 doi:10.20944/preprints202203.0145.v1

Article

Data-Driven EEG Band Discovery with Decision Trees

Shawhin Talebi *, John Waczak, Bharana Fernando, Arjun Sridhar, David J. Lary

Hanson Center for Space Sciences, University of Texas at Dallas, Richardson, TX 75080, USA
* Correspondence: shawhin.talebi@utdallas.edu

Abstract: Electroencephalography (EEG) is a brain imaging technique in which electrodes are placed on the scalp. EEG signals are commonly decomposed into frequency bands called delta, theta, alpha, and beta.While these bands have been shown to be useful for characterizing various brain states, their utility as a one-size-fits-all analysis tool remains unclear. We present a two-part data-driven methodology for objectively determining the best EEG bands for a given dataset in this paper. First, a decision tree is used to estimate the optimal frequency band boundaries for reproducing the signal's power spectrum for a predetermined number of bands. The optimal number of bands is then determined using an Akaike Information Criterion (AIC)-inspired quality score that balances goodness-of-fit with a small band count. Data-driven EEG band discovery may aid in objectively capturing key signal components and uncovering new indices of brain activity.

Keywords: Electroencephalography (EEG); EEG Bands; Decision Tree, Machine Learning

1. Introduction

The electrical activity produced by the brain was discovered by Richard Caton. Hans Berger later demonstrated that this activity could be recorded directly from the scalp [1]. This technique for measuring brain activity is called electroencephalography (EEG). It consists of an array of electrodes placed on the scalp that record fluctuations in electric potential arising from the activity of synchronized neural populations [2,3].

A popular method of analyzing EEG is spectral analysis. This consists of decomposing signals onto a frequency basis (Figure 1) and grouping frequencies into spectral bands (i.e. frequency ranges). Commonly used bands are: delta, theta, alpha, and beta [4]. EEG bands correspond to brain phenomena in specific brain areas and contexts. For example, alpha activity from occipital regions (i.e. visual cortex) in relaxed, awake animals track with eye closures [5]. During sleep, alpha band activity is observed at sleep onset, also called sleep spindles (7 – 14 Hz), and delta waves (1 – 4 Hz) appear in deep sleep stages [5].

Despite the widespread use of established spectral bands (e.g. delta, theta, alpha, and beta), there are two potential concerns with the current approach. First, there is significant variability in band boundaries across studies, as shown in Figure 2. This disagreement may be a result of a variety of factors such as hardware, filtering, and experimental task [6]. Second, ideal band definitions may depend on individual characteristics such as: age, genetics, personality, and task performance [7].

These concerns motivate the use of data-driven approaches for discovery of optimal EEG band boundaries. Such an approach tailors EEG bands to a specific experimental context and population in an automated way. Here, we present a method that makes use of decision trees, a popular machine learning framework. Optimal bands are inferred for an input EEG power spectrum. Through this method, suitable EEG bands can be derived in a flexible yet objective manner, which may provide informative and interpretable indices using EEG.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

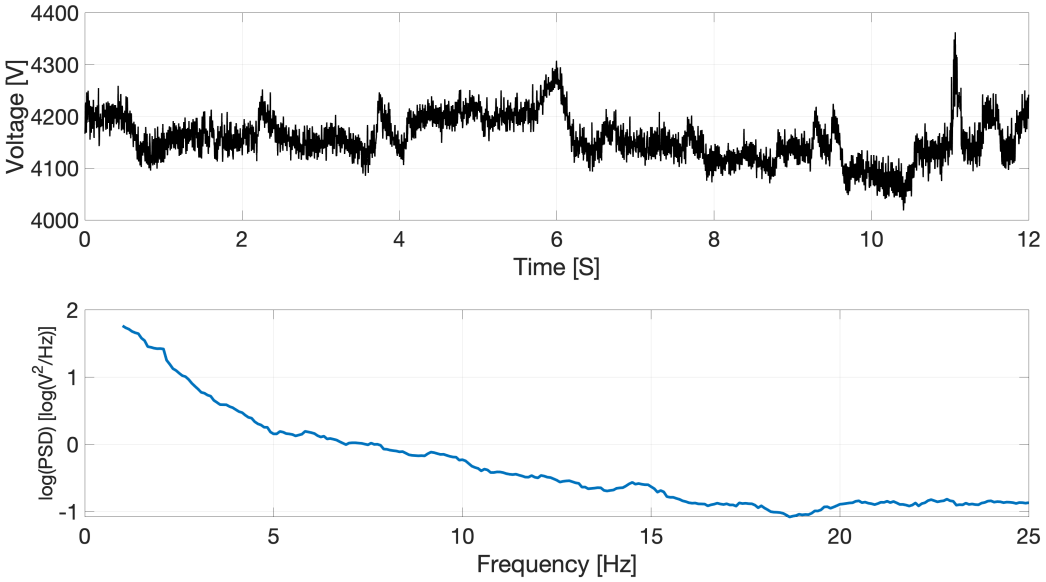

Figure 1. (**Top**) Example EEG time series signal. (**Bottom**) EEG signal's corresponding frequency spectrum, where the natural logarithm of the signal's power spectral density is plotted against frequency.

