IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Cross-domain MLP and CNN Transfer Learning for Biological Signal Processing: EEG and EMG

JORDAN J. BIRD[1], JHONATAN KOBYLARZ[2], DIEGO R. FARIA[1], ANIKÓ EKÁRT[1], AND EDUARDO P. RIBEIRO[2]

[1]School of Engineering and Applied Science, Aston University, Birmingham, United Kingdom
[2]Department of Electrical Engineering, Federal University of Parana, Curitiba, Brazil

Emails: {birdj1, d.faria, a.ekart}@aston.ac.uk[1], jhonatankobylarz@gmail.com[2], edu@eletrica.ufpr.br[2]

JJB and JK are co-first authors

ABSTRACT In this work, we show the success of unsupervised transfer learning between Electroencephalographic (brainwave) classification and Electromyographic (muscular wave) domains with both MLP and CNN methods. To achieve this, signals are measured from both the brain and forearm muscles and EMG data is gathered from a 4-class gesture classification experiment via the Myo Armband, and a 3-class mental state EEG dataset is acquired via the Muse EEG Headband. A hyperheuristic multi-objective evolutionary search method is used to find the best network hyperparameters. We then use this optimised topology of deep neural network to classify both EMG and EEG signals, attaining results of 84.76% and 62.37% accuracy, respectively. Next, when pre-trained weights from the EMG classification model are used for initial distribution rather than random weight initialisation for EEG classification, 93.82%(+29.95) accuracy is reached. When EEG pre-trained weights are used for initial weight distribution for EMG, 85.12% (+0.36) accuracy is achieved. When the EMG network attempts to classify EEG, it outperforms the EEG network even without any training (+30.25% to 82.39% at epoch 0), and similarly the EEG network attempting to classify EMG data outperforms the EMG network (+2.38% at epoch 0). All transfer networks achieve higher pre-training abilities, curves, and asymptotes, indicating that knowledge transfer is possible between the two signal domains. In a second experiment with CNN transfer learning, the same datasets are projected as 2D images and the same learning process is carried out. In the CNN experiment, EMG to EEG transfer learning is found to be successful but not vice-versa, although EEG to EMG transfer learning did exhibit a higher starting classification accuracy. The significance of this work is due to the successful transfer of ability between models trained on two different biological signal domains, reducing the need for building more computationally complex models in future research.

INDEX TERMS Applied Machine Learning, Biological Signal Processing, EEG, EMG, Knowledge Adaptation, Neural Networks, Transfer Learning

I. INTRODUCTION

It is no secret that the hardware requirements of Deep Learning are far outgrowing the average consumer level of resource availability, even when a distributed processing device such as a GPU is considered [1]. In addition to this, limited data availability often hampers the machine learning process. It is for these reasons that researchers often find similar domains to transfer the learning between, effectively saving computational resources through said similarities by applying cross-domain interpretation. By doing so, once impossible tasks become possible, despite limited resources. A well-known example is VGG (Visual Geometry Group), a set of 16 and 19 hidden-layer Convolutional Neural Networks (CNNs) which have been trained to the extreme on a large image dataset [2]. Useful recognisable features from images such as points, lines, curves, and geometric shapes can be transferred over to a differing CNN task since these features always exist within the domain. Thus, cross-domain transfer learning is enabled in order to interpret new data [3], [4].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

Electrical biological signals show a similarly non-juxtapose pattern of behaviour [5], [6], and thus the domain-transfer may be possible, although it is currently not yet well-understood. *If it is possible, then to what extent and effects are those possibilities?*

Here, we study for the first time whether cross-domain transfer learning can impact the classification ability of models when trained on Electroencephalographic (brainwave) and Electromyographic (muscular wave) data. This is performed through the transfer of initial weights via the best models of each, and learning is continued from this initial starting point. When compared to the classical method of random weight distribution initialisation, we argue that knowledge can be transferred from EMG to EEG and vice-versa, succesfully. We also compare the results to a model fine-tuned by ImageNet weights in order to discern that useful domain-related knowledge is actually being transferred rather than simply general image rules, which could be learnt from any range of sources and domains.

With better classification results come higher impact applications. In the domain of Human-Robot Interaction, the control of prosthetic devices [7]–[9], enabling telepresence within settings such as care assistance [10], [11], as well as within hazardous settings such as bomb disposal [12], and remote environments [13], as well as risk of potential injury [14]–[16] are just a few of many possible fields that successful knowledge transfer could potentially advance, through both improved classification ability and lower computational expense required to train models.

The most notable scientific contributions of this work are the following:

1) The collection of an original EMG dataset of hand gestures gathered from the left and right forearms.
2) Derivation of a strong set of neural network hyper-parameters through an evolutionary search algorithm, via a multi-objective fitness function towards the best interpretation and classification ability of both EEG and EMG data.
3) Successful transfer of knowledge between the two domains through unsupervised transfer learning, enabling increased classification ability of the neural networks when weights are transferred between them as opposed to traditional random initial weight distribution. Better starting abilities, learning curves, and asymptotes of the network learning process are observed when knowledge is transferred.
4) To the authors' knowledge, cross-domain transfer learning is performed between differing biological signals (EEG and EMG) for the first time.

