Correlation between Situational Awareness and EEG signals

Jan Luca Kästle\textsuperscript{a}, Bani Anvari\textsuperscript{b}, Jakub Krol\textsuperscript{b}, Helge A Wurdemann\textsuperscript{b}

\textsuperscript{a}Centre for Transport Studies, Department of Civil, Environmental and Geomatic Engineering, University College London, UK, email: b.anvari@ucl.ac.uk
\textsuperscript{b}Department of Mechanical Engineering, University College London, UK

Abstract

An important aspect in safety-critical domains is Situational Awareness (SA) where operators consolidate data into an understanding of the situation that needs to be updated dynamically as the situation changes over time. Among existing measures of SA, only physiological measures can assess the cognitive processes associated with SA in real-time. Some studies showed promise in detecting cognitive states associated with SA in complex tasks using brain signals (e.g. electroencephalogram/EEG). In this paper, an analytical methodology is proposed to identify EEG signatures associated with SA on various regions of the brain. A new data set from 32 participants completing the SA test in the PEBL is collected using a 32-channel dry-EEG headset. The proposed method is tested on the new data set and a correlation is identified between the frequency bands of $\beta$ (12 – 30 Hz) and $\gamma$ (30 – 45 Hz) and SA. Also, activation of neurons in the left and right hemisphere of the parietal and temporal lobe is observed. These regions are responsible for the visuo-spatial ability and memory and reasoning tasks. Among the presented results, the highest achieved accuracy on test data is 67%.

Keywords: Situational Awareness, electroencephalograph (EEG), Psychology Experiment Building Language (PEBL) test, Random Forests, Decision Trees, Brain Computer Interface

1. Introduction

Understanding how people are acquainted with their environment, referred to as Situational Awareness (SA), is crucial in complex tasks that involve processing multiplexed information by a human operator. In order to reduce the cognitive load and undesirable and potentially fatal incidents in safety-critical domains, there is a need to identify the lack of SA in a timely manner \cite{1,2,3}. Safety risks have emerged, for instance, from drivers handling highly automated vehicles that dramatically increase in numbers on our roads \cite{3}. Drivers are provided with increasingly sophisticated autonomy features (e.g. cruise control, lane keeping ability). Each level of autonomy (defined in the J3016 “Levels of Driving Automation” standard by SAE International \cite{4}) requires a different level of the driver’s SA. During a single journey, drivers are then challenged to engage in and change between different levels of autonomy without any training causing safety concerns.

The most cited approach for defining SA’s processing trajectory is established by Endsley \cite{5}, where SA is divided into three levels: SA Level 1 associated with the subjects’ perception of their environment; SA Level 2 associated with their comprehension; and SA Level 3 associated with projection into a future state. It is noted that sufficient SA at lower levels is required, to reach adequate SA at subsequent levels. Only if SA Level 3 is achieved, the subject can adequately act to meet their objectives \cite{5}.

Existing measures of SA are categorised into subjective and objective measures. In majority of previous studies, SA is assessed through task performance questionnaires where the results are mapped onto a certain score for each level of SA \cite{5–8}. Subjective measures can be inaccurate as the subjects may misstate their responses \cite{9}. On the other hand, physiological measures give an objective understanding of the subject’s reflexive emotional response which can be more accurate in the assessment and analysis of SA \cite{10}. One of the most promising physiological signals are brain activity recordings, as they simultaneously exhibit signals for all sensory inputs and therefore provide a deep insight into a subject’s physiological response. A widely used sensor to measure brain activity is the electroencephalograph (EEG), measuring brain activity based on voltage fluctuations \cite{11}. Compared to other sensors available for measuring brain activity (e.g. MEG, fMRI and fNIRS), EEG is non-invasive \cite{12}, easy to apply \cite{13}, has a high temporal resolution (up to 1000 Hz \cite{14}), and is portable \cite{15}. However, mapping physiological data to SA is challenging due to high measurement noise and difficulty of disentangling signals into specific stimuli \cite{15}.