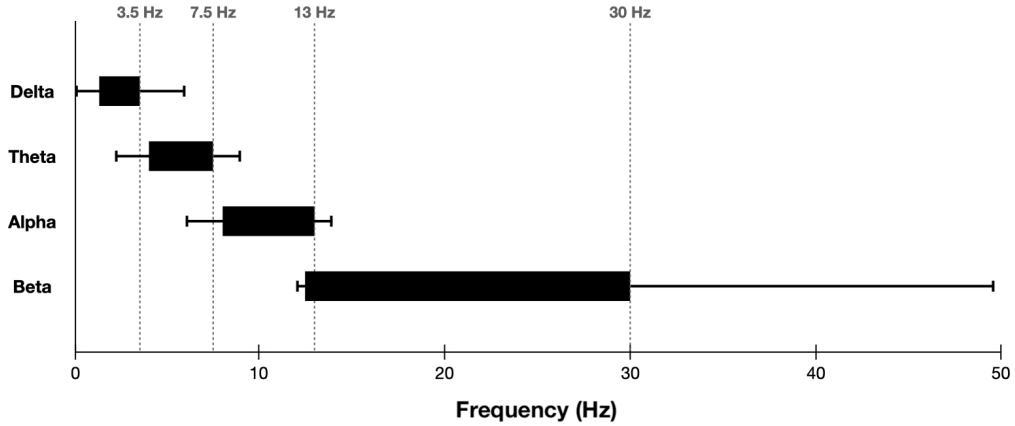

Figure 2. Box plot illustrating variability between EEG band boundaries across studies. Boxes indicate typical frequency range of each band. Whiskers represent smallest and largest band edges observed across studies. Plot adapted from figure in [6].

2. Methods

2.1. Intuition

This method generates optimal frequency bands based on an EEG signal's spectral information. The task is framed as a regression problem, and a decision tree is used to solve it. The final decision tree partitions frequency values into bins which produce the best estimation of the input signal's log power spectrum.

There are two main benefits to using a decision tree in this context. First, due to the structure of decision tree regression, frequency values are grouped into true bins. In other words, frequency values in a discovered band are adjacent, which may not be guaranteed by other regression techniques. Second, is ease-of-use. There are many efficient and ready-to-use implementations of decision tree optimization across many computational frameworks [8–11].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

2.2. Decision Trees

Decision trees are a widely-used and intuitive machine learning approach. Typically, they are used to solve prediction problems. That is, identifying a discrete target class (classification) or estimating a continuous target value (regression) from a set of predictor variables [12].

Data can be used to *grow* decision trees in an optimization process called training. Training requires a training dataset, which consists of predictor variables labeled with target values. A standard strategy for training a decision tree is recursively partitioning data via a greedy search method. The search determines the gain from each splitting option, and then chooses the one that provides the greatest gain [12,13]. Splitting options are the observed predictor variable values in the training dataset. Gain is determined by the split criterion e.g. Gini impurity or mean squared error (MSE).

For example, in a regression task, data records are recursively split into two groups such that the weighted average MSE of the target value is minimized from the resulting groups. This splitting procedure can continue until all data partitions are pure, meaning every data record in a given partition corresponds to a single target value. Although this implies decision trees can be perfect estimators, such an approach would result in overfitting. Therefore, the trained decision tree would not perform well on data sufficiently different than the training dataset.

One way to combat the overfitting problem is hyperparameter tuning. Hyperparameters are values that constrain the growth of a decision tree. Common decision tree hyperparameters are: maximum number of splits, minimum leaf size, and number of splitting variables. The key result of setting decision tree hyperparameters is to limit the tree's size, which can help avoid predictions only suitable to the training dataset. In this work, we use decision tree hyperparameters to control the number of discovered frequency bands.

2.3. Band Discovery with Decision Trees

Optimal EEG frequency bands can be estimated using the decision tree framework. Here, *optimal* means the frequency groupings that best reproduce an input signal's log power spectrum for a set number of bands. To achieve this goal, a decision tree is used to solve a regression problem in the usual way. A visual overview of the method is shown in Figure 3.

We use a single predictor variable (frequency) to estimate a single target variable (natural logarithm of the power spectral density). The decision tree then splits frequency values into subgroups and assigns each subgroup a single target value estimation. A greedy search of the decision tree parameter space yields frequency splits that best reproduce target values. Thus, through this optimization process we automatically obtain the optimal member-adjacent frequency bands for a predefined band count.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

