ACCEPTED MANUSCRIPT

A study on CNN image classification of EEG Signals represented in 2D and 3D

To cite this article before publication: Jordan J. Bird *et al* 2021 *J. Neural Eng.* in press https://doi.org/10.1088/1741-2552/abda0c

A Study on CNN Image Classification of EEG Signals represented in 2D and 3D

Jordan J. Bird[1], **Diego R. Faria**[2], **Luis J. Manso**[3], **Pedro P.S. Ayrosa**[4], **and Anikó Ekárt**[5]

[1,2,3]Aston Robotics Vision and Intelligent Systems Lab (ARVIS Lab), Aston University, United Kingdom

E-mail: {birdj1[1], d.faria[2], l.manso[3]}@aston.ac.uk

[4] Universidade Estadual de Londrina, Londrina, Brazil

E-mail: ayrosa@uel.br

[5] School of Engineering and Applied Science, Aston University, United Kingdom

E-mail: a.ekart@aston.ac.uk

August 2020

Abstract.

Objective: The novelty of this study consists of the exploration of multiple new approaches of data pre-processing of brainwave signals, wherein statistical features are extracted and then formatted as visual images based on the order in which dimensionality reduction algorithms select them. This data is then treated as visual input for 2D and 3D CNNs which then further extract 'features of features'.

Approach: Statistical features derived from three electroencephalography datasets are presented in visual space and processed in 2D and 3D space as pixels and voxels respectively. Three datasets are benchmarked, mental attention states and emotional valences from the four TP9, AF7, AF8 and TP10 10-20 electrodes and an eye state data from 64 electrodes. 729 features are selected through three methods of selection in order to form 27x27 images and 9x9x9 cubes from the same datasets. CNNs engineered for the 2D and 3D preprocessing representations learn to convolve useful graphical features from the data.

Main results: A 70/30 split method shows that the strongest methods for classification accuracy of feature selection are One Rule for attention state and Relative Entropy for emotional state both in 2D. In the eye state dataset 3D space is best, selected by Symmetrical Uncertainty. Finally, 10-fold cross validation is used to train best topologies. Final best 10-fold results are 97.03% for attention state (2D CNN), 98.4% for Emotional State (3D CNN), and 97.96% for Eye State (3D CNN).

Significance: The findings of the framework presented by this work show that CNNs can successfully convolve useful features from a set of pre-computed statistical temporal features from raw EEG waves. The high performance of K-fold validated algorithms argue that the features learnt by the CNNs hold useful knowledge for classification in addition to the pre-computed features.

Classification of EEG Signals represented in 2D and 3D 2

1. Introduction

Recent advances in consumer facing technologies have enabled machines to have non-human skills. Inputs which once mirrored one's natural senses such as vision and sound have been expanded beyond the natural realms [1]. An important example of this is the growing consumerist availability of the field of electroencephalography (EEG) [2,3]; the detection of thoughts, actions, and feelings from the human brain. To engineer such technologies, researchers must consider the actual format of the data itself as input to the machine or deep learning models, which subsequently develop the ability to distinguish between these nominal thought patterns. Usually, this is either statistically 1-Dimensional or temporally 2-Dimensional since there is an extra consideration of time and sequence. Due to the availability of resources in the modern day, a more enabled area of research into a new formatting technique is graphical representation, i.e., presenting the 1-Dimensional mathematical descriptors of waves in multiple spatial dimensions in order to form an image or model in 3D space. This format of data can then be further represented by feature maps from convolutional operations. With preliminary success of the approach, a deeper understanding must be sought in order to distinguish in which spatial dimension brainwave signals are most apt for interpretation. With the classical method of raw wave data being used as input to a CNN in mind, dimensionality reduction is especially difficult given the often blackbox-like nature of a CNNs internal feature extraction processes [4]. In this work, we extract statistical temporal features from the waves which serve as input to the CNN, which allows for direct control of input complexity since dimensionality reduction can be used to choose the best n features within the set with the task in mind. Reduction of a CNN topology, whether that be network depth or layer width, gives less control over which features are and are not computed. Given the technique of feature extraction as input to the CNN, and thus the aforementioned direct control of input complexity, reduction of CNN complexity reduces the number of *'features of features'* computed; that is, all the chosen input attributes are retained.

The remainder of this report is structured as follows. Firstly, the remainder of this section outlines the scientific contributions of this work. In Section 2, technical background and related scientific works are presented and discussed. Following the background and related works, Section 3 then provides details of the methodology of the experiments performed during this study. Section 4 then reports the results of the experiments, along with comparison to related state-of-the-art scientific knowledge. Finally, Section 6 provides an outline for suggestions of future work and presents the final notes and conclusions from the study.

1.1. Scientific Contributions

In this work, an experimental framework is presented in which evolutionary optimisation of neural network hyperparameters is applied in conjunction with a visual data pre-processing technique preliminarily explored in a previous work. During the previous

study [5], a 2D CNN was succesfully applied to a 2D image representation of EEG features with a dimensionality reduction algorithm on a 4-channel EEG dataset. In this work, we explore visual data reshaping in 2 and 3 dimensions in order to form pixel image and voxel cube representations of statistical features extracted from electrical brain activity, through which 2D and 3D CNN convolve *'features of features'*. In addition, we also explore multiple methods of dimensionality reduction and describe their relationships to both the general classification ability of the model as well as the reshaping technique. In comparison to previous works on both attention (concentrating/relaxed) and emotional (positive/negative), many of the techniques explored in this study produce competitive results. Finally, the application to other EEG devices is shown by the application of the method to an open-source dataset. We apply the three 2D and 3D approaches to classification to a 64-channel EEG dataset acquired from an OpenBCI device, which achieves 97.96% 10-fold mean classification accuracy on a difficult binary problem (Eyes open/closed), arguing that the approach is dynamically applicable to BCI devices of higher resolution and for problems other than the frontal lobe activity classification in the first two experiments. This both suggests some future work with other devices, as well as collaboration between research fields in order to build on and improve the framework further.

2. Background and Related Works

In this section, the technical philosophies of the related Scientific fields are outlined, as well as important works that are related to the experiments carried out throughout this paper.

2.1. Electroencephalography

Electroencephalography is the process of using electrodes applied to the cranium in order to measure electrical signals produced by the brain [6, 7] due to the nervous oscillations caused by certain hormonal balances such as serotonin, dopamine and noradrenaline. Electrodes can be placed invasively or subdurally under the skull and directly on to the brain itself [8]. Other electrodes are able to read bioelectrical signals from on the surface of the head and are thus less invasive; via either Electro-Gel wet electrodes or simply placed dry electrodes [9]. The signal strength of the raw electrical data is recorded sequentially, producing what is known as a 'brainwave'.

The Muse EEG headband is comprised of four dry electrodes placed on the TP9, AF7, AF8 and TP10 placements. Muse operates an on-board artefact separation algorithm in order to remove the noise from the recorded data [10]. The muse streams over Bluetooth Low Energy (BLE) at around 220Hz, which we reduce to 150Hz in order to make sure that all data collected is uniform. Muse has been used in various Brain-computer interface projects since its introduction in May 2014. They have been

Classification of EEG Signals represented in 2D and 3D 4

particularly effective for use in neuroscientific research projects, since the data is of relatively high quality and yet the device is both low-cost and easy to use since it operates dry electrodes. This was shown through an exploration into Bayesian binary classification [11]. Sentiment analysis via brainwave patterns has been performed in a process of regression in order to predict a user's level of enjoyment of a performed task [12,13]. The works were shown to be effective for the classification of enjoyment of a mobile phone application. The Muse produces bipolar readings from the four electrodes with the AFz placement as a reference. According to the technical specifications‡, the signals are oversampled and then downsampled to yield a the output, and the sampling has a 2uV (RMS) noise. The noise is suppressed via the Driven-Right-Leg/Reference feedback configuration using the AFz sensor. A Notch filter of 50Hz is applied to the raw waves since the experiment was performed in the United Kingdom.

Attention state classification is a widely explored problem for statistical, machine and deep learning classification [14,15]. Common Spatial Patterns (CSP) benchmarked at 93.5% accuracy in attention state classification experiments, suggesting it is possibly one of the strongest state-of-the-art methods [16]. Researchers have found that binary classification is often the easiest problem for EEG classification, with Deep Belief Networks (DBN) and Multilayer Perceptron (MLP) neural networks being particularly effective [17–19]. The best current state-of-the-art benchmark for classification of emotive EEG data achieves scores of around 95% classification accuracy of three states, via the Fisher's Discriminant Analysis approach [20]. The study noted the importance of the prevention of noise through introducing non-physical tasks as stimuli rather than those that may produce strong electromyographic signals. Stimuli to evoke emotions for EEG-based studies are often found to be best with music [21] and film [22,23].