EEG is commonly used to acquire physiological measurement data in order to predict subject’s mental state, especially to detect workload, mental fatigue, sleepiness, and drowsiness in different environments \cite{17,19}. However, one of the subject’s mental states, which has not yet been extensively studied using physiological measures is SA.

This study focuses on using brain activity as an objective measure to identify two classes of SA (high SA and low SA) of a person, and proposes an analytical methodology for identifying spatial areas and frequency bands associated with SA in the brain (see Figure 1).
Figure 1: Overview of the procedure used in this study to identify the spatial locations in a brain and frequency ranges which can be associated with Situation Awareness (SA).

The contributions of this paper are as follows:

• A new data set was collected using a 32-channel dry-EEG headset from 32 participants completing the well established Psychology Experiment Building Language (PEBL) SA test.

• A new analytical methodology is proposed to identify EEG signatures associated with SA levels (see Figure 1). In the first stage, labels are defined using the PEBL SA test data. Then, the EEG data is preprocessed. In the third stage, features are extracted using Independent Component Analysis and Principal Component Analysis to determine new temporal signals and transforming them to frequency space with Fast Fourier Transform and periodogram. The obtained frequency characteristics allow classifying experimental runs into two categories of high and low SA in the fourth stage.

• Using our methodology on the new data set, the most important spatial areas and frequency ranges related to SA are identified.

Section 2 provides technical details of the experimental part of the procedure, and the detailed description of the employed algorithms for data processing are given in Section 3. The results are presented in Section 4 followed by the discussion and concluding remarks in Sections 5 and 6.

2. Experimental procedure

2.1. PEBL’s Situational Awareness Test

The PEBL is a framework for psychological assessments [20], which contains a dynamic visual tracking test based on the Situational Awareness Global Assessment Technique (SAGAT) proposed by Endsley and Garland [21], referred to as SA test. In the SA test, five animals of different shapes and colours are moving continuously over a grid. The participants are asked to monitor the animals’ positions, types and direction of movement within the grid, and answer certain questions after some of the animals disappear from the screen. The SA tests are evaluated according to the accuracy metrics:

1. In order to test participant’s perception (SA Level 1), they are asked to mark the last location of all five animals and the average position error is recorded (see Figure 2a).

2. To test the degree of participant’s comprehension of the situation (SA Level 2), they are asked to identify the type of animal located on two given positions (see Figure 2b). The output includes two binary values specifying if the animals are identified correctly.

3. To evaluate the degree of participant’s projection (SA Level 3), they are asked to specify the most recent movement direction of one of the animals, and the angular error is recorded.

A full experiment consists of five blocks and 90 runs that can be used for detailed analysis: Block 0 is a practice session with three sequential runs of the same assessment. Data from this block is not taken into account for further analysis. Blocks 1-3 consist of 15 consecutive runs of the same assessment and the participants are aware of the questions to follow while watching the simulations (45 runs in total). Block 4 features 15 runs of all three assessments in a random order and the participant are unaware of the questions to follow (45 runs in total).

2.2. Participants

Thirty two subjects (13 female & 19 male) between ages of 18 and 39 participated in the study (Mean = 27.66, SD = 5.46). All participants reported good health and were not sleep-deprived while carrying out the experiment.

2.3. EEG Equipment and Data Acquisition

A 32 channel g.tec g.Nautilus wireless EEG headset with gold-plated dry electrodes was used to measure the brain activity of all participants with a frequency of 250 Hz (see top left image in Figure 1). The EEG channels were placed according to the International 10-20 system [11]. A high-pass filter is applied to remove slow frequency drift (> 2 Hz) related to environmental influences. Also, frequencies between 48 Hz and 52 Hz are removed using a notch filter in order to avoid interfaces with the power line. The brain activity was recorded throughout the whole duration of the SA test.
3. Methodology

The structure of this section is as follows: the data obtained from the SA test is divided into two categories of high and low SA in Section 3.1. Subsequently, the SA test labels and EEG data are time synchronised. The data slices with timestamps corresponding to assessment instances are retrieved in Section 3.2. The results are used to obtain spatio-spectral features in Section 3.3. Finally, two algorithms based on decision trees are employed to classify the extracted features into two categories of high and low SA in Section 3.4.

