Application of Random Forest to classify EEG data of mTBI patients and control adults obtained during a Visuospatial Working Memory Task

Cruz, W.,[1] Cavanagh, J.F.,[2] Lin, C.Y.[1]

[1]*National ChengKung University, NCKU* [2]University of New Mexico, UNM

Introduction

The combination of electroencephalographic (EEG) recording and cognitive experimental tasks provides an excellent tool for studying human neural dynamics. EEG provides high temporal resolution time series data sampled across multiple scalp locations that produces large amounts of data, and some of these datasets are freely available in repositories on the internet (Fig. 1).

Figure 1. Open-science neuroinformatics database repository.

Most of the time the analysis of such data requires the implementation of many signal extraction methods that are further used to characterize the psycho physiological phenomena being studied. Sometimes the large amount of data collected for one experiment is not fully used because most of the researchers prefer to take a confirmatory and deductive approach, thus focusing their attention in a rather narrow number of features but unintentionally overlooking other important latent features. The utility of the random forest algorithm was investigated as a computational framework for extracting the most relevant features from an EEG data set obtained from an online repository (RRID: ds003523) with data of mild Traumatic Brain Injury (mTBI) patients and Healthy Controls (HC) during a Visuospatial Working Memory (VSWM) Task .

Random forest is a supervised ML method mostly used for classification purposes; the algorithm generates a set number of decision trees, each of which is made based on different subsets of data extracted from the training set. These subsets are selected following a random sampling approach; this iterative process and the number of decision trees computed are further used to reach a classification consensus, and the most common output is selected as the most relevant model which is usually the one that contains the most relevant features for the purpose of classifying the cases or instances being considered. Therefore it is possible to use EEG signal recording along with this classification framework to characterize the neural dynamics and predict both, performance and diagnosis.

Advantages of Analyzing EEG data with ML algorithms

Machine Learning techniques include different computational frameworks that are capable of mining large datasets, some advantages of using these frameworks when analyzing EEG datasets are:

- **Identify** relevant questions concerning EEG data.
- **Discover** new knowledge through pattern recognition and mathematical modeling.
- **Applicability** in both medical and social sciences.
- **Predict** the performance in a task.
- **Diagnose** patients based on their neural activity patterns.

Methods

The data used was obtained from an OpenNeuro repository but the subjects that conformed the final groups for the present analysis, mild Traumatic Brain Injury patients (mTBI, n = 27) and Healthy Controls (HC, n =27), were matched using demographic variables and their score in a Visuospatial Working Memory (VSWM) task, thus ensuring that both groups did not differ significantly by age (p = 0.67), sex (p = 0.58), nor by hit ratio (p= 0.97). The EEG epoched signals covering three memory phases (i.e. Baseline, Encoding, Retention) were analyzed to extract 5 frequency components from each of the 63 scalp sites. The EEG data was labelled by group and separated as either correct or incorrect for classification purposes.

Visuospatial Working Memory (VSWM) task

In the VSMW task, participants had to perform a yes-no recognition task (see Figure 2) and were asked to respond whether a location defined by a square containing a question mark (i.e. probe) had been occupied by a red dot in the preceding visual array. Participants were told to ignore the yellow dots. Altogether there were three conditions, either showing three targets or three red dots (*Condition 3*), showing three targets and two distractors (*Condition 3 + 2*), or showing 5 targets (*Condition 5*)

