PART III RESEARCH PROJECT

EEG BASED DETECTION OF DRIVER DROWSINESS USING MACHINE LEARNING

May 13, 2019

Candidate number: 8221T
University of Cambridge
Department of Physics

SUPERVISED BY DR TRISTAN BEKINSCHTEIN

Abstract

The goal of this project is to serve as a proof of concept of the feasibility of tracking variations in alertness from information extracted from EEG signals. In particular, EEG data was recorded from 20 subjects performing a driving simulation task. The data was pre-processed and then epoched into 45s segments, with the degree of drowsiness quantified by an appropriate behavioural index, namely the average reaction time within the epoch (local mean reaction time). Subsequently, a total of 67 different features were extracted for each epoch from four distinct sectors - eye blinks, the power spectral density, entropy and inter-electrode coherences. Significant correlations between individual features and the behavioural index of driving performance is shown to be ubiquitous in the different sectors, demonstrating the direct link between fluctuations in vigilance and the dynamics of brain activity. Through feature dimensionality reduction techniques it is further demonstrated that features in each sector exhibited significant contributions to the final predictive model, with those of the blinks being the most prominent. Furthermore, an SVM regression predictive model was developed using a Gaussian kernel and deployed on the datasets that were suitable for analysis, yielding a net RMSE of $0.24 \pm 0.05s$. The work of this project has direct use in Brain Computer Interface applications for use as a driver drowsiness warning system.

Contents

1 Introduction

Driver drowsiness is a major issue and is believed to be a primary causal factor in automotive related accidents. Consequently, there has been a recent surge of research into BCI (Brain Computer Interface) technology, applications of which include warning systems that are designed to prevent such accidents (Lin et al., 2010). An increase in drowsiness is characterised by the decrease in a person's information processing speed and memory capacities (Eoh et al., 2005), such reductions in alertness can be caused by monotonous yet attention-demanding activities (Jun at al., 2005). There are numerous methods which may be used in determining driver drowsiness, including eye tracking via computer vision systems (Devi, Bajaj, 2008), Electrocardiography (ECP) and methods that employ other biomedical signal processing. The most prominent method and the one that will be employed in this project is Electroencephalography (EEG), a non-invasive electrophysiological monitoring method that records the electrical activity of the brain.

Despite the efficacy of EEG in analysing brain activity, contaminations and artifacts are ubiquitous in EEG data. These artifacts can arise from both internal sources such as eye movement, blinks and muscle noise, and external sources such as line noise and channel noise. Consequently, rigorous pre-processing techniques are required to extract the component of the signal derived from brain activity. In the field of cognitive neuroscience there are discrete frequency bands canonically defined as delta [0.5 – 4 Hz], theta [4 – 8 Hz], alpha [8 – 13 Hz], beta [13 – 20 Hz] and gamma [13 – 40 Hz][1]. Changes in the distribution of power contained within these frequency bands have been shown to be a manifestations of changes in states of consciousness. Therefore, we can characterise transitions in drowsiness by identifying the corresponding shifts in the power contained within and between these bands. To be able have an objective measure of variations in alertness requires the definition of a metric (or gold standard) that quantifies a level of alertness, the metric that will be used in this paper is reaction times. However, the use of reaction times as a gold standard has its limitations. Firstly, inter – subject variations restrict its use to a relative measure of alertness, rather than absolute, meaning that comparisons of drowsiness between subjects is uninformative. Secondly, if we assume there is a one to one mapping from a cognitive state (i.e. state of arousal that is well defined by the brain state) to a probability distribution function that governs the spread of reaction times, then the sparsity of the sampling means that reaction times as a metric are at best, an approximation of drowsiness (discussed further in section 2.3).

Since its emergence as a learning method Support Vector Machine's (SVM) have been used extensively in EEG related studies. They have predominantly been used for classification tasks, such as labelling single trials with a discrete degree of drowsiness. For example, in Yeo et al. (2009), two of the authors who were trained in interpreting the EEG signal classed individual trials as alert, drowsy or asleep. Using feature extraction in the frequency domain in conjunction with an SVM classifier allowed them to reach classification accuracies of 99.3%. The attraction of biomedical signal processing studies to SVM lies in its generalisation capacity independent of input space dimension (Cristianini and Shawe-Taylor, 2000). This is especially useful in EEG data where independently derived features are often redundant due to the redundancy that is inherent in the data [1].

In this project, SVM regression together with a range of features extracted from both the frequency

[1] Note that the definition of gamma in this project deviates from the canonical definition.

and time domain of the multi-channel EEG data are used to predict reaction times.

2 Analysis

2.1 Outline

In this study, raw EEG data was acquired from the Department of Experimental Psychology of the University of Cambridge. The data consisted of 20 subjects each performing an hour-long driving simulation in which they controlled the position of the car using a joystick whilst 64 electrodes were attached to their scalp. Throughout the experiment white blocks would appear on the screen in quasi-regular time intervals, the white block would then move across the screen until the subject pressed a button on the joystick to confirm they had registered the white block. Both of these types of events were marked on the EEG data such that their reaction times could be calculated. In addition to the reaction propagators, every 5 minutes the subject received a questionnaire based on the Karolinska Sleepiness Scale [2] which spans 9 levels, from 1 – extremely alert, to 9 – very sleepy. The sparsity of this data, dubious reliability of objective self-assessment and susceptibility to confirmation bias rendered this data unsuitable for detailed analysis, although it was useful as a validation for the model predictions.

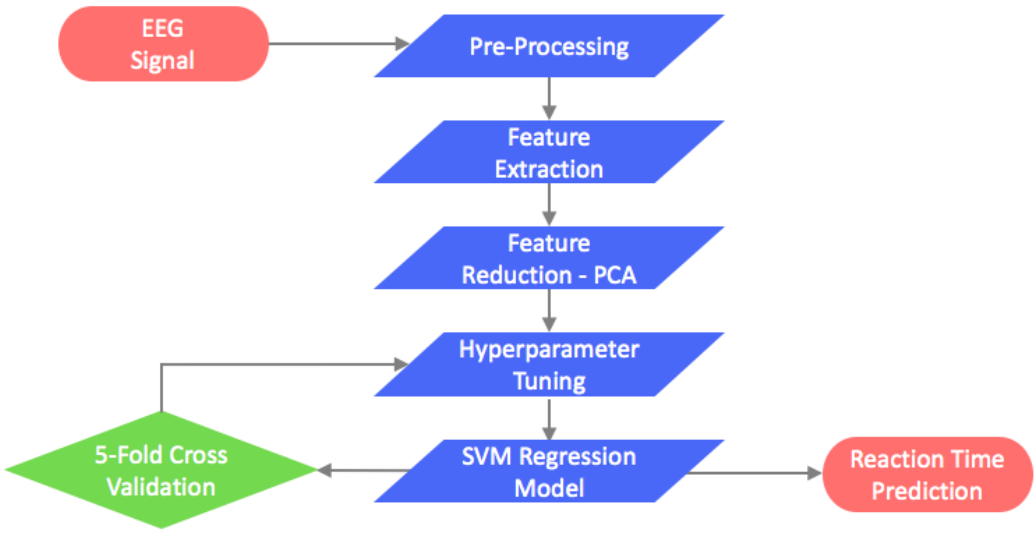

Figure 1: *Analysis pipeline for the EEG signals.*

2.2 Pre-processing

Before the EEG signals can be analysed a series of pre-processing steps must be conducted to remove internal and environmental sources of noise. This was performed via the following pipeline, which makes extensive use of EEGLAB, an open source MATLAB toolbox for analysing EEG data.