OpenBCI, used in the 64-channel extension of this study, is an open-source Brain-computer interface device, which has the ability to interface with standard Electroencephalographic [24], Electromyographic [25], and Electrocardiographic [26] electrodes. OpenBCI with selected electrodes has seen 95% classification accuracy of sleep states when discriminative features are considered by a Random Forest model in the end-to-end system Light-weight In-ear BioSensing (LIBS) [27]. In this study, OpenBCI data is used to detect eye state, that is, whether or not the subject has opened or closed their eyes. In addition to the obvious nature of muscular activity around the eyes, according to Brodmann's Areas, the visual cortex is also an indicator of visual stimuli [28, 29], and thus a higher resolution EEG is recommended for full detection. In [30], researchers achieved an accuracy of 81.2% of the aforementioned states through a Gaussian Support Vector Machine trained on data acquired from 14 EEG electrodes. It was suggested that with this high accuracy, the system could be potentially used in the automatic switching of autonomous vehicle states from manual driving to autonomous, in order to prevent a fatigue-related accident. A related work

‡ Additional technical detail on the Muse can be found at http://developer.choosemuse.com/hardware-firmware/hardware-specification

Classification of EEG Signals represented in 2D and 3D 5

found that K-Star clustering enabled much higher classification accuracies of these states to around 97% [31], but it must be noted that only one subject was considered and thus generalisation and further use beyond the subject would be considered difficult when generalisation works are considered [32, 33]; in this study, ten subjects are considered. In a similar dataset as seen in this work, researchers found that K-Nearest Neighbour classification (where $k = 3$) could produce a classification accuracy of 84.05% [34]. In the classification problem of the states of eyes open and closed (a binary classification problem), a recent work found that statistical classification via 7-nearest neighbours of the data following temporal feature extraction achieved a mean accuracy of 77.92% [35]. The study extracted thirteen temporal features and found that wave kurtosis was a strong indicator for the autonomous inference of the two states.

2.1.1. Statistical Extraction of EEG for Deep Machine Learning Due to the temporal nature of the EEG waves, single point measures rarely harbor any useful classification accuracy and thus make weak datasets. In this work, statistical features are extracted through a sliding time-window approach [5, 36, 37] (https://github.com/jordan-bird/eeg-feature-generation). The EEG signal is divided into a sequence of windows of length one second, with consecutive windows overlapping by 0.5 seconds, e.g., $[(0s - 1s), (0.5s - 1.5s), (1s - 2s), \dots]$). Each time window is further halved and quartered, which are used to extract additional features.

In this work the following statistical features were generated for each time window via the process that can be observed in Algorithm 1 as in the previous aforementioned works, where $\mathbf{y}_k = [y_{k1}, \dots, y_{kN}]$, within which K are vectors of paired observations [5, 36, 37]:

- Considering the full time window:
  - The sample mean and sample standard deviation of each signal [38]:

$$\bar{y}_k = \frac{1}{N} \sum_{i=1}^{N} y_{ki} \tag{1}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y}_k)^2}{n - 1}} \tag{2}$$

  - The sample skewness and sample kurtosis of each signal [39]:

$$g_{1,k} = \frac{\sum_{i=1}^{N} (y_{ki} - \bar{y}_k)^3}{N s_k^3}, \tag{3}$$

$$g_{2,k} = \frac{\sum_{i=1}^{N} (y_{ki} - \bar{y}_k)^4}{N s_k^4} - 3. \tag{4}$$

  - The maximum and minimum value of each signal.

Classification of EEG Signals represented in 2D and 3D 6

Result: Features extracted from raw data for every w_t

User defined the size of the sliding window $w_t = 1s$;

Input: raw wave data;

Initialisation of variables $init = 1$, $w_t = 0$;

while *getting sequence of raw data from sensor ($> 1min$)* **do**

    if *init* **then**

        $prev\_lag = 0$;

        $post\_lag = 1$;

    end

    $init = 0$;

    for *each slide window ($w_t - prev\_lag$) to ($w_t + post\_lag$)* **do**

        Compute mean of all w_t values $y_1, y_2, y_3...y_n$; $\bar{y}_k = \frac{1}{N} \sum_{i=1}^{N} y_{ki}$;

        Compute asymmetry and peakedness by 3rd and 4th order moments skewness and kurtosis $g_{1,k} = \frac{\sum_{i=1}^{N}(y_{ki}-\bar{y}_k)^3}{N s_k^3}$ and $g_{2,k} = \frac{\sum_{i=1}^{N}(y_{ki}-\bar{y}_k)^4}{N s_k^4} - 3$;

        Compute the max and min value of each signal $w_{max}^t = \max(w_t)$ and $w_{min}^t = \min(w_t)$;

        Compute sample variances $K \times K$ matrix $\mathbf{S}$ of each signal **Compute** sample covariances of all signal pairs, $s_{k\ell} = \frac{1}{N-1} \sum_{i=1}^{N} (y_{ki} - \bar{y}_k)(y_{\ell i} - \bar{y}_\ell)$; $\forall\ k, \ell \in [1, K]$;

        Compute Eigenvalues of the covariance matrix $\mathbf{S}$, $\boldsymbol{\lambda}$ solutions to: $det(\mathbf{S} - \boldsymbol{\lambda}\mathbf{I}_K) = 0$, where $\mathbf{I}_K$ is the $K \times K$ identity matrix, and $det(\cdot)$ is the determinant of a matrix;

        Compute the upper triangular elements of the matrix logarithm of the covariance matrix $\mathbf{S}$, where the matrix exponential for $\mathbf{S}$ is defined via Taylor expansion $e^{\mathbf{B}} = \mathbf{I}_K + \sum_{n=1}^{\infty} \frac{\mathbf{S}^n}{n!}$, then $\mathbf{B} \in C^{K \times K}$ is a matrix logarithm of $\mathbf{S}$;

        Compute magnitude of frequency components of each signal via Fast Fourier Transform (FFT), magFFT(w_t);

        Get the frequency values of the ten most energetic components of the FFT, for each signal, getFFT(w_t, 10);

    end

    $w_t = w_t + 1s$;

    $prev\_lag = 0.5s$; $post\_lag = 1.5s$;

    Output Features Fw_t extracted within the current w_t

end

Algorithm 1: Algorithm to extract features from raw biological signals.

Classification of EEG Signals represented in 2D and 3D   7

- The sample variances $K \times K$ matrix $\mathbf{S}$ of each signal, plus the sample covariances of all signal pairs [38]:

$$s_{k\ell} = \frac{1}{N-1} \sum_{i=1}^{N} (y_{ki} - \bar{y}_k)(y_{\ell i} - \bar{y}_\ell);$$

$$\forall \, k, \ell \in [1, K]$$

(5)

- The eigenvalues of the covariance matrix [40] $\mathbf{S}$, which are the $\boldsymbol{\lambda}$ solutions to:

$$det(\mathbf{S} - \boldsymbol{\lambda}\mathbf{I}_K) = 0 \tag{6}$$

where $\mathbf{I}_K$ is the $K \times K$ identity matrix, and $det(\cdot)$ is the determinant of a matrix.
- The upper triangular elements of the matrix logarithm of the covariance matrix [41, 42] of the covariance matrix $\mathbf{S}$: where the matrix exponential for $\mathbf{S}$ is defined via Taylor expansion,

$$e^{\mathbf{B}} = \mathbf{I}_K + \sum_{n=1}^{\infty} \frac{\mathbf{S}^n}{n!}, \tag{7}$$

then $\mathbf{B} \in C^{K \times K}$ is a matrix logarithm of $\mathbf{S}$.:
- The magnitude of the frequency components of each signal, obtained using a Fast Fourier Transform (FFT).
- The frequency values of the ten most energetic components of the FFT, for each signal.

• With the above in mind, the following are calculated in regards to the 0.5s windows:
- The change in the sample means and in the sample standard deviations between the first and second half-windows, for all signals.
- The change in the maximum and minimum values between the first and second half-windows, for all signals.

• And finally, for the 0.25s windows:
- The sample mean of each each quarter-window, plus all paired differences of sample means between the quarter-windows, for all signals.
- The maximum (minimum) values of each quarter-window, plus all paired differences of maximum (minimum) values between the quarter-windows, for all signals.

Additionally, each data object is also given the features calculated in the previous window, bar those that would be identical. This allows for further temporal consideration. This data then follows the below process of attribute selection in order to reduce the number of attributes to one that can be reshaped into squares and cubes, in order to form the objects for the CNN to process. Note that not all features are specific to EEG, given that the algorithm is a general purpose feature extraction process for

Classification of EEG Signals represented in 2D and 3D 8

temporal wave data. Due to this, it is thus important to perform feature selection in order to isolate generated features that are useful for the specific problem in mind - in this case, features from this large set that may be useful for an EEG classification problem.

2.2. Attribute Selection

Attribute selection, or dimensionality reduction, is the process of reducing the dataset by features in order to simplify the learning process. Importantly, it is the focus of discarding weaker elements in order to simplify the process but at the smallest cost of classification ability [43–45]. In neural networks, for an example, large input datasets greatly increase the number of hyperparameters to be tuned by the optimisation algorithms and thus the computational resources required [46]. The three methods of feature selection chosen due to the findings of literature review are One Rule, Kullback-Leibler Divergence, and Symmetrical Uncertainty.

One Rule feature selection is the scoring of an attribute based on how well it can be branched to classify data based on the singular attribute [47]. *Kullback-Leibler Divergence*, or *Relative Entropy*, is the measure of how a feature set's probability distribution differs from another [48,49]. Finally, *Symmetrical Uncertainty* is the rating of attribute classification ability based on a mutual dependence, or lack thereof [50].

2.3. CNN and Visual Space Learning

Convolutional Neural Networks (CNN) are a form of Artificial Neural Network (ANN) which perform autonomous feature extraction from attributes based on their spatial positioning [51]. To perform this, data is convolved in order to form new maps from the original data, of which the connections to an interpretation Multilayer Perceptron (MLP) are considered parameters for loss-reducing optimisation [52]. The spatially-aware focus of pooling is inspired by the operations of the biological photo-receptors [53, 54]. The size of the window for this is known as the *'kernel'* and is a manual hyperparameter set pre-training, as well as the layers of convolutional operations themselves.