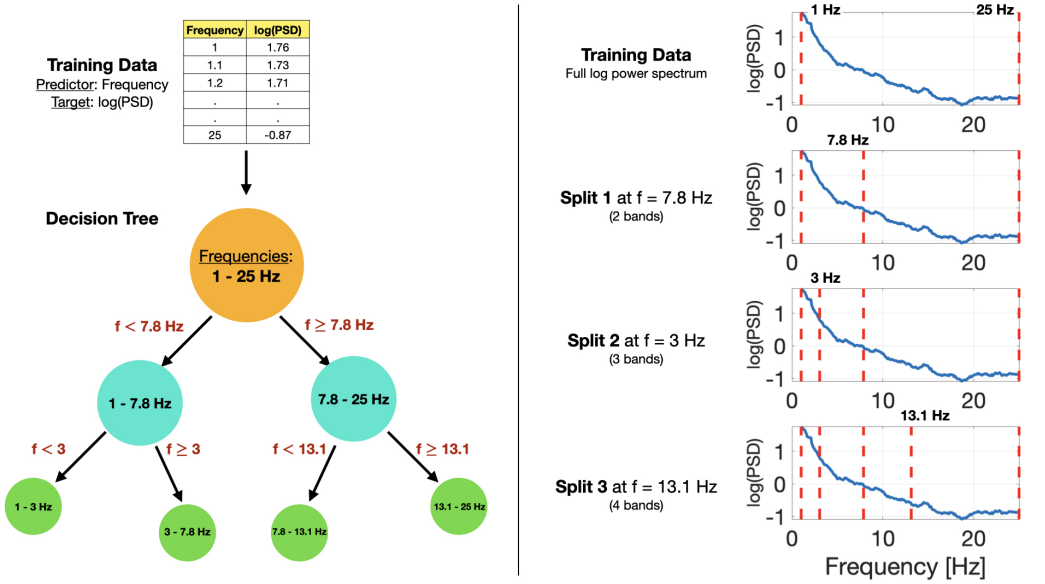

Figure 3. (**Left**) Visual summary of a decision tree partitioning frequency values based on the natural logarithm of the power spectral density. (**Right**) Visualization of decision tree splits of frequency values with power spectra.

2.4. Quality Score for Band Boundaries

Although decision tree optimization can be leveraged to identify optimal EEG frequency bands, this method requires the number of bands to be predetermined. Instead of choosing a band count manually, here we describe an objective data-driven strategy. The choice of band count is framed as an optimization problem, where we define an objective that can be optimized with respect to the band count.

One choice of objective is the r^2 regression score. In this context, the r^2 value corresponds to how well a set of decision tree derived EEG band boundaries reproduce an underlying power spectrum. While the decision tree optimization strategy described previously will ensure band boundaries are optimal for a given number of bands, different choices of band count will correspond to different r^2 values. An example of this is shown in Figure 4, where the r^2 regression scores of several different choices of band count are plotted for the same dataset.

However, the r^2 score is a problematic objective choice, since it strictly increases with the number of bands. Therefore, the maximum regression score would correspond to the largest possible number of bands i.e. a frequency "band" for every observed frequency value. One simple solution is to introduce an objective that incorporates both the r^2 regression score and a penalty for the number of bands. This is the goal of popular measures such as the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) [14]. Taking inspiration from AIC, we construct an empirically derived quality score (QS) to help choose a model that balances the best regression score while limiting the number of bands.

AIC is a measure of model quality, where smaller values imply better models [14,15]. It is defined in terms of the maximum value of the likelihood function for the model, L, and the number of parameters in the model, k.

$$AIC = -\log L^2 + 2k$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

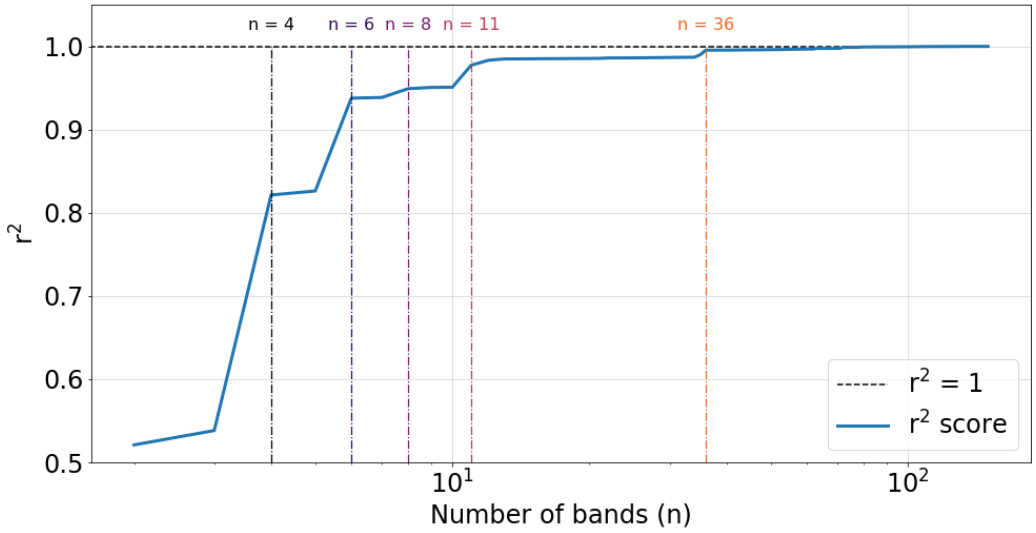

Figure 4. r^2 regression scores plotted against the number of frequency bands included in the decision tree model. The data used to derive these bands and r^2 values is the artificial data described in Case Study 1 in Section 3.1. Colored dashed vertical lines highlight large jumps in r^2 and are labeled by the corresponding number of bands.

The quality score (QS) we employ closely resembles AIC with two modifications. First, in lieu of the squared maximum likelihood value, we use the r^2 regression score. Since r^2 values are between $[0, 1]$, the first term in the QS equation below will be between $[0, \infty)$, however this range is not very large in practice e.g. for $r^2 \geq 0.135$, the first term is approximately between $[0, 2]$. Second, we divide the second term by N, where N is the maximum number of bands, or equivalently, the total number of observed frequency values. This ensures the second term in the equation below takes values in the range $[0, 2]$.

$$QS = -\log r^2 + 2k/N$$

QS provides a way to compare EEG band boundaries in way that accounts for both goodness-of-fit and band count. It will typically take values between 0 and 2, where smaller values correspond to better models. By computing the QS for every possible band count, we can choose the best EEG band boundaries as the choice with the smallest QS.

Although QS takes inspiration from AIC, a theoretically grounded quantity, it's derivation is empirical, therefore it may not be most suitable for all applications. Furthermore, there are countless other objective choices to optimize band count. The decision tree method described in Section 2.3 is independent of this band count optimization step, and thus can be enhanced by a variety of choices.