3.1. Label definition with PEBL SA test data

In the first stage, samples corresponding to each type of assessment explained in Section 2.1 are supposed to be labelled into two classes of high and low SA. Assuming that position and angular errors corresponding to both degrees of participants’ SA are normally distributed, a Gaussian mixture model with two components is fitted over all the data points obtained from the tests on SA Level 1 and 3 (see Figure 2). The threshold between the two components is used to distinguish high and low SA here.

The output of the assessment of SA Level 2 consists of two binary values specifying if the animals were identified correctly or not. Since the probability of correctly identifying at least one of the two animals by choosing them randomly is high (40%), only two correct answers are classified as high SA. All other runs are set to low SA.

3.2. Construction of data samples

Both EEG and SA test data are timestamped using one computer, and the EEG data is mapped onto the results of the SA tests. The maximal offset between the data points recorded from EEG and the SA test is 2 ms, which is based on the temporal resolution of the EEG equipment (250 Hz). The EEG data covering the time prior to the moment when a question was posed and only when the animals were moving on the screen, was isolated. Assuming that participants’ SA is independent between the runs, the outcome of each test is treated as an individual sample. In total 2880 labelled samples (32 trials × 90 runs) are acquired.

The dimension of the retained data $y$ for a single sample is $y \in \mathbb{R}^{p \times N}$, where $p$ corresponds to the number of EEG channels and $N$ is the number of time instances ($N = 250\tau$). In this study, we have 32 channels, the frequency of data acquisition is 250 Hz and $\tau$ is time duration when the animals move prior to posing a question in seconds such that $\tau \sim U(2,4)$.

3.3. Feature extraction using ICA and PCA

3.3.1. Component Analysis

The cognitive processes are not associated with only one part of the brain as a single stimuli affects multiple spatial locations of a brain. Also, the EEG signal of a single channel consists of a superposition of different stimuli or sources. The goal is to extract features which are related primarily to different SA levels and map them on various regions of the brain. In order to decompose the superposed signal into its constituent stimuli and provide a spatial imprint of each source, Independent Component Analysis (ICA) [22] and Principal Component Analysis (PCA) [23] are used here.

For $p$ EEG channels and $N$ time instances, both ICA and PCA aim to compute matrix $W \in \mathbb{R}^{p \times p}$ so that the EEG measurements $y$ can be expressed as a linear combination of the underlying source signals $s \in \mathbb{R}^{p \times N}$, such that

$$y = Ws. \quad (1)$$

In (1), the $i$-th row of $y$ corresponds to a measurement from a single channel, whilst the $i$-th row of $s$ corresponds to time series associated with single source or stimuli. Equation (1) approximates a single channel measurement as a weighted summation of all identified sources. Also, the imprint the $i$-th source has on brain locations can be examined using the magnitude of the values in $i$-th column of $W$, which, in essence, gives spatial structure associated with a specific stimulus. The higher the magnitude of the coefficient in $W$, the larger is the impact of a stimuli on the corresponding brain location.

Both ICA and PCA aim to find values of $W$ and $s$, which best approximate given $y$. Despite similar linear formulation, the two algorithms use different objectives to identify an appropriate form of $s$. ICA relies on the assumption that the signals are
non-Gaussian and statistically independent [22]. The decorrelation technique of PCA outputs sources such that it minimises the reconstruction error if $r$ ($r < p$) components are used, i.e., $\min_{s,W} ||y - Ws||_F$, where $W \in \mathbb{R}^{p \times r}$, $s \in \mathbb{R}^{r \times N}$ and $|| \cdot ||_F$ denotes the Frobenius norm [23]. The principal components can be ranked based on the explained variance, and for EEG measurements, this results in the first principal components showing areas of the brain with largest changes in activation.