2.2.1 Filtering and downsampling

The raw EEG data which was sampled at 1024 Hz, was downsampled to 256 Hz to reduce the size of the datasets. Given that the majority of information pertaining to brain activity lies between 0.5 and 40 Hz along with the significant line noise at 50 Hz, a finite impulse response (FIR) band pass filter was applied between 0.5 and 45 Hz. Nyquist's criterion states that for a band limited function the sampling rate must be larger than or equal to twice the highest frequency component present in the signal. Therefore, these pre-processing steps ensure that the signal is immune to aliasing.

2.2.2 Interpolation of bad channels and artifact removal

Individual electrodes may be defective or have a poor connection to the scalp, resulting in a significantly higher impedance and lower signal to noise ratio. Bad channels were identified by evaluating channel statistics such as the standard deviation and kurtosis (measure of the 'tailedness' of a distribution) to assess their deviations from Gaussianity (empirically bad channels have non-Gaussian distributions). Electrodes with a kurtosis more than 5 standard deviations from the mean were identified as bad channels. The continuous EEG data was then scrolled through and any unidentified bad channels were marked. These channels were then interpolated using spherical spline interpolation. The continuous data then was scrolled through and sections of the data containing high amplitude transient artifacts were identified by inspection and removed.

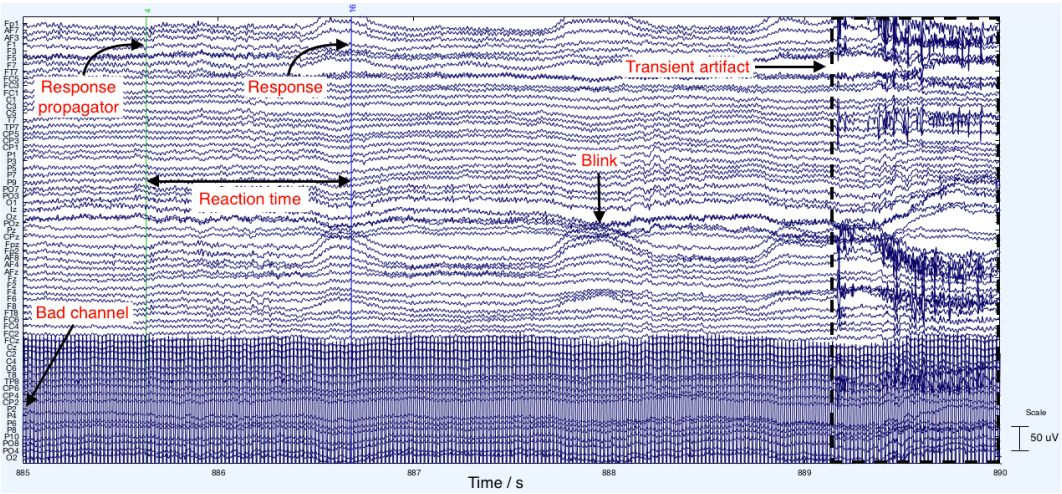

Figure 2: *Annotated raw EEG data for subject 5, illustrating the manifestation of a bad channel (P2) and high amplitude transient artifact.*

2.2.3 Independent Component Analysis (ICA)

In the final stage of pre-processing an independent component analysis (ICA) decomposition was performed on each dataset using the Infomax algorithm [3]. ICA is a technique for finding the independent sub-components of a signal. It assumes that the subcomponents are both non-Gaussian and statistically independent [2], where the independent signals are found by max-

[2] However the net signal is Gaussian, since by the central limit theorem the distribution of a sum of independent random variables tends towards a Gaussian distribution.

imising entropy. If we consider an N dimensional input vector $\mathbf{X}(t) = (x_1(t), ..., x_n(t))$, ICA finds an un-mixing matrix $\mathbf{W}(t)$ such that the unmixed components $\mathbf{Y}(t) = (y_1(t), ..., y_n(t))$ of the linear transform $\mathbf{Y}(t) = \mathbf{W}(t)\mathbf{X}(t)$ are maximally independent. The efficacy of ICA's application to EEG signals has been demonstrated in [4] provided the measured EEG signal is a linear, instantaneous superposition of the source signals. Therefore, using ICA, it is possible to identify artifacts from their characteristic signatures in the component scalp map and power spectrum. These artifacts can then be then removed, with the remaining components being superposed and projected back to the time domain.

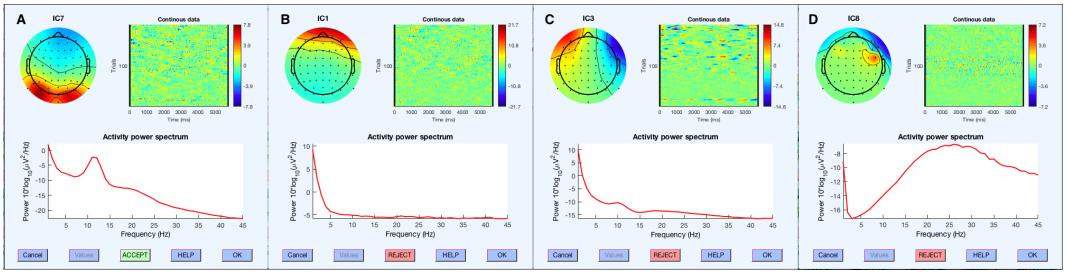

Figure 3: *For each independent component of subject 18, the scalp topography (top left), continuous signal activity (top right) and component power spectrum (bottom) are shown. Component A corresponds to a brain component, and is identifiable from the characteristic* $1/f$ *power spectrum modulated by peak(s) derived from rhythmic brain activity. The scalp topography also appears dipolar because the electric fields generated from the brain activity can be accurately modelled by an equivalent current dipole (ECD). Components B and C are derived from blinks and lateral eye movement respectively. They are characterised by their highly concentrated power at frequencies below 5 Hz and the scalp topographies suggest ECD's near the eyes (eye movement also creates an electric field that appears dipolar). Component D captures the effects of an electric field generated from muscle activity and is distinguished by the concentration of power in higher frequencies and focusing of the scalp topography outside of the skull.*

2.3 EEG epoching

For a given cognitive state, it is assumed that this maps to a well-defined probability distribution for the reaction times at that time. Consequently, in order to accurately constrain the mean and characteristic width of the distribution sufficiently many samples of the distribution are required. Given that reaction times were registered at a rate of 5 min^{-1} meant that epochs on the order of 1 minute were required to accurately constrain the true mean of the reaction times. However, the selection of epoch length also requires consideration of the timescale of variations in alertness, the choice of epoch length therefore requires a compromise between having sufficiently many samples to accurately estimate the local mean reaction time, and the level of alertness being quasi-stationary over the timescale of the epoch. Epoch lengths of 45s were selected following experimentation and research conducted in Gilden et al, 1995, which showed that behavioural performance and neural activity both exhibit a characteristic $1/f$ decay, suggesting that the largest fluctuations in the cognitive state occur over timescales of minutes, rather than seconds.

2.4 Feature Extraction

Feature extraction is a crucial stage of all supervised machine learning problems, for every epoch (observation) a feature vector is extracted. These feature vectors are collated over all epochs

together with the target outputs, which are the means of the reaction times within each epoch (called the local mean reaction times) are then used to train a model which is then capable of estimating the expected local mean reaction time, and therefore the alertness of the driver. A total of 67 features are extracted across the time and frequency domain of the signal.
The four primary sectors of feature extraction are:

1. Blinks
2. Power Spectral Density (PSD)
3. Spectral Entropy
4. Coherence

2.4.1 Blinks

It has been shown in Ingre et al, 2006, that there are strong correlations between the duration of blinks and driver fatigue as measured by the Karolinska Sleepiness Scale. Furthermore, in Caffier et al. 2003 they establish the presence correlations between drowsiness and the variance, skewness, and kurtosis of the blink duration distribution.
In this project, blinks were extracted and analysed (before ICA pruning) using an EEGLAB plugin called 'BLINKER' [5]. Firstly, using only the frontal electrodes the Blinker algorithm applies a band pass filter in the interval [1, 20] Hz, thereafter it identifies the time sections of the signal for which the signal amplitude is 1.5 standard deviations above the mean, these intervals then form the candidates for potential blinks. To assess whether the candidates can be classed as blinks for further analysis, the algorithm then calculates the signal to noise ratio for every blink and rejects those outside a pre-established threshold. Each of the remaining blinks then have a straight line fit to the ascending and descending slopes of the blink peak using linear regression to form a tent, this allows the blink to be fully characterised .