Visual Space learning, is the process of projecting data as a matrix and convolving with the above methods, but on unconventional graphical data formatted as such. Visual space learning in EEG is a relatively new approach, with most simply considering signal strengths interpolated where the centroid is relative to the electrode placement location [55, 56]. Recently, the static statistical descriptions of brainwaves have been found to be extremely effective when formed as an image and convolved to feature maps [5]. The preliminary method of graphical 2D Euclidean Space representations of brainwave signals is to be expanded further in these studies.

Classification of EEG Signals represented in 2D and 3D   9

2.4. Evolutionary Topology Search

Result: Array of best solutions at final generation
initialise *Random solutions*;
for *Random solutions : rs* **do**
    test accuracy of *rs*;
    set accuracy of *rs*;
end
set solutions = Random Solutions;
while *Simulating* **do**
    for *Solutions : s* **do**
        parent2 = roulette selected Solution;
        child = breed(s, parent2);
        test accuracy of *child*;
        set accuracy of *child*;
    end
    Sort *Solutions* best to worst;
    for *Solutions : s* **do**
        if *s index > population size* **then**
            delete s;
        end
    end
    increase maxPopulation by growth factor;
    increase maxNeurons by growth factor;
end
Return *Solutions*;

Algorithm 2: Evolutionary Algorithm for ANN optimisation [57].

Deep Evolutionary Multilayer Perceptron, or DEvoMLP is an approach to hyper-heuristically optimising a Neural Network topology through evolutionary computation [57, 58]. Networks are treated as individual organisms in the process where their classification ability dictates their fitness metric, thus it is a single-objective algorithm. The pseudocode for the algorithm is given in Algorithm 2. The process to combine two networks follows the aforementioned work, where the depth of the hidden layers is decided by selecting one of the two parents at random or mutation at a 5% chance. Then, for each layer, the number of neurons is decided by selecting the nth layer of either parent at random (provided both parent networks have an nth layer), again a 5% mutation chance dictates a random mutation resulting in the number of neurons being a random number between 1 and *maxNeurons*. To give an example of a process within the algorithm, a neural network $i, 64, 32, 16, o$ (where i are the input neurons, and o are the output neurons) which has three hidden layers of neurons $(64, 32, 16)$ and a second

Classification of EEG Signals represented in 2D and 3D   10
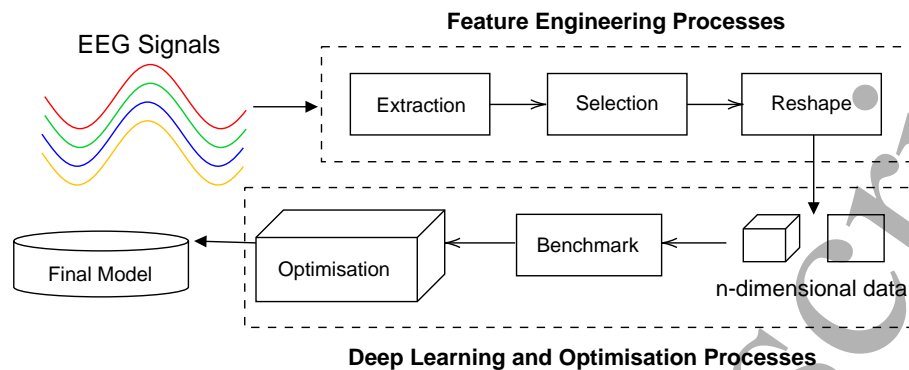

Figure 1. Overview of the Methodology. EEG Signals are Processed into 2D or 3D data Benchmarked by a 2D or 3D CNN. Three Different Attribute Selection Processes are Explored. Finally, the Best Models have their Interpretation Topologies Optimised Heuristically for a Final Best Result.

neural network $i, 100, 10, o$ are chosen as the two candidates to breed and create a neural net offspring. If, in this example, parent 2 is chosen to provide depth to the offspring, then the offspring topology would be i, x, y, o, and neuron counts x and y now need to be chosen. Layer x may be chosen from parent 1 and y from parent 2, creating an offspring neural network topology $i, 64, 10, o$ which has two hidden layers of 64 and 10 neurons respectively. Layers x and y could have both been chosen from parent 1 which would result in the offspring $i, 64, 32, o$ since it had the hidden depth of 2 from parent 2. Indeed, the breeding process can, and does, produce an offspring that is identical to one of the parents. Since we already know this fitness value, a random solution is generated instead.

Thus, after simulation, the goal of the DEvo algorithm is to derive a more effective neural network topology for the given dataset. The algorithm is implemented due to neural network hyperparameter tuning being a non-polynomial problem [59]. It is, of course, extremely complex; a ten population roulette breeding simulation executed for ten generations would produce 120 neural networks to be trained, since eleven are produced every generation. Resource usage is extreme for the simulation, but the final result gives a network topology apt for the given data, and can this finding can thus be used in other experiments.

3. Method

In this section, the method of these experiments are described. A diagram of the process described in this study can be seen in Fig. 1. Two datasets for the experiment are sourced from a previous study [36] which made use of the aforementioned Muse headband (TP9, AF7, AF8, TP10), see Section 2 for technical detail. Firstly, the *'attention state'* dataset (https://www.kaggle.com/birdy654/eeg-brainwave-dataset-mental-state), which is collected from four subjects; two

Classification of EEG Signals represented in 2D and 3D 11

Table 1. Class labels for the data belonging to the three datasets

Dataset	No. Classes	Labels
Concentration State	3	Relaxed, Neutral, Concentrating
Emotional State	3	Negative, Neutral, Positive
Eye State	2	Closed, Open

male, two female, at an age range of 20-24. The subjects under stimuli were either relaxed, concentrating, or from lack of stimuli, neutral. Three minutes per state are recorded for each subject, giving a total of thirty-six minutes of EEG brainwave data. The concentrating class is stimulated by the 'shell game' wherein the subjects must concentrate to follow the movement of a ball hidden under one of three cups which are switched around. The relaxed state is induced with classical music and is recorded several moments after the exercise begins, and the neutral state is finally recorded free of any stimuli.

In the second experiment, the *'Emotional State'* dataset (https://www.kaggle.com/birdy654/eeg-brainwave-dataset-feeling-emotions) is acquired. To gather this data, six minutes of EEG data are recorded from two subjects of ages 21 and 22. negative or positive emotions are evoked via film-clip stimuli, and finally a stimulus-free 'neutral' class of EEG data is also recorded. Similarly to dataset 1, this gives a total of thirty-six minutes of EEG brainwave data equally belonging to one of the three classes. Unlike the first and third datasets, this experiment focuses on classification of a more limited subject-set given that there are only two subjects involved. There were three film clips that were intended to evoke a positive emotional response; La La Land from Summit Entertainment, Slow Life from BioQuest Studios, and Funny Dogs from MashupZone. Likewise, there were three clips that were intended to evoke a negative emotional response; Marley and Me from Twentieth Century Fox, Up from Walt Disney Pictures, and My Girl from Imagine Entertainment. Note that different forms of positive and negative valence are collected - for the positive, an upbeat musical and dance number, clips of marine life performing feats of nature, and clips of dogs performing interesting and funny activities. For the negative emotion-evoking film clips, these dealt with the final moments spent with a beloved pet, the loss of a loved one after a long marriage, and finally a child attempting to grasp the concept of death. Also note that subjects involved knew that the negative clips were from movies, and this may have impacted the data.

With the subject-limited dataset (emotions) and relatively less limited dataset (concentration), a third dataset is explored in order to benchmark the algorithms when a large subject-set is considered. The dataset is sourced from a BCI2000 EEG device [60–62]. This data describes a multitude of tasks performed by 109 subjects for one to three minutes with 64 EEG electrodes. A random subset of 10 people is taken due to the computational complexity requirements, thus the experiments are focused on datasets of 2, 4, and 10 subjects in order to further compare performance. In this work,

Classification of EEG Signals represented in 2D and 3D 12

each subject had their EEG data recorded for 2 minutes (two 1 minute sessions) for each class. Thus, in total, a dataset was formed of 40 minutes in length - 20 minutes for each class, made up from ten individuals. Classes are reduced from the large set to a binary classification problem, due to the findings of literature review into the behaviours of binary classification in Brain-machine Interaction. The classes chosen are *"Eyes Open"* and *"Eyes Closed"*, since these two tasks require no physical movement from the subjects and thus noise from EMG interference is minimal. Table 1 gives detail on the number of classes in the dataset as well as their class labels.

Mathematical temporal features are subsequently extracted via the aforementioned method in Section 2.

As of the time of writing, the first two datasets (which were collected by the authors for previous works) have not been used in experiments by other authors while the third, from the ML repository, is popular in several recent publications. The aforementioned concentrating and emotional EEG datasets have been explored on the Kaggle cloud computing platform by other data scientists, but results remain unpublished as of yet within academic works.

Firstly, a reduction of dimensionality of the datasets is performed. The chosen number of attributes is 729; this is due to 729 being a square and a cube number and thus therefore being directly comparable in both 2D and 3D space. 729 features thus are reformatted into a square of 27x27 features for 2-dimensional space classification, as well as a cube of 9x9x9 features for 3-dimensional space classification. Each of the attributes in descending rank of their values assigned by the feature ranking algorithms are given as the order (see Future Work for plans to improve on this as a combinatorial optimisation problem), to which each row of the image is filled from left to right, top to bottom. This process is repeated for the 3D process for 9x9 squares which are repeated 9 times to produce the third axis. Alternatives of 64 and 1000 are discarded; firstly, 64 in previous work has been shown to be a relatively weak set of attributes, and larger datasets outperformed such a number by far. Secondly, 1,000 in preliminary exploration showed numerous weak attributes selected. Reduced data is then normalised between values of 0 to 255 in order to correlate to a pixel's brightness value for an image. Note that the CNN for learning will further normalise these values to the range of 0 to 1 by dividing them by 255. The order of the visual data is dictated by the dimensionality reduction algorithms from left to right, with the most useful feature selected by the algorithm in the upper left and the least useful in the lower right (and front to back for 3D). The CNN then extracts *'features of features'* by convolving over this reshaped data.