Data from all participants is standardised such that:

$$ z = \frac{(y - \mu)}{\sigma} \quad (2) $$

Here, $\mu$ is the mean and $\sigma$ is the standard deviation of the samples to approximate a standard normally distributed data set. For both ICA and PCA, the values of $W_{ICA}$ and $W_{PCA}$ are computed using the standardised data from all participants, $z$, whilst $s_{ICA}$ and $s_{PCA}$ are calculated based on $z$ corresponding to each sample, e.g. $s_{PCA} = W_{PCA}^{-1}z$.

### 3.3.2. Spectral transformation

The most commonly used EEG features are identified based on the Power Spectral Density (PSD) of different EEG bands which gives information about the amplitude distribution of different frequency bands across the brain [24][25]. The PSDs of five EEG bands $\delta$ (0-4 Hz), $\theta$ (4-8 Hz), $\alpha$ (8-12 Hz), $\beta$ (12-30 Hz), and $\gamma$ (30-45 Hz) are of most interest as they are demonstrated to provide statistically significant information in multiple applications [26][27].

In this paper, two different approaches are used to extract frequency information for different brain regions: a Fast Fourier Transform (FFT) and periodogram with a window function. The FFT is based on Discrete Fourier Transform (DFT) of the original time series which outputs $X_\omega \in \mathbb{C}$ giving information about amplitude and phase of the oscillations at frequency $\omega$. The PSD can be approximated using a magnitude of the output $|X_\omega|^2$. Periodogram estimates PSD by taking the Fourier transform of the auto correlation function of the original time series [28]. In order to reduce the sensitivity of the output to small variations in time series, a sequence is multiplied with window functions to obtain multiple shorter subsequences, and the periodogram is computed for each of them. The final output is computed by calculating the average. Here, the flat top window is used to construct window functions. Although, periodogram preprocessed with a window function is known to be more robust to noise, some useful information might be lost through averaging. Consequently, both datasets obtain using FFT and periodogram are considered in further analysis.

The relationships between different brain areas are included in $s_{ICA}$ and $s_{PCA}$ which are obtained by taking a weighted sum of time series corresponding to multiple channels. Both FFT and periodogram are applied to $z$, $s_{ICA}$ and $s_{PCA}$ to give $|X_\omega|^2$. $\omega$ is assigned with discrete values for $\omega \leq 125$ Hz ($\omega = 0.2k$, where $k \in \mathbb{Z}^+$).

The total magnitude of each band is defined as a sum of $|X_\omega|^2$ corresponding to constituent frequencies, e.g. $|X_\omega|^2 = \sum_{\omega \in \omega} |X_\omega|^2$. Additionally, the power spectrum ratio is computed as in Equation (3) [12]:

$$ PSR = \frac{|X_\alpha|^2}{(|X_\alpha|^2 + |X_\delta|^2)} \quad (3) $$

This procedure is applied to $z$, $s_{ICA}$ and $s_{PCA}$ resulting in 576 features (6 frequency features $\times$ 32 time series $\times$ 3 types of signals) for both FFT and periodogram.

#### 3.3.3. Outlier detection and removal

The EEG measurements recorded using dry electrodes are prone to noise and artefacts which create outliers among features. Many machine learning algorithms are sensitive to outliers during the training phase, as a single outlier can affect the classification model. Therefore, anomalous values have to be detected and subsequently replaced or removed before the data set is fed into the training algorithm.

If the samples with big magnitudes are compared with corresponding time series, as shown in Figure 3 they usually correspond to high amplitude oscillations characterising artefacts. Instead of discarding samples where at least one of the features has an unusually large value, outlier identification is performed individually for each feature. This is based on the assumption that an artefact does not necessarily impact all frequency ranges and all EEG channels equally. In fact, in the example presented in Figure 3 the PSD is affected by an artefact only for frequencies smaller than $\approx 30$ Hz.