2.5. Software Implementation

This two-part technique is implemented using the Sci-Kit learn Python library, a popular and free machine learning software [8]. Our code is open-source and publicly available at the GitHub repository: https://github.com/mi3nts/decisiontreeBinning. Although Python is used for our implementation, other statistical software packages can be readily used to implement this method [9–11].

3. Results

In the following subsections we explore two case studies which apply this data-driven method for EEG frequency band discovery to an artificial and open-source experimental dataset, respectively. A Python script to reproduce both case studies are freely available at the GitHub repository: https://github.com/mi3nts/decisiontreeBinning.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

3.1. Case Study 1: Artificial Data

As a first demonstration of the method we produce an artificial EEG power spectrum as shown in Figure 5. The spectrum consists of the characteristic 1/f shape for EEG signals with added white noise.

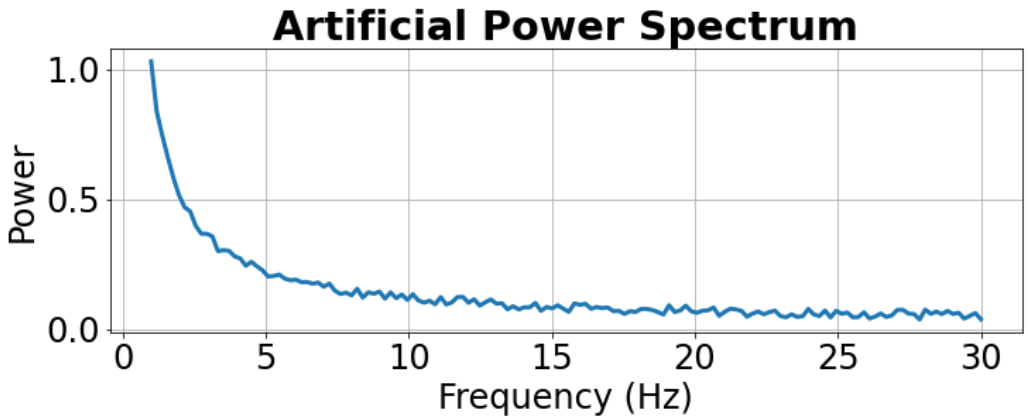

Figure 5. Artificial power spectra for initial demonstration of data-driven method. The spectrum consists of the characteristic 1/f curve for EEG signals with added white noise.

The results of applying decision tree based band discovery to the artificial power spectrum is depicted in Figure 6 for 5 different choices of band count. In each plot, the true power spectrum is shown as a solid blue line, the decision tree estimated spectrum is plotted as a dashed orange line, and the discovered band boundaries are indicated by dashed vertical red lines. The plots are titled according to the number of bands and r^2 (coefficient of determination) regression score. The r^2 score indicates how well the discovered bands reproduce the original spectrum. As a comparison, the typical boundaries of the delta, theta, alpha, and beta bands are shown at the bottom of Figure 6 [6]. The r^2 score of the standard bands is computed by comparing the average power value within each band with the true values.

The greedy search algorithm used in decision tree regression preserves band boundaries when new bands are added. In Figure 6, for example, 7.3 Hz is a band edge in every case (i.e. from 2 bands to 6 bands). It is interesting to note, the discovered 4 bands case is nearly identical to the typical delta, theta, alpha, and beta band boundaries according to [6]. Thus, it may be that the typical band boundaries are a good representation of this characteristic power spectrum.

In Figure 6, as more bands are added, the r^2 regression score increases. A diagrammatic representation of this observation is shown in Figure 4, where model regression scores are plotted against the number of bands. Since there are 150 unique frequency values in this first artificial dataset, the maximum number of bands is 150. Colored dashed vertical lines indicate band choices which exhibit a large jump in the r^2 score.

Since the r^2 score strictly increases with the number of bands, using it as an objective from which to choose the band count would always result in a “band" for every observed frequency value. However, the AIC-inspired quality score (QS) defined in Section 2.4 does not suffer from this issue. This is illustrated in Figure 7, which plots QS against the number of bands. A minimum value is observed at 6 bands, implying the best choice of band count for this spectrum is 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

