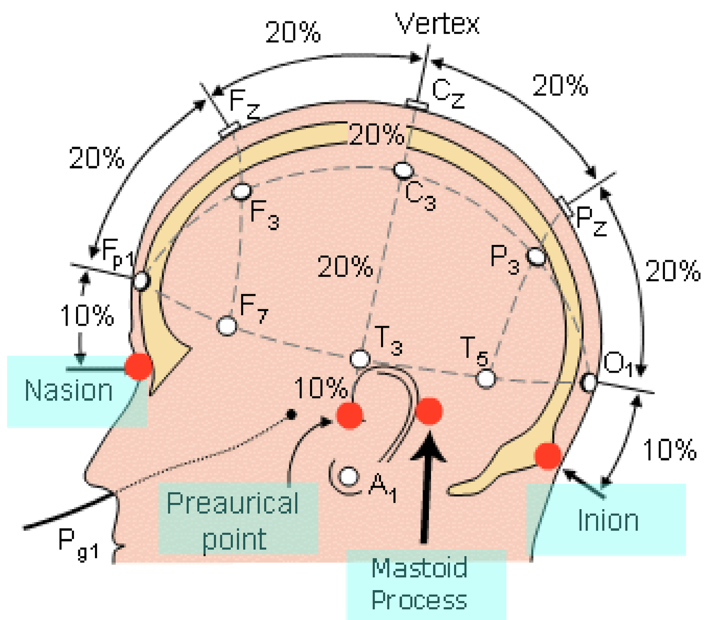
The Diagnosis of Psychogenic Non-epileptic Seizures using Machine Learning

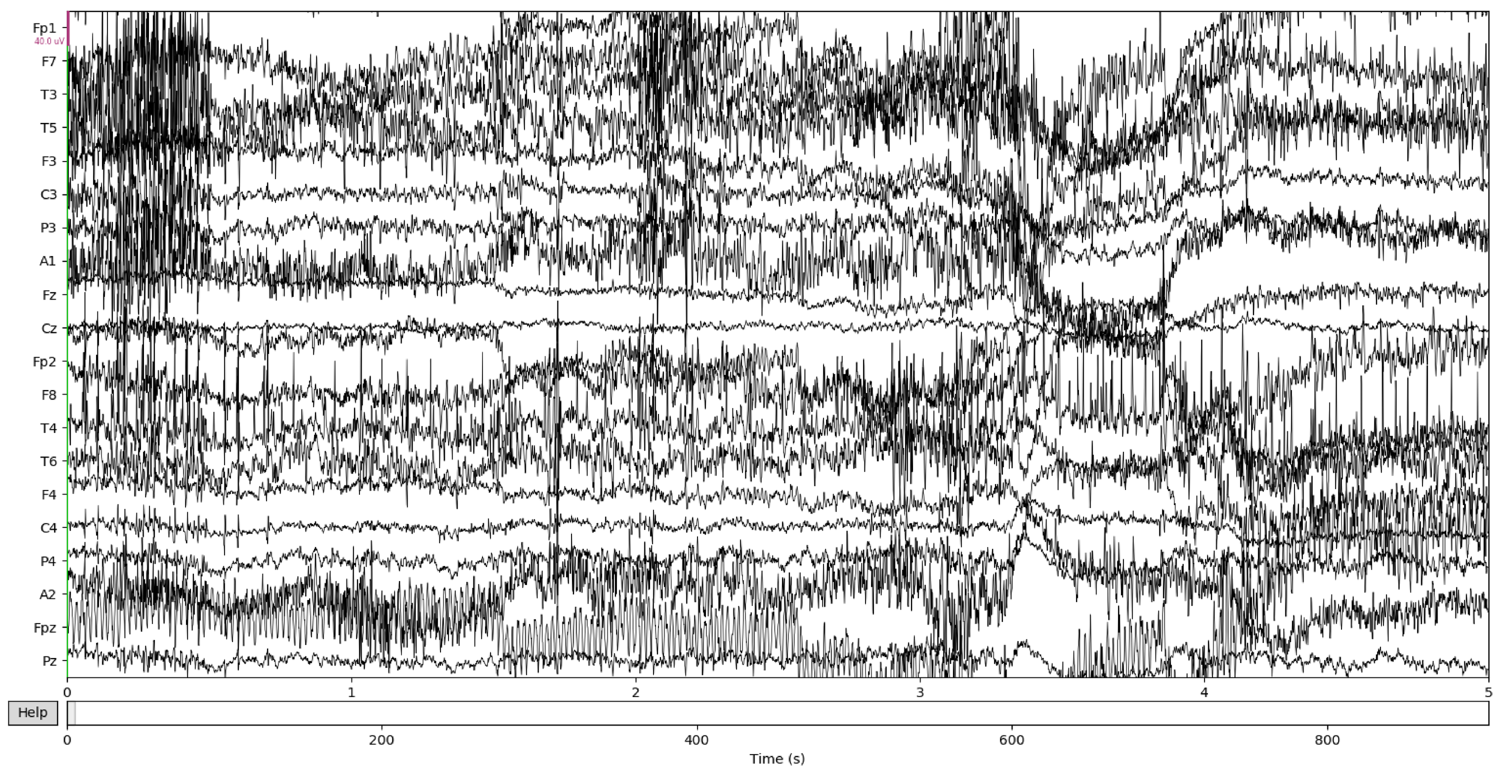
C.H.L. Hinchliffe[1], M. Yogarajah[2], L. Tang[3], D. Abasolo [1]