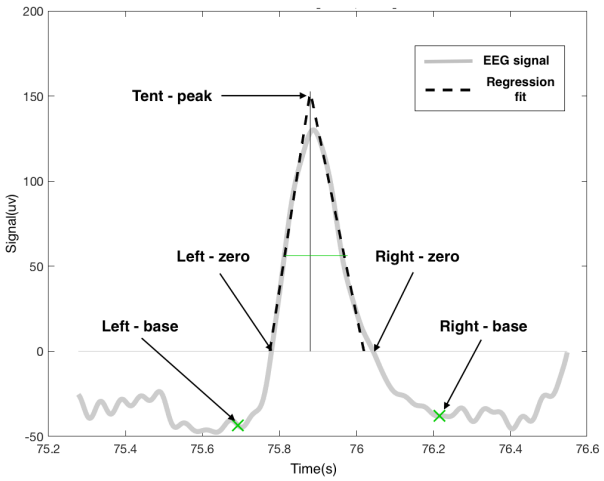

Figure 4: *A blink recorded from a frontal electrode (FP2) of subject 18 with straight lines fit using linear regression,* $R^2_{left} = 0.991$ *and* $R^2_{right} = 0.992$.

Following the characterisation of the blinks identified in the dataset, the following 9 features were extracted:

1. Frequency
2. Duration (mean and variance)
3. Closing time (mean and variance)
4. Reopening time (mean and variance)
5. Amplitude Velocity Ratio (mean and variance)

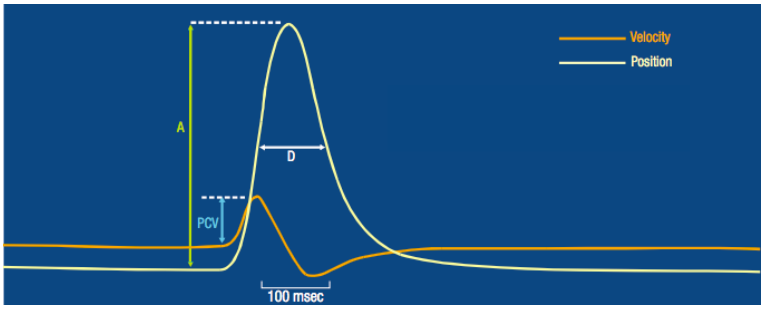

Figure 5: *The amplitude (A), duration at half amplitude (D) and peak closing velocity (PCV) of a normal blink. The amplitude velocity ratio of a blink is given by A/PCV. [Picture credits: Johns MW, 2003 [6]]*

Each one of features 2-4 was calculated using the zero crossing of the ascending and descending slopes, along with the zero crossings given by the linear regression fit of the slopes. The data from the method which gave the highest Spearman's rank correlation with the local mean reaction time was then retained as a feature for the final model. No assumption was made on the linearity (or lack of it) of the relationship between the blink properties and the reaction times, therefore the use of Spearman's rank is appropriate, since it only seeks to quantify the degree to which to variables are monotonically related.

2.4.2 Power Spectral Density (PSD)

The brain is a complex non-linear dynamic system, meaning the signal it generates is also non-linear and non-stationary [3]. Therefore, analysis of the dynamics of the power spectral density requires the use of non-linear techniques, such as time-frequency analyses, which essentially amounts to performing time localised Fast Fourier transforms over multiple time intervals within each epoch. All of the following PSD analyses was performed using only the parietal electrodes ('Pz' , 'P7' & 'P8') which exhibited the strongest correlations with reaction times.
Therefore, transforming the signal into the time-frequency representation yields

$$\tilde{F}(\omega_i) = \sum_{t=1}^{T} f(t) exp\left(\frac{-2\pi i(\omega_i - 1)(t-1)}{T}\right) \tag{1}$$

Such that the power spectral density is given by

$$P(\omega_i) = |\tilde{F}(\omega_i)|^2 \tag{2}$$

Now, if we divide this by the total power contained within all frequencies bands, then the power spectral density can be considered as a discrete probability distribution function (and switching $\omega_i \rightleftharpoons f_i$ for clarity)

[3] Brain signals are quasi-stationary on the timescale of 0.25 seconds [7].

$$p(f_i) = \frac{P(f_i)}{\sum_{i=1}^{N} P(f_i)} \tag{3}$$

Where $P(f_i)$ is the power contained within the frequency bin i, and the sum is over all frequency bins in the range [0.5, 40] Hz . The power spectral density is then further concatenated to form a set of 20 features consisting of the relative power contained within (or equivalently, the probability of lying within) 2 Hz bins spanning the range [0.5, 40] Hz.
Additional features are extracted from the canonically defined frequency bands: delta, theta, alpha, beta and gamma. The relative power contained within these bands is calculated as before, along with two further features inspired by the work of Yeo et al. 2009,
the Spectral Centroid

$$\mu_f = \langle f \rangle \tag{4}$$

and the Spectral Variance

$$\sigma_f^2 = \langle f^2 \rangle - \langle f \rangle^2 \tag{5}$$

Where

$$\langle f^m \rangle = \sum_{i=1}^{N} p(f_i) f_i^m \tag{6}$$

Mathematically, they are equal to the mean and variance of the frequency calculated using the discrete probability distribution derived from the normalised power spectral density.
Finally, in the alpha frequency band the dynamics of the local maxima (peaks) were explored via the implementation of a custom designed curve fitting algorithm. For each epoch, the algorithm identifies the peaks in the data according to a set of criterion related the peak prominence (how much a peak stands out relative to its background) and the peak separation (to prevent a single peak being misidentified as two). Subsequently, the four most prominent peaks were selected and fit using the following function

$$P_{fit}(f) = a + \frac{b}{f} + \sum_{k=1}^{4} c_k exp\left(\frac{(f-d_k)^2}{2e_k^2}\right) \tag{7}$$

This function can be interpreted as the 1/f transient arrhythmic brain activity modulated by a series of Gaussians which capture the mean and characteristic width of the peaks associated with the rhythmic brain activity. Next, the area within the full width at half maximum (FWHM), which quantifies the power within the peak (peak power), is calculated for the four peaks fit by the Gaussians and evaluates to

$$\int_{d_k-\sqrt{2ln2}e_k}^{d_k+\sqrt{2ln2}e_k} c_k e^{-(f-d_k)^2/2e_k^2} df \propto c_k e_k \tag{8}$$

Where the proportionality constant is irrelevant since the features are normalised before being fed to the regression model. Finally, the dominant peak is identified as the one with the largest peak power, this power along with its location are then used as features. They are termed the dominant

peak power (DPP) and dominant peak frequency (DPF).