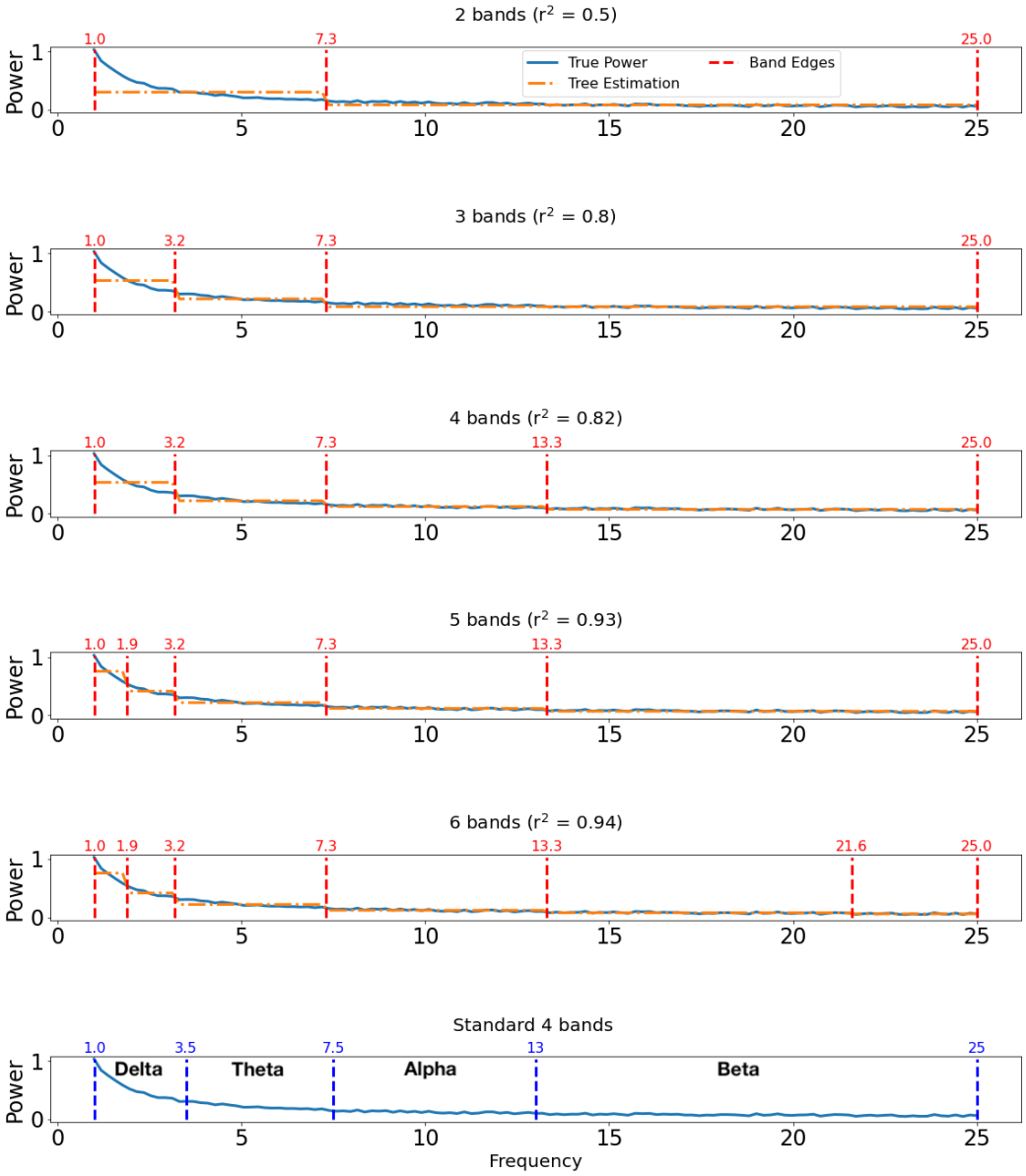

Figure 6. Band comparisons for artificial power spectrum. The true power spectra are plotted with solid blue lines, predicted spectra are plotted with dashed orange lines, and band boundaries are indicated by dashed vertical red lines. The plots are titled according to their number of bands and R^2 regression score. For comparison, typical values of the standard 4 bands (delta, theta, alpha, and beta) according to [6] are shown at in the bottom plot.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

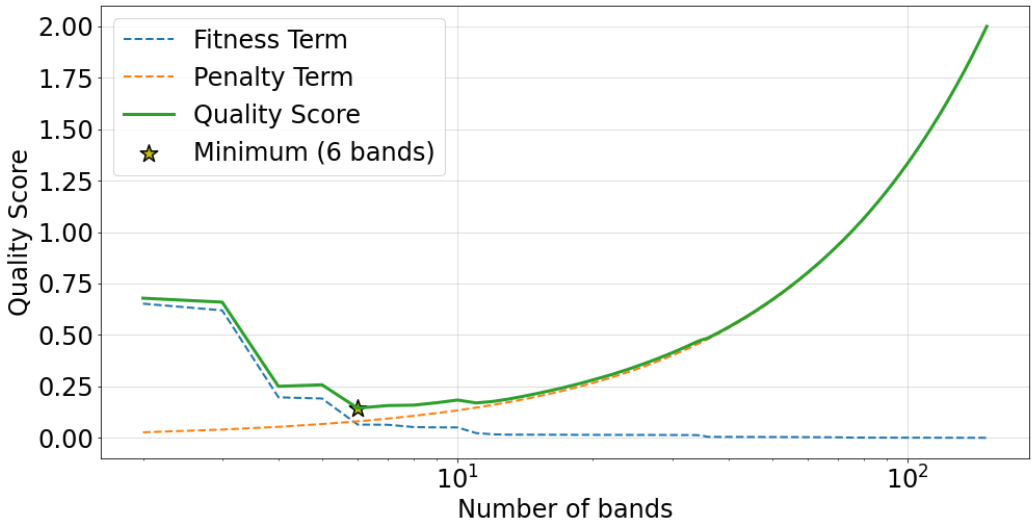

Figure 7. Empirically derived quality score (QS) plotted against number of bands for case study of an artificially generated power spectrum. The r^2-based fitness term in QS is shown as a dashed blue line, the band count penalty term is plotted as a dashed orange line, QS is plotted as a green line, and the minimum QS value is indicated by a yellow star.

The top plot in Figure 8 outlines the discovered bands employing a quality score (QS) minimization strategy. The plot title indicates the number of bands (6), r^2 regression score (0.94), and the quality score of the band definitions (0.14). The bottom plot in Figure 8 similarly outlines the standard bands, titled with the same metrics. The QS of the standard bands is computed using the QS equation in Section 2.4 with k=4. Although the discovered bands include more parameters, the QS is about half than that for the standard bands, thus it is a better characterization of the underlying spectrum based on this objective.