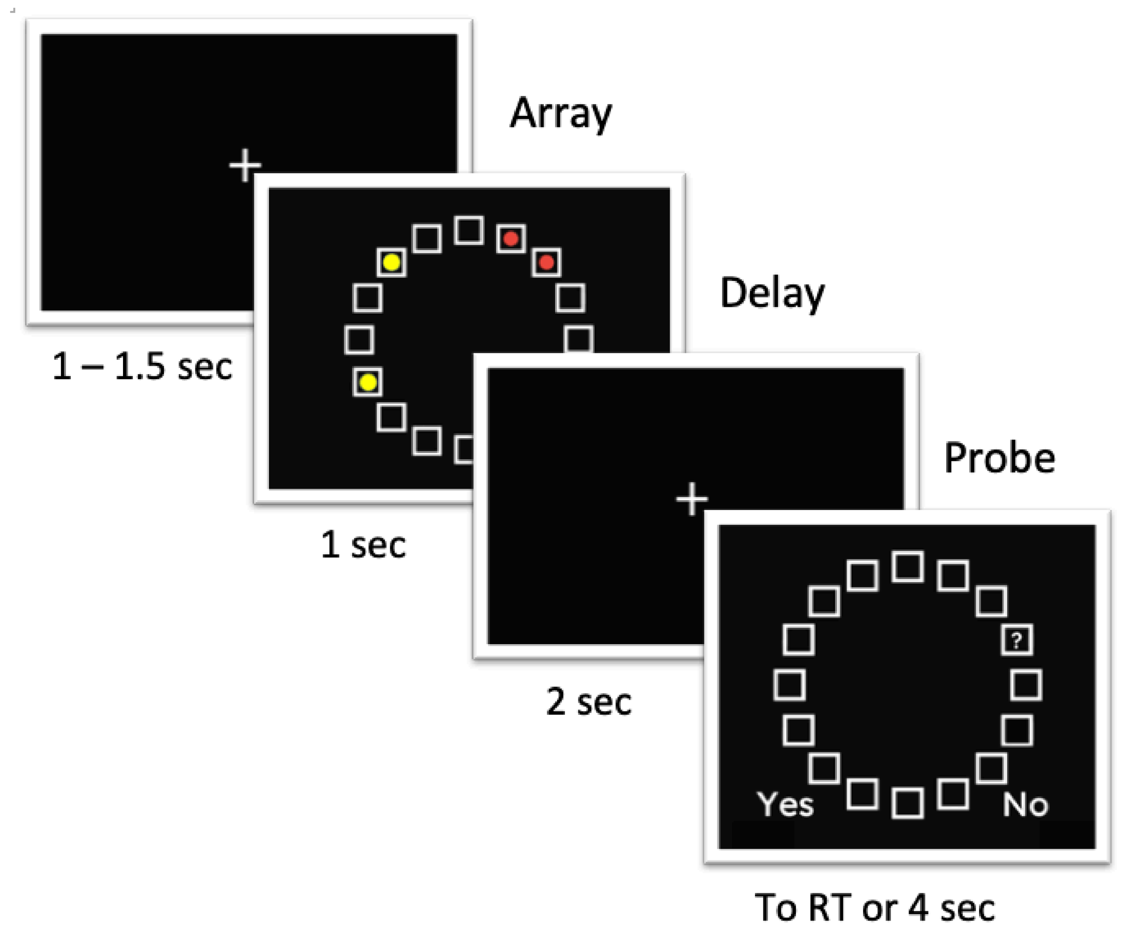

Figure 2. Visuospatial Working Memory (VSWM) task.

EEG data

EEG activity was recording using a 64 channel cap and the electrodes were located according to the 10-20 system. Additional electrodes were placed above and below the right orbit (VEOG) while simultaneously monitoring EKG activity. Continuous EEG was monitored and reference to the *Fz* channel and acquired at a 1000 Hz sampling rate. The VSWM task was administered using MATLAB. Total EEG set up time was in average 30 minutes, and the VSWM task included a practice block and an experimental block with 150 trials.

After the EEG signals were fully collected, the data of each subject was first loaded into EEGLab (Delorme and Makeig, 2004), then the EKG and VEOG were removed leaving only the channels that recorded brain activity. Following each subject's data was decomposed by ICA using the picard algorithm and the artifact components were removed with the SASICA module

- **Sed consequat** id ante vel efficitur. Praesent congue massa sed est scelerisque, elementum mollis augue iaculis.
  - In sed est finibus, vulputate nunc gravida, pulvinar lorem. In maximus nunc dolor, sed auctor eros porttitor quis.
  - Fusce ornare dignissim nisi. Nam sit amet risus vel lacus tempor tincidunt eu a arcu.
  - Donec rhoncus vestibulum erat, quis aliquam leo gravida egestas.