Secondly, with the reduced data reshaped to both squares and cubes, classification is performed by Convolutional Neural Networks operating in 2D and 3D space. In the previous study [5], as in this work, the order of attributes represented visually are selected by the feature selection algorithms. Scoring is applied by each algorithm and attributes are sorted in descending order, and this is then reshaped into 27×27 square

Classification of EEG Signals represented in 2D and 3D 13

Table 2. Pre-optimisation Network Architecture for Preliminary Experiments [5]

Layer	Output	Params
Conv2d (ReLu)	(0, 14, 14, 32)	320
Conv2d (ReLu)	(0, 12, 12, 64)	18496
Max Pooling	(0, 6, 6, 64)	0
Dropout (0.25)	(0, 6, 6, 64)	0
Flatten	(0, 2304)	0
Dense (ReLu)	(0, 512)	1180160
Dropout (0.5)	(0, 512)	0
Dense (Softmax)	(0, 3)	1539

or $9 \times 9 \times 9$ cube. Visual representation, thus, is performed in three different ways, dependent on the scores applied by the three feature selection methods in this study. This is discussed as a point for further exploration in the Future Work section of this study.

In this stage, topology of networks is simply selected based on the findings of previous experiments (see Section 2). Preliminary hyperparameters from previous work are given as a layer of 32 filters from a kernel of length and width of 3, followed by a layer of 64 filters from a kernel of the same dimensions, a dropout of 0.25 before the outputs are flattened and interpreted by a layer of 512 ReLu neurons. These kernels are to be extended into a third dimension matching the length and width of the windows for the 3D experiments. A generalised view of the network pre-optimisation can be seen in Table 2.

The selected methods of feature selection were those observed in previous experiments as strong algorithms for EEG classification. These are Kullback-Leibler Divergence (Information Gain), One Rule, and Symmetrical Uncertainty. Model training takes place on an NVidia GTX980Ti Graphical Processing Unit, with its implementation in TensorFlow. All models are trained via a 70/30 training/test split for 100 epochs, with a batch size of 64. The loss metric of the models is defined as categorical cross-entropy:

$$CE = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}), \tag{8}$$

where M is the number of class labels (3 or 2 in these cases), y is a binary indication of a correct prediction (1 or 0), and p is the predicted probability of observation o of class c. The entropy of each class within the testing split is calculated and added for a final, overall result. In this case, this is the entropy of the three classes of attention state - relaxed, neutral, and concentrating. Complexity of training when considering epochs, examples, no. features, no. neurons is $O(n^2)$, computational cost is variable based on the hardware used (e.g. if parallelisation is possible) and software (e.g. the method in which the version of the libraries use), times to execute are noted via the hardware

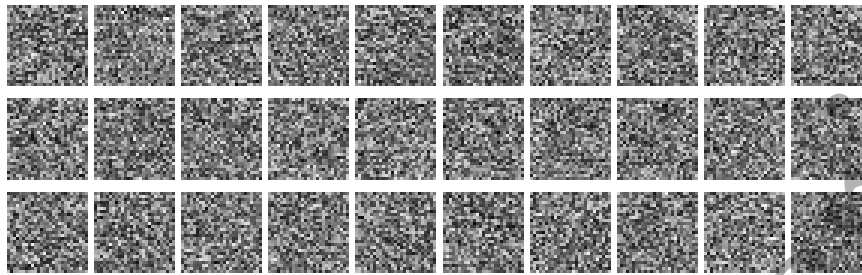*Classification of EEG Signals represented in 2D and 3D* 14

Figure 2. Thirty Samples of attention state EEG Data Displayed as 27x27 Images. Row one shows Relaxed Data, Two shows Neutral Data, and the Third Row Shows Concentrating Data.
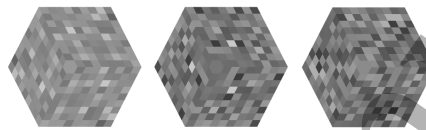

Figure 3. Three attention state Samples Rendered as 9x9x9 Cubes of Voxels. Leftmost Cube is Relaxed, Centre is Neutral, and Rightmost Cube represents Concentrating Data.
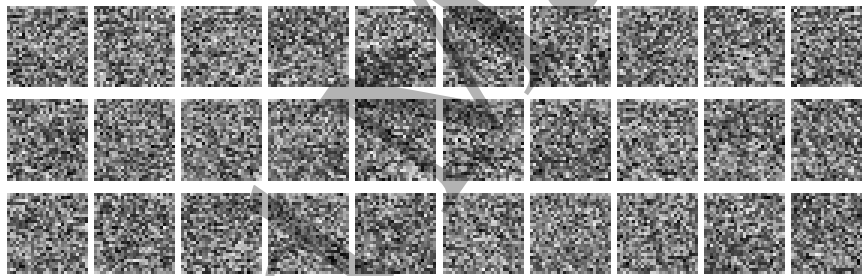

Figure 4. Thirty Samples of Emotional State EEG Data Displayed as 27x27 Images. Row one shows Negative Valence Data, Two shows Neutral Data, and the Third Row Shows Positive Valence Data.
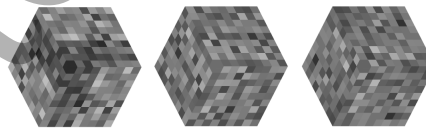
Figure 5. Three Emotional State Samples Rendered as 9x9x9 Cubes of Voxels. Leftmost Cube is Negative Valence, Centre is Neutral, and Rightmost Cube represents Positive Valence Data.

given above on a clean operating system; the evolutionary topology search for the smaller datasets executed for approximately an hour, whereas the larger dataset took one day for the search algorithm to complete. In terms of the final CNN training process, the smaller datasets need only several minutes for the CNN to train since convergence for this data was relatively fast, but the larger dataset was observed to take 24 minutes to finish training. For unseen data prediction, a forward pass has the complexity of $O(n)$.

Samples of visually rendered attention states can be seen in Figures 2 and 3. The

Classification of EEG Signals represented in 2D and 3D   15

examples in these figures show how the data looks when rendered as square images for the 2D CNN and as cubes of voxels for input to the 3D CNN. Note that within the cubes, a large difference between relaxed and the other two states can be observed where it seemingly contains lower values (denoted by lighter shades of grey). In comparison to the 2D representations, it is visually more difficult to discern between the classes, which may also be the case for the CNN when encountering these two forms of data as input. Firstly, figure 2 shows thirty samples of attention state data as 27x27 images whereas figure 3 shows the topmost layer of 9x9x9 cubes rendered for each state. Likewise, examples of the emotions dataset reshaped within 2D and 3D space can be seen in Figures 4 and 5. This process is followed for each and every data point in the set respectively for either a 2D or 3D Convolutional Neural Network. Following this, the DEvo algorithm as described in Section 2.4 is executed upon the best 2D and 3D combinations of models in order to explore the possibility of a better architecture. A population size of 10 are simulated for 10 generations. Hyperparameter limits are introduced as a maximum of 5 hidden layers of up to 4096 neurons each. Networks train for 100 epochs. The goal of optimisation are the interpretation layers that exist after the CNN operations. Following this, the best sets of hyperparameters for each dataset are used in further experiments. During these experiments, the networks are retrained but rather than the 70/30 train/test split used previously, the value of $k = 10$ is selected. Hyperparameters for each 2D and 3D network are those that were observed to be best in the previous heuristic search, this is performed due to the intense resource usage that a heuristic search of a problem space when k-fold cross validation is considered (and would thus be impossible). These experiments are performed due to the risk of overfitting during hyperparameter optimisation when a train/test split is used, due to hyperparameters possibly being overfit to the 30% of testing data, even though a dropout rate of 0.5 is implemented.

Following the experiments on K-Fold Cross Validation, the trained models are then applied to further unseen data through Leave One Subject Out Cross Validation. This is performed by training the model on all the data except for one subject $(n-1)$, and then attempting to predict the class labels of the data collected from the remaining individual in order to examine the extent of cross-subject generalisation. This is performed for all subjects, individual results are considered as well as an overall mean and standard deviation of the set of results attained via the validation process.

The final step of the method of this experiment is to compare and contrast with related studies that use these same datasets.

4. Results

4.1. Attention state Classification

Feature Selection Firstly, attribute selection for the attention state dataset is performed. Overviews of these processes can be seen in Table 3. Selection via Information

Classification of EEG Signals represented in 2D and 3D 16

Table 3. Datasets Produced by Three Attribute Selection Techniques for the attention state Dataset, with their Minimum and Maximum Kullback-Leibler Divergence Values of the 729 Attributes Selected

Selector	Max KBD	Min KBD
Kullback-Leibler Divergence	1.225	0.278
One Rule	0.621	0.206
Symmetrical Uncertainty	1.225	0.233

Table 4. Benchmark Scores of the Pre-optimised 2D CNN on the attention state Selected Attribute Datasets

Dataset	Acc. (%)	Prec.	Rec.	F1
Kullback-Leibler Divergence	91.29	0.91	0.91	0.91
One Rule	93.89	0.94	0.94	0.94
Symmetrical Uncertainty	85.06	0.85	0.85	0.85

Table 5. Benchmark Scores of the Pre-optimised 3D CNN on the attention state Selected Attribute Datasets

Dataset	Acc. (%)	Prec.	Rec.	F1
Kullback-Leibler Divergence	91.52	0.92	0.92	0.92
One Rule	93.62	0.94	0.94	0.94
Symmetrical Uncertainty	85.2	0.85	0.85	0.85

gain selected the attribute with the highest KBD, with a value of 1.225, its minimum KBD was also the highest at 0.278. Interestingly, the OneRule approach selected much lower KBDs of maximum 0.621 and minimum 0.206 values. The Symmetrical Uncertainty dataset was relatively similar to KBD in terms of maximum and minimum selected values.