The outliers can be identified using statistical or algorithmic approaches (e.g. Isolation Forest). In this study both approaches are used. In a statistical method, the threshold $\Delta$ is defined based on properties of the sample distribution, such as standard deviation, and the values larger than $\Delta$ are denoted as outliers. Three statistical outlier detection methods are used here with $\Delta$ set to: 1) $\Delta = \mu + \sigma$, 2) $\Delta = \mu + 2\sigma$ and 3) $\Delta = \bar{x}$, where $\mu$, $\sigma$ and $\bar{x}$ are mean, standard deviation and median of a feature $x$.

In the Isolation Forest (IF) approach which is used here, a split value between the minimal and maximal value is randomly chosen creating two subsets of the data set. This procedure is
repeated until all subsets consist of a single data point or only of identical data points. The maximum amount of splits required to separate a data point is used as a parameter to denote outliers, since anomalous points need fewer splits to be isolated [29].

Once the outliers are identified, they can be replaced with substitute values in the procedure referred to as imputation. In this study, the outliers are imputed with a missing value token (NaN), 0, $\mu$, and max $x$. Separate classification models are calculated for each of the possible imputations and data with no outlier replacement.

3.4. Classification with Random Forest and Boosted Trees

Two algorithms based on Decision Trees are used to classify EEG data: Random Forest (RF) and Boosted Trees (BT). Decision tree is a flowchart-like model in which a node represents a conditional test, branches depict an outcome of a test, and the leaf (lowest branch of the tree) is associated with one of the classes. The given input is propagated along the tree based on the result of conditional statements at the nodes until it reaches one of the leaves. The decision tree-based algorithms allow highly complex, non-linear decision boundaries between different categories. Also, they are easily interpretable and allow identification of the most statistically significant features where these characteristics are prevalent among conditional splits in the top nodes of the model [30].

When the number of nodes in a decision tree is left unconstrained, the algorithm tends to overfit the training data set. Thus, RF is used here to create multiple decision trees using a subset of the original data set. New data points are classified by all created trees based on a majority vote. The algorithm relies on the assumption that even though constituent trees might overfit to the training data, the generalisation of the model improves the results when the average is taken.

The BT algorithm addresses the underfitting issue of a model, where the classification accuracy is low on the training data. An example of such a model, is a shallow decision tree with small number of nodes. Shallow decision trees are referred to as weak learners and their accuracy is slightly better than classifier which chooses the output randomly. The underlying idea is to construct many weak learners and combine them in order to boost the accuracy of the overall model and at the same time prevent overfitting [31].

In both algorithms, it is possible to determine which features have a higher impact on the classification. One criterion is to calculate how much the impurity of the training data (i.e. ratio of multiple classes in each leaf) is reduced due to split along a specific feature. For example, if a feature separates a set of samples into subsets with no label overlap, this characteristic would be considered more important. Furthermore, the criterion is proportional to the amount of samples that are segregated, i.e. the feature which well isolates $n$ samples has less impact than the feature which is able to segregate $n+1$ samples. Thus, features which determine a conditional split in the top nodes have significantly higher impact on the impurity in the subsequent trees and usually are deemed more dominant [30].

4. Results

4.1. PEBL data classification

The histogram of the average error for 1056 runs in the assessment of SA Level 1 is shown in Figure [31]. The threshold between the two components of the fitted Gaussian mixture is 0.49 hpx which is used to distinguish high and low SA. Consequently, 76.2% of the data set are labelled as high SA and 23.8% are labelled as low SA.

In the assessment of SA Level 2, 64.6% of the participants provided two correct answers and their data points are labelled as high SA, while the remaining 35.4% are labelled as low SA.

The results of the assessment on SA Level 3 are shown in Figure [32]. The threshold between the two components of the fitted Gaussian mixture is 37.43° which splits the data set into 62.3% high SA and 37.7% low SA.

![Figure 4: Two-component Gaussian mixtures are fitted on the position and angular errors in order to identify the threshold for high and low SA.](image)

4.2. Component Analysis

The PCA is conducted using data from all participants and based on the number of EEG channels. 32 principal components are computed. The three most important components for all participants and a single participant are shown in Figures [5](a) and [5](b) respectively.
Component 1  Component 2  Component 3
(a) All participants

Component 1  Component 2  Component 3
(b) Single participant

Figure 5: The magnitude of coefficients in three most important components identified through PCA for (a) all participants and (b) one participant.

