenario, the F1 score started from a peak value of 0.771 and fluctuated no lower than 0.751 as variables were included. For the stepwise selection-based scenario, the F1 score increased from 0.633 to 0.759 as variables were included. For the random scenario, the F1 score increased from 0.632 to 0.759 as more variables were included.

Thus, all three model performance metrics (prediction accuracy, AUROC, and F1 score) suggested that the top ranked variables by our ranking algorithm were more useful for ES prediction than the low ranked variables. Compared to the stepwise selection-based scenario, the top three variables selected by our ranking algorithm were more informative and achieved much higher validation accuracy, AUROC, and F1 score. In addition, models with the top eight variables by the overall composite rankings would achieve the best prediction

performance, while adding more variables to the model did not help improve model performance.

## Discussion

ES are common and associated with unfavorable neurobehavioral outcomes in children.<sup>8,11,15,23,24</sup> However, CEEG is resource-intensive,<sup>40,41</sup> so evidence-based strategies for targeting CEEG resources to patients at greatest risk for experiencing ES are needed. We previously used standard logistic regression methodology to develop an ES prediction model that incorporated three clinical and two EEG variables. It had an AUROC of 0.80, and at a 0.10 cutoff selected to emphasize sensitivity (to avoid missing patients experiencing ES), implementation would limit 29% of patients from undergoing CEEG while failing to identify 8% of patients with ES.<sup>33</sup> Applying the model in a separate validation cohort yielded similar performance characteristics.<sup>43</sup> Unfortunately, our data indicate that neither application of machine learning techniques, including advanced neural network approaches, nor incorporation of many potential predictor variables improved model performance compared to standard logistic regression. Model performance was similar using the standard logistic regression and all three machine learning techniques when assessed using several measures (prediction accuracy, AUROC, and F1 score). Further, the ranking of most variables was generally similar across the different approaches, including the standard logistic regression approach. Finally, when variables were added in the order by our ranking algorithm or by stepwise regression, the top variables yielded optimal performance by all performance metrics. Adding additional variables did not improve and sometimes worsened model performance. Overall, even with advanced machine learning techniques and inclusion of additional potential predictor variables, model performance reached a threshold with a validation accuracy of 0.72-0.78, and other accuracy assessments indicated a similar threshold effect.

These data indicate that standard logistic regression techniques using a small number of variables<sup>33,43</sup> may have achieved the most effective and parsimonious prediction models for determining ES risk. Novel approaches may be necessary to improve the value of CEEG through more targeted use. Several approaches may be beneficial. First, ES prediction in more homogeneous cohorts may be necessary. Second, value might be improved by instituting a more tailored approach to CEEG duration rather than targeting CEEG initiation to specific patient populations. This kind of strategy would entail initiation of CEEG broadly across patients but reduce CEEG duration in patients at lowest risk for ES. Studies in both adults and children indicate the value of this approach.<sup>32,58</sup> Our data indicated that the top-ranked overall composite variable was epileptiform discharges in the initial 30 minutes, which would be detectable with only a brief screening EEG. Thus, even in models intending to select patients for CEEG initiation, widespread use of a screening EEG may be beneficial, thereby making CEEG duration the key variable to hone for more cost-effective, tailored approaches.

According to the overall composite ranking which combined the three machine learning technique rankings, the top five variables for ES prediction were: (1) epileptiform discharges in the initial 30 minutes, (2) clinical seizures prior to CEEG initiation, (3) sex, (4) age

dichotomized at 1 year, and (5) epileptic encephalopathy. The clinical and EEG variables associated with higher risk of ES are concordant with prior studies identifying risk factors for seizures to be: epileptiform discharges,<sup>2,5,6,12,17,31</sup> the presence of clinically-evident seizures prior to CEEG initiation,<sup>6,7,11-13,17,31</sup> and younger age (often dichotomized at one-year or two-years).<sup>6,11,13,14,16,31,42</sup> The finding that sex is a key predictor is a new finding and may warrant further investigation. Acute encephalopathy category was ranked quite low (twelfth) despite being identified as a variable impacting ES incidence in prior studies.<sup>7,12,14,31</sup> Applying this variable in clinical prediction models is sometimes difficult due to overlapping conditions and diagnostic uncertainty at the time of presentation in some patients. Thus, it may be advantageous that our data indicate it is not a key variable for ES prediction. Similarly, prior studies have identified EEG background category as a predictor of ES,<sup>2,5,31</sup> but it was only the eighth ranked variable in our study. However, since some CEEG would be required to identify epileptiform discharges, the top ranked variable, removing EEG background from prediction models would not preclude the need for at least a brief screening EEG. It is also possible that EEG background is ranked lower because it may a better predictor of ES over longer periods of time, whereas epileptiform discharges in the acute period are more predictive initially.