Classification The classification abilities of the 2D CNN can be seen in Table 4. The strongest 2D CNN was that applied to the One Rule dataset, achieving 93.89% classification ability.

The classification abilities of the 3D CNN can be seen in Table 5. The strongest 3D CNN was that applied to the One Rule dataset, which achieved 93.62% classification ability.

In comparison, results show that the 2D CNN was slightly superior with an overall score of 93.89% as opposed to a similar score achieved by the 3D CNN benchmarking in at 93.62%. Both superior results came from the dataset generated by One Rule selection, even though its individual selections were much lower in terms of their relative entropy when compared to the other two selections, which were much more difficult to classify.

Classification of EEG Signals represented in 2D and 3D   17

Table 6. Datasets Produced by Three Attribute Selection Techniques for the Emotional State Dataset, with their Minimum and Maximum Kullback-Leibler Divergence Values of the 729 Attributes Selected

Dataset	Max KBD	Min KBD
Kullback-Leibler Divergence	1.058	0.56
One Rule	0.364	0.107
Symmetrical Uncertainty	0.364	0.168

Table 7. Benchmark Scores of the Pre-optimised 2D CNN on the Emotional State Selected Attribute Datasets

Dataset	Acc. (%)	Prec.	Rec.	F1
Kullback-Leibler Divergence	98.22	0.98	0.98	0.98
One Rule	97.28	0.97	0.97	0.97
Symmetrical Uncertainty	97.12	0.97	0.97	0.97

Table 8. Benchmark Scores of the Pre-optimised 3D CNN on the Emotional State Selected Attribute Datasets

Dataset	Acc. (%)	Prec.	Rec.	F1
Kullback-Leibler Divergence	97.28	0.97	0.97	0.97
One Rule	96.97	0.97	0.97	0.97
Symmetrical Uncertainty	97.12	0.97	0.97	0.97

4.2. Emotional State Classification

Feature Selection Table 6 shows the range of relative entropy for the results feature selection algorithms on the emotional state dataset. Similarly to the attention state dataset, the KBD selection technique had much higher values in its selection, also as previously seen, the One Rule selector preferred smaller KBD attributes. Unlike the previous attribute selection process though, was that the Symmetrical Uncertainty this time bares far more similarity to the One Rule process whereas in the attention state experiment it closely followed that of the KBD process.

Table 7 shows the results for the 2D CNN on the datasets generated for emotional state. The best model was that of which was trained on the KBD dataset, achieving a very high accuracy of 98.22%.

Classification Table 8 shows the results for the 3D CNN when trained on datasets of selected attributes for the emotional state dataset. The best model was trained on the KBD dataset of features, which achieved 97.28% classification accuracy.

In comparison, the most superior method of data formatting for emotional state EEG dataset is in two dimensions, but very scarcely with a small difference of 0.94%. Unlike the attention state experiment, the best data in both instances on this experiment seemed to be those selected by their relative entropy. 2D One Rule and 3D relative

Classification of EEG Signals represented in 2D and 3D   18
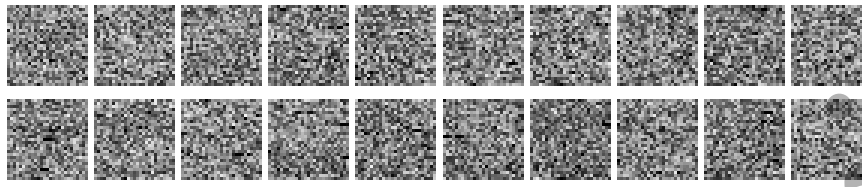

Figure 6. Twenty Samples of Eye State EEG Data Displayed as 27x27 Images. Row one shows Eyes Open, Row Two shows Eyes Closed.
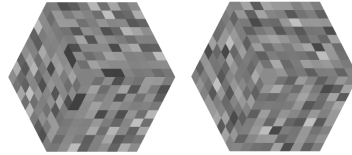

Figure 7. Two Eye State EEG Samples Rendered as 9x9x9 Cubes of Voxels. Left Cube is Eyes Open and Right is Eyes Closed.

Table 9. Attribute Selection and the Relative Entropy of the Set for the Eye State Dataset

Selector	Max KBD	Min KBD
Kullback-Leibler Divergence	0.349	0.102
One Rule	0.349	0.025
Symmetrical Uncertainty	0.349	0.0597

entropy achieved the same score, likewise the 2D and 3D Symmetrical Uncertainty experiments also achieved the same score.

4.3. Extension to 64 EEG Channels

For an extended final experiment, the processes successfully explored in this article are applied to a dataset of a differing nature. The whole process is carried out in the given order. Details of the dataset and experimental process can be found in Section 3.

Figures 6 and 7 show samples of eye state data in both 2D and 3D. Table 9 shows the attribute selection processes and the relative entropy of the gathered sets. As could be logically conjectured, all of the feature selectors found much worth (0.349) in the log covariance matrix of the *Afz* electrode, located in the centre of the forehead. Closely following this in second place for all feature selectors (0.3174) was the log covariance matrix of the *Af4* electrode, placed to the right of the *Afz* electrode. Interestingly, as well as this data which is arguably electromyographical in origin, many features generated from the activities of Occipital electrodes *O1*, *Oz* and *O2* were considered useful for classification, these electrodes are place around the area of the brain that receives and processes visual information from the retinae, the visual cortex. With this in mind, it is logical to conjecture that such a task will produce strong binary classification accuracies since feature selection has favoured areas around the eyes themselves and the cortex within which visual signals are processed.

Classification of EEG Signals represented in 2D and 3D 19

Table 10. Benchmark Scores of the Pre-optimised 2D and 3D CNN on the Eye State Selected Attribute Datasets

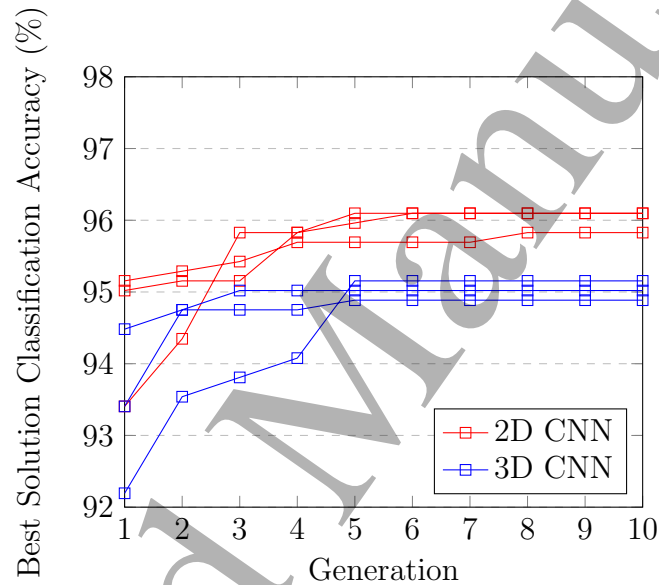| Dims	Dataset	Acc. (%)	Prec.	Rec.	F1
2D	Kullback-Leibler Divergence	97.03	0.97	0.97	0.97
	One Rule	95.34	0.95	0.95	0.95
	Symmetrical Uncertainty	96.89	0.97	0.97	0.97
3D	Kullback-Leibler Divergence	96.05	0.96	0.96	0.96
	One Rule	94.49	0.95	0.95	0.95
	Symmetrical Uncertainty	97.46	0.97	0.97	0.97

Figure 8. Evolutionary Improvement of DEvoCNN for the attention state Classification Problem

Table 10 shows the comparison of results for the 2D and 3D CNNs on the Eye State dataset. As would be expected, very high classification accuracies are considered since the eyes and visual cortex both feature in the 64-channel OpenBCI EEG. Unlike the prior experiments, the 3D CNN on a raster cube prevails over its 2D counterpart when Symmetrical Uncertainty is used for feature selection at a score of 97.46% classification accuracy. As observed previously, other than this one model, all 2D models outperform the 3D alternative.