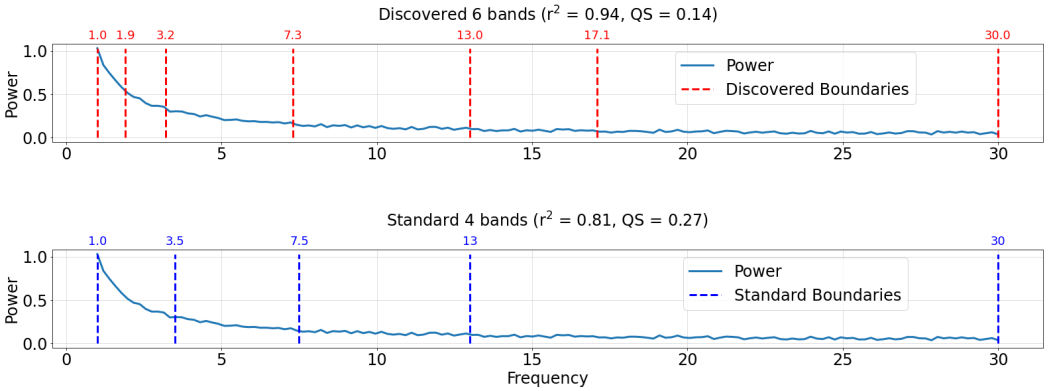

Figure 8. Comparison of discovered and standard bands for case study of an artificially generated power spectrum. Plots are titled by the number of bands, r^2 regression score, and the quality score of the respective band boundaries. True power spectrum is plotted as a solid blue line. (**Top**) Discovered bands using the proposed decision tree method employing a minimum quality score (QS) technique. Discovered band boundaries are indicated by dashed vertical red lines. (**Bottom**) Typical standard band boundaries taken from review by Newsom [6]. Standard band boundaries are indicated by dashed vertical dark blue lines.

3.2. Case Study 2: Experimental Data

We evaluate the band discovery method on experimental data from the PhysioNet dataset: EEG During Mental Arithmetic Tasks [16,17]. EEG data were collected monopolarly using the Neurocom EEG 23-channel system (Ukraine, XAI-MEDICA). The electrodes were placed on the scalp according to the International 10/20 montage. A 30 Hz cut-off frequency

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

high-pass filter and a 50 Hz power line notch filter were used. The data are artifact-free segments of 60 seconds. In preprocessing, Independent Component Analysis (ICA) was used to eliminate artifacts (eyes, muscles, and cardiac). For this case study, the baseline EEG recording from Subject 00 is used. Occipital electrodes (O1 and O2) are averaged to produce an aggregate occipital EEG signal.

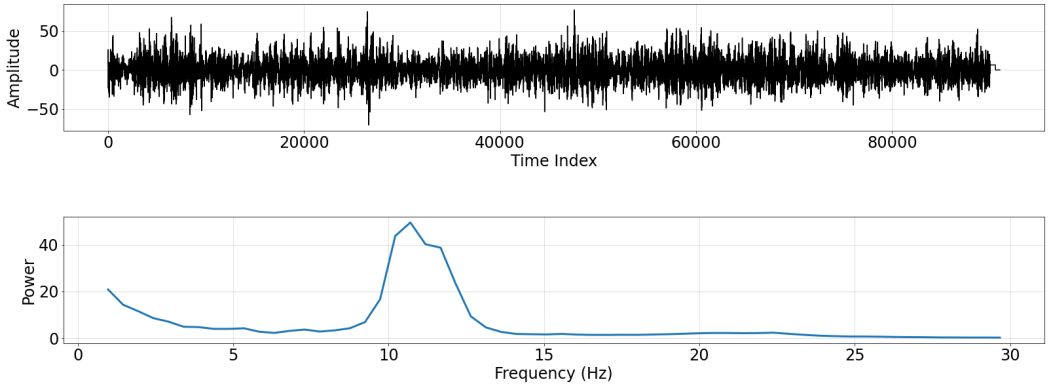

Figure 9. (**Top**) Time series of aggregated occipital EEG signal. (**Bottom**) Power spectral density plotted against frequency for aggregated occipital EEG signal plotted from approximately 1 – 30 Hz.

The aggregated occipital time series signal and its corresponding power spectrum are shown in Figure 9. Due to the signal preprocessing scheme used here, the power spectrum does not follow the typical $1/f$ spectrum. Nevertheless, an alpha rhythm peak is observed. Although this experimental power spectrum is characteristically different than the previous artificial spectrum, the EEG bands discovered by our data-driven will automatically adapt to it.

We repeat the two-part strategy from Case Study 1. First, we derive band boundaries using the proposed decision tree strategy for every possible choice of band count (i.e. 2 to 60 bands). Second, we use the quality score (QS) to identify the best number of bands. The quality score (QS) is plotted against the number of bands in Figure 10. The minimum QS value occurs for the 6 bands case.