The remainder of this article is structured as follows: In Section II important background on Artificial Neural Networks and the domains of EEG and EMG are described, and also the concept of Transfer Learning is introduced. In Section III the experimental setup is described. The MLP and CNN transfer learning experiments are described in Sections IV and V, and subsequent findings are then presented in Section VI. Finally, future work and conclusions are presented in Section VII.

II. BACKGROUND

In this section, the important scientific concepts as the basis of this work, as well as notable state-of-the-art research within Transfer Learning and Biological Signal Processing are considered.

A. MULTILAYER PERCEPTRON ARTIFICIAL NEURAL NETWORKS

A Multilayer Perceptron (MLP) is an Artificial Neural Network (ANN) trained via validation, backpropagation of errors [17] and a subsequent gradient descent optimisation algorithm [18] in order to perform a classification or regression prediction task [19]. The task in the context of this work is to label a wave, i.e., what class the wave data belong to, based on statistical descriptions of the wave behaviour. The goal of the learning process is to reduce the error rate of the output of the network when compared to the ground truth; the loss function minimised in the experiments reported here is the cross-entropy loss [20], [21] of the networks:

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

where M is the number of classes (3 for EEG, 4 for EMG), y is a binary indicator of whether the prediction that class c is the class of observed data o is correct, and finally p is the probability that aforementioned data o belongs to the class label c.

Learning by backpropagation is performed through a gradient descent optimisation algorithm that drives the updating of the weights within the neural network. In this study, the Adaptive Moment Estimation (ADAM) algorithm is applied [22]. Inspired by RMSProp [23] and Momentum [24], the main steps of ADAM are the following:

1) The exponentially weighted average of past gradients, v_{dW} is calculated.
2) The exponentially weighted averages of the squares of past gradient, s_{dW} is calculated.
3) The bias towards zero in the previous are corrected, deriving $v_{dW}^{corrected}$ and $s_{dW}^{corrected}$ respectively.
4) The network parameters are updated through the following process:

$$
\begin{aligned}
v_{dW} &= \beta_1 v_{dW} + (1-\beta_1)\frac{\partial \mathcal{J}}{\partial W} \\
s_{dW} &= \beta_2 s_{dW} + (1-\beta_2)\left(\frac{\partial \mathcal{J}}{\partial W}\right)^2 \\
v_{dW}^{corrected} &= \frac{v_{dW}}{1-(\beta_1)^t} \\
s_{dW}^{corrected} &= \frac{s_{dW}}{1-(\beta_2)^t} \\
W &= W - \alpha \frac{v_{dW}^{corrected}}{\sqrt{s_{dW}^{corrected}} + \varepsilon},
\end{aligned}
\tag{2}
$$

in which β_1 and β_1 are tunable hyperparameters, $\frac{\partial \mathcal{J}}{\partial W}$ is a cost gradient of the current network layer to be tuned, W is a matrix of weights, α is the learning rate, and ε is a small value introduced to prevent the possibility of division by zero.

B. CONVOLUTIONAL NEURAL NETWORKS

A Convolutional Neural Network (CNN) is a Deep Learning algorithm capable of collecting an input matrix and ascribing weights and bias in parallel under the constraints of a predictive problem [25], [26], resulting in specific features. A *Convolutional* layer performs a dot product between two matrices, where one matrix is the set of learnable parameters and the other one is known as a kernel, producing an *Activation Map*, as shown below:

$$
G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k], \tag{3}
$$

where the input matrix is f and the kernel is denoted as h.

It has been previously shown that CNNs succeed with Biological Signal interpretation. Using EEG-based mental state, Tripathi and Acharya [27] performed two different classifications using both Deep (DNN) and Convolutional (CNN) Neural Networks, where DNN achieved 75.78% accuracy for valence using the 4040 values as input vector, and CNN achieved 81.41% accuracy in two dimensions, respectively. Following the Convolutional layers, interpretation layers are used to learn from convolutional inputs through the same process as MLP (see Subsection II-A). Additionally, filters are employed to extract statistics from the input matrix and a single interpretation layer concludes the classification into K possible labels.

C. EEG AND EMG

In this subsection, the necessary scientific background on Electroencephalography (EEG) and Electromyography (EMG) is provided.

EEG is the process of measuring and recording electrical signals produced by the brain through electrodes placed either upon or within the cranium, that is, the nervous impulses produced by the neurological structure of the brain [28, p. 31]. Since the activities of a person can largely be summed up by brain activity [29], to classify an EEG pattern would provide useful information, and is thus used as a point of input and control for a Brain Computer Interface (BCI).

BCIs have been successfully operated in a variety of clinical situations such as evaluation of seizures, epilepsy, confused states and coma [30]. Classification of the brain activity preceding a stroke showed that a stroke could be predicted before actual occurrence based on abnormalities in brain activity [31], through a Random Forest algorithm applied for statistical classification of extracted feature sets. Successful rehabilitation of motor functions within stroke patients is aided in a process of measurement and classification of brain activity along with robotic feedback [32]. Classification of brain activity has also been successful, similarly to stroke, in the preemptive detection of epileptic seizures before they occur [33], [34].

