REAL-TIME APPLICATION OF MACHINE LEARNING ALGORITHMS TO ELECTROENCEPHALOGRAPHY FOR PREDICTING BRAIN STATES

Mikheil Oganesyan, Nickolai Vysokov, Anna Tarasenko, Illya Tarasenko

INTRODUCTION

Most existing methods of processing EEG data rely on carrying out artefact removal, pre-processing, feature generation and selection in a manual, offline fashion. Thus, any for of real-time closed-loop/neurofeedback application becomes difficult to carry out.

In this presentation, we report and demonstrate a sample pipeline for processing EEG data obtained from consumer-grade equipment and for predicting the state of a subject's brain in real time, with no human involvement.

The methods focus on maximising performance while minimising computational complexity, with a goal of porting the technology to low power embedded devices.

EXPERIMENT SETUP

For the LIVE DEMO, an OpenBCI EEG recording system is connected to BrainPatch electrodes consisting of saline-soaked foam with a layer of conductive silicone (see poster 760.28/L38).

Recording electrodes were positioned at O1 and O2 with reference at T3, and T4 was used as a bias. We have collected data from healthy volunteers where they were asked to keep their eyes open for 30s and then closed for 30s. The features (band powers) were extracted and used to train the classifier (see pipeline). The outcome of the training was the ability to reliably predict the state of the visual system as demonstrated by the probability plot on the next slide.

For experiments with concentration, OpenBCI EEG system with dry electrodes using 8 channels (Fp1, Fp2, C3, C4, P7, P8, O1, O2) was used to collect brain signals. Earlobes were used as bias and reference. Healthy volunteers were asked to sit idle for 1 min and then perform a memory task (find matching pairs) for 1 minute. The features were extracted and used to train the classifier (see pipeline). The outcome of this was the ability to predict brain with moderate reliability (see Results tab).

PIPELINE DESCRIPTION

Preprocessing Stage:

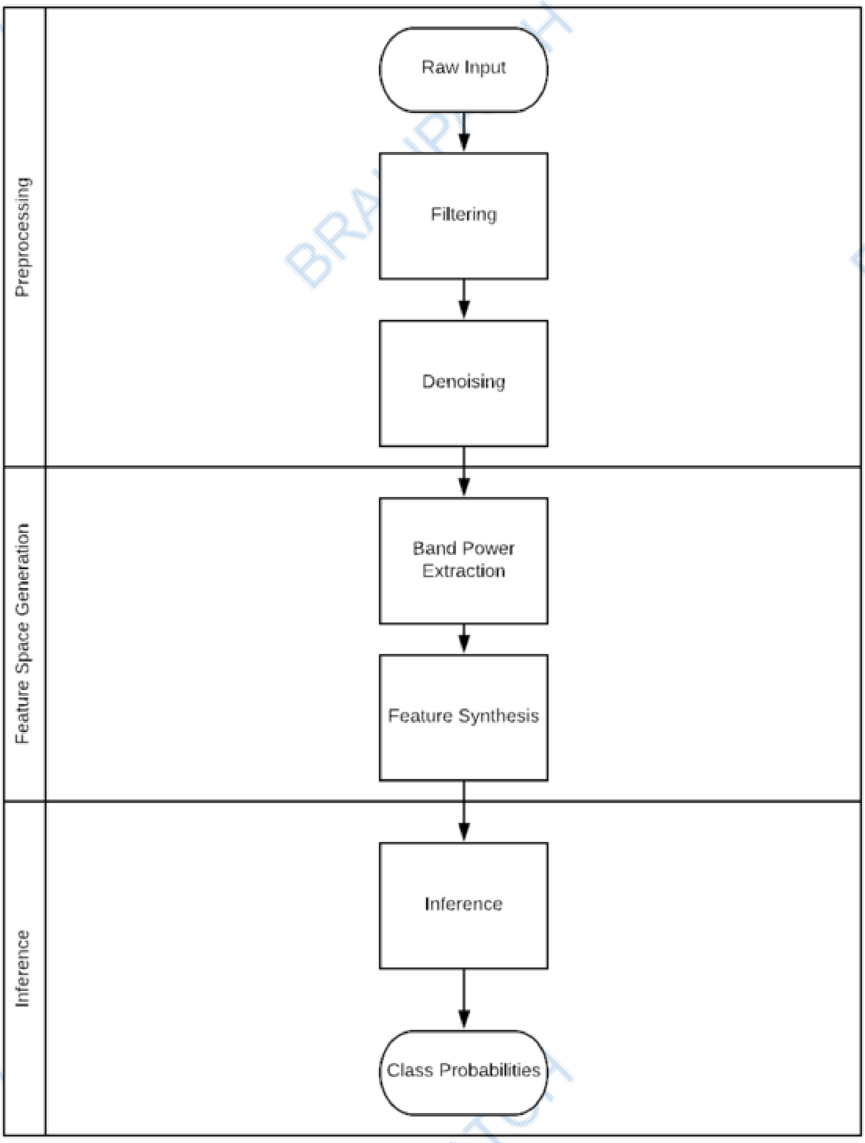
- The signal is bandpass filtered between 1Hz and 97Hz to remove the DC offset and high level harmonics of the power line.
- The notch filter is applied around 60Hz to remove the power line noise.
- Artefacts such as blinks can be optionally removed
- Dynamic median filter can be optionally applied for removing spike noise (eg: electrode popping). The median filter is applied in a dynamic fashion to only remove artefacts with abnormally high power, thus making it ideal for removing spike and electrode popping artefacts.