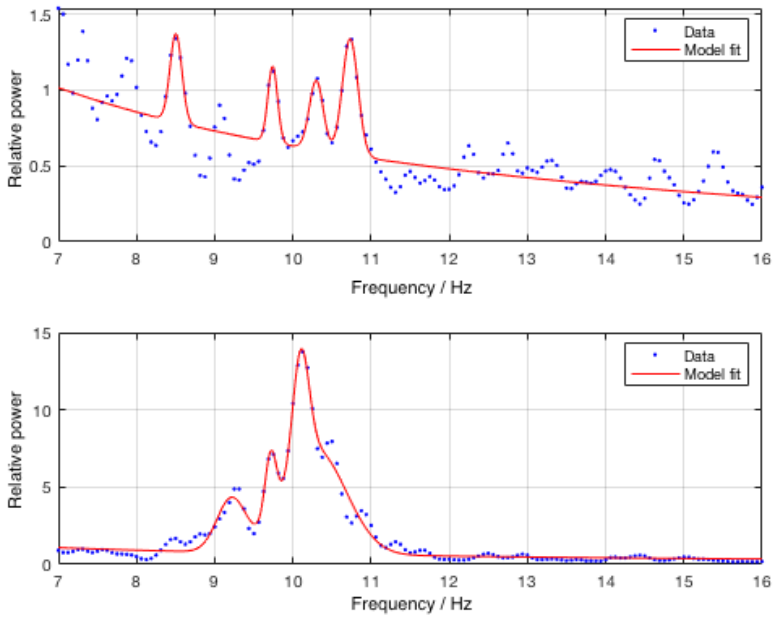

Figure 6: *Power spectrum in the alpha band with the peak-finder function fit for subject 6 for two distinct epochs in which the reaction times were* 0.8*s (upper) and* 1.8*s (lower).*

2.4.3 Entropy

Using the probability distribution previously derived from the normalisation of the power spectral density, it as possible to calculate the Shannon entropy as

$$H = -\frac{1}{log_2(N)} \sum_{i=1}^{N} p(f_i) log_2(p(f_i)) \tag{9}$$

This method of computing the Shannon entropy is known as the spectral entropy. It is calculated using the full frequency range (0.5 – 40 Hz) and can be interpreted as a measure of the irregularity or unpredictability in a signal. The normalisation can be interpreted as the maximal spectral entropy of white noise, uniformly distributed in the frequency domain, this imposes the constraint that $0 < H < 1$. A value of $H = 0$ corresponds to a completely regular signal whereas a value of $H = 1$ corresponds to a signal with maximal irregularity.

2.4.4 Coherence

For every epoch, the time-frequency coherence (TFC) [8] was computed in the electrodes in the frontal, temporal, parietal, and occipital regions of the brain within the canonical frequency bands. The time-frequency coherence is defined as

$$C_{xy}(t,f) = \frac{|S_{xy}(t,f)|^2}{S_{xx}(t,f)S_{yy}(t,f)} \tag{10}$$

Where t and f denote the time (which epoch) and frequency band. S_{xx} and S_{yy} refer to the auto power spectral densities of signal x and y whilst, S_{xy} corresponds to the cross power spectral density between signals x and y. In this context, coherence is a measure of the amplitude-phase relation between two distinct regions of the brain. For example, for a given time and frequency band, if the coherence between the signals from two channels is high (≈ 1), it implies the amplitude ratio and phase shift between the two signals is approximately constant [9].

2.5 Feature reduction - PCA

Some form of feature selection / reduction technique was required to remove the redundant features, since their inclusion had the potential to degrade the regression model. Traditional feature selection methods such as stepwise models, sequential feature selection and neighbourhood component analysis (NCA) were considered, however they all showed inconsistencies in their prioritisation and selection of features. The most likely reason for this was the fact that features from all sectors exhibited high Spearman's rank cross correlations, which is again a consequence of the inherent redundancy in the EEG data. Therefore, principle component anaylsis (PCA) was used to project to the d-dimensional feature space onto a lower n-dimensional subspace which best captured the variance of the data. The advantages of this approach were two-fold: the removal of redundant features and the extraction of non-mutual information between features exhibiting strong inter-correlations, as oppose to their omission (and by consequence, the complete loss of non-mutual information). Furthermore, the significant intrinsic error in the local mean reaction time lends itself to a simpler model in which the output is governed by n ($\ll d$) features which best capture the coarse grain cognitive transitions. The PCA was performed using singular value decomposition of which there is more detail in the appendix.

3 Results and discussion

At the early stages of this project, feature extraction was conducted using only basic power spectral density features such as the dynamics of the relative power in the canonically defined frequency bands. Examination of the feature correlations with reaction times demonstrated that only 4 of the 20 datasets had features which showed statistically significant correlations (at the 1% significance level). The proposed solution to this was to conduct a far more extensive and broad feature extraction process, operating in both the time and frequency domain. Unfortunately, no further statistically significant correlations were found in any of the features belonging to the aforementioned datasets. Therefore, for completeness the subsequent architecture is built and tested on the 4 datasets in which the features showed non-negligible correlations with the reaction times. Further potential explanations for the poor performance of the remaining 16 datasets is detailed in section 4.

3.1 Feature analysis

The subsequent figures summarise the relationship between the features and reaction times, with the degree of correlation quantified by the Spearman's rank correlation coefficient, ρ. The aim is to find features which show the strongest correlations with reaction times such that they

can be used to train a predictive model. In figures where no reference is made to the dataset, correlation analysis was conducted by first performing soft normalisation[4] of feature vectors and reaction times, then concatenating the scaled features and target output before evaluating the net Spearman's rank ρ value. This accounts for the intrinsic variability in both the reaction time and baseline brain activity across subjects. To give an indication of the significance of different ρ values, a $\rho = 0.15$ corresponds to a P value of 0.01 (i.e. Spearmans rank correlations higher than 0.15 are statistically significant to 1%).

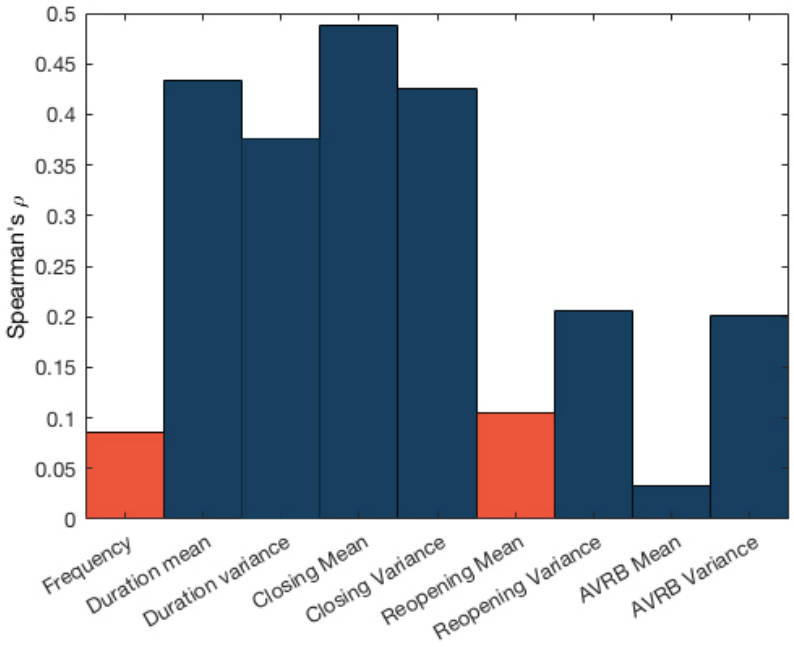

Figure 7: *Absolute Spearman's rank ρ values for the blink features, where orange corresponds to negative correlation and blue to positive.*

As shown in figure 8, some of the blink features demonstrate the strongest correlations with reaction times of all of the features, in particular the mean and variance of the duration and closing time. The fact that both the mean and variance of these particular features both show such high correlations is further evidence of them being a true behavioural indicator of drowsiness. These results agree with the literature which declares the properties of blinks to be amongst the best behavioural markers of transitions in alertness [10].