[1]Centre for Biomedical Engineering, Department of Mechanical Engineering Sciences, University of Surrey, Guildford, GU2 7XH [2]Atkinson Morley Wing, St George's Hospital, London, SW17 0QT [3]Department of Computer Science, University of Surrey, Guildford, GU2 7XH

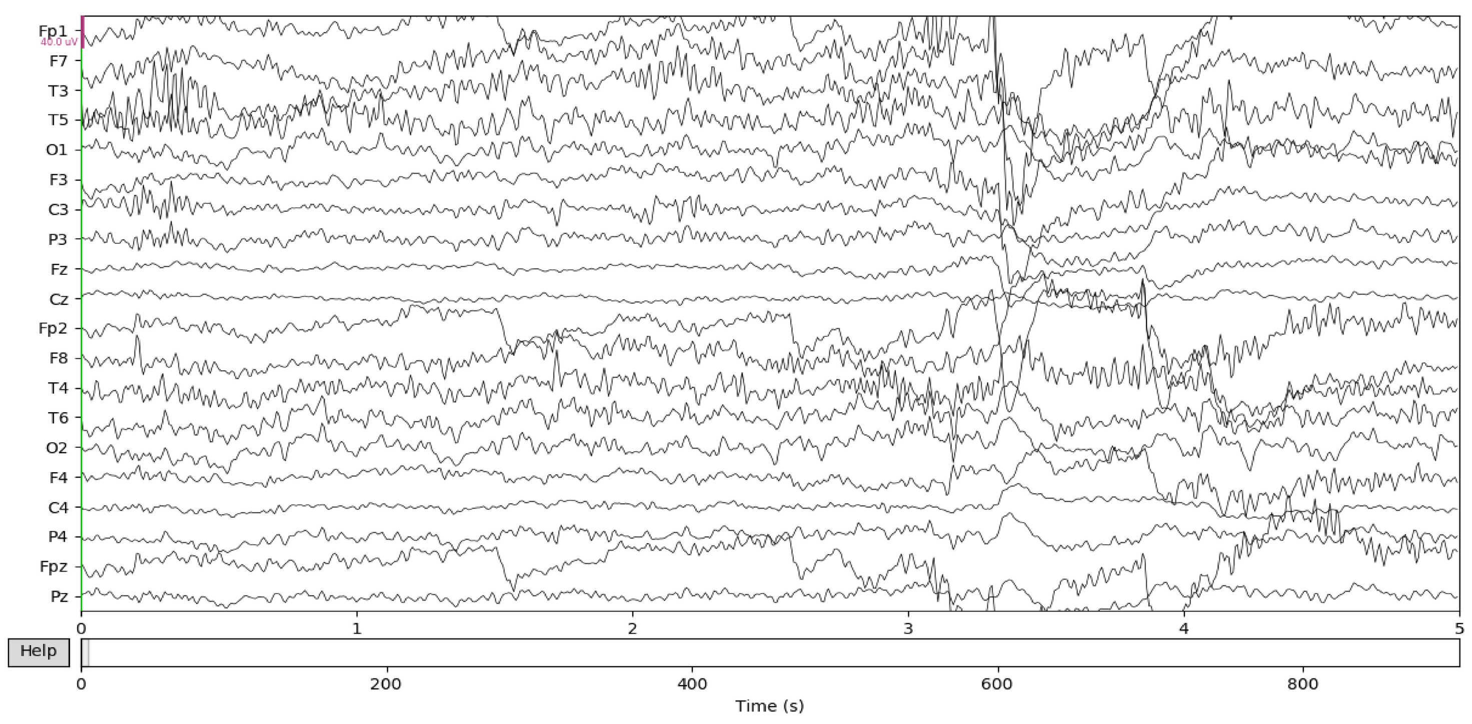
Hypothesis: machine learning is a viable option for the diagnosis of PNES using electroencephalogram (EEG) recordings during interictal recordings.

