1 Introduction

Transcranial magnetic stimulation (TMS) is a common, non-invasive experimental technique used to evoke action potentials in cortical regions of the brain. In particular, researchers often target the motor cortex and measure the motor evoked potential (MEP) aroused by the stimulation. When the purpose of the experiment is to inquire on neuroplasticity, such stimulations are performed under the *paired pulse paradigm*.

The *paired pulse paradigm* (or *double pulse paradigm*) consists in eliciting a series of two temporally proximate pulses (in the order of miliseconds). The evoked potentials of the double-stimulation are compared to those of single test stimulations, and their relative amplitude is taken as a proxy of neuroplasticity in the brain. The time separating each of the paired pulses is termed *interstimulus interval* (ISI). It is the general case that low intervals (4 or 5 miliseconds) produce intracortical inhibition, with the evoked potentials of paired stimulations being generally lower than those of single pulses. Greater intervals, on the other hand, tend to produce facilitation.

In the context of this paper, we shall term any coefficient that serves to represent the proportional relationship between the potentials evoked by paired and test pulses a *measure of relative amplitude*. Measures of relative amplitude in neuroscience are generally computed at the subject level. This is reasonable, since hypotheses generally deal with differences across subject groups. However, TMS experimental results are pulse-specific evoked potentials, and transforming them to subject- or group-specific measures implies down-scaling the data resolution.

The goal of this paper is to provide pulse-specific measures of relative amplitude. This is, coefficients that represent the relative amplitude of each individual paired pulse with respect to the set of test pulses in an experimental session. The purpose of this endeavor is to keep data resolution at its highest, which on its turn allows for the implementation of data-driven artificial intelligence in the analysis of the experimental results. Thus, from a computer science perspective, our objective is constrained to the sphere of feature engineering.

We will show how pulse-specific measures of relative amplitude allow for otherwise unfeasible computational analyses of TMS data, such as the use of machine learning models for the detection of different pulse-response patterns among different groups of clinical subjects. In particular, we will show they allow for a machine learning classifier to correctly determine whether a subject belongs to one of four clinical categories, based only on its evoked potentials and across different inter-stimulus intervals, with an accuracy of up to 90%.

2 Relative amplitude features

For simplicity, we will deal with the hypothetical situation in which a single ISI was used for paired stimulations. All of our results generalize to different inter-stimulus intervals.

2.1 Definitions

Let k be the number of experimental subjects in some subject group $\mathcal{G}$, to each of whom n paired stimulations and m test stimulations were elicited.

Definition 1 *Let $\boldsymbol{P}^{n\times k}$, $\boldsymbol{T}^{m\times k}$ be matrices representing the paired and test potentials evoked across each of the k subjects, such that*

$$\boldsymbol{P} := \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{T} := \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1k} \\ t_{21} & t_{22} & \cdots & t_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mk} \end{bmatrix}$$

2.2 The ρ and δ features

Definition 2 *Let $x \in \boldsymbol{P}_{\star i}$ be the MEP of a single paired stimulation elicited on the ith experimental subject, and $\boldsymbol{t} = \boldsymbol{T}_{\star i}$ the vector containing the MEPs of all test pulses elicited on that subject. Then we define two pulse-specific relative amplitude measures,*

$$\rho(x) := \frac{mx}{\sum_{i=1}^{m} t_j} \tag{1}$$

$$\delta(x) := \frac{x}{m} \sum_{j=1}^{m} \frac{1}{t_j} \tag{2}$$

Remark 1 $\forall x : x \in \mathbb{R}^+ : \delta(x) \geq \rho(x)$. *(For a proof of this property, consult the appendix.)*

Notice that $\rho(x)$ is the proportion between the potential x, evoked by a paired stimulation, with respect to the average potential of single test stimulations. On the other hand, $\delta(x)$ is the average proportion of x with respect to each single test pulse.

Each feature improves the performance of a random forests model in a similar degree, as will be shown later. This corroborates that both capture different but complementary information —as it is intended in their formulations. In particular, ρ can be a useful representation of the relative importance of each double pulse in relation to the overall distribution of the test pulses in the subject. It measures the deviation of the double pulse x with respect to the average test pulse. On the other hand, δ is a measure of the proportionality of a double pulse with respect to different values in the (certainly exponential) distribution of the test pulses in the subject.

Since the neural response to transcranial stimulations follows a distribution β very close to exponential (see **Statistics** section), we experimented with the inclusion of the features above applied to the logarithmically transformed MEPs. The inclusion of both $\rho(x), \rho(\ln(x))$ for every x improves the performance of a random forest model significantly, and the same occurs for δ, as can be seen in the **Empirical results** section.