EMG is a measure of the electrical potential difference between two points whose origin are individual or groups of muscle fibres [35]. Similarly to EEG, the activity of the muscle can largely be summed up by electrical impulses produced, and can thus form a point of control in a Muscle-Computer Interface (muCI) [36]. Similarly to the Muse headband operated in many EEG studies, due to its consumer-friendliness and future potential based on its low-cost yet high-performing nature, the Myo armband is a prominent device used in muCI systems, frameworks, and applications. For example, researchers collaborating from multiple fields found that accurate gesture classification could lead to a new standard for *New Interfaces for Musical Expression* (NIME) [37].

In the Human-Machine Interaction community, Myo has been succesfully employed in Sign Language Recognition, with the classification of 20 Brazilian Sign Language letters with 96-99% accuracy [38]. Following Myo's proprietary system of classification boasting around 83% accuracy, researchers found that through the application of K-Nearest Neighbour (KNN) and Dynamic Time Warping (DTW) algorithms, the classification of five Myo gestures could be improved to around 86% (+3%) [39].

D. TRANSFER LEARNING

Transfer Learning, as the name suggests, consists in transferring something learnt in one problem or task to another. Oftentimes, Transfer Learning is the application of a model trained on source data to unseen data of the same domain called target data [40]. The model trained on the source data can be further trained on the target data, before its deployment on the target data. Cross-domain transfer learning is a similar application of a pre-trained model from one domain to another domain of different nature; for example, in this study, models trained on two different datasets of biological waves from the brain and forearm muscles are applied to one another's data for further training.

Transfer of knowledge is considered successful from one domain to the other when the starting point, the learning curves and asymptotes are higher than those of the traditional

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

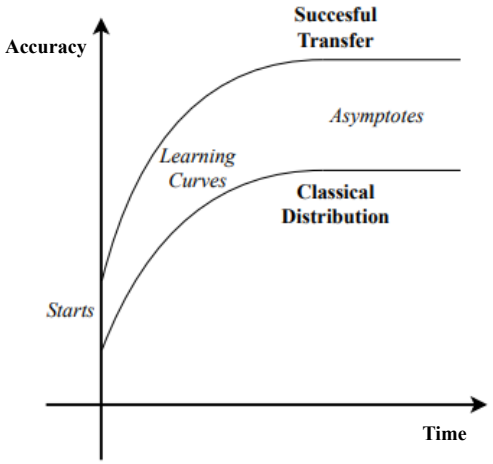

FIGURE 1: Example of a successful Transfer Learning experiment. Transfer Learning (top line) has a higher starting point, steeper curve, and higher asymptote in comparison to learning via random weight distribution (bottom line).

source-train source-classify approach [41]. A visual representation of a successful Transfer Learning experiment can be seen in Figure 1, where the starting point is higher for transfer learning compared to random distribution, and subsequently the learning curve is also steeper and the asymptote is higher.

Generally, there are two main reasons for the application of Transfer Learning [40]. Firstly, pre-trained models and computational resources have become easily accessible [42], there are countless available models trained over many hours on extremely powerful hardware. Examples include VGG [2], Inception [43], and MobileNet [44]. Secondly, the lack of a large enough dataset for learning is often negated by transferring previous knowledge to the domain at hand [40].

Pan and Yang [45] define three main types of Transfer Learning as follows:

1) *Inductive Transfer Learning* is knowledge transfer when the source and target domains are identical but a new task is to be learned. For example, if five EMG gestures are classified and further learning enables the model to learn to recognise additional gestures, based on the current knowledge, then inductive transfer learning takes place.
2) *Unsupervised Transfer Learning* is the transfer of knowledge between two differing domains and likewise differing tasks. In this study, unsupervised transfer learning finds success in sharing knowledge between EEG and EMG domains through the mental state and gesture recognition tasks.
3) *Transductive Transfer Learning* is the process of sharing knowledge between differing domains but for the same task. For example, if an EEG headband is to be calibrated to a subject's data (a slightly different domain) to complete the same mental state recognition task, then transductive transfer learning takes place.

Recently, many Transfer learning techniques have been applied successfully in real-world problems, for example, cancer subtype discovery [46], building-space optimisation [47], [48], text-mining [49], [50], and reinforcement-learning for videogame AI [51], [52].

Cross-domain transfer learning has been given relatively little attention in the field of biological signal processing, with research almost exclusively opting for same-domain personalisation, or calibration. EEG and EMG signals are excellent candidates for cross-domain transfer learning, given their similarities, yet this idea has not been investigated. *In this study we aim to fill this gap and establish cross-domain transfer learning between EEG and EMG domains.*

It has been shown that models do not generalise well between subjects, thus there is a need for transfer learning to achieve accurate classification results [53], [54].