The results in figure 10 suggest that spectral power shifts are dominated by the alpha frequency band, such that reductions in alertness correspond to power transitioning from the other bands into the alpha band. Furthermore, we identify the delta band's statistical properties as being significant features, meaning that the power spectral density within the delta band is highly dynamic and a strong correlate of drowsiness.

[4] Soft normalisation entails subtracting the mean and dividing by the standard deviation of a given vector.

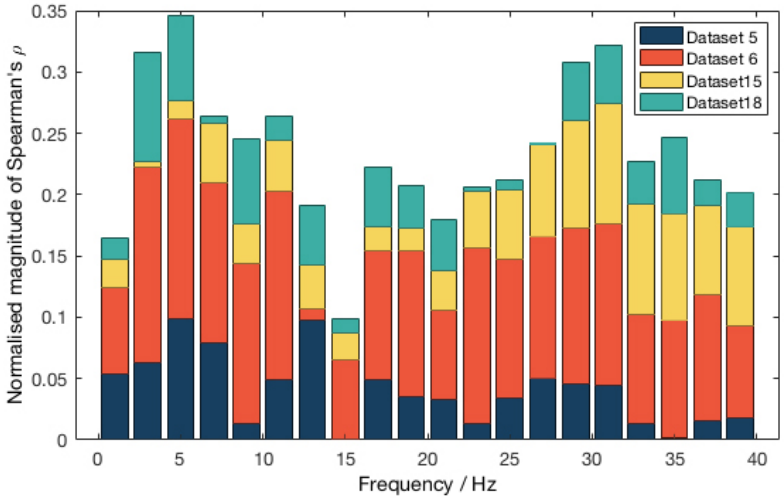

Figure 8: *Normalised absolute Spearmans rank correlation coefficient, ρ, for the PSD features of the relative power in 2 Hz bins for all 4 datasets.*

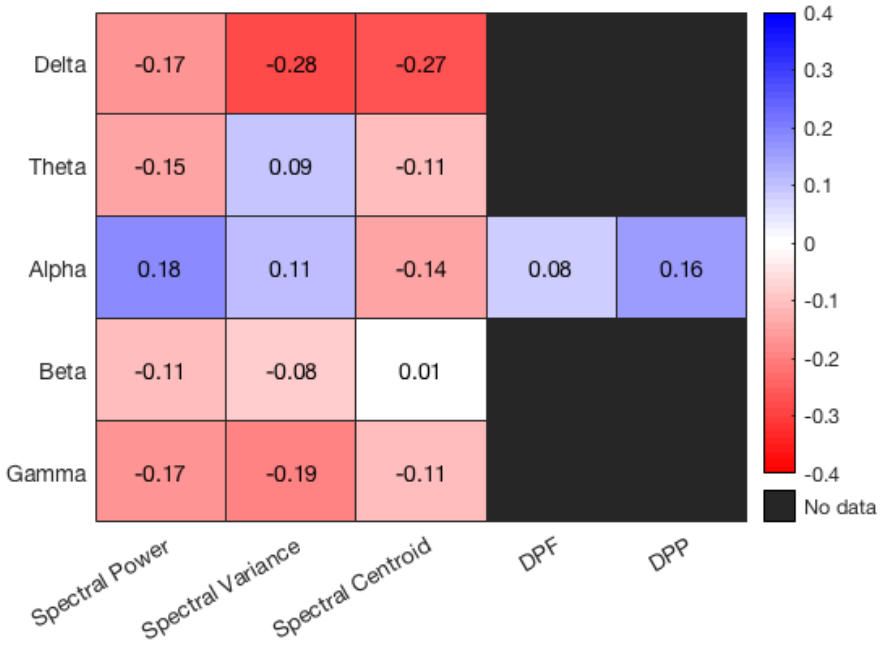

Figure 9: *Spearman's ρ values for the statistical features derived from the PSD discrete probability distribution function and dominant peak identification extracted for each canonical frequency band.*

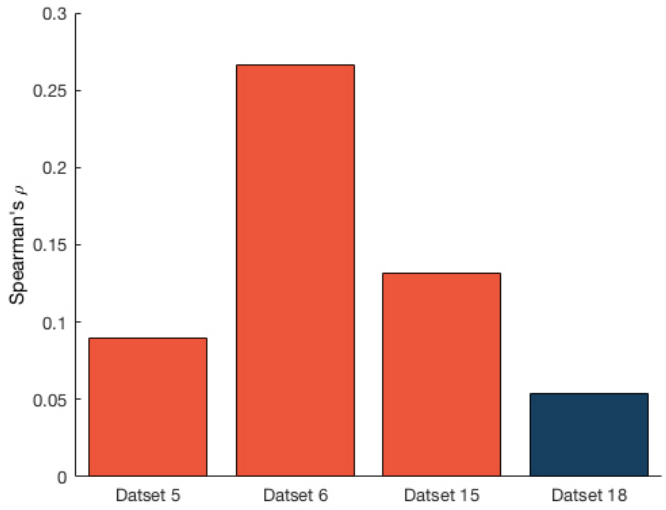

Figure 10: *Spearman's ρ values for the spectral entropy for each of the 4 datasets, with orange corresponding to a negative correlation and blue to a positive correlation. The predominantly negative correlations of entropy with reaction times are indicative of a state of heightened alertness being synonymous with a brain state of higher complexity.*

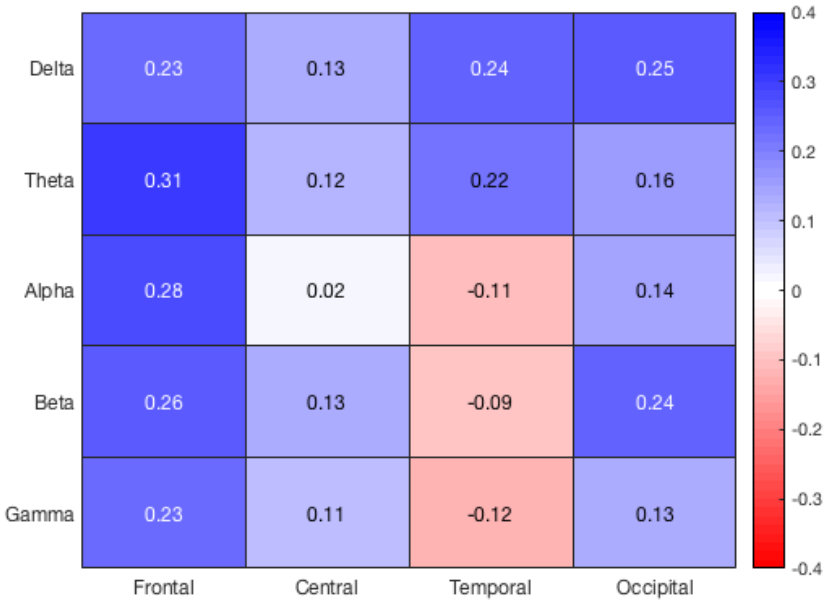

Figure 11: *Spearman's ρ values for the coherences at each region of the brain for every frequency band. The frontal lobe, followed by the temporal and occipital lobes are the most informative. The dominance of positive correlations here suggests that the partial amplitude-phase locking of spatially distinct localised brain activity is conducive to a state of reduced alertness.*

3.2 Feature importance

The relative contribution of the features across the different sectors was calculated using the PCA loadings and variance explained from a singular value decomposition of the four datasets concatenated. This ensured that the generalisability of features across different subjects could be best captured.