There are several limitations to this study. Some of the limitations relate to the overall cohort and study design have been noted previously<sup>31-33,43</sup> and are summarized below. First, the study was conducted at one center. While the overall ES incidence and risk factors were consistent with prior studies, generalizability may be enhanced by replication in multi-center studies. Second, we did not test our models in an independent cohort due to limited sample size and low prevalence of seizures. For this type of study, it would be ideal to have an independent dataset that is not used for model building or validation to test for the final prediction models. In practice, when sample size is large and the prevalence of the outcome is relatively high, one approach is to divide the data into training, validation, and model testing datasets. Third, although EEG data used standardized terminology<sup>47,48</sup> with good inter-rater reliability<sup>49</sup> and commonly used definitions for seizures<sup>1,16,23,53</sup> and background categories,<sup>11,23,24,42,50,51</sup> scoring was performed by a single electroencephalographer. We evaluated the same set of potential predictor variables as previously studied using the standard logistic regression. Inclusion of many variables using data mining approaches, potentially including clinical, EEG, laboratory, and imaging variables not initially expected to improve ES prediction, would be expected to yield superior model performance. However, the results from our ranking algorithm suggest that many variables in this data set provide redundant information, and the top five variables ranked by our algorithm capture the most representative pattern for ES prediction. Last, we tested how the sample size of the training dataset would affect the model performance by randomly selecting 216 samples as validation data and from the remaining data randomly selecting 100, 200, 300, 400, and 500 samples to form the training data to fit the model and then evaluate the model performance. As shown in Supplement Figure 1, the model's accuracy and F1 score increased as the training sample size increased, which suggested that increasing the sample size could potentially improve the performance of these models.

Overall, the ranking algorithm was helpful in deriving a parsimonious model for ES prediction with optimal performance. However, application of state-of-the-art machine



learning models did not substantially improve model performance compared to prior logistic regression models. Thus, with the currently available samples and variables in this data, there may be a limit to the model performance that can be achieved by ES prediction models even with advanced techniques and inclusion of numerous potential predictor variables guided by a ranking algorithm. To further improve the ES prediction, we may need to collect more samples and variables that provide additional information.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Financial Disclosures:

Jian Hu – Reports no disclosures.

France W. Fung – Reports no disclosures.

Alexis A. Topjian – Reports no disclosures.

Rui Xiao – Funding from NIH (NICHD) U54-HD086984.

Darshana Parikh - Reports no disclosures.

Maureen Donnelly - Reports no disclosures.

Lisa Vala - Reports no disclosures.

Nicholas S. Abend – Funding from NIH (NINDS) K02NS096058.

## References

1. Abend NS, Wusthoff CJ, Goldberg EM, Dlugos DJ. Electrographic seizures and status epilepticus in critically ill children and neonates with encephalopathy. *Lancet Neurol* 2013;12:1170–9. [PubMed: 24229615]
2. Jette N, Claassen J, Emerson RG, Hirsch LJ. Frequency and predictors of nonconvulsive seizures during continuous electroencephalographic monitoring in critically ill children. *Arch Neurol* 2006;63:1750–5. [PubMed: 17172615]
3. Tay SK, Hirsch LJ, Leary L, Jette N, Wittman J, Akman CI. Nonconvulsive status epilepticus in children: clinical and EEG characteristics. *Epilepsia* 2006;47:1504–9. [PubMed: 16981867]
4. Shahwan A, Bailey C, Shekerdemian L, Harvey AS. The prevalence of seizures in comatose children in the pediatric intensive care unit: a prospective video-EEG study. *Epilepsia* 2010;51:1198–204. [PubMed: 20163439]
5. Abend NS, Topjian A, Ichord R, et al. Electroencephalographic monitoring during hypothermia after pediatric cardiac arrest. *Neurology* 2009;72:1931–40. [PubMed: 19487651]
6. Williams K, Jarrar R, Buchhalter J. Continuous video-EEG monitoring in pediatric intensive care units. *Epilepsia* 2011;52:1130–6. [PubMed: 21671924]
7. Greiner HM, Holland K, Leach JL, Horn PS, Hershey AD, Rose DF. Nonconvulsive status epilepticus: the encephalopathic pediatric patient. *Pediatrics* 2012;129:e748–55. [PubMed: 22331332]
8. Kirkham FJ, Wade AM, McElduff F, et al. Seizures in 204 comatose children: incidence and outcome. *Intensive Care Med* 2012;38:853–62. [PubMed: 22491938]
9. Arango JI, Deibert CP, Brown D, Bell M, Dvorchik I, Adelson PD. Posttraumatic seizures in children with severe traumatic brain injury. *Childs Nerv Syst* 2012;28:1925–9. [PubMed: 22843174]