4.4. Hyperheuristic Optimisation of Interpretation Topology

In this section, the best networks for the three datasets are evolutionarily optimised in an attempt to improve their abilities through augmentation of interpretation network structure and topology, the dense layers following the CNN. Figures 8, 9, and 10 show the evolutionary simulations for the improvement of the interpretation of networks for

Classification of EEG Signals represented in 2D and 3D   20
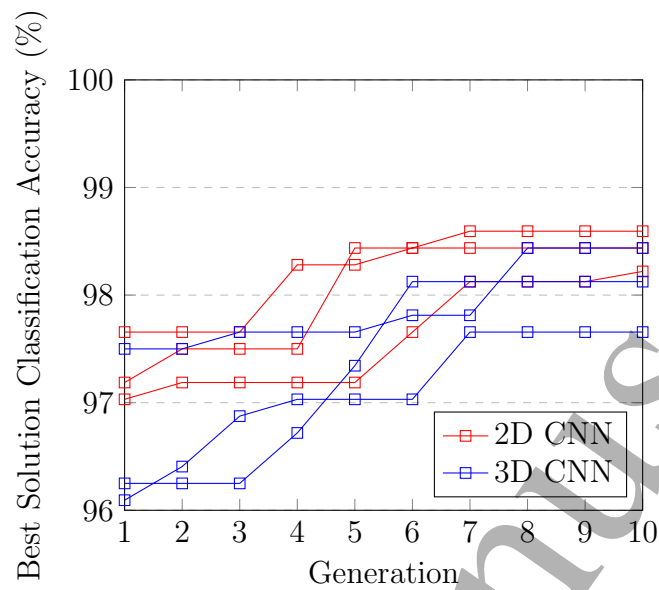

Figure 9. Evolutionary Search of Network Topologies for the Emotional State Classification Problem
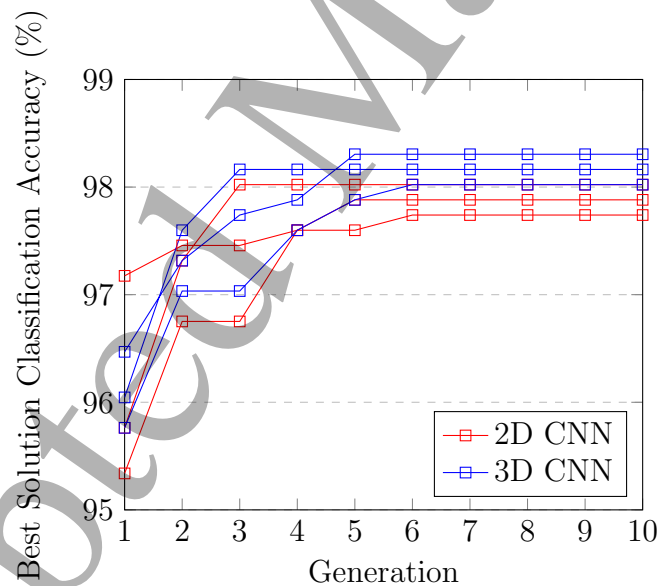
Figure 10. Evolutionary Search of Network Topologies for the Eye State Classification Problem

Attention, Emotional, and Eye State datasets respectively. For the deep hidden layers following the CNN structure detailed in 2, the main findings were as follows:

- Attention state: The best network was found to be a 2D CNN with three hidden interpretation layers $(2705, 3856, 547)$, which achieved 96.1% accuracy. The mean accuracy scored by 2D CNNs was 96%. These outperformed the best 3D network with 5 interpretation layers $(3393, 935, 2517, 697, 3257)$ which scored 95.15%, with a mean performance of 95.02%.

Classification of EEG Signals represented in 2D and 3D 21

Table 11. Benchmark Scores of the Pre and Post-optimised 2D and 3D CNN on all Datasets (70/30 split Validation). Model gives Network and Best Observed Feature Extraction Method. (Other ML metrics omitted and given in previous tables for readability)

Experiment	Model	Accuracy (%)
Attention State	2D CNN, Rule Based	93.89
	3D CNN, Rule Based	93.62
	2D DEvoCNN, Rule Based	**96.1**
	3D DEvoCNN, Rule Based	95.15
Emotional State	2D CNN, KLD	98.22
	3D CNN, KLD	97.28
	2D DEvoCNN, KLD	**98.59**
	3D DEvoCNN, KLD	98.43
Eye State	2D CNN, KLD	97.03
	3D CNN, Symm. Uncertainty	97.46
	2D DEvoCNN, KLD	98.02
	3D DEvoCNN, Symm. Uncertainty	**98.3**

- Emotional State: The best network was found to be a 2D CNN with two hidden interpretation layers $(165, 396)$, which achieved 98.59% accuracy. The mean accuracy scored by 2D CNNs was 98.41%. Close to this was the best 3D network with 1 interpretation layer (476) which scored 98.43%, with a mean performance of 98.07%.

- Eye State: The best network was found to be a 3D CNN with three hidden interpretation layers $(400, 2038, 1773)$ which achieved 98.31% classification accuracy. The mean accuracy scored by 3D CNNs was 98.16%. The best 2D network was 98.02%, with a mean performance of 97.88%.

Table 11 shows the overall results gained by the four methods applied to the three datasets, from the findings of the two previous experiments. The best results for 2D and 3D CNNs are taken forward in the following section in order to perform cross validation. It can be observed that the DEvoCNN approach slightly improved on all networks, but the findings in the first experiment carry over in that the best dimensional-awareness remain so even after evolutionary optimisation.

Figures 11, 12 and 13 show the confusion matrices for the concentration, emotions, and eye state unseen data respectively. Most errors in the concentration dataset arise from relaxed data being misclassified as neutral data which was also observed to occur vice versa, albeit limitedly. The small number of mistakes from the emotions dataset occurred when misclassifying negative as positive and vice versa, the neutral class was classifed perfectly. In the eye state dataset, eyes closed were the most misclassified data at 0.97 to 0.03.

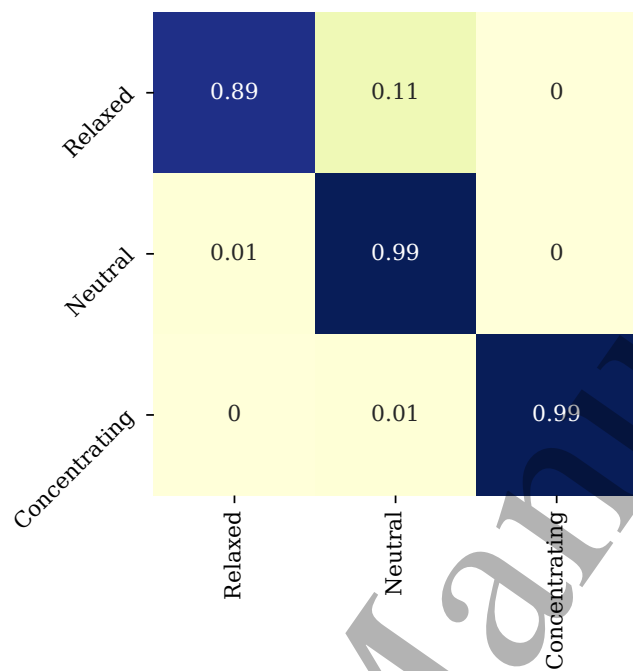*Classification of EEG Signals represented in 2D and 3D*

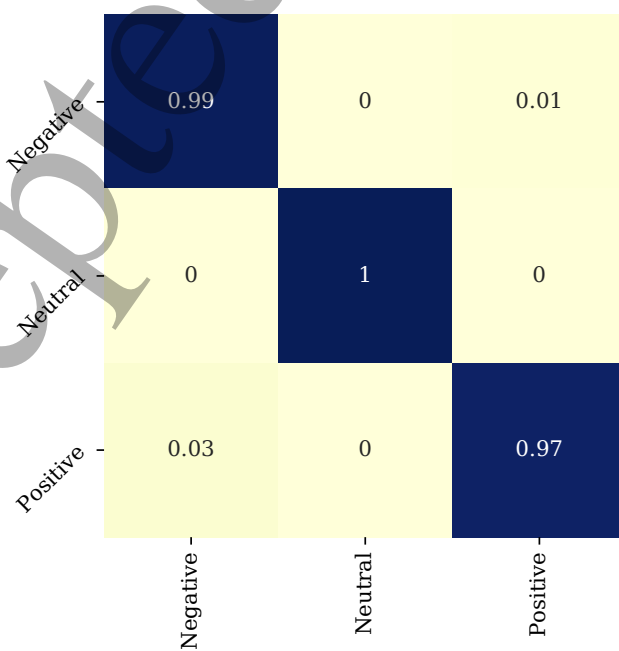**Figure 11.** Normalised confusion matrix for the unseen concentration data.

Figure 12. Normalised confusion matrix for the unseen emotions data.

Classification of EEG Signals represented in 2D and 3D 23

Figure 13. Normalised confusion matrix for the unseen eye state data.

Table 12. Final Benchmark Scores of the Post-optimised Best 2D and 3D CNN on all Datasets via K-fold cross validation.

Experiment	Model	Acc. (%)	Std.	Prec.	Rec.	F1
Attention State	2D CNN	97.03	1.09	0.97	0.97	0.97
	3D CNN	95.87	0.82	0.96	0.96	0.96
Emotional State	2D CNN	98.09	0.55	0.98	0.98	0.98
	3D CNN	98.4	0.53	0.98	0.98	0.98
Eye State	2D CNN	97.33	0.79	0.97	0.97	0.97
	3D CNN	97.96	0.44	0.98	0.98	0.98

4.5. K-fold Cross Validation of Selected Hyper-parameters

In this section, the best sets of hyperparameters for each dataset are used in further experiments where each model is benchmarked through 10-fold cross validation.

Table 12 shows the mean accuracy of networks when training via 10-fold cross validation. As was alluded to through the simpler data split experiments, the best models for the first two datasets were found when the data was arranged as a 2-Dimensional grid of pixels whereas the best model for the eye state dataset was in 3D with both a higher accuracy and lower standard deviation of per-fold accuracies. Standard deviation was relatively low between folds, all below 1% except for the 2D CNN attention state model which has a standard deviation of 1.09%.