Figure 11 compares the bands discovered by applying the proposed band discovery strategy to the experimental data (top plot), the optimal bands from Case Study 1 (middle plot), and the standard EEG band boundaries from [6] (bottom plot). The discovered bands from the experimental data (top plot) outperforms the other band choices, with both a significantly higher r^2 score and lower (better) QS. The poor performance of the discovered bands from Case Study 1 highlights the value of tailoring EEG bands to specific datasets. Additionally, the bands discovered from the experimental data isolate spectral features. For example, the peak in power spectral density between 10 and 12 Hz is partitioned into a dedicated band.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2022 doi:10.20944/preprints202203.0145.v1

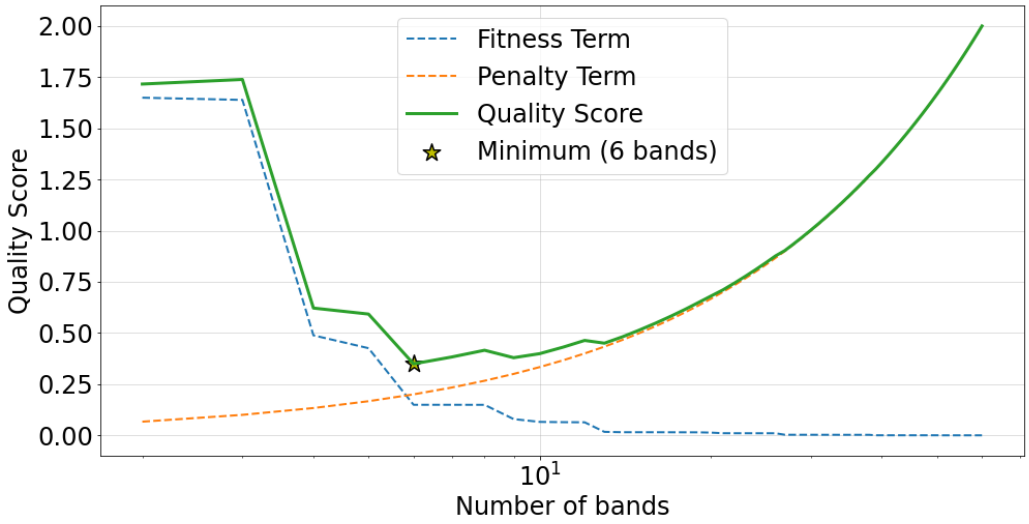

Figure 10. Empirically derived quality score (QS) plotted against number of bands for case study of experimental EEG data. The r^2-based fitness term in QS is shown as a dashed blue line, the band count penalty term is plotted as a dashed orange line, QS is plotted as a green line, and the minimum QS value is indicated by a yellow star.

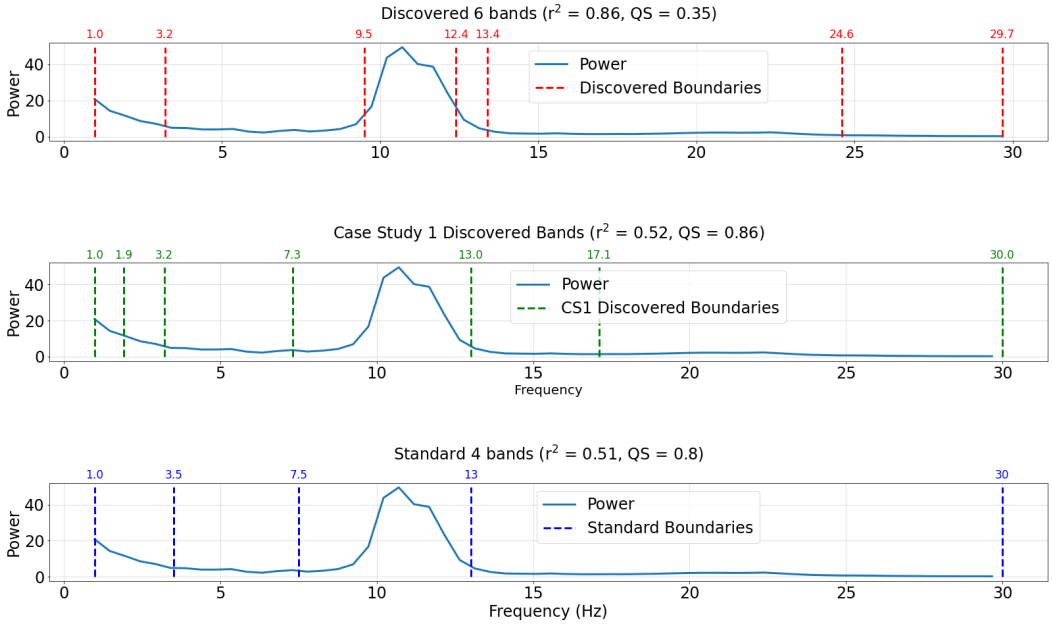

Figure 11. Comparison of discovered and standard bands for case study of experimental EEG data. Plots are titled by the r^2 regression score and the quality score of the respective band boundaries. True power spectrum is plotted as a solid blue line. (**Top**) Discovered bands using the proposed decision tree method employing a minimum quality score (QS) technique. Discovered band boundaries are indicated by dashed vertical red lines. (**Middle**) Discovered bands derived from artificial power spectrum in Case Study 1. Discovered band boundaries from Case Study 1 are indicated by dashed vertical green lines. (**Bottom**) Typical boundaries of standard bands are taken from review by Newsom [6]. Standard band boundaries are indicated by dashed vertical dark blue lines.

4. Discussion

The proposed method for automated EEG band discovery has two key strengths. First, the method provides a way to determine frequency bands that are representative of the underlying power spectrum while keeping the number of bands to a minimum. Second, the method is readily accessible since it is based on decision tree optimization,

 doi:10.20944/preprints202203.0145.v1

which has many efficient and ready-to-use implementations [8–11]. Additionally, we made our implementation of the technique open-source and publicly available (https://github.com/mi3nts/decisiontreeBinning).