A block containing some math

Nullam non est elit. In eu ornare justo. Maecenas porttitor sodales lacus, ut cursus augue sodales ac.

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

Interdum et malesuada fames $\{1, 4, 9, \ldots\}$ ac ante ipsum primis in faucibus. Cras eleifend dolor eu nulla suscipit suscipit. Sed lobortis non felis id vulputate.

A heading inside a block

Praesent consectetur mi $x^2 + y^2$ metus, nec vestibulum justo viverra nec. Proin eget nulla pretium, egestas magna aliquam, mollis neque. Vivamus dictum $\mathbf{u}^\intercal\mathbf{v}$ sagittis odio, vel porta erat congue sed. Maecenas ut dolor quis arcu auctor porttitor.

Another heading inside a block

Sed augue erat, scelerisque a purus ultricies, placerat porttitor neque. Donec $P(y \mid x)$ fermentum consectetur $\nabla_x P(y \mid x)$ sapien sagittis egestas. Duis eget leo euismod nunc viverra imperdiet nec id justo.

Nullam vel erat at velit convallis laoreet

Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Phasellus libero enim, gravida sed erat sit amet, scelerisque congue diam. Fusce dapibus dui ut augue pulvinar iaculis.

First column	Second column	Third column	Fourth
Foo	13.37	384,394	α
Bar	2.17	1,392	β
Baz	3.14	83,742	δ
Qux	7.59	974	γ

Table 1. A table caption.

Donec quis posuere ligula. Nunc feugiat elit a mi malesuada consequat. Sed imperdiet augue ac nibh aliquet tristique. Aenean eu tortor vulputate, eleifend lorem in, dictum urna. Proin auctor ante in augue tincidunt tempor. Proin pellentesque vulputate odio, ac gravida nulla posuere efficitur. Aenean at velit vel dolor blandit molestie. Mauris laoreet commodo quam, non luctus nibh ullamcorper in. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos.

Nulla varius finibus volutpat. Mauris molestie lorem tincidunt, iaculis libero at, gravida ante. Phasellus at felis eu neque suscipit suscipit. Integer ullamcorper, dui nec pretium ornare, urna dolor consequat libero, in feugiat elit lorem euismod lacus. Pellentesque sit amet dolor mollis, auctor urna non, tempus sem.

References

Aust, F. and Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. R package version 0.1.0.9997.

Baddeley, A. D. and Hitch, G. (1974). Working memory. In *Psychology of learning and motivation*, volume 8, pages 47–89. Elsevier.

Barth, M. (2021). *tinylabels: Lightweight Variable Labels*. R package version 0.2.1.

Boonstra, T. W., Powell, T. Y., Mehrkanoon, S., and Breakspear, M. (2013). Effects of mnemonic load on cortical activity during visual working memory: linking ongoing brain activity with evoked responses. *International journal of psychophysiology*, 89(3):409–418.

Borchers, H. W. (2021). *pracma: Practical Numerical Math Functions*. R package version 2.3.3.

Brandt, E., Wilson, J. K., Rieger, R. E., Gill, D., Mayer, A. R., and Cavanagh, J. F. (2020). Respiratory sinus arrhythmia correlates with depressive symptoms following mild traumatic brain injury. *Journal of Psychophysiology*.

Broadway, J. M., Frank, M. J., and Cavanagh, J. F. (2018). Dopamine d2 agonist affects visuospatial working memory distractor interference depending on individual differences in baseline working memory span. *Cognitive, Affective, & Behavioral Neuroscience*, 18(3):509–520.

Broadway, J. M., Rieger, R. E., Campbell, R. A., Quinn, D. K., Mayer, A. R., Yeo, R. A., Wilson, J. K., Gill, D., Fratzke, V., and Cavanagh, J. F. (2019). Executive function predictors of delayed memory deficits after mild traumatic brain injury. *cortex*, 120:240–248.