A highly promising proposal [55] consists of a two-step ensemble of filter-bank classification of EEG data via two models, one for the original dataset, and another for a small dataset collected from a new subject. The baseline classification ability for nine individual subjects improves by approximately 10%. Similarly, the kernel principal component analysis (kernel PCA) approach in leads to an improvement from 58.95% to 79.83% (+20.88) classification accuracy when transfer learning from the original dataset is performed for a new subject [56].

Similarly to EEG, transfer learning in EMG is most often concerned with cross-subject learning rather than cross-domain application [57]. Researchers gathered and combined datasets of EMG data measured from a total of 36 subjects via the Myo armband (as also used in this study). The dataset was split into sets of 19 and 17 subjects. Transfer learning of learnt features of a Convolutional Neural Network led to a classification accuracy improvement of 3.5%. Though this improvement is small, the achieved accuracy is the state-of-the-art for the dataset.

Transfer Learning in EMG has been successful in calibrating to electrode shift, change of posture, and disturbances due to sweat and fatigue [58] through small calibration recordings that subsequently require fewer than 60 seconds of training time. Prahm et al. found that the exercises increased in accuracy after disturbance from 74.6% to 97.1%.

Motivated by the small successes of cross-subject transfer learning within EEG and EMG domains independently, as well as the similar nature and behaviour of these biological signals, *we propose to explore the potential of applying learnt knowledge from one biological signal domain to the other and vice versa.*

III. EXPERIMENTAL SETUP

In this section,we describe the data acquisition, feature extraction, topology selection, learning, and transfer learning

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

(a) Muse EEG Headband (InteraXon)

(b) MYO EMG Armband (Thalmic Labs)

FIGURE 2: The devices used to acquire data for this experiment.

experiments.

In terms of hardware, all models are trained on TensorFlow via an NVidia GTX980Ti Graphics Processing Unit. For topology selection, our previously proposed DEvo [59] algorithm is executed for 15 generations. Hard limits of a maximum of 5 hidden layers and 512 neurons were set. Evolutionary topology optimisation allowed for 100 epochs of training and transfer learning was observed with 30 epochs of training. These values were chosen based on the observation that in preliminary experiments there was little or no further improvement after these numbers of generations and epochs, respectively. Training validation is enabled through 10-fold cross validation, where the ten folds are shuffled. Other hyperparameters that were chosen were the ReLu activation function for the hidden layers, and the ADAM optimisation algorithm [22] for tuning of weights during training.

A. DATA ACQUISITION

Two datasets are used in this experiment, EEG and EMG. Here we describe the acquisition of the EEG and EMG datasets. The devices used to acquire datasets can be seen in Figure 2.

The EEG dataset was collected in a previous study [60]. This dataset was obtained from four subjects, two male and two female. The subjects performed three tasks, while the sensors were recording the data. The tasks involved three different states of brain activity: concentration, relaxation and neutral. The EEG data were acquired using the *Muse Headband*, which is a commercial EEG device with four dry electrodes (TP9, AF7, AF8 and TP10). The EMG dataset was gathered via the *Myo armband*, a commercial electromyograph monitoring device with 8 dry electrodes for the measurement of electrical current produced by the skeletal muscles within the arm. The *Myo armband* is capable of recording EMG data at sample rate of 200Hz. In addition, it also has a nine axis Inertial Measurement Unit (IMU) operating at a sample rate of 50Hz. For this project, only the EMG data are used and so the inertia of the arm is not considered.

Ten subjects contributed to the EMG dataset, six male and four female all aged 22-40. The subjects performed four different gestures for 60 seconds each, and the sensors recorded EMG data produced by the muscles in the forearm. The gestures performed were, clenching and relaxation of the fist, spreading and relaxation of fingers, swiping right, and swiping left. The observations were performed twice, once for the right arm and once for the left. Figure 3 shows the experimental setup of a subject wearing the Myo armband.

FIGURE 3: Data collection from a male subject via the Myo armband on their left arm.

B. FEATURE EXTRACTION

Biological signal processing encounters issues in classification due to their non-stationary, nonlinear and random nature. Since the wave class is observed over time, temporal observations must be made rather than single points. For this reason, short segments of the wave are considered and a statistical feature matrix is generated based on observation of the temporal segment. Previous studies identified a set of statistics to be extracted from EEG brainwaves [59], [60]. In this study, a set of statistics was extracted from both the EMG and EEG signals of a temporal nature. Initially, the sampling rate was reduced to a uniform 200Hz based on fast Fourier transformations along a given axis due to the variable nature of a Bluetooth Low Energy (BLE) connection. The signal was assumed to be periodic since a Fourier method is used. This led to a realistic down-sampling since the dominant energy was concentrated in the range of 20-500Hz as observed in [60], even though the frequency range of the EEG sensor was superior to this. The Muse headband performed Notch filtering at 50Hz since the experiment was performed in the United Kingdom.