Feature Space Generation Stage:

- Theta (4-7Hz), Alpha (8-14Hz), Beta (15-32Hz), Gamma (32-90Hz) bands are extracted using bandpass filters from a windowed signal chunk of 1000 samples in length (sampled at 250Hz). For the live processing task, a sliding window is utilised.
- Log-variance of each extracted component is calculated to yield band power.
- Various combination features such as Alpha/Gamma ratio, as well as Beta/(Alpha + Theta) ratio are calculated.

Inference:

Inference was done using a random forest classifier. Random forest algorithm works by 'bagging' outputs of many decision trees, thus making it highly generalisable. Being based on decision trees, the model makes it possible to gain insight into preferred features and hence the underlying physiological prinicples of the task at hand. Finally, being a light weight model, it can be easily ported to embedded devices.

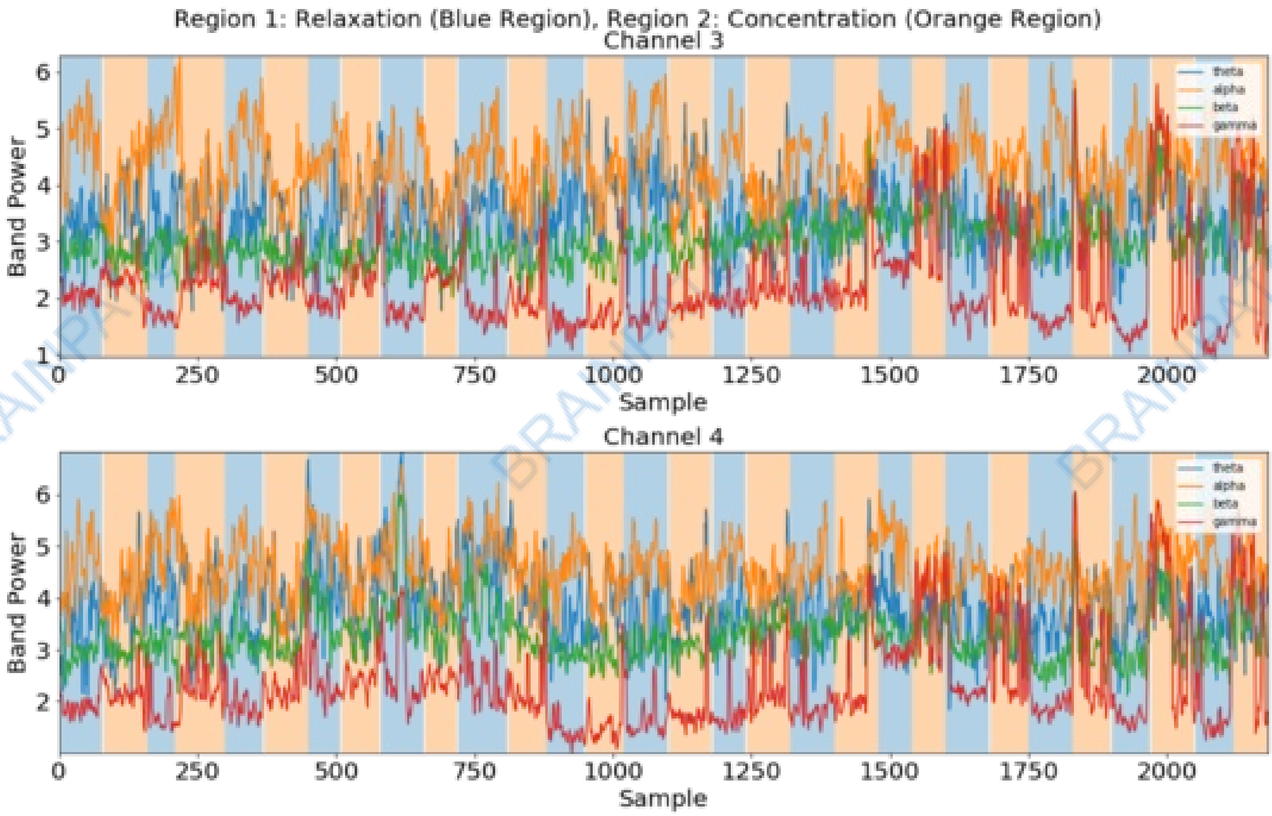


BAND POWERS

The feature space based on two sample channels is shown below. Notice the considerable increase in gamma activity during period of high concentration (shaded orange regions) compared to idleness (shaded blue regions). Furthermore, there is a visible and consistent reduction in alpha power during periods of high concentration.

However, there are also more subtle changes in the beta and theta band activity which are not visible to the naked eye in the plot, especially compared to the alpha/gamma activities. Research has shown that the ratio of beta/(alpha + theta) was found to be a good indicator of concentration.

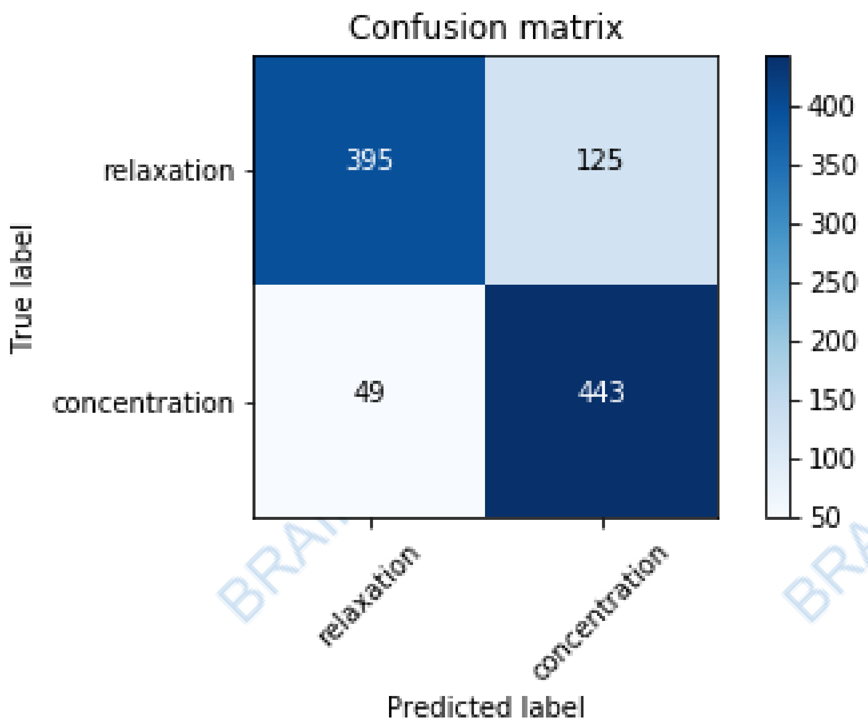
The above findings were confirmed in our experiments with the aforementioned ratios being hevaily favoured by the Random Forest classifier when making the decision.