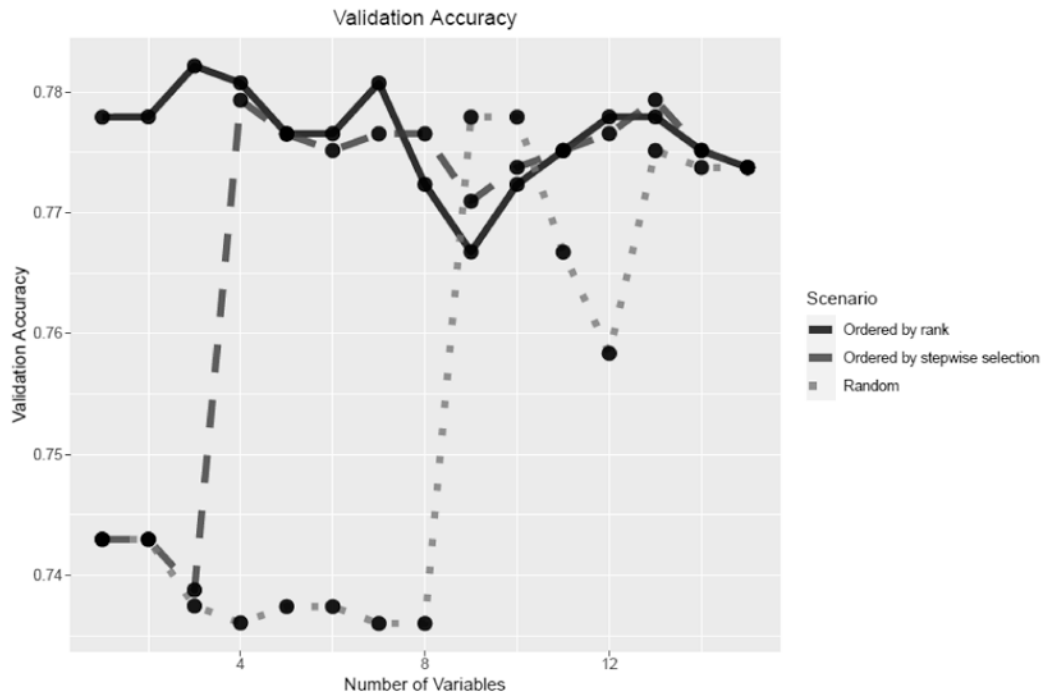
10. Piantino JA, Wainwright MS, Grimason M, et al. Nonconvulsive seizures are common in children treated with extracorporeal cardiac life support. *Pediatr Crit Care Med* 2013;14:601–9. [PubMed: 23823196]
11. Abend NS, Arndt DH, Carpenter JL, et al. Electrographic seizures in pediatric ICU patients: cohort study of risk factors and mortality. *Neurology* 2013;81:383–91. [PubMed: 23794680]
12. McCoy B, Sharma R, Ochi A, et al. Predictors of nonconvulsive seizures among critically ill children. *Epilepsia* 2011;52:1973–8. [PubMed: 22003955]
13. Schreiber JM, Zelleke T, Gaillard WD, Kaulas H, Dean N, Carpenter JL. Continuous video EEG for patients with acute encephalopathy in a pediatric intensive care unit. *Neurocrit Care* 2012;17:31–8. [PubMed: 22565632]
14. Arndt DH, Lerner JT, Matsumoto JH, et al. Subclinical early posttraumatic seizures detected by continuous EEG monitoring in a consecutive pediatric cohort. *Epilepsia* 2013;54:1780–8. [PubMed: 24032982]
15. Payne ET, Zhao XY, Frndova H, et al. Seizure burden is independently associated with short term outcome in critically ill children. *Brain* 2014;137:1429–38. [PubMed: 24595203]
16. Abend NS, Gutierrez-Colina AM, Topjian AA, et al. Nonconvulsive seizures are common in critically ill children. *Neurology* 2011;76:1071–7. [PubMed: 21307352]
17. Gold JJ, Crawford JR, Glaser C, Sheriff H, Wang S, Nespeca M. The role of continuous electroencephalography in childhood encephalitis. *Pediatr Neurol* 2014;50:318–23. [PubMed: 24507696]
18. Vlachy J, Jo M, Li Q, et al. Risk Factors for Seizures Among Young Children Monitored With Continuous Electroencephalography in Intensive Care Unit: A Retrospective Study. *Front Pediatr* 2018;6:303. [PubMed: 30374434]
19. Sansevere AJ, Duncan ED, Libenson MH, Loddenkemper T, Pearl PL, Tasker RC. Continuous EEG in Pediatric Critical Care: Yield and Efficiency of Seizure Detection. *J Clin Neurophysiol* 2017;34:421–6. [PubMed: 28430674]