It must be said that the ρ function is not an entirely new contribution. Traditionally, the group-level relative amplitude measure is conceived as the average, across all subjects in a group, of the average paired response divided by the average test response. This means that, if we let S_i be the average relative amplitude of the ith subject, then S_i has been traditionally defined as

$$S_i = \frac{\frac{1}{n}\sum_{j=1}^{n} x_{ji}}{\frac{1}{m}\sum_{j=1}^{m} t_{ji}} = \frac{\rho(x_{1i}) + \cdots + \rho(x_{ni})}{n}$$

as it is easy to see from decomposing the sum in the numerator. In other words, the traditionally used measure of relative amplitude, at the subject level, has always been the average ρ in a subject, albeit it was never defined explicitly.

2.3 The weighted variants ρ_w, δ_w

As stated earlier, action potentials evoked by transcranial magnetic stimulations follow a Gamma distribution that is very close to the exponential. (see **Statistics** section). Experimentation has shown that the inclusion of weighted variants of the features defined above greatly improves the performance of a random forests model (see **Empirical results**).

The use of weighted instead of arithmetic averages may be useful in dealing with the excessive influence of outliers or highly spread out points in the feature. For example, the weight vectors may be computed using the MAD or the inverse-variance of each x.

In our **Empirical results**, we show the influence of including inverse-variance variance weights to the model. But the general formulation of these alternative features, for any desired weight vector, is

$$\rho_w(x) := \frac{xm\sum_{j=1}^{m} w_j}{\sum_{j=1}^{m} t_j w_j} \tag{3}$$

$$\delta_w(x) := \frac{\frac{x}{m}\sum_{j=1}^{m} \frac{w_j}{t_j}}{\sum_{j=1}^{m} w_j} \tag{4}$$

where $\mathbf{w}$ is some appropriate weight vector.

3 Statistics

Writing note 1 *The description of the experimental process in this section is poor. Also I do not know the total number of subjects, experiments are still being conducted. Need input from the lab on these matters.*

We used data collected across N =? subjects at the *Laboratory for the Study of Slow-wave sleep activity*, University of Pennsylvania, with H =? healthy controls and D =? diagnosed with major depressive disorder (MDD). Transcranial magnetic stimulation of the motor cortex was elicited to them after a night of baseline sleep and after a night of slow-wave disruption (SWD) sleep. Motor evoked potentials were measured via an electrode (?) placed on the subjects' thumb (?). In the slow-wave disruption session, an auditory stimulus with sufficient strength to interrupt the normal occurrence of slow-wave sleep, yet not strong enough to wake the subjects, was elicited. This experimental setting produces four distinct categories, two depending on the subject group and two on the type of sleep session underwent, as shown in the table below.

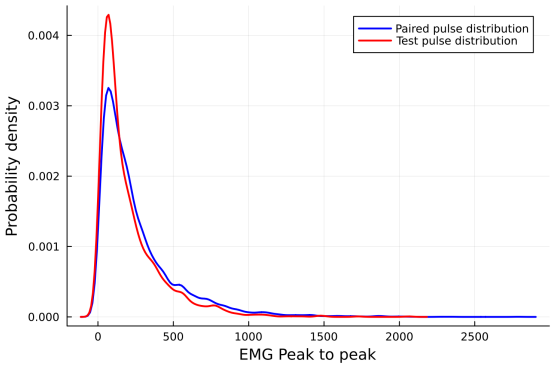
	Baseline	Slow-wave disruption
Healthy control	HC BL	HC SWD
Major depressive disorder	MDD BL	MDD SWD

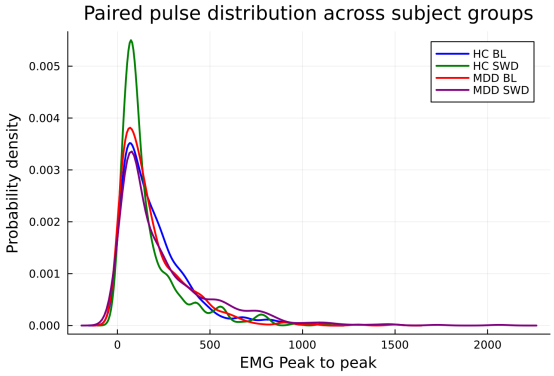
In our statistical analyses, we have contemplated test and paired pulses separately, since they are different type of stimulations. We have found the peak-to-peak EMG to follow an exponential distribution in both test and paired pulses. This is true when taking different inter-stimulus intervals in consideration as well as when observing the distribution of distinctly spaced paired pulses.

Although distributions were always exponential, the β parameter of said distributions varied across subject groups and session types.