5. Future Works

Since the presented approach is agnostic to *how* the power spectrum is generated, it can readily be applied to other power spectra (e.g. audio signals, hyperspectral imaging). Hyperspectral imaging, for instance, captures images with layers beyond the standard red, green, and blue. This provides a power spectrum for each pixel of a hyperspectral image. Using the proposed band discovery method, interesting spectral features in hyperspectral images can be detected in a self-supervised way.

Additionally, this method can be applied to other types of predictor variables. For example, using time as the predictor variable and a time varying quantity (e.g. heart rate) as the target, temporal epochs will be discovered, as opposed to frequency bands.

6. Conclusions

EEG serves as a window to underlying neural processes. Spectral analysis of EEG examines the oscillations in electric potentials arising from the brain. Despite the widespread use of established delta, theta, alpha, and beta bands for EEG, their boundaries vary widely across studies, which may be a result of variations in experimental details and participant differences. This motivates the use of objective and data-driven approaches to EEG band discovery.

In this work, we leveraged the readily available optimization of a decision tree for regression to discover EEG bands most appropriate for a given dataset and predetermined number of bands. The best choice of band count was then determined using an AIC-inspired quality score. We applied the presented method to both artificial and open-source experimental data. Discovered bands isolated spectral features into dedicated bands and outperformed the standard band definitions. Data-driven EEG band discovery may provide new indices of neural activity which can adapt to a variety of experimental and subject characteristics.

Author Contributions: Conceptualization, S.T. and J.W.; methodology, S.T., J.W, and D.J.L.; software, S.T.; validation, S.T. and J.W.; formal analysis, S.T..; investigation, S.T.; data curation, S.T.; writing—original draft preparation, S.T.; writing—review and editing, S.T., J.W., B.F., A.S., D.J.L.; visualization, S.T.; supervision, D.J.L.; project administration, S.T. and D.J.L.

Funding: This research was funded by the following grants: The US Army (Dense Urban Environment Dosimetry for Actionable Information and Recording Exposure, U.S. Army Medical Research Acquisition Activity, BAA CDMRP Grant Log #BA170483). EPA 16th Annual P3 Awards Grant Number 83996501, entitled Machine Learning Calibrated Low-Cost Sensing. The Texas National Security Network Excellence Fund award for Environmental Sensing Security Sentinels. SOFWERX award for Machine Learning for Robotic Teams. Support from the University of Texas at Dallas Office of Sponsored Programs, Dean of Natural Sciences and Mathematics, and Chair of the Physics Department are gratefully acknowledged. The authors acknowledge the OIT-Cyberinfrastructure Research Computing group at the University of Texas at Dallas and the TRECIS CC* Cyberteam (NSF 2019135) for providing HPC resources that contributed to this research (https://utdallas.edu/oit/departments/circ/, accessed 02/22/2022.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EEG	Electroencephalography
BIC	Bayesian Information Criterion
AIC	Akaike Information Criterion

References

1. Mulert, C.; Lemieux, L. EEG - fMRI: Physiological basis, technique, and applications. *EEG - fMRI: Physiological Basis, Technique, and Applications* **2010**, pp. 1–539. doi:10.1007/978-3-540-87919-0.
2. Jackson, A.F.; Bolger, D.J. The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. *Psychophysiology* **2014**, *51*, 1061–1071. doi:10.1111/psyp.12283.
3. Cohen, M.X. Where Does EEG Come From and What Does It Mean?, 2017. doi:10.1016/j.tins.2017.02.004.
4. Louis, E.K.S.; Frey, L.C. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*; 2016; pp. 1–95.
5. Da Silva, F.L. EEG: Origin and Measurement. *EEG - fMRI: Physiological Basis, Technique, and Applications* **2009**, pp. 19–38. doi:10.1007/978-3-540-87919-0_2.
6. Newson, J.J.; Thiagarajan, T.C. EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies. *Frontiers in Human Neuroscience* **2019**, *0*, 521. doi:10.3389/FNHUM.2018.00521.
7. Cohen, M.X. A data-driven method to identify frequency boundaries in multichannel electrophysiology data. *Journal of Neuroscience Methods* **2021**, *347*, 108949. doi:10.1016/J.JNEUMETH.2020.108949.
8. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
9. MathWorks. fitctree, 2021.
10. DataCamp. rpart: Recursive Partitioning and Regression Trees, 2021.
11. DecisionTree.jl Documentation, 2021.
12. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and regression trees*; 2017; pp. 1–358. doi:10.1201/9781315139470.
13. Kotsiantis, S.B. Decision trees: a recent overview. *Artificial Intelligence Review 2011 39:4* **2011**, *39*, 261–283. doi:10.1007/S10462-011-9272-4.
14. Salkind, N. Bayesian Information Criterion. *Encyclopedia of Measurement and Statistics* **2013**, pp. 1–3. doi:10.4135/9781412952644.n4
15. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* **1974**, *19*, 716–723. doi:10.1109/TAC.1974.1100705.
16. Zyma, I.; Tukaev, S.; Seleznov, I.; Kiyono, K.; Popov, A.; Chernykh, M.; Shpenkov, O. Electroencephalograms during Mental Arithmetic Task Performance. *Data 2019, Vol. 4, Page 14* **2019**, *4*, 14. doi:10.3390/DATA4010014.
17. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **2000**, *101*, e215–e220, [https://www.ahajournals.org/doi/pdf/10.1161/01 doi:10.1161/01.CIR.101.23.e215.