20. Sanchez Fernandez I, Sansevere AJ, Gainza-Lein M, Buraniqi E, Tasker RC, Loddenkemper T. Time to continuous electroencephalogram in repeated admissions to the pediatric intensive care unit. *Seizure* 2018;54:19–26. [PubMed: 29182970]
21. Sanchez SM, Arndt DH, Carpenter JL, et al. Electroencephalography monitoring in critically ill children: current practice and implications for future study design. *Epilepsia* 2013;54:1419–27. [PubMed: 23848569]
22. Sanchez Fernandez I, Abend NS, Arndt DH, et al. Electrographic seizures after convulsive status epilepticus in children and young adults. A retrospective multicenter study. *Journal of Pediatrics* 2014;164:339–46.
23. Topjian AA, Gutierrez-Colina AM, Sanchez SM, et al. Electrographic Status Epilepticus is Associated with Mortality and Worse Short-Term Outcome in Critically Ill Children. *Crit Care Med* 2013;41:215–23. [PubMed: 23164815]
24. Wagenman KL, Blake TP, Sanchez SM, et al. Electrographic status epilepticus and long-term outcome in critically ill children. *Neurology* 2014;82:396–404. [PubMed: 24384638]
25. Hasbani DM, Topjian AA, Friess SH, et al. Nonconvulsive electrographic seizures are common in children with abusive head trauma. *Pediatr Crit Care Med* 2013;14:709–15. [PubMed: 23842589]
26. O'Neill BR, Handler MH, Tong S, Chapman KE. Incidence of seizures on continuous EEG monitoring following traumatic brain injury in children. *J Neurosurg Pediatr* 2015;16:167–76. [PubMed: 25955809]
27. Vaewpanich J, Reuter-Rice K. Continuous electroencephalography in pediatric traumatic brain injury: Seizure characteristics and outcomes. *Epilepsy Behav* 2016;62:225–30. [PubMed: 27500827]
28. Gwer S, Idro R, Fegan G, et al. Continuous EEG monitoring in Kenyan children with non-traumatic coma. *Arch Dis Child* 2012;97:343–9. [PubMed: 22328741]
29. Ostendorf AP, Hartman ME, Friess SH. Early Electroencephalographic Findings Correlate With Neurologic Outcome in Children Following Cardiac Arrest. *Pediatr Crit Care Med* 2016;17:667–76. [PubMed: 27164188]