	Mean	Median
HC BL	196.77	149.36
HC SWD	165.98	99.76
MDD BL	187.56	127.69
MDD SWD	247.57	151.07

Each of the means in the above table correspond to the estimated β parameter of the distribution of the paired pulses on each subject group.





Each observation in the data we used was a specific experimental observation resulting of an individual transcranial stimulation. The original features of the data were:

1. An *EMG* variable with the EMG peak-to-peak of each observation.
2. A *Label* categorical feature, which encoded the group of the subject of each observation, was the target variable.
3. An *ISI* feature that econded the inter-stimulus interval of each pulse. A value of -1 indicated that the given pulse was a test pulse (no inter-stimulus interval).

To these features, we included (separately and then together) the engineered features ρ and δ, which were computed for each observation using the Julia programming language. The code used to compute the features is publicly available (at ...).

4 Empirical results

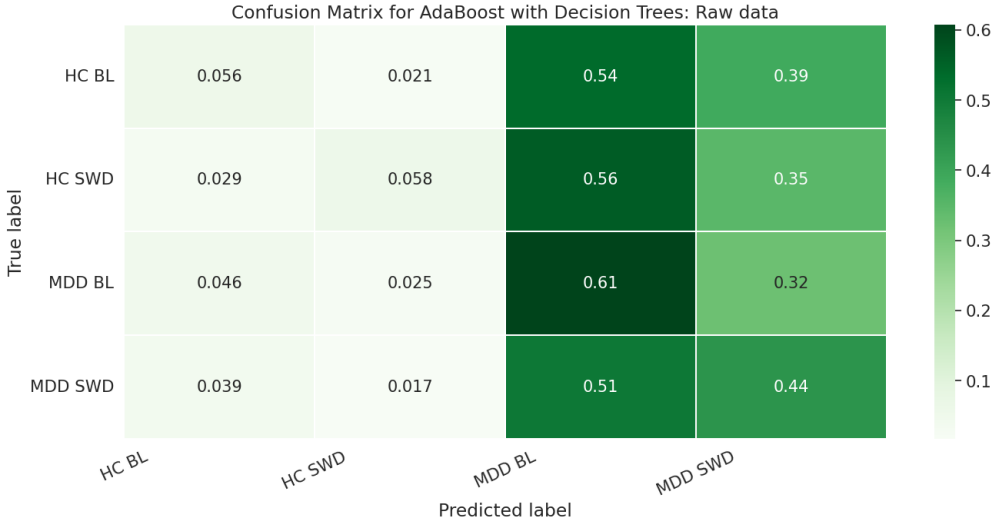
The objective was to evaluate whether the inclusion of the engineered, pulse-specific measures of relative amplitude improved the performance of a machine learning model, and in what degree. In order to do this, we set a random forests model with the task of classifying every observation with its appropriate label. In order words, the model was set out to infer, based on the properties of each transcranial stimulation response, the group of the subject upon which the stimulation was elicited, as well as the type of sleep session after which it occurred.

We trained the model first on the data without pulse-specific measures of relative amplitude (this is, with only the three features listed in the **Statistics** section). We then trained the exact same model on the data with the inclusion of δ but not ρ, and the inclusion of ρ but not δ. We then trained the same model with both engineered features, and lastly with both features and also their weighted variantes, using inverse-variance weights.

4.1 Raw data

When trained on the raw data, without the inclusion of pulse-specific relative amplitude measures, the model's accuracy was of $\approx 34.2\%$. However, the confusion of the model shows the errors are concentrated

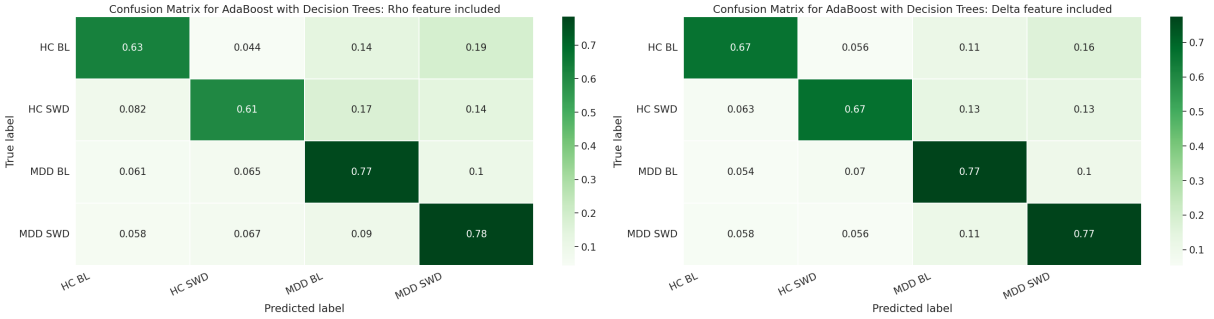
on the healthy control categories, while categorization of diagnosed subjects was substantially better. This implies major depressive subjects show statistically significant and distinct patterns in their TMS responses in comparison to healthy controls, and may be considered evidence for depression-induced differences in neuroplasticity.