Both EEG and EMG signals are non-stationary and random waves and it is for this reason that feature extraction must be performed when a non-temporal learning process is followed. Feature extraction was performed by introducing sliding windows of length of 1 second at an overlap of 0.5 seconds to segment the data. Algorithm 1 describes the handcrafted features extracted in this work. Considering how we split the 1 second window, the two 0.5 second half-windows produced due to the differences between two 1 second windows, we compute: (i) the change in both the sample means

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

and in the sample standard deviations between the first and second half-window; (ii) the change in both the maximum and minimum values between the first and second half-windows. Then, considering the two 0.25 second quarter-windows produced due to further differences between computed windows, we then compute (i) the sample mean of each quarter-window; (ii) all paired differences of sample means between the quarter-windows; (iii) the maximum (minimum) values of each quarter-window, plus all paired differences of maximum (minimum) values between the quarter-windows. A Discrete Fourier Transform is performed on the time windows and all resultant values are treated as attributes.

Result: Features extracted from raw data for every w_t
User defined the size of the sliding window $w_t = 1s$;
Input: sequence of raw data (EEG or EMG);
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 the 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 the asymmetry and waves peakedness represented by 3^{rd} and 4^{th} 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 maximum and minimum value of each signal $w_{max}^t = \max(w_t)$ and $w_{min}^t = \min(w_t)$;

    Compute the sample variances $K \times K$ matrix $\mathbf{S}$ of each signal, plus the 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 the Eigenvalues of the covariance matrix $\mathbf{S}$, which are the $\boldsymbol{\lambda}$ 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 $\mathbf{B} \in \mathbb{C}^{K\times K}$ is a matrix logarithm of $\mathbf{S}$;

    Compute the magnitude of the frequency components of each signal, obtained using a 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: Feature extraction algorithm for a sequence of data (EEG or EMG signals).

A change in attribute values is also considered via a *lag window*, in which each window is passed the previous extracted value vector from the preceding window. The first window does not receive this vector since no window preceded it. Due to the redundancies of the maximum, mean, and minimum values of quarter-windows due to the overlaps, these values are not considered within lag window attribute vectors.

The features extend beyond the usual values read by clinicians due to the lower quality of the Muse headband as opposed to what can be expected from clinical EEG, with dimensionality reduction often then used in order to select a set of useful features from a large dataset in which weaker features may exist. In this work, identification of weak features has to be left to the deep learning algorithms since identical inputs are needed for the networks in order to enable the possibility of transfer learning (preventing a mismatch in topology, or neuron representation etc.). This is a limitation discussed in Section VII, and future experiments are outlined in order to produce a synchronised dimensionality reduction algorithm between the two datasets concurrently in the form of a combinatorial optimisation problem. Dimensionality reduction was not performed here since it was not needed for the exploration of transfer learning between biological signal domains.

A vector representation of the wave behaviour is formed through the above process and 988 numerical features are created. Since the same process is followed for the same number of electrodes, the datasets describing EMG and EEG have the same dimensionality and thus transfer learning is possible with these features as inputs to an ANN.

IV. METHOD I: MLP TRANSFER LEARNING

A. DERIVATION OF BEST MLP TOPOLOGY

Although many studies focus on grid search of topologies [61]–[63], this study applies a multi-objective evolutionary algorithm in order to select the best neural network architecture for both classification problems. We apply an evolutionary algorithm instead of a classical grid search for two main reasons [64]–[66]:

1) Evolutionary search allows for exploration within promising areas of the problem space at a finer level. Previous experiments, such as speech recognition [67], found complex best solutions for the problem, e.g. a combination of three deep layers of 599, 1197, and 436 neurons. Including such multiples within a grid search would increase computational complexity of the search beyond realistic possibility.
2) With multi-objective optimisation through mean accuracy via equal scalarisation (see equation 4), the algorithm was able to search for a best solution for both of the problems rather than having to be executed twice, followed by statistical analyses to calculate a best topology.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

Deep Evolutionary Optimisation (DEvo) is an evolutionary algorithm, inspired by Darwinian evolution in Nature, used to search the problem space of neural network topology in order to select a best-performing structure of hidden layers and neurons i.e. a selection of network hyperparameters. DEvo has been successfully used in benchmark problems [68] as well as in phoneme classification [67] and EEG classification [59].

```
Result: Output the best neural network configurations discovered
        per generation
User defined population size p, maximum number of hidden layers
  h, and maximum number of neurons n
Initialise p random solutions with hidden layers in the range [1..h],
  each hidden layer is assigned [1..n] neurons ;
Train all neural networks through forward-pass and
  backpropagation;
Calculate fitness F of all solutions;
Sort solutions by F;
while Termination not met do
    for For each solution parent1 do
        Select a random solution parent2;
        Select network depth d randomly from parent1, parent2 or
          mutate
        while d not met do
            Randomly select layer from parent1, parent2 or
              mutate
            if NOT mutate then
                if Selected layer exists within selected parent
                  then
                    Amend layer to solution offspring
                else
                    if Selected layer exists within other parent
                      then
                        Amend layer to solution offspring
                    else
                        Assign random value;
                        \\since none exist
                    end
                end
            end
        end
    end
    Calculate fitness F of all solutions;
    Sort solutions by F;
    Keep the top p solutions from all parents and offspring;
    Amend best solution to best solutions b;
end
Output all best solutions b of each generation;
```

Algorithm 2: Generalised process of the evolutionary search of Neural Network topology

As illustrated by the flow diagram in Figure 4, the general idea of single-objective *DEvo* is as follows:

1) Create an initial population of random solutions.
2) Simulate the following until a user-defined termination:
    a) Select parent networks for the current generation of the evolutionary cycle.
    b) Optionally, alter the depth or width of the network via random mutation, to prevent premature convergence to local minima within the population.
    c) Train the neural networks by performing forward-pass and backpropagation of errors through a given optimisation function for a user-defined number of epochs, and calculate the fitness.