30. Sanchez Fernandez I, Sansevere AJ, Guerriero RM, et al. Time to electroencephalography is independently associated with outcome in critically ill neonates and children. *Epilepsia* 2017;58:420–8. [PubMed: 28130784]
31. Fung FW, Jacobwitz M, Parikh DS, et al. Development of a model to predict electroencephalographic seizures in critically ill children. *Epilepsia* 2020;61:498–508. [PubMed: 32077099]
32. Fung FW, Fan J, Vala L, et al. EEG Monitoring Duration to Identify Electroencephalographic Seizures in Critically Ill Children. *Neurology* 2020;95:e1599–e608. [PubMed: 32690798]
33. Fung FW, Jacobwitz M, Vala L, et al. Electroencephalographic seizures in critically ill children: Management and adverse events. *Epilepsia* 2019;60:2095–104. [PubMed: 31538340]
34. Abend NS, Sanchez SM, Berg RA, Dlugos DJ, Topjian AA. Treatment of electrographic seizures and status epilepticus in critically ill children: a single center experience. *Seizure* 2013;22:467–71. [PubMed: 23601851]
35. Williams RP, Banwell B, Berg RA, et al. Impact of an ICU EEG monitoring pathway on timeliness of therapeutic intervention and electrographic seizure termination. *Epilepsia* 2016;57:786–95. [PubMed: 26949220]
36. Sanchez SM, Carpenter J, Chapman KE, et al. Pediatric ICU EEG monitoring: current resources and practice in the United States and Canada. *J Clin Neurophysiol* 2013;30:156–60. [PubMed: 23545766]
37. Brophy GM, Bell R, Claassen J, et al. Guidelines for the evaluation and management of status epilepticus. *Neurocrit Care* 2012;17:3–23. [PubMed: 22528274]
38. Herman ST, Abend NS, Bleck TP, et al. Consensus statement on continuous EEG in critically ill adults and children, part I: indications. *J Clin Neurophysiol* 2015;32:87–95. [PubMed: 25626778]
39. Herman ST, Abend NS, Bleck TP, et al. Consensus statement on continuous EEG in critically ill adults and children, part II: personnel, technical specifications, and clinical practice. *J Clin Neurophysiol* 2015;32:96–108. [PubMed: 25626777]
40. Gutierrez-Colina AM, Topjian AA, Dlugos DJ, Abend NS. EEG Monitoring in Critically Ill Children: Indications and Strategies. *Pediatric Neurology* 2012;46:158–61. [PubMed: 22353290]
41. Abend NS, Topjian AA, Williams S. How much does it cost to identify a critically ill child experiencing electrographic seizures? *J Clin Neurophysiol* 2015;32:257–64. [PubMed: 25626776]
42. Yang A, Arndt DH, Berg RA, et al. Development and validation of a seizure prediction model in critically ill children. *Seizure* 2015;25:104–11. [PubMed: 25458097]
43. Fung FW, Parikh DS, Jacobwitz M, et al. Validation of a model to predict electroencephalographic seizures in critically ill children. *Epilepsia* 2020.
44. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of clinical epidemiology* 2008;61:344–9. [PubMed: 18313558]
45. Children's Hospital of Philadelphia, EEG Monitoring Clinical Pathway. 2020. (Accessed November 19, 2020, at <https://www.chop.edu/clinical-pathway/eeg-monitoring-clinical-pathway>.)
46. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics* 2009;42:377–81. [PubMed: 18929686]
47. Kane N, Acharya J, Beniczky S, et al. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. *Clin Neurophysiol Pract* 2017;2:170–85. [PubMed: 30214992]
48. Hirsch LJ, LaRoche SM, Gaspard N, et al. American Clinical Neurophysiology Society's Standardized Critical Care EEG Terminology: 2012 version. *J Clin Neurophysiol* 2013;30:1–27. [PubMed: 23377439]
49. Abend NS, Gutierrez-Colina A, Zhao H, et al. Interobserver reproducibility of electroencephalogram interpretation in critically ill children. *J Clin Neurophysiol* 2011;28:15–9. [PubMed: 21221016]
50. Abend NS, Wiebe DJ, Xiao R, et al. EEG Factors After Pediatric Cardiac Arrest. *J Clin Neurophysiol* 2018;35:251–5. [PubMed: 29438177]