However, it is important to note that the proxy for neuroplasticity in the double pulse paradigm is precisely the relative amplitude of double pulses with respect to test pulses. The raw data lacks any measure of relative amplitude. Besides, regardless of the fact that the model above suggests a critical difference in the response patterns between subject groups, its accuracy is still poor.

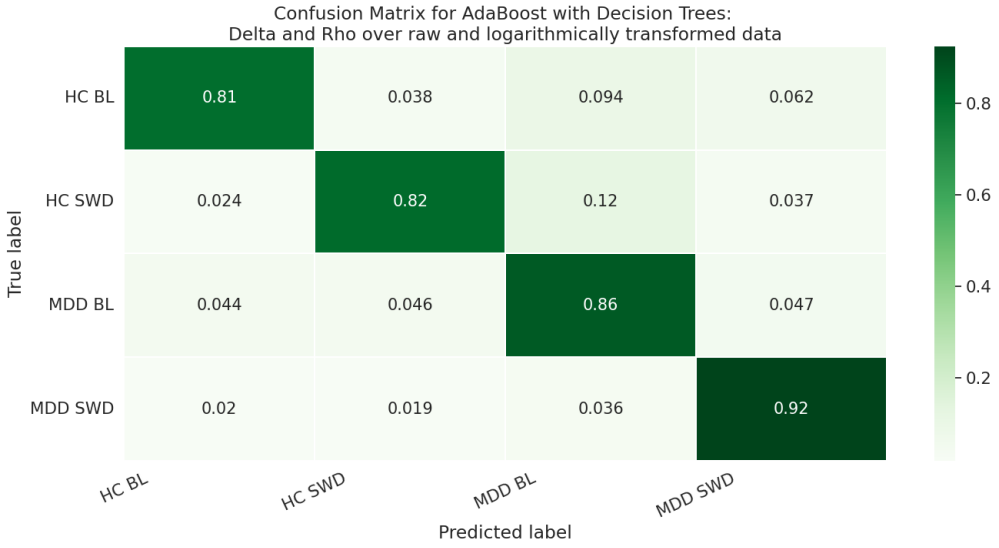
4.2 Engineered data

With the inclusion of the ρ feature, accuracy increases by a factor of ≈ 2.1 to 72.6%. The inclusion of the δ feature alone increased it, in comparison to the raw data, had a more or less equivalent impact, increasing accuracy to 73.4%.

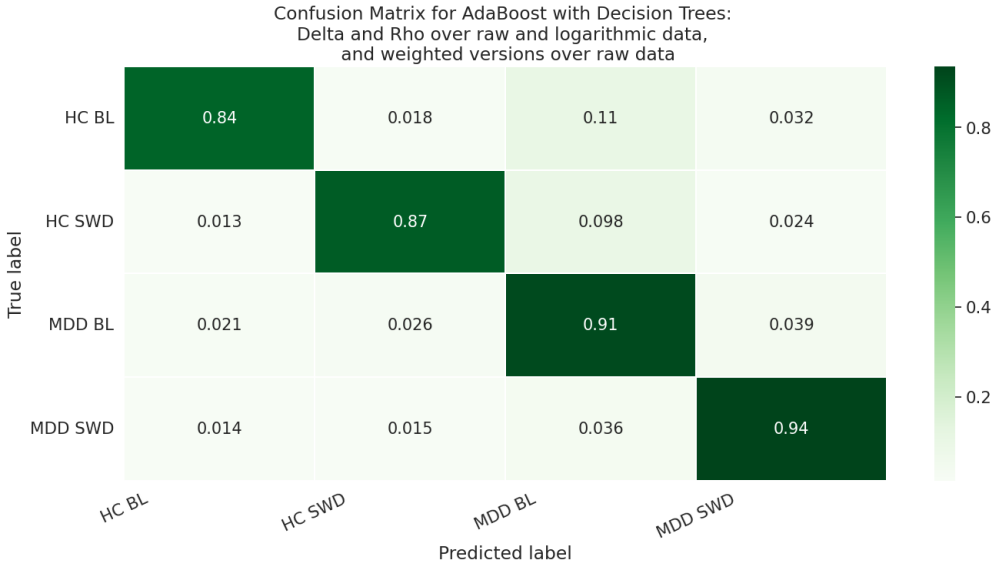


The model still shows a higher accuracy when classifying diagnosed subjects, but the overall accuracy across all categories was significantly improved.