The most dominant PCA structures in Figures 5 represent the areas of largest changes in the brain activity. It can be seen that the activation patterns for a single participant differ from the average response, which indicates the effect of individual brain plasticity that can encumber classification accuracy.

ICA is conducted using data from all participants. Contrary to PCA, the identified sources are not ranked in ICA, hence no clear criterion of importance can be used to choose a few dominant components. Therefore, the discussion on the spatial characteristics of the generated ICA components is not included in this paper.

4.3. Outlier detection

The ratio of values identified as outliers for 576 features using FFT and periodogram are shown in Figure 6 using boxplots. It can be seen that the smallest proportion of samples is identified as outliers when $\Delta = \mu + 2\sigma$ and the highest when $\Delta = 2x$. Also, a smaller percentage of samples are identified as outliers when applying periodogram. Since a more uniform distribution of values are identified as outliers using IF, the output of this algorithm is used for further analysis. This also allows to retain a higher percentage of the original data.

Figure 6: Percentage of values identified as outliers among 576 features using FFT and periodogram.

Obtained after using periodogram, the parameters are assigned as $M = 100$, $\eta = 0.1$ and $l = 3$

For RF, (1) the number of trees, (2) the minimum amount of samples in a leaf, (3) the minimum required amount of samples in a leaf to conduct another split and increase the depth, and (4) the number of layers in the trees are examined and only the number of trees is considered as the hyperparameter. The other three parameters are left unconstrained leading to creation of decision trees with high number of nodes. This approach provides a balance between low bias of individual decision trees and variance reduction when the number of decision trees is increased. The number of trees is set to 1000 to maximise the test accuracy.

Table 1 shows the train accuracy, test accuracy, precision and recall for RF and BT, and features identified using FFT and periodogram. Train accuracy and test accuracy describe the percentage of correctly classified samples for the training data set and testing data set respectively. Precision shows the ratio of correctly classified features (true positives, TP) over all retrieved features (including false positives, FP) as in (4):

$$\text{Precision} = \frac{TP}{TP + FP}$$ (4)

Recall describes the ratio of correctly classified features (TP) over all relevant items (including false negatives, FN) as in (5):

$$\text{Recall} = \frac{TP}{TP + FN}$$ (5)

It can be seen that the highest test accuracy of 67% is achieved for BT with outliers replaced by mean values. This method also features the highest value for precision (66.5%), whereas the highest recall is reported for BT with outliers replaced by zeros (85.4%). The maximum values for test accuracy, precision and recall are slightly lower for features extracted with FFT (66.8%, 65.8%, 87%). The highest reported accuracy is achieved for BT with outliers replaced by a maximal value (66.8%). For RF, all reported test accuracies are lower compared to BT. Additionally, no analysis has been carried out with outliers imputed with NaN values, as they are not supported by the used RF algorithm.
5. Discussion

The division of SA test data into two categories with the threshold set to the intersection of two PDF curves constituting Gaussian mixture resulted in data imbalance where two thirds of samples were classified as high SA and one third as low SA (see Figure 4). Since no participants reported any tiredness or fatigue prior to the test, a good overall SA is expected. Fluctuation of SA within a single run can be explained with momentary distractions and shifts in concentration of the subject. For the assessment of SA Level 1, the Gaussian mixture curve fits the data set well (error in Cumulative Distribution Function (CDF) is 7.49%), however, the two clusters are not well separated with mean values of PDF\(_{High\ SA}\) and PDF\(_{Low\ SA}\) being close to each other. The Gaussian mixture also fits well the data points corresponding to the assessment for SA Level 3 (error in CDF is 5.09%), and the difference between the mean values of the PDFs is large indicating that the data set can be easily separated into two data sets. The approach of assigning labels to the results of the SA Level 2’s test also show a similar ratio between high and low SA for the assessment of SA Levels 1 and 3.