Classification of EEG Signals represented in 2D and 3D 24

Table 13. Leave one Subject Out (Unseen Data) for the Concentration State Dataset

Subject left out	*1*	*2*	*3*	*4*	*Mean*	*Std.*
Accuracy (%)	84.33	86.27	81.91	89.66	85.54	0.03

Table 14. Leave one Subject Out (Unseen Data) for the Emotions Dataset

Subject left out	*1*	*2*	*Mean*	*Std.*
Accuracy (%)	91.18	84.71	87.95	0.03

Table 15. Leave one Subject Out (Unseen Data) for the Eye State Dataset (individual 109 subjects removed for readability purposes)

Subject left out	*Mean*	*Std.*
Accuracy (%)	83.8	3.44

Table 16. Comparison of the best concentration dataset model (2D CNN) to other models

Model	Acc. (%)	Std.	Prec.	Rec.	F1
2D CNN	97.03	1.09	0.97	0.97	0.97
Extreme Gradient Boosting	93.62	0.01	0.94	0.94	0.94
Random Forest	91.64	0.02	0.92	0.92	0.92
KNN(10)	86.03	0.03	0.87	0.86	0.86
Decision Tree	84.65	0.02	0.85	0.85	0.85
AdaBoost Long Short-Term Memory [58]	84.44	0.02	0.85	0.85	0.85
Long Short-Term Memory [58]	83.84	0.03	0.84	0.84	0.84
Deep Neural Network [58]	79.81	0.02	0.8	0.8	0.8
Linear Discriminant Analysis	79.44	0.02	0.81	0.79	0.8
Support Vector Classifier	77.46	0.02	0.78	0.78	0.77
Quadratic Discriminant Analysis	74.27	0.02	0.74	0.74	0.73
Naive Bayes	52.18	0.03	0.53	0.52	0.47

4.6. Leave One Subject Out Validation of Selected Hyperparameters

Tables 13, 14 and 15 show the leave one subject out results for each of the three datasets with the best CNN model. The model is trained on all subjects except for one, and classifies the data belonging to that left out subject.

5. Discussion

Tables 16, 17 and 18 show comparisons of the best models found in this study to other machine learning models. Although the top mean scores were noted to be the CNNs found in this study, their deviance is relatively high. In some cases such as in the emotions and eye state datasets for example, the CNN only slightly outperforms a Random Forest which is far less computationally expensive to execute in comparison.

Classification of EEG Signals represented in 2D and 3D 25

Table 17. Comparison of the best emotions dataset model (3D CNN) to other statistical models

Model	Acc. (%)	Std.	Prec.	Rec.	F1
3D CNN	98.4	0.53	0.98	0.98	0.98
Extreme Gradient Boosting	98.38	0.01	0.98	0.98	0.98
Random Forest	98.36	0.01	0.98	0.98	0.98
AdaBoost Long Short-Term Memory [58]	97.06	0.01	0.97	0.97	0.97
Long Short-Term Memory [58]	96.86	0.01	0.97	0.97	0.97
Deep Neural Network [58]	96.11	0.02	0.96	0.96	0.96
Decision Tree	94.98	0.02	0.95	0.95	0.95
Linear Discriminant Analysis	93.9	0.02	0.94	0.94	0.94
KNN(10)	92.64	0.01	0.93	0.93	0.93
Support Vector Classifier	92.03	0.01	0.93	0.92	0.92
Quadratic Discriminant Analysis	77.35	0.11	0.82	0.78	0.77
Naive Bayes	65.24	0.04	0.65	0.65	0.63

Table 18. Comparison of the best eye state dataset model (3D CNN) to other statistical models

Model	Acc. (%)	Std.	Prec.	Rec.	F1
3D CNN	97.96	0.44	0.98	0.98	0.98
AdaBoost Long Short-Term Memory	97.87	0.04	0.98	0.98	0.98
Long Short-Term Memory	97.87	0.04	0.98	0.98	0.98
Extreme Gradient Boosting	97.95	0.01	0.98	0.98	0.98
Deep Neural Network	97.91	0.01	0.98	0.98	0.98
Random Forest	97.9	0.01	0.98	0.98	0.98
KNN(10)	94.82	0.01	0.95	0.95	0.95
Linear Discriminant Analysis	94.32	0.01	0.94	0.94	0.94
Support Vector Classifier	92.75	0.02	0.93	0.93	0.93
Decision Tree	90.79	0.02	0.91	0.91	0.91
Quadratic Discriminant Analysis	83.12	0.02	0.84	0.83	0.83
Naive Bayes	66.61	0.03	0.7	0.67	0.65

It is also worth noting that the CNN, for these datasets, seemingly outperforms Long Short Term Memory Networks and Multilayer Perceptrons.

6. Conclusion and Future Works

As discussed at the start of this paper, 729 features were selected in order to directly compare 2D and 3D visual space for EEG classification, since 729 can be used to make both a perfect square and cube. Experiments show the superiority of the 2-Dimensional approach and there are of course many more numbers within the bounds of the attribute set that make only a perfect square, 1273 to be exact. If cube comparison is discarded, image size should be explored in order to explore whether there is a better set of results totalling either more or fewer than the 729 chosen. The feature extraction for the 64-channel dataset produces 23,488 attributes and thus further studies into this can

Classification of EEG Signals represented in 2D and 3D   26

attempt to compare different sized images and cubes due to the abundance of features. Furthermore, the method of reshaping to 2D and 3D through order of their feature selection scores was performed in a relatively simple fashion for purposes of preliminary exploration. In future studies, due to the success found in this work, the method of reshaping and ordering of the attributes within the shape will be studied considering the reshape method an additional network hyperparameter. This presents a combinatorial optimisation problem that should be further explored and solved in order to present more scientifically sound methods for reshaping. In addition, in future, it would be useful to explore other methods of feature extraction using the CNN model. In this work, we compare our approach to statistical models which also have the features as input - it is well documented in the field that features must be extracted from the raw signals when non-temporal learning methods are to be performed [63–65]. Otherwise, low classification accuracies are often encountered and thus models with little use that cannot classify unseen data. Although this would not be possible with the raw signal domain, the raw signals may be more useful for convolutional neural networks to learn from in future benchmarking experiments. Another limitation of this study is that unseen data was restricted to both holdout test sets and unseen subjects, in future a further dataset should be collected in order to enable testing on a larger amount of unseen data.

In this work, models were explored with a train/test split and finally benchmarked with k-fold cross validation. Ron Kohavi [66] argued that data splits are usually inferior to k-fold cross-validation, which is further inferior to leave one out cross-validation where a model is trained for each and every data point ($k = datapoints$). Since this would require the availability of an extensive amount of computational resources before this experiment, it is now feasible to take the best models in this work ahead and attempt leave-one-out cross validation. As previously described, the main limitation of this study is the method of reshaping, three methods were explored which were dictated by the score metrics of three different dimensionality reduction techniques. In future, a combinatorial optimisation algorithm could be used with CNN classification metrics as a function fitness to optimise. Future work could specifically explore the affects of reshaping on CNNs operating in different numbers of spatial dimensions and thus then how this may be useful for future tasks. The techniques were applied generally to four and 64-channel EEG recordings, thus applied to datasets of much different width (given that temporal techniques are extracted from each electrode), and future would could explore if differing successful techniques could be applied with either a task or electrode count in mind. Datasets with larger numbers of subjects and leave-one-subject-out testing could also be explored in future works in order to discern whether these models improve the ability of unseen subject classification or whether calibration is required.

To finally conclude, initially, nine preliminary deep learning experiments were carried out twice for three EEG datasets. Three in 2-Dimensional space and three in 3-Dimensional space and compared. In cases of attention and emotional state, the 2D CNN outperformed the 3D CNN when rule-based and entropy-based feature selection

Classification of EEG Signals represented in 2D and 3D   27

is performed respectively. On the other hand, for eye state with a 64-channel EEG, the 3D CNN produced the best accuracy when feature were selected via their Symmetrical Uncertainty. The best 2D and 3D models for each were then taken forward for topology optimisation, and finally, to prevent overfitting, said topologies were validated using 10-fold cross validation. Final results show that the data preprocessing methods first shown retained their best overall score, but all were improved upon after topology optimisation and subsequent k-fold cross validation.

7. References

[1] Lana EP, Adorno BV, Tierra-Criollo CJ. Detection of movement intention using EEG in a human-robot interaction environment. Research on Biomedical Engineering. 2015;31(4):285–294.

[2] Cassani R, Banville H, Falk TH. MuLES: An open source EEG acquisition and streaming server for quick and simple prototyping and recording. In: Proceedings of the 20th International Conference on Intelligent User Interfaces Companion; 2015. p. 9–12.

[3] Maskeliunas R, Damasevicius R, Martisius I, Vasiljevas M. Consumer-grade EEG devices: are they usable for control tasks? PeerJ. 2016;4:e1746.

[4] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM computing surveys (CSUR). 2018;51(5):1–42.

[5] Ashford J, Bird JJ, Campelo F, Faria DR. Classification of EEG Signals Based on Image Representation. In: UK Workshop on Computational Intelligence. Springer; 2019. .

[6] Swartz BE. The advantages of digital over analog recording techniques. Electroencephalography and clinical neurophysiology. 1998;106(2):113–117.

[7] Coenen A, Fine E, Zayachkivska O. Adolf Beck: A forgotten pioneer in electroencephalography. Journal of the History of the Neurosciences. 2014;23(3):276–286.

[8] Shah AK, Mittal S. Invasive electroencephalography monitoring: Indications and presurgical planning. Annals of Indian Academy of Neurology. 2014;17(Suppl 1):S89.