As stated earlier, the inclusion of the δ and ρ features computed over the logarithmically transformed EMG peak-to-peak improved the model's accuracy. Concretely, accuracy was increased to 86%, with the following confusion matrix.



At last, if to the previous model we add also the weighted versions of δ and ρ, using inverse-variance weights, we obtained an accuracy of 89.8%, with the following confusion matrix.



5 Discussion

The previous results show that the inclusion of pulse-specific relative amplitude measures greatly improve the accuracy of a random forests model trained over TMS data for subject classification. More importantly, it becomes clear that, via the engineered features δ and ρ, researchers can attain evidence in favour or against hypotheses pertaining to differences among subject groups. Particularly, in showing that pulse-specific measures of relative amplitude are significant information to determine the sleep session and group of a subject, and being relative amplitude measures a proxy for neuroplasticity, the previous model provides evidence in favour of sleep-modulated, depression-induced differences in neuroplasticity.

The use of machine learning models over neuroscientific observations is not only a promising tool

in the production of evidence. There is a diagnostic potential that is still to be evaluated. Indeed, if machine learning models can detect the neural patterns that distinguish, under specific experimental conditions, healthy control from diagnosed subjects, such models can potentially be implemented in the diagnosis process as powerful clinical tool.

Although long and serious scientific effort is still required to appraise the diagnostic potential of machine learning models, we believe our results allow for a certain amount of conservative optimism on the matter.

In short, pulse-specific relative amplitude features are successful in making machine learning models applicable to TMS experimental data. Thus, they can play an important part in future research by allowing for new ways of analyzing and extracting meaningful information of TMS results.

6 Appendix

Proof 1. In **Remark 1**, we observed the following property:

$$\forall x : x \in \mathbb{R}^+ : \delta(x) \geq \rho(x).$$

Such property can be proven via induction. Firstly, recall that

$$\delta(x) = \frac{x}{m}\sum^{m}\frac{1}{t_j}$$
$$\rho(x) = \frac{xm}{\sum^{m} t_j}$$

Let $S_1^m = \sum^m \frac{1}{t_j}, S_2^m = \sum^m t_j$. We operate under the assumption that $t_i \in \mathbb{R}^+$. It is the case that

$$\frac{x}{m}\sum^{m}\frac{1}{t_j} \geq \frac{xm}{\sum^{m} t_j}$$
$$S_2^m S_1^m \geq m^2$$

This holds for $m = 1$, since $\frac{1}{t_1} + t_1 \geq 1 \iff 1 + t_1^2 \geq t_1$. So we may assume $S_1^k S_2^k \geq k^2$. We now set out to show that

$$S_1^{k+1} S_2^{k+1} \geq (k+1)^2$$

This can be proven as follows.

$$S_1^{k+1} S_2^{k+1} \geq (k+1)^2$$
$$(S_1^k + \frac{1}{t_{t+1}})(S_2^k + t_{k+1}) \geq k^2 + 2k + 1$$
$$S_1^k S_2^k + t_{k+1} S_1^k + \frac{1}{t_{k+1}} S_2^k + 1 \geq k^2 + 2k + 1$$
$$S_1^k S_2^k + t_{k+1} S_1^k + \frac{1}{t_{k+1}} S_2^k \geq k^2 + 2k$$

We know $S_1^k S_2^k \geq k^2$ and then it suffices to show $t_{k+1} S_1^k + \frac{S_2^k}{t_{k+1}} \geq 2k$. To prove this, simply observe that

$$\frac{1}{t_{k+1}}\sum_{j=1}^{m} t_j + t_{k+1}\sum_{j=1}^{m}\frac{1}{t_j} \geq 2k$$
$$\left(\frac{t_1}{t_{k+1}} + ... + \frac{t_k}{t_{k+1}}\right) + \left(\frac{t_{k+1}}{t_1} + ... + \frac{t_{k+1}}{t_k}\right) \geq 2k$$
$$\iff \overbrace{a + \frac{1}{a} + b + \frac{1}{b} + ... + n + \frac{1}{n}}^{2k \text{ terms}} \geq 2k$$

We have $\min f = 2$ for $f(x) = x + \frac{1}{x}$ for $x \in \mathbb{R}^+$. Then $\min(a + \frac{1}{a} + ... + n + \frac{1}{n}) = 2k$ for $a, ..., n \in \mathbb{R}^+$, which concludes the demonstration.