Introduction: Psychogenic non-epileptic seizures (PNES) are the result of a functional disorder that superficially resembles epilepsy but is not caused by epileptic activity in the brain[1]. PNES is as prevalent as multiple sclerosis[2, 3] and is, on average, misdiagnosed for seven years[4]. Hypothesis: machine learning is a viable option for the diagnosis of PNES using electroencephalogram (EEG) recordings during interictal recordings. Current gold standard methods of diagnosis rely on the recording of a seizure during EEG recordings, and the absence of ictal epileptic activity[5].





Preprocessing:
1. Interpolate bad channels
2. Filter – band-pass Hamming window, cut-off frequencies at 0.5 and 40 Hz
3. Segment into 10-s non-overlapping epochs, excluding labelled artefacts
4. Remove noisy epochs with the AutoReject python package
5. Baseline adjust using the mean



Materials: This study used seizure-free activity EEG recordings from seven patients with PNES (aged: 40.7±9.25 years, 2 male) and seven with epilepsy (aged: 44.4±7.69 years, 5 male). Each patient contributed ~15 minutes of EEG.

Method: As shown in the flow chart, the signals were filtered and split into 10-second, non-overlapping epochs. From which, various features selected using Random Forest and AdaBoost feature importances were extracted from each channel. A range of machine learning algorithms, listed in the legend of Figure 1, were compared. Due to the limited dataset, the data were separated into training and test sets using leave-one-subject-out and repeating for every subject. The parameters of the selected algorithms were then tuned using grid search with 10-fold cross-validation and principle component analysis was applied to the features. These fine-tuned models were then evaluated using the test subject. The evaluation metrics for each subject were averaged for