The data set containing features extracted using a combination of periodogram with window function mostly achieves a lower ratio of outliers compared to the one employing FFT (see Figure 6). This result is expected since the preprocessing of data using window function is designed to ensure that the resulting PSD is more robust to slight changes of timeseries and consequently is less sensitive to noise. In the methodology where \(\Delta = 2\), for some features over 60% of the samples are excluded, drastically reducing the amount of useful information. This can be explained by examining the distributions of features which in most cases are similar to log-normal distribution, where the median is always smaller (sometimes significantly) than the mean.

Despite artefacts being known to have a high impact on a broad range of applications when EEG measurements are used, the outliers do not impact the accuracy of the classification in this study. This is due to the fact that models based on decision trees are significantly less sensitive to high absolute values both during training and testing, when compared to methods where the model contains terms directly proportional to the input (e.g. logistic regression). In models based on decision tress, the output is decided only based on the condition of an input being greater than a certain threshold value. Thus, for example, replacing a value in the data set with another value at the upper range of the distribution, e.g. the maximum of the retained samples, results in the same output as the original high magnitude input.

It is notable that despite the train accuracy of 100% for classification with RF (see Table 1), the model gives a comparable performance on the test set to BT algorithm for example (train accuracy between 75 – 90%). This is because despite each constituent decision tree overfitting the data, once the average results based on “vote of majority” is used, the output is more generalisable. Furthermore, models with alternative values of hyperparameters do not provide better results.

The obtained RF model gives us insights into the most influential features. The identified six most important features are shown in Figure 7. Four of these features lie in the \(\beta\) band constituting of higher frequencies (12 – 30 Hz) which indicates an increase of attention and alertness of the subject with high activation [33]. The other presented features lie in \(\gamma\) band (30 – 45 Hz), responsible for cognitive functioning and information processing [34]. Figure 7 shows both left (components 2 and 6) and right hemisphere (components 1, 3, 4, and 5) are responsible for SA.

![Figure 7: The magnitude of coefficients in spatial structures corresponding to six statistically most significant features in the classifier with the highest test accuracy.](image)

<table>
<thead>
<tr>
<th>Test</th>
<th>BT NaN</th>
<th>BT 0</th>
<th>BT (\mu)</th>
<th>BT max (x)</th>
<th>BT outlier</th>
<th>RF 0</th>
<th>RF (\mu)</th>
<th>RF max (x)</th>
<th>RF outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train accuracy</td>
<td>0.901</td>
<td>0.753</td>
<td>0.878</td>
<td>0.772</td>
<td>0.791</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Test accuracy</td>
<td>0.651</td>
<td>0.655</td>
<td>0.670</td>
<td>0.653</td>
<td>0.653</td>
<td>0.635</td>
<td>0.660</td>
<td>0.656</td>
<td>0.663</td>
</tr>
<tr>
<td>Precision</td>
<td>0.652</td>
<td>0.644</td>
<td>0.665</td>
<td>0.649</td>
<td>0.646</td>
<td>0.630</td>
<td>0.650</td>
<td>0.656</td>
<td>0.655</td>
</tr>
<tr>
<td>Recall</td>
<td>0.807</td>
<td>0.854</td>
<td>0.826</td>
<td>0.826</td>
<td>0.839</td>
<td>0.842</td>
<td>0.848</td>
<td>0.811</td>
<td>0.839</td>
</tr>
<tr>
<td>Test accuracy</td>
<td>0.822</td>
<td>0.902</td>
<td>0.771</td>
<td>0.852</td>
<td>0.750</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Precision</td>
<td>0.633</td>
<td>0.627</td>
<td>0.622</td>
<td>0.668</td>
<td>0.653</td>
<td>0.649</td>
<td>0.644</td>
<td>0.649</td>
<td>0.658</td>
</tr>
<tr>
<td>Recall</td>
<td>0.629</td>
<td>0.630</td>
<td>0.615</td>
<td>0.652</td>
<td>0.643</td>
<td>0.644</td>
<td>0.638</td>
<td>0.649</td>
<td>0.658</td>
</tr>
<tr>
<td>Test accuracy</td>
<td>0.839</td>
<td>0.804</td>
<td>0.863</td>
<td>0.870</td>
<td>0.854</td>
<td>0.832</td>
<td>0.839</td>
<td>0.814</td>
<td>0.807</td>
</tr>
</tbody>
</table>