To obtain an additional and independent measure of feature importance for comparison, NCA was used on the concatenated datasets to find the feature weights. In particular, NCA was run a thousand times with the performance assessed via a loss function which used the mean square error between the predicted and target value. Then, the mean feature weights were extracted and taken to be the relative feature importance. NCA can be used as feature selection technique which seeks to learn a distance metric via a linear transformation of the data such that the model performance is maximised in the transformed space.

The feature significance of the relative power in the 2 Hz frequency bins indicates that they contain information that is not captured by the canonical frequency bands, even though the focus of many related papers is solely on these bands.

NCA suggests a feature importance heavily biased towards the blink feature sector, whilst the allocation of feature importance suggested by the PCA eigenvectors and eigenvalues is more evenly distributed. A potential explanation for this disparity is that whilst both techniques involve transformations to lower dimensional spaces in which the variance is maximised, unlike PCA, NCA also attempts to cluster these data points. The importance of the blink features is further supported by the dominance of their Spearman's rank correlations over the other features.

A reason why the blinks might be such a significant feature is the fact that their extraction and analysis requires no pre-processing and they are easily distinguishable in the EEG signal (see figure 2). Whereas all of the other features require extensive pre-processing, therefore allowing the entrance of many sources of noise and errors that may convolute the signal and attenuate the feature correlations.

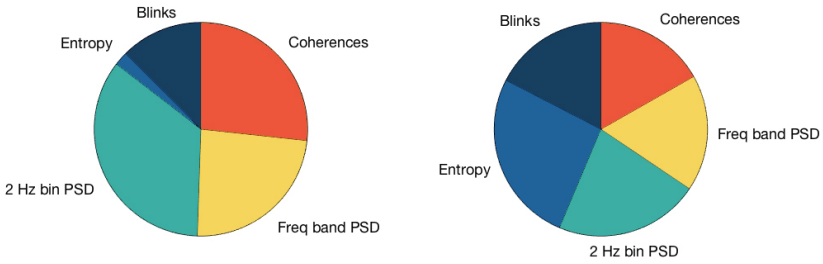

Figure 12: *PCA relative feature sector importance (left) and the relative feature sector importance normalised by the number of features in the sector (right).*

Whilst a number of BCI studies have successfully produced predictive algorithms using even as few as one electrode, they are predominantly based on analysing and tracking the dynamics of the PSD from electrodes placed on the parietal/occipital lobes. The use of three electrodes, one on the occipital lobe and two symmetrically placed on the frontal lobe would facilitate the capture of blink and (frontal) coherence based features, which have just been demonstrated to be the most

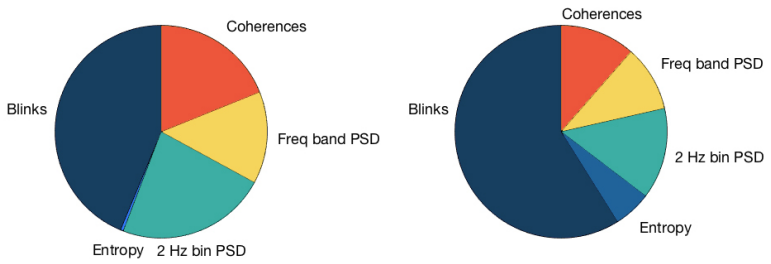

Figure 13: *NCA relative feature sector importance (left) and the relative feature sector normalised by the number of features in the sector (right).*

important. Importantly this would accomplish the same as what eye-tracking based vigilance monitors do, in addition to acquiring the information associated with entropy and the PSD.

3.3 Model evaluation

An evaluation of the model performance first requires the definition of a suitable metric. A conventional index used for regression predictor models is the root mean square error (RMSE) which is given by

$$RMSE = \frac{\sqrt{\left(\sum_{i=1}^{N} RT_i - \widehat{RT_i}\right)^2}}{N} \tag{11}$$

Where RT_i are the predicted reaction times, $\widehat{RT_i}$ are the observed reaction times and N is the number of observations.

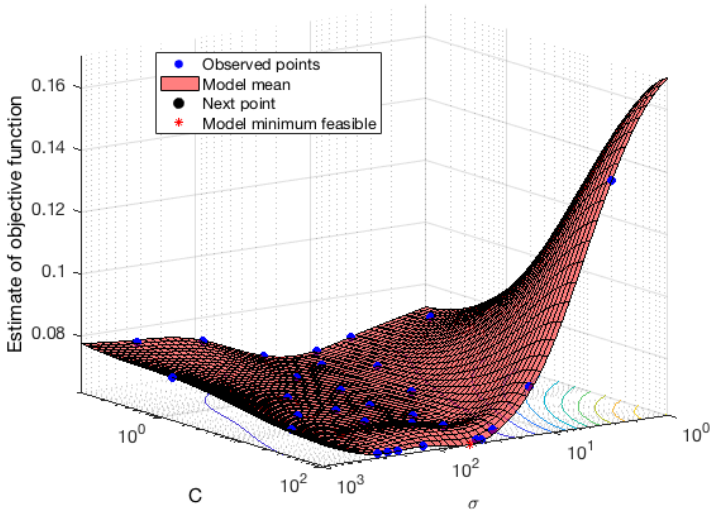

Figure 14: *Bayesian optimisation of the hyperparameter pair (σ, C), demonstrating the strategic sampling of regions around the local minima of the objective function.*

To obtain the optimal model performance the hyperparameter pair (σ, C) was optimised on the concatenation of all 4 datasets using Bayesian optimisation. Bayesian optimisation iteratively builds a probabilistic functional model of the objective function (the function to be minimised) that guides an informed sampling of the hyperparameter phase space. The strategic sampling of the phase space facilitates a faster convergence to the optimal hyperparameter pair compared to the heuristic grid and random search algorithms. To minimise overfitting of the hyperparameters to the training data, the model performance was evaluated within 5-fold cross validation. This entailed randomly partitioning the training data into 5 distinct folds, then iteratively training the model on 4 of the folds and evaluating the performance on the 5th fold until each fold had been used for evaluation once, then taking the average of the reported errors.

Deploying the final model on each of the 4 datasets individually and evaluating the performance once again via 5-fold cross validation yielded an RMSE of $0.24 \pm 0.05s$. The predicted and actual reactions times are illustrated in figures 16 and 17 for two of the datasets. Whilst the SVR model developed does demonstrate an ability to track the coarse grain fluctuations in alertness, it was not able to follow high frequency fluctuations in reaction times across neighbouring epochs. This is predominantly a manifestation of the intrinsic error in the reaction times due to insufficient sampling of the instantaneous probability distribution. Another significant contribution to the error is derived from the features extracted, and thus the EEG signal itself, which required extensive pre-processing to isolate the brain signal from the various sources of noise. The inability to extract purely the brain component, along with the limitations of EEG only providing a partial representation of the true brain activity will have compounded to form a significant error.