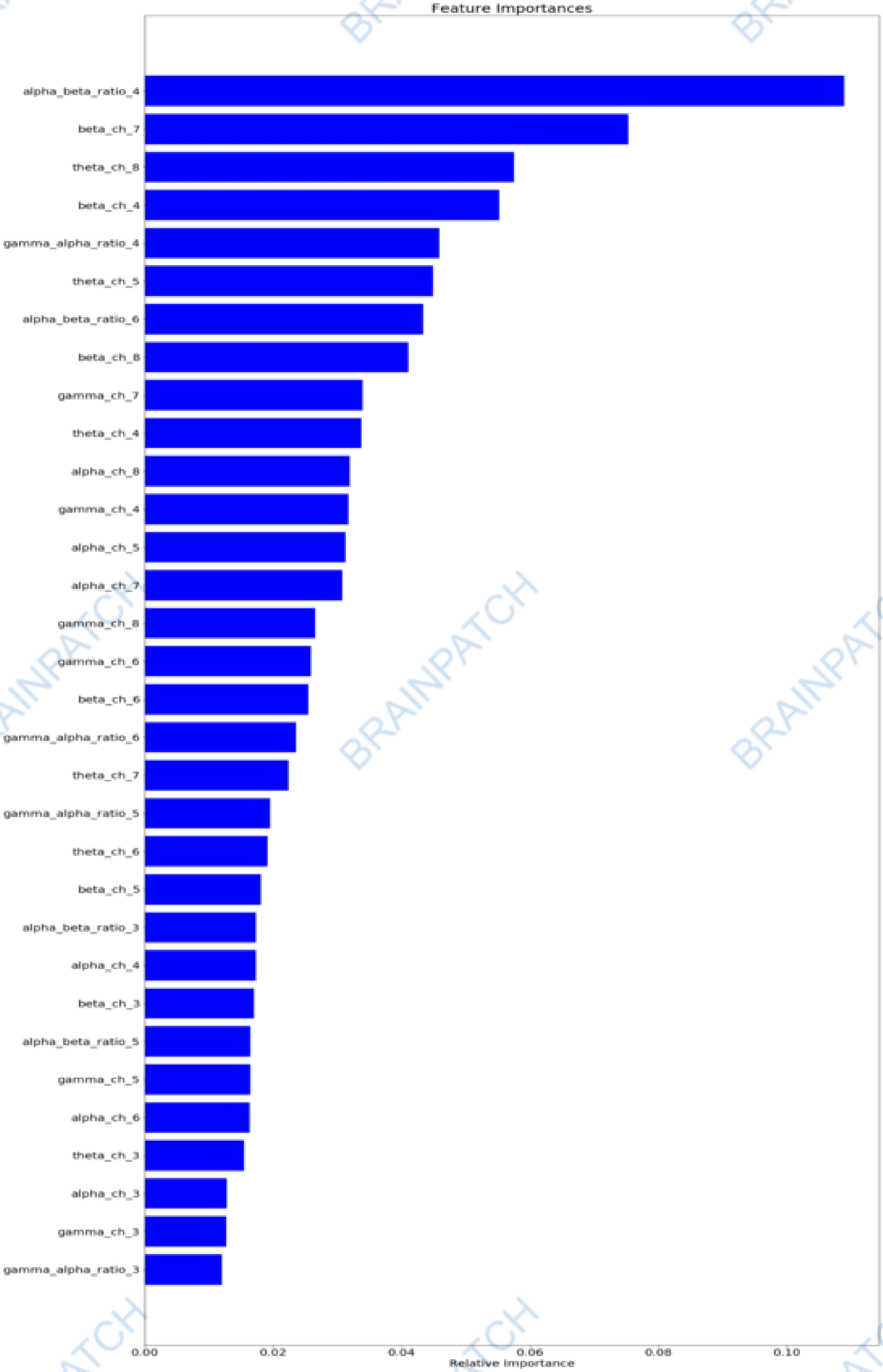


CLASSIFIER PERFORMANCE

The classifier was evaluated on a withheld test set. In order to ensure the highest level of rigour, the test set was not subsampled from the training set. Instead, it was recorded in separate trials, sampled from 5 volunteers.

The performance of the classifier reached an F1 score of 0.83 and an accuracy score 0.82. The confusion matrix below shows the tendency of the model to give false positive results. This is most likely due to wildly differing neural behaviour patterns among different individuals. Training a fully generalised model may come with a significant cost in accuracy, thus favouring models customised to individual users in commercial application.




Feature Importances
alpha_beta_ratio_4
beta_ch_7
theta_ch_8
beta_ch_4
gamma_alpha_ratio_4
theta_ch_5
alpha_beta_ratio_6
beta_ch_8
gamma_ch_7
theta_ch_4
alpha_ch_8
gamma_ch_4
alpha_ch_5
alpha_ch_7
gamma_ch_8
gamma_ch_6
beta_ch_6
gamma_alpha_ratio_6
theta_ch_7
gamma_alpha_ratio_5
theta_ch_6
beta_ch_5
alpha_beta_ratio_3
alpha_ch_4
beta_ch_3
alpha_beta_ratio_5
gamma_ch_5
alpha_ch_6
theta_ch_3
alpha_ch_3
gamma_ch_3
gamma_alpha_ratio_3
0.00
0.02
0.04
0.06
0.08
0.10
Relative importance