[9] Taheri BA, Knight RT, Smith RL. A dry electrode for EEG recording. Electroencephalography and clinical neurophysiology. 1994;90(5):376–383.

[10] Oliveira AS, Schlink BR, Hairston WD, König P, Ferris DP. Induction and separation of motion artifacts in EEG data using a mobile phantom head device. Journal of neural engineering. 2016;13(3):036014.

[11] Krigolson OE, Williams CC, Norton A, Hassall CD, Colino FL. Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. Frontiers in neuroscience. 2017;11:109.

[12] Abujelala M, Abellanoza C, Sharma A, Makedon F. Brain-ee: Brain enjoyment evaluation using commercial eeg headband. In: Proceedings of the 9th acm international conference on pervasive technologies related to assistive environments. ACM; 2016. p. 33.

[13] Plotnikov A, Stakheika N, De Gloria A, Schatten C, Bellotti F, Berta R, et al. Exploiting real-time EEG analysis for assessing flow in games. In: 2012 IEEE 12th International Conference on Advanced Learning Technologies. IEEE; 2012. p. 688–689.

[14] Chai TY, Woo SS, Rizon M, Tan CS. Classification of human emotions from EEG signals using statistical features and neural network. In: International. vol. 1. Penerbit UTHM; 2010. p. 1–6.

[15] Tanaka H, Hayashi M, Hori T. Statistical features of hypnagogic EEG measured by a new scoring system. Sleep. 1996;19(9):731–738.

[16] Li M, Lu BL. Emotion classification based on gamma-band EEG. In: Engineering in medicine and biology society, 2009. EMBC 2009. Annual international conference of the IEEE. IEEE; 2009. p. 1223–1226.

[17] Zheng WL, Zhu JY, Peng Y, Lu BL. EEG-based emotion classification using deep belief networks. In: Multimedia and Expo (ICME), 2014 IEEE International Conference on. IEEE; 2014. p. 1–6.

Classification of EEG Signals represented in 2D and 3D 28

[18] Ren Y, Wu Y. Convolutional deep belief networks for feature extraction of EEG signal. In: 2014 International Joint Conference on Neural Networks (IJCNN). IEEE; 2014. p. 2850–2853.

[19] Li K, Li X, Zhang Y, Zhang A. Affective state recognition from EEG with deep belief networks. In: 2013 IEEE International Conference on Bioinformatics and Biomedicine. IEEE; 2013. p. 305–310.

[20] Bos DO, et al. EEG-based emotion recognition. The Influence of Visual and Auditory Stimuli. 2006;56(3):1–17.

[21] Lin YP, Wang CH, Jung TP, Wu TL, Jeng SK, Duann JR, et al. EEG-based emotion recognition in music listening. IEEE Transactions on Biomedical Engineering. 2010;57(7):1798–1806.

[22] Wang XW, Nie D, Lu BL. Emotional state classification from EEG data using machine learning approach. Neurocomputing. 2014;129:94–106.

[23] Koelstra S, Yazdani A, Soleymani M, Mühl C, Lee JS, Nijholt A, et al. Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos. In: International Conference on Brain Informatics. Springer; 2010. p. 89–100.

[24] Suryotrisongko H, Samopa F. Evaluating OpenBCI spiderclaw V1 headwear's electrodes placements for brain-computer interface (BCI) motor imagery application. Procedia Computer Science. 2015;72:398–405.

[25] Buchwald M, Jukiewicz M. Project and evaluation EMG/EOG human-computer interface. Przeglad Elektrotechniczny. 2017;93.

[26] Apiwattanadej T, Zhang L, Li H. Electrospun polyurethane microfiber membrane on conductive textile for water-supported textile electrode in continuous ECG monitoring application. In: 2018 Symposium on Design, Test, Integration & Packaging of MEMS and MOEMS (DTIP). IEEE; 2018. p. 1–5.

[27] Nguyen A, Alqurashi R, Raghebi Z, Banaei-Kashani F, Halbower AC, Vu T. LIBS: a lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. GetMobile: Mobile Computing and Communications. 2017;21(3):31–34.

[28] Jacobs KM. Brodmann's areas of the cortex. Encyclopedia of Clinical Neuropsychology. 2011;p. 459–459.

[29] Finney EM, Fine I, Dobkins KR. Visual stimuli activate auditory cortex in the deaf. Nature neuroscience. 2001;4(12):1171.

[30] Karuppusamy NS, Kang BY. Driver fatigue prediction using eeg for autonomous vehicle. Advanced Science Letters. 2017;23(10):9561–9564.

[31] Rösler O, Suendermann D. A first step towards eye state prediction using eeg. Proc of the AIHLS. 2013;.

[32] Tu W, Sun S. A subject transfer framework for EEG classification. Neurocomputing. 2012;82:109–116.

[33] Zheng WL, Lu BL. Personalizing EEG-based affective models with transfer learning. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. AAAI Press; 2016. p. 2732–2738.

[34] Sabancı K, Koklu M. The classification of eye state by using kNN and MLP classification models according to the EEG signals. International Journal of Intelligent Systems and Applications in Engineering. 2015;3(4):127–130.

[35] Sinha N, Babu D, et al. Statistical feature analysis for EEG baseline classification: Eyes Open vs Eyes Closed. In: 2016 IEEE region 10 conference (TENCON). IEEE; 2016. p. 2466–2469.

[36] Bird JJ, Manso LJ, Ribiero EP, Ekart A, Faria DR. A Study on Mental State Classification using EEG-based Brain-Machine Interface. In: 9th International Conference on Intelligent Systems. IEEE; 2018. .

[37] Bird JJ, Ekart A, Buckingham CD, Faria DR. Mental Emotional Sentiment Classification with an EEG-based Brain-Machine Interface. In: The International Conference on Digital Image and Signal Processing (DISP'19). Springer; 2019. .

[38] Montgomery DC, Runger GC. Applied Statistics and Probability for Engineers. John Wiley &

Classification of EEG Signals represented in 2D and 3D 29

Sons; 2010.

[39] Zwillinger D, Kokoska S. CRC Standard Probability and Statistics Tables and Formulae. Chapman & Hall; 2000.

[40] Strang G. Linear algebra and its applications. Brooks Cole; 2006.

[41] Chiu TY, Leonard T, Tsui KW. The matrix-logarithmic covariance model. Journal of the American Statistical Association. 1996;91(433):198–210.

[42] Haber HE. Notes on the Matrix Exponential and Logarithm; 2019. online. Available from: http://scipp.ucsc.edu/~haber/webpage/MatrixExpLog.pdf.

[43] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. vol. 112. Springer; 2013.

[44] Kohavi R, John GH. Wrappers for feature subset selection. Artificial intelligence. 1997;97(1-2):273–324.

[45] John GH, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem. In: Machine Learning Proceedings 1994. Elsevier; 1994. p. 121–129.

[46] LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436.

[47] Mejía-Lavalle M, Sucar E, Arroyo G. Feature selection with a perceptron neural net. In: Proceedings of the international workshop on feature selection for data mining; 2006. p. 131–135.

[48] Kullback S, Leibler RA. On information and sufficiency. The annals of mathematical statistics. 1951;22(1):79–86.

[49] Kullback S. Information theory and statistics. Courier Corporation; 1997.

[50] Yu L, Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proceedings of the 20th international conference on machine learning (ICML-03); 2003. p. 856–863.

[51] Ciresan DC, Meier U, Masci J, Gambardella LM, Schmidhuber J. Flexible, high performance convolutional neural networks for image classification. In: Twenty-Second International Joint Conference on Artificial Intelligence; 2011. .

[52] Cireşan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. arXiv preprint arXiv:12022745. 2012;.

[53] Nave R. HyperPhysics. Georgia State University, Department of Physics and Astronomy; 2000.

[54] Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology. 1968;195(1):215–243.

[55] Abhang PA, Gawali BW. Correlation of EEG images and speech signals for emotion analysis. British Journal of Applied Science & Technology. 2015;10(5):1–13.

[56] Gevins A, Smith ME, McEvoy L, Yu D. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cerebral cortex (New York, NY: 1991). 1997;7(4):374–385.

[57] Bird JJ, Ekart A, Faria DR. Evolutionary Optimisation of Fully Connected Artificial Neural Network Topology. In: SAI Computing Conference 2019. SAI; 2019. .

[58] Bird JJ, Faria DR, Manso LJ, Ekart A, Buckingham CD. A Deep Evolutionary Approach to Bioinspired Classifier Optimisation for Brain-Machine Interaction. Complexity. 2019;2019. Available from: https://doi.org/10.1155/2019/4316548.

[59] Knuth DE. Postscript about NP-hard problems. ACM SIGACT News. 1974;6(2):15–16.

[60] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Transactions on biomedical engineering. 2004;51(6):1034–1043.

[61] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov PC, Mark R, et al. Components of a new research resource for complex physiologic signals. PhysioBank, PhysioToolkit, and Physionet;.

[62] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215–e220.

[63] Subasi A. EEG signal classification using wavelet feature extraction and a mixture of expert

Classification of EEG Signals represented in 2D and 3D 30

model. Expert Systems with Applications. 2007;32(4):1084–1093.

[64] Azlan WAW, Low YF. Feature extraction of electroencephalogram (EEG) signal-A review. In: 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES). IEEE; 2014. p. 801– 806.

[65] Krishnan S, Athavale Y. Trends in biomedical signal feature extraction. Biomedical Signal Processing and Control. 2018;43:41–63.

[66] Kohavi R, et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai. vol. 14. Montreal, Canada; 1995. p. 1137–1145.