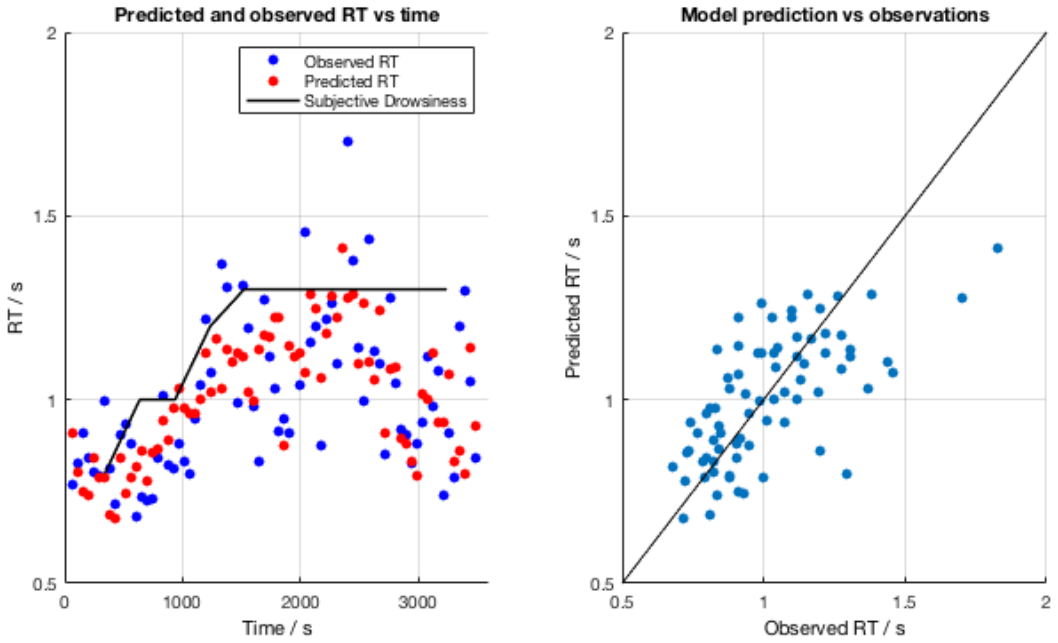

Figure 15: *Predicted and observed reaction times with subjective drowsiness ratings for dataset 6, yielding* $RMSE = 0.194 \pm 0.004s$.

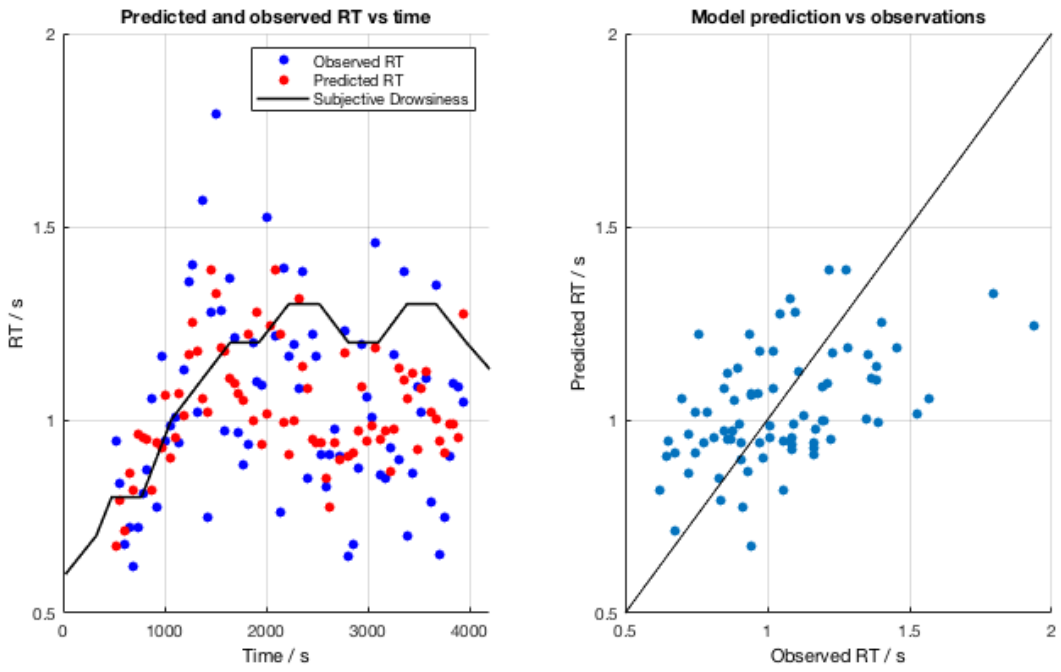

Figure 16: *Predicted and observed reaction times with subjective drowsiness ratings for dataset 15, yielding* $RMSE = 0.233 \pm 0.007s$.

4 Limiting factors and further work

The lack of significant correlations in the other datasets could be attributed to the negligible variability in their reaction times, meaning that they could have maintained a constant level of alertness throughout. Moreover, the datasets showed significant amounts of general noise and transient artifacts were pervasive; this caused ICA to fail completely for some datasets.

One of the primary limitations of this project was the sparsity of the reaction time data, which in turn demanded excessively long epochs that were susceptible to intrinsic variations in cognitive state. Further analysis using the architecture developed in this project would benefit greatly from the use of a behavioural measure with a sufficiently high sampling rate such that the epoch length could be reduced and the fine grain variations in alertness could be more accurately constrained. Ideally, this would involve a measure of driving error, a reaction time sampler operating at a minimum rate of 10 min^{-1} along with the subjugation of participants to a mild form of sleep deprivation prior to the experiment. This would increase the probability of subjects exhibiting larger scale fluctuations in alertness.

Additional features that are worthwhile investigating lie predominantly in the discrete probability distribution function (PDF) found from the PSD. Whilst the first and second moments of the PDF were computed (μ_f and σ_f^2) it is possible that higher order moments also contain non-mutual information. For example, the 3rd and 4th moments of the distribution which correspond to the skewness and kurtosis, and are given by:

$$\Gamma_3 = \langle f^3 \rangle + 3\langle f^2 \rangle \langle f \rangle + \langle f \rangle^3 \tag{12}$$

$$\Gamma_4 = \langle f^4 \rangle + 4\langle f^3 \rangle \langle f \rangle + 3\langle f^2 \rangle^2 + 6\langle f^2 \rangle \langle f \rangle^2 + \langle f \rangle^4 \tag{13}$$

5 Conclusion

This project has demonstrated the feasibility of using EEG signals to track the coarse grain fluctuations in alertness using an SVM regression model. Features across each of the sectors exhibited statistically significant correlations with the behavioural index of drowsiness. The blink features (namely blink duration and closing time) show great promise for further investigation since they were shown to be the best predictors of alertness out of all of the features. Furthermore, their negligible pre-processing requirements and minimal computation time renders them an ideal candidate for driver drowsiness BCI applications. However, the absence of any correlations in 16 of the 20 datasets suggests that the implementation of these methods on further independent data with a more accurate behavioural index is required to properly assess the individual feature importance and obtain a more accurate predictive model.

6 Appendix

6.1 SVM regression

Support Vector Machines are a class of supervised machine learning algorithms that can be applied to regression or classification problems, with regression being the variant considered in this project (Support Vector Regression). SVR aims to construct an optimal hyperplane coupled to two equidistant planes defining a tolerance, using a loss function that dictates a penalty of zero within the tolerance planes and scales linearly with separation from the hyperplane thereafter. This type of loss function is termed epsilon insensitive and the subsequent goal of $\epsilon-$SVR is to find a small w such that $f(x) = \langle w, X \rangle + b$ has a maximum deviation of epsilon from the target value of y_i, where $X = \{(x_i, y_i)\}$ and $i = [1, 2, ..., N]$.

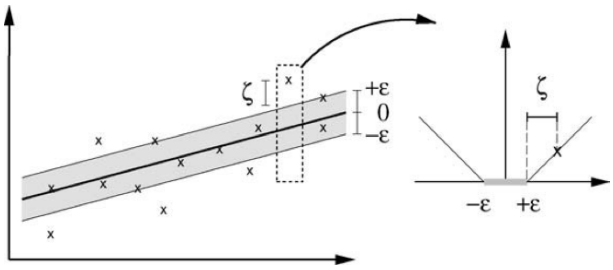

Figure 17: *A 2D slice of a linear SVM hyperplane together with the boundary layers to form a 'soft' margin (left) and the ϵ-insensitive loss function (right). [Picture credits: Scholkopf and Smola, 2004 [11]]*

A formulation for this is proposed in [8] as

$$\min\left(\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right) \tag{14}$$

$$\text{subject to} \begin{cases} y_i - w \cdot x_i - b \leq \epsilon + \xi_i \\ w \cdot x_i + b \leq y_i + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{15}$$

Where, ξ_i and ξ_i^* are slack variables. The constant $C > 0$ determines the penalty imposed on observations which violate the margin, and involves a balance between the flatness of f and the amount to which deviations larger than epsilon are tolerated. In cases where a linear hyperplane is insufficient, the SVR model is obtained by replacing the inner product $\langle x_i, x_j \rangle$ with a non-linear kernel function, $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, which performs a mapping to a higher dimensional space where a linear hyperplane may be found. A Guassian kernel was used in this project and is defined as:

$$k(x_i, x_j) = exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) \tag{16}$$

Where the hyperparameter σ controls the curvature (granularity) of the hyperplane.