A more detailed process of the evolutionary search can be observed in Algorithm 2. In particular, the re-combinations of two parent networks is shown in detail.

Since the search must derive a *'best of both worlds'* solution for both the EMG and EEG problems, a new fitness function is introduced to score a solution:

$$F(s) = 0.5\frac{A(EMG)}{100} + 0.5\frac{A(EEG)}{100}, \tag{4}$$

where $A(EMG)$ and $A(EEG)$ are the mean accuracy scores of the networks when trained with EMG and EEG data respectively through shuffled 10-fold cross validation. Equal weights are allocated to the two components as EEG and EMG training are equally important. Only hidden layers are to be optimised, therefore the input and output layers of the network are simply hard-coded.

B. BENCHMARKING OF TRANSFER LEARNING

For transfer learning, the following process is followed:

1) A neural network with randomly distributed weights is trained to classify the EMG dataset.
2) A neural network with randomly distributed weights is trained to classify the EEG dataset.
3) The best weights from the EMG network are applied to a third neural network, which is then trained to classify the EEG dataset.
4) Mirroring step 3, the best weights from the EEG network (step 2) are initialised to a fourth neural network, which is then trained to classify the EMG dataset.

The four networks are then compared. EEG to EMG-EEG and EMG to EEG-EMG in order to discern whether knowledge has been transferred. If higher starts, curves, and asymptotes are observed, then knowledge is considered succesfully transferred between the two domains.

V. METHOD II: CNN TRANSFER LEARNING

A. REPRESENTING BIOLOGICAL WAVES AS IMAGES

In order to generate a square matrix, after the feature extraction process, the final 28 attributes are removed from each dataset. This is done because 961 is the closest square number within the attribute set (31x31) and the final attributes are chosen in order to retain identical inputs to the networks for both datasets. After normalisation of all attributes between the values of 0 and 255, they are then projected as 31px square images. Examples of waves projected into visual space can be observed in Figures 5 and 6. Though padding would be applied in the situation where a square reshape is not possible (if square input is considered), this is not needed in this experiment since 961 attributes are selected (31x31 reshape).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

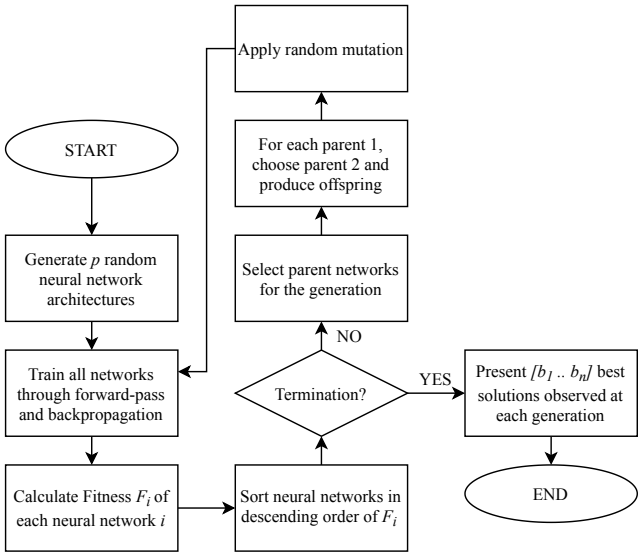

FIGURE 4: Flow Diagram of the evolutionary search of Neural Network topology. Population size is given as p and fitness calculation is given as F. Set $\{b1..b_n\}$ denotes the best solution presented at each generation

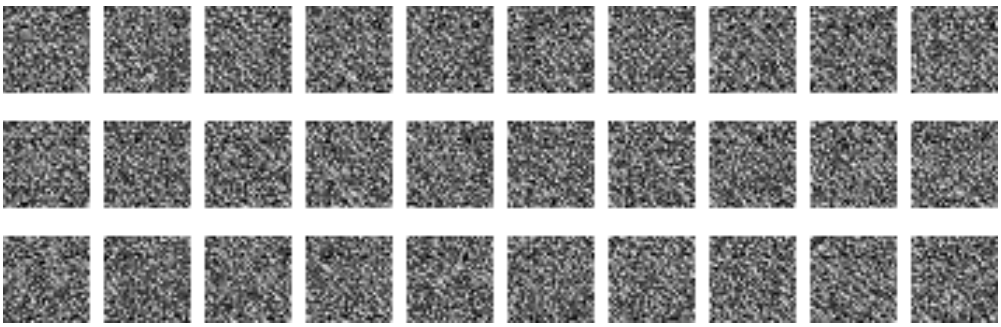

FIGURE 5: 30 Samples of EEG as 31x31 Images. Top row shows relaxed, middle row shows neutral, and bottom row shows concentrating.