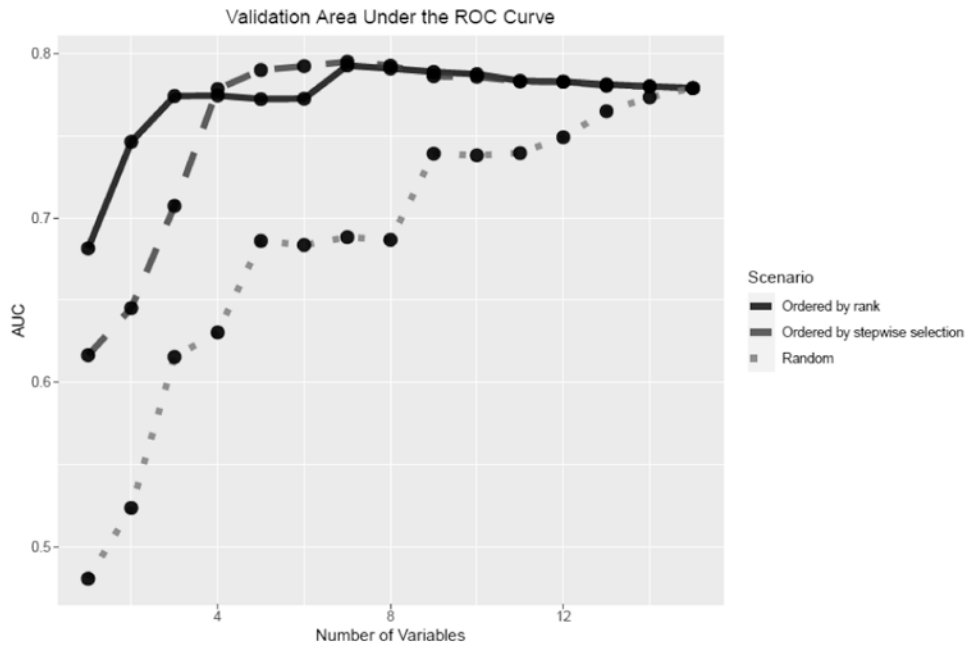
51. Abend NS, Xiao R, Kessler SK, Topjian AA. Stability of Early EEG Background Patterns After Pediatric Cardiac Arrest. *J Clin Neurophysiol* 2018;35:246–50. [PubMed: 29443794]
52. Fung FW, Jacobwitz M, P DS, et al. Development of a model to predict electroencephalographic seizures in critically ill children. *Epilepsia* In press.
53. Beniczky S, Hirsch LJ, Kaplan PW, et al. Unified EEG terminology and criteria for nonconvulsive status epilepticus. *Epilepsia* 2013;54 Suppl 6:28–9.
54. Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002;2:18–22.
55. Tibshirani R Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 1996;58:267–88.
56. Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. *arXiv* 2019.
57. Fukushima K, Miyake S. Neocognitron: a self - organizing neural network model for a mechanism of visual pattern recognition. In: *Competition and Cooperation in Neural Nets*. Berlin, Heidelberg: Springer; 1982: 267–285. In: Amari S, Arbib MA, eds. *Competition and Cooperation in Neural Nets Lecture Notes in Biomathematics*. Berlin: Springer; 1982:267 - 85.
58. Struck AF, Osman G, Rampal N, et al. Time-dependent risk of seizures in critically ill patients on continuous electroencephalogram. *Ann Neurol* 2017;82:177–85. [PubMed: 28681492]