6.2 PCA

Let the nxp matrix containing the feature vectors be given by $\mathbf{X}$, where the rows corresponds to the observations and the columns to the variables (features). Since the matrix $\mathbf{X}$ is in general non-square ($n \neq p$), a singular value decomposition must be applied. This consists of first centering and scaling the columns of $\mathbf{X}$ such that the mean and standard deviation are 0 and 1 respectively. Subsequently, a transformation is found such that $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^* \tag{17}$$

Where $\mathbf{U}$ is an nxn unitary matrix, Σ is an nxp rectangular diagonal matrix and $\mathbf{V}^*$ is a pxp unitary matrix. If the diagonal elements of Σ are sorted in order of descending magnitude, we can then write $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^* \approx \mathbf{U_r}\Sigma_\mathbf{r}\mathbf{V_r^*} \tag{18}$$

Which can be interpreted as taking the first r principle components of the SVD, with $\mathbf{V_r^*}$ containing the loadings of the SVD, therefore the reduced feature space of a feature matrix $\mathbf{X}$ is given by $\mathbf{XV_r^*}$.

References

[1] Mervyn VM Yeo, Xiaoping Li, Kaiquan Shen, and Einar PV Wilder-Smith. Can svm be used for automatic eeg detection of drowsiness during car driving? *Safety Science*, 47(1):115–124, 2009.

[2] Torbjörn Åkerstedt and Mats Gillberg. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2):29–37, 1990.

[3] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[4] Tzyy-Ping Jung, Colin Humphries, Te-Won Lee, Scott Makeig, Martin J McKeown, Vicente Iragui, and Terrence J Sejnowski. Extended ica removes artifacts from electroencephalographic recordings. In *Advances in neural information processing systems*, pages 894–900, 1998.

[5] Kelly Kleifges, Nima Bigdely-Shamlo, Scott E Kerick, and Kay A Robbins. Blinker: Automated extraction of ocular indices from eeg enabling large-scale analysis. *Frontiers in neuroscience*, 11:12, 2017.

[6] MW Johns et al. The amplitude-velocity ratio of blinks: a new method for monitoring drowsiness. *Sleep*, 26(SUPPL.), 2003.

[7] Wlodzimierz Klonowski. Everything you wanted to ask about eeg but were afraid to get the right answer. *Nonlinear Biomedical Physics*, 3(1):2, 2009.

[8] Langford B White and Boualem Boashash. Cross spectral analysis of nonstationary processes. *IEEE Transactions on Information Theory*, 36(4):830–835, 1990.

[9] Ramesh Srinivasan, William R Winter, Jian Ding, and Paul L Nunez. Eeg and meg coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. *Journal of neuroscience methods*, 166(1):41–52, 2007.

[10] Philipp P Caffier, Udo Erdmann, and Peter Ullsperger. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European journal of applied physiology*, 89(3-4):319–325, 2003.

[11] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[12] Chin-Teng Lin, Che-Jui Chang, Bor-Shyh Lin, Shao-Hang Hung, Chih-Feng Chao, and I-Jan Wang. A real-time wireless brain–computer interface system for drowsiness detection. *IEEE transactions on biomedical circuits and systems*, 4(4):214–222, 2010.

[13] Thiago LT da Silveira, Alice J Kozakevicius, and Cesar R Rodrigues. Automated drowsiness detection through wavelet packet analysis of a single eeg channel. *Expert Systems with Applications*, 55:559–565, 2016.

[14] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J Mckeown, Vicente Iragui, and Terrence J Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178, 2000.

[15] Simo Monto, Satu Palva, Juha Voipio, and J Matias Palva. Very slow eeg fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. *Journal of Neuroscience*, 28(33):8268–8272, 2008.

[16] James Stuart Peter Macdonald, Santosh Mathan, and Nick Yeung. Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus eeg alpha oscillations. *Frontiers in psychology*, 2:82, 2011.

[17] Michael Ingre, Torbjörn Åkerstedt, Björn Peters, Anna Anund, and Göran Kecklund. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. *Journal of sleep research*, 15(1):47–53, 2006.

[18] Chin-Teng Lin, Chun-Hsiang Chuang, Chih-Sheng Huang, Shu-Fang Tsai, Shao-Wei Lu, Yen-Hsuan Chen, and Li-Wei Ko. Wireless and wearable eeg system for evaluating driver vigilance. *IEEE Transactions on biomedical circuits and systems*, 8(2):165–176, 2014.

[19] Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161, 1997.

[20] Ibtissem Belakhdar, Walid Kaaniche, Ridha Djmel, and Bouraoui Ouni. A comparison between ann and svm classifier for drowsiness detection based on single eeg channel. In *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pages 443–446. IEEE, 2016.

[21] Sridhar R Jagannathan, Alejandro Ezquerro-Nassar, Barbara Jachs, Olga V Pustovaya, Corinne A Bareham, and Tristan A Bekinschtein. Tracking wakefulness as it fades: micro-measures of alertness. *NeuroImage*, 176:138–151, 2018.

[22] Hong J Eoh, Min K Chung, and Seong-Han Kim. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics*, 35(4):307–320, 2005.

[23] SF Liang, CT Lin, RC Wu, YC Chen, TY Huang, and TP Jung. Monitoring driver's alertness based on the driving performance estimation and the eeg power spectrum analysis. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 5738–5741. IEEE, 2006.

[24] L De Gennaro, Michele Ferrara, and M Bertini. The boundary between wakefulness and sleep: quantitative electroencephalographic changes during the sleep onset period. *Neuroscience*, 107(1):1–11, 2001.

[25] Mitul Kumar Ahirwal and Narendra D Londhe. Power spectrum analysis of eeg signals for estimating visual attention. *International Journal of computer applications*, 42(15):22–25, 2012.

[26] Ibtissem Belakhdar, Walid Kaaniche, Ridha Djemal, and Bouraoui Ouni. Single-channel-based automatic drowsiness detection architecture with a reduced number of eeg features. *Microprocessors and Microsystems*, 58:13–23, 2018.

[27] Abdul-Bary Raouf Suleiman, Toka Abdul-Hameed Fatehi, et al. Features extraction techniques of eeg signal for bci applications. *Faculty of Computer and Information Engineering Department College of Electronics Engineering, University of Mosul, Iraq*, 2007.

[28] Tzyy-Ping Jung, Scott Makeig, Magnus Stensmo, and Terrence J Sejnowski. Estimating alertness from the eeg power spectrum. *IEEE transactions on biomedical engineering*, 44(1):60–69, 1997.

[29] Abdulhamit Subasi and M Ismail Gursoy. Eeg signal classification using pca, ica, lda and support vector machines. *Expert systems with applications*, 37(12):8659–8666, 2010.

[30] David L Gilden, Thomas Thornton, and Mark W Mallon. 1/f noise in human cognition. *Science*, 267(5205):1837–1839, 1995.