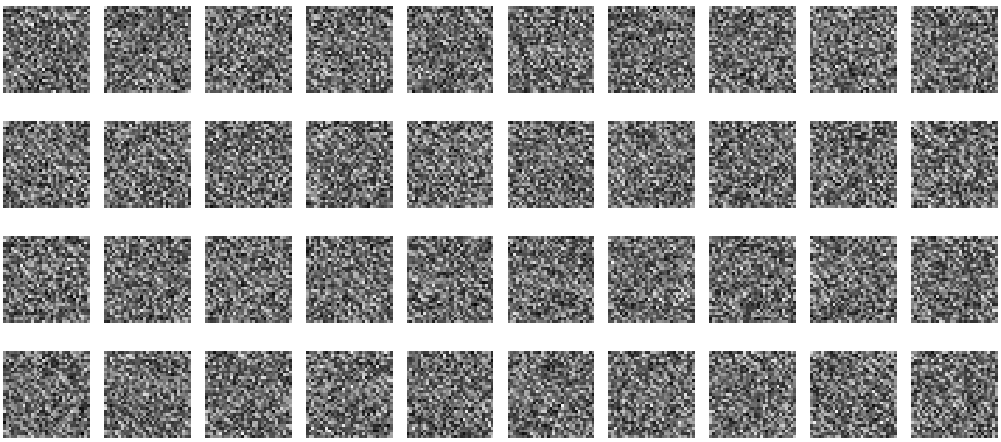

FIGURE 6: 40 Samples of EMG as 31x31 Images. Top row shows close fist, second row shows open fingers, third row shows wave in and bottom row shows wave out.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

TABLE 1: Network topology and parameters found to be best in a previous work [69]

Layer	Output	Parameters
Conv2d (ReLu)	(0, 14, 14, 32)	320
Conv2d (ReLu)	(0, 12, 12, 64)	18,496
Max Pooling	(0, 6, 6, 64)	0
Dropout (0.25)	(0, 6, 6, 64)	0
Flatten	(0, 2304)	0
Dense (ReLu)	(0, 512)	1,180,160
Dropout (0.5)	(0, 512)	0
Dense (Softmax)	(0, 3)	1,539

B. DERIVATION OF HYPERPARAMETERS

Due to the computational costs of hyperparameter optimisation used in the MLP experiment when a CNN is considered, CNN topology is instead inspired by Ashford et al.'s work [69]. In this work, the topology observed in Table 1 was found to be best after empirical exploration. Dropout is introduced in order to prevent overfitting after two convolutional operations (and subsequent pooling) and again after an interpretation layer of 512 densely connected neurons.

C. BENCHMARKING OF TRANSFER LEARNING

The benchmark of the CNN transfer learning follows the same process as detailed in Section IV-B, except the weight transfer applies to input, convolutional, and hidden interpretation layers.

The hypothesis of this experiment ie. that transfer learning has occurred cross-domain, not simply through deep learning, is tested by comparison to a popular pre-trained model. For this purpose, the ResNet50 architecture and weights [70] are used when trained on the ImageNet dataset. This architecture is chosen based on its aptitude for smaller images as opposed to the previously mentioned VGG16 model, more fitting to the nature of the images generated by the algorithm. The experiments are given unlimited time to train in order to explore this, with an early stop executing after 10 epochs with no observed improvement of validation accuracy. Other model hyperparameters are identical to their transfer learning counterparts.

VI. EXPERIMENTAL RESULTS

In this section, the results from the two experiments are discussed. Firstly, an MLP network topology is derived through the previously described DEvo method before transfer learning capabilities are benchmarked. Initially, the models are trained starting from random weight distribution (baseline) in order to provide the baseline. Secondly, the model trained on EMG dataset is used to transfer knowledge to a model training to classify the EEG dataset and then vice-versa. These are then compared to their baseline non-transfer learning counterparts. This is carried out a second time with Convolutional Neural Networks (without evolutionary search) where signals have been projected as raster images.

The MLP experiments are presented and discussed in Subsection VI-A and the CNN experiments are then presented and discussed in Subsection VI-B.