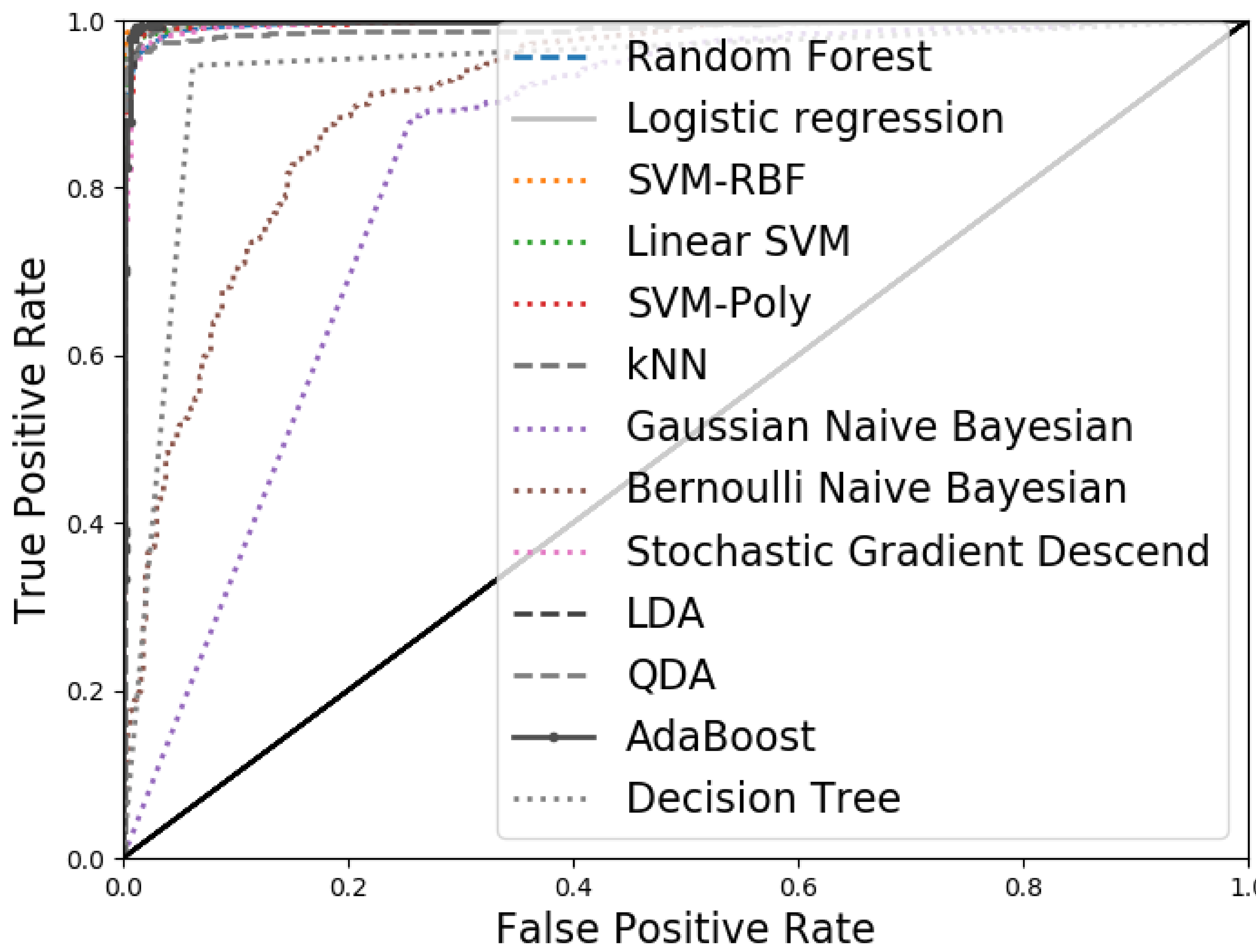


Figure 1 ROC-AUC plot comparing the machine learning algorithms before fine tuning

Extract features from every channel

Features from each channel:
- Energy of five frequency bands
- Higuchi fractal dimension
- Line length
- Spectral Hjorth complexity
- Number of maxima
- Lempel-Ziv complexity
- Correlation coefficients
- Skewness
- Hurst exponent
- Number of zero crossings

Separate six subjects for the train dataset

Use ten-fold cross-validation for a grid search to optimise the model parameters

Test fine-tuned models

Test the final model tuned using the test dataset, evaluate using the precision, recall, accuracy and F1-score

Separate one subject for the test dataset

Repeat, using a different subject for the test set every time, for every subject

Table 1 Average evaluation metrics from leave-one-subject-out fine-tuned models

Machine Learning Algorithm	Precision (%)	Recall (%)	Accuracy (%)	F1_score (%)
LDA	88.65	**87.92**	**88.42**	**88.08**
Logistic regression	86.73	85.13	86.10	85.76
SVM-RBF	**91.07**	80.29	86.06	85.23
Stochastic Gradient Descend	89.90	79.89	85.58	83.73
AdaBoost	81.37	69.10	76.84	74.67
Random Forest	86.43	63.88	77.01	72.85
kNN	73.62	65.73	70.86	69.00

Results: Initial experimentation with the whole dataset showed that both Naïve Bayesian methods and the decision tree were the weakest of all methods and were excluded. Leave-one-subject-out showed that LDA was the best classifier, with an average accuracy of 88.42%±0.10, followed by logistic regression, then SVM-RBF with 86.10%±0.08 and 86.06%±0.10 respectively, as seen in Table 1. Unfortunately, the dataset was too small to be certain of a superior technique but did show that machine learning can be a reliable tool.

Conclusion: This pilot study has shown that machine learning can be used for the differential diagnosis of PNES and epilepsy from EEGs alone, thus reducing the costs of the current gold standard method. However, with limited data and over-fit models, more investigation with a significantly larger number of subjects is required.

References

1. Brown RJ, Reuber M. Clin. Psychol. Rev. 2016; 47:55-70. 2. Benbadis SR, Hauser WA. Seizure. 2000; 9:280-281. 3. Bayly J et al. Epilepsia. 2013; 54(8):1402-1408. 4. Reuber M et al. Neurology. 2002; 58(3): 493-495. 5. Devinsky O, Gazzola D, LaFrance WC. Nat. Rev. Neurol. 2011; 7:210–220. Image credit: 10/20 system | Polysomnography Study Guide n.d. https://sleeptechstudy.wordpress.com/category/1020-system/ (accessed June 12, 2020).