### Highlights

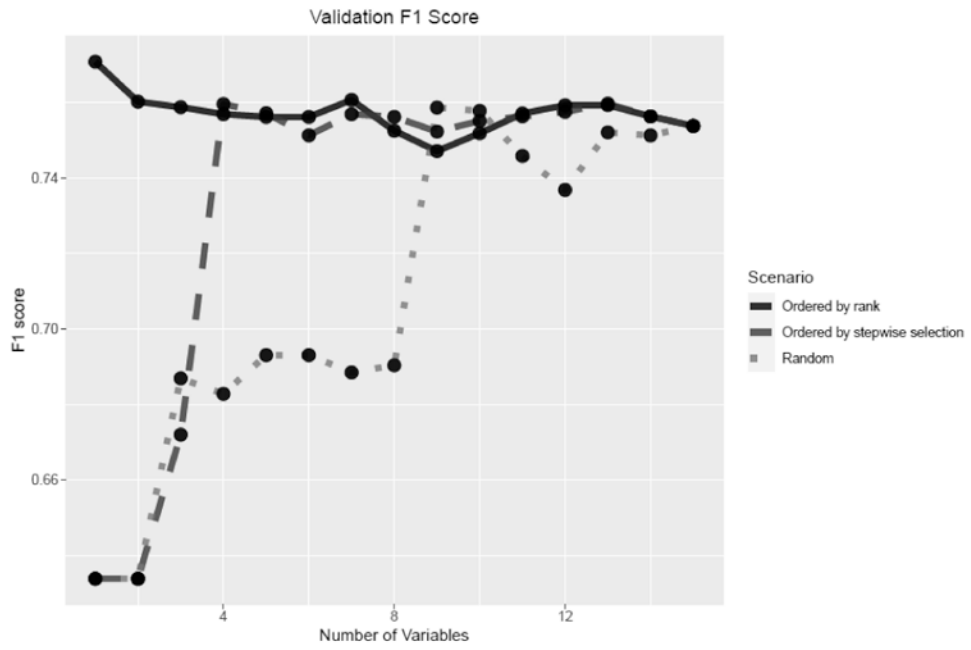
- The goal of this study was to determine whether machine learning techniques would enhance our ability to incorporate key variables into a parsimonious model with optimized prediction performance for electroencephalographic seizure (ES) prediction in critically ill children. We analyzed data from a prospective observational cohort study of 719 consecutive critically ill children with encephalopathy who underwent clinically-indicated continuous EEG monitoring (CEEG). We implemented and compared three state-of-the-art machine learning methods for ES prediction: (1) random forest; (2) Least Absolute Shrinkage and Selection Operator (LASSO); and (3) Deep Learning Important FeaTures (DeepLIFT). We developed a ranking algorithm based on the relative importance of each variable derived from the machine learning methods. Compared to the stepwise selection-based approach in logistic regression, the top variables selected by our ranking algorithm were more informative, and adding additional variables did not improve and sometimes worsened model performance. Overall, the ranking algorithm was helpful in deriving a parsimonious model for ES prediction with optimal performance. However, application of the machine learning models did not substantially improve model performance compared to prior logistic regression models. Thus, to further improve the ES prediction, we may need to collect more samples and variables that provide additional information.



**Figure 1.** Validation accuracy in scenarios in which variables were added in the order by ranking algorithm, stepwise selection, or randomly.



**Figure 2.** Validation using area under the receiver operating characteristic curve (AUC) in scenarios in which variables were added were added in the order by ranking algorithm, stepwise selection, or randomly.



**Figure 3.** Validation using the F1 score in scenarios in which variables were added in the order by ranking algorithm, stepwise selection, or randomly.



**Table 1.**

Summary of clinical and EEG variable associations with electroencephalographic seizures. Data are presented as counts (percentages).

Variable		Total	Electrographic Seizures	No Electrographic Seizures
		<b>719</b>	<b>184 (26%)</b>	<b>535 (74%)</b>
Clinical	Age			
	<1 Year	144 (20%)	47 (26%)	97 (18%)
	1 Year	575 (80%)	137 (75%)	438 (82%)
	Sex			
	Male	417 (58%)	118 (64%)	299 (56%)
	Female	302 (42%)	66 (36%)	236 (44%)
	Race			
	White	329 (46%)	81 (44%)	248 (46%)
	Black or African American	208 (29%)	54 (29%)	154 (29%)
	Asian	22 (3%)	7 (4%)	15 (3%)
	Other	3 (0.4%)	1 (0.5%)	2 (0.4%)
	Unknown	157 (22%)	41 (22%)	116 (22%)
	Ethnicity			
	Not Hispanic or Latino	604 (84%)	156 (85%)	448 (84%)
	Hispanic or Latino	100 (14%)	26 (14%)	74 (14%)
	Unknown	15 (2%)	2 (1%)	13 (2%)
	Acute Encephalopathy Category			
	Acute Non-Structural	159 (22%)	22 (12%)	137 (26%)
	Acute Structural	350 (49%)	79 (43%)	271 (51%)
	Epilepsy	210 (29%)	83 (45%)	127 (24%)
	Epileptic Encephalopathy			
	No	608 (85%)	127 (69%)	481 (90%)
	Yes	111 (15%)	57 (31%)	54 (10%)
	Prior Developmental Delay/Intellectual Disability			
	No	344 (48%)	69 (37%)	275 (51%)
	Yes	375 (52%)	115 (63%)	260 (49%)
	Prior Diagnosis of Epilepsy			
	No	455 (63%)	82 (45%)	373 (70%)
	Yes	264 (37%)	102 (55%)	162 (30%)
	Clinical Seizures Prior to CEEG			
No Seizures	241 (34%)	30 (16%)	211 (39%)	
Seizures or Status Epilepticus	478 (66%)	154 (84%)	324 (61%)	
Mental Status				
No Coma	589 (82%)	159 (86%)	430 (80%)	
Coma	130 (18%)	25 (14%)	105 (20%)	
Mental Status				