Table 1: Train accuracy, test accuracy, precision, and recall for all outlier imputations for BT/RF and periodogram/FFT
Also, all four main lobes, frontal lobe (component 2), parietal lobe (components 5 and 6), occipital lobe (components 1 and 3), and temporal lobe (component 4) are activated in the presented features in Figure 7. All of the identified brain regions are known to carry out tasks needed to acquire SA. The right hemisphere is known to positively correlate with vigilance for simple tasks [35], and intuition [36] whereas logical and linear thinking are attributed to the left hemisphere [37]. However, results on brain lateralisation are not universally agreed on [38]. The frontal lobe in the brain is, among others, responsible for concentration, spatial abilities, and short-term memory, while tasks carried out in the parietal lobe include spatial and visual perception, and memory tasks [39]. The occipital lobe interprets vision, e.g. colour, light, and movement, and the temporal lobe sequences and organises visual and auditory input and carries out reasoning tasks [39].

Looking at a similar study that used a 128-channel Geodestic Sensor Net (GSN) with semi-wet electrodes, loss of SA showed a strong activation of neurons in the orbitofrontal cortex [40]. The frequency band is not stated in this paper. Using 128-channel GSN EEG sensors, 7 electrodes are placed on the orbitofrontal cortex which are prone to artefacts induced by eye movement or blinking. Also, semi-wet electrodes allow for higher quality signal collection in comparison to dry electrodes. We did not obtain similar results in our study which can be explained by using dry electrodes and having two electrodes covering the areas close to the orbitofrontal cortex (FP1 and FP2).

In [41], activation in $\theta$, $\alpha$, and $\beta$ band decreased with loss of SA. The results are based on data collected using a 14-channel semi-wet EEG headset with 10 participants and for a task-specific application. This is in line with the results of our study, as we found a correlation between $\beta$ band and SA. An earlier investigation [24] also revealed a positive correlation between the power spectrum in the $\theta$ band and SA, however, their sample size with eight participants is small and no brain regions are identified.

6. Conclusions and Future Work

A novel analytical methodology for correlating physiological signals using EEG to Situational Awareness was presented in this paper. A new data set from 32 participants completing the SA test in the PEBL framework was collected using a 32-channel dry-EEG headset. A correlation has been found between the $\beta$ (12 – 30 Hz) and $\gamma$ (30 – 45 Hz) frequency bands and SA. The observed activation of neurons occurred in the four main lobes of the brain (frontal, parietal, occipital and temporal cortex). The combination of these frequencies and brain regions are responsible for concentration and visuo-spatial abilities, which are known to be important in order to build up SA. The highest achieved accuracy of the classifier is 67%.

In future work, a higher number of participants will be invited to conduct the experiment in order to construct separate classifiers for each of SA levels. Thus, the impact of the amount of SA for each level could also be assessed. Furthermore, the proposed method relies on the assumption that data can be split into high and low SA using a Gaussian mixture. Other methods incorporating different underlying distributions needs to be explored including division of the SA test results into larger number of classes to differentiate between degree of various SA Levels. Finally, different machine learning algorithms, such as Support Vector Machine and neural networks, could be assessed for improving the classification accuracy.

Acknowledgement

This work is supported by the Engineering and Physical Sciences Research Council (grant numbers EP/R037795/1 and EP/S01800X/1). The authors also express appreciation to the g.tec’s team, especially Emlyn Clay and Matt Pitkén for their fantastic support with the equipment.

This publication is using a new dataset which is available at: https://rdr.ucl.ac.uk/. Please contact b.anvari@ucl.ac.uk for any further inquiries regarding the data.

References