Variable		Total	Electrographic Seizures	No Electrographic Seizures
		<b>719</b>	<b>184 (26%)</b>	<b>535 (74%)</b>
	Baseline	611 (85%)	31 (17%)	77 (14%)
	Worse than Baseline	108 (15%)	153 (83%)	458 (86%)
	Sedatives			
	None	360 (50%)	114 (62%)	246 (46%)
	Intermittent Dosed	36 (5%)	6 (3%)	30 (6%)
	Continuous Infusion	323 (45%)	64 (35%)	259 (48%)
EEG	EEG Background Initial 30 Minutes			
	Normal – Asleep	241 (34%)	36 (20%)	205 (38%)
	Slow and Disorganize	370 (51%)	117 (64%)	253 (47%)
	Discontinuous	38 (5%)	16 (9%)	22 (4%)
	Burst-Suppression	17 (2%)	8 (4%)	9 (2%)
	Attenuated and Featureless	53 (7%)	7 (4%)	46 (9%)
	Variability or Reactivity on Day 1			
	No	224 (31%)	70 (38%)	154 (29%)
	Yes	495 (69%)	114 (62%)	381 (71%)
	Epileptiform Discharges in the Initial 30 Minutes			
	None	563 (78%)	95 (52%)	468 (87%)
	Present	156 (22%)	89 (48%)	67 (13%)

**Table 2.**

Summary of the model performance using five-fold cross validation. Data are presented as mean (range).

Model	Training accuracy	Validation accuracy	AUROC	F1 score
Logistic Regression*	0.757 (0.726-0.773)	0.716 (0.643-0.762)	0.769 (0.741-0.815)	0.729 (0.662-0.773)
Random Forest	0.963 (0.958-0.967)	0.740 (0.699-0.764)	0.706 (0.676-0.732)	0.732 (0.688-0.751)
LASSO Regression	0.786 (0.773-0.794)	0.785 (0.748-0.826)	0.781 (0.736-0.822)	0.762 (0.721-0.802)
DeepLIFT	0.962 (0.940-0.972)	0.751 (0.706-0.812)	0.716 (0.680-0.753)	0.748 (0.721-0.760)

Notes: AUROC, area under receiver operating characteristic curve; DeepLIFT, Deep Learning Important Features; LASSO, Least Absolute Shrinkage and Selection Operator.

\*The logistic regression values are derived from a prior publication.<sup>31</sup>

**Table 3.**

Rankings for all variables obtained from Random Forest, LASSO, and DeepLIFT. An overall composite ranking is calculated based on the three methods weighted by their validation accuracies, and it is compared to the order from stepwise selection algorithm in logistic regression.

Variable	Model			Overall Composite Ranking	Stepwise Selection Ranking
	Random Forest	LASSO	DeepLIFT		
Epileptiform Discharges in the Initial 30 Minutes	1	1	2	1	4
Clinical Seizures Prior to CEEG	3	2	1	2	1
Sex	2	7	6	3	5
Age	5	6	4	4	2
Epileptic Encephalopathy	8	4	5	5	6
Sedatives	4	9	7	6	7
Prior Diagnosis of Epilepsy	12	3	8	7	14
EEG Background Initial 30 Minutes	9	5	9	8	3
Mental Status (Coma)	10	13	3	9	11
Variability or Reactivity on Day 1	6	12	10	10	13
Prior Developmental Delay / Intellectual Disability	7	11	11	11	10
Acute Encephalopathy Category	13	8	13	12	9
Ethnicity	14	10	12	13	8
Race	11	14	14	14	12
Mental Status (Worse than Baseline)	15	15	15	15	15