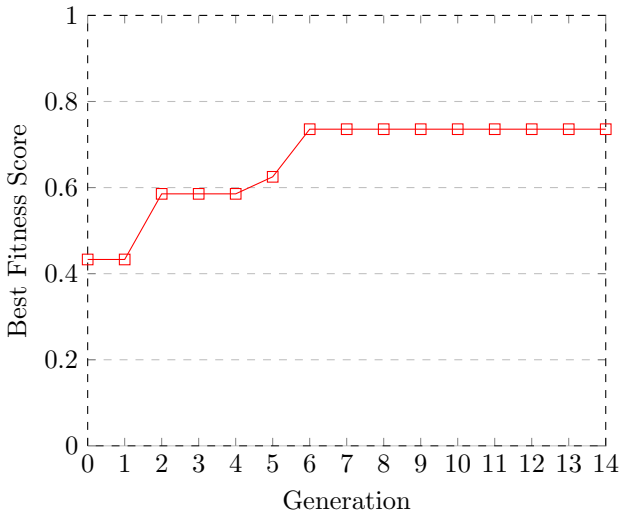

FIGURE 7: Highest (best) fitness observed per generation of the combined and normalised fitnesses of EEG and EMG data classification. The two fitness components are considered equally weighted to produce the same topology in order to allow direct transfer of weights.

TABLE 2: Comparison of the MLP Training Processes of EMG and EEG with random weight distribution compared to weight transfer learning between EMG and EEG

Experiment	Training Accuracy (%)		
	Epoch 0	*Final Epoch*	*Best Epoch*
EMG	62.84	83.57	84.76
EEG	54.7	62.1	62.73
Transfer Learning (EEG to EMG)	**65.22 (+2.38)**	85	**85.12 (+0.36)**
Transfer Learning (EMG to EEG)	**84.95 (+30.25)**	93.28	**93.82 (+29.95)**

A. EXPERIMENT 1: MLP TRANSFER LEARNING

1) Hyperparameter Selection for Initial Random Distribution Learning

Figure 7 shows the fitness evolution (Equation 4) of neural network topologies for the two datasets, where each point is the combined mean fitness for EEG and EMG and the best topology. The best result was found to be a network of 5 hidden layers, with neuron counts $206, 226, 298, 167, 363$ respectively at a combined fitness of 0.74. This network topology is thus taken forward in the experiments towards transfer learning capability between the networks of EEG to EMG and vice-versa.

2) MLP Transfer Results

Finally, the transfer learning experiment is executed following the process described in section IV-B. Figure 8 and Table 2 detail the learning processes of both EMG and EEG as well as the transfer learning experiments, from one

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

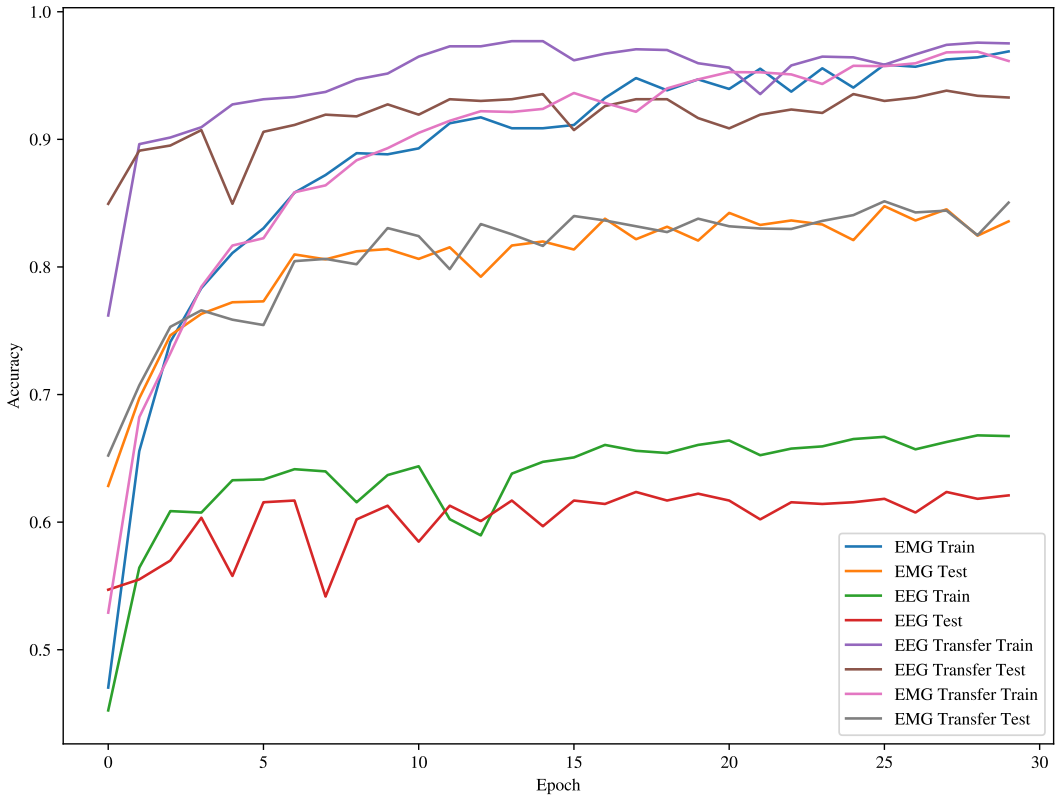

FIGURE 8: Test and Training Accuracies of EMG, EEG, and transfer between EMG and EEG. *'EEG Transfer' denotes EMG to EEG and likewise for 'EMG Transfer'.*

domain to the other and vice-versa. Transfer learning was most successful when EMG data was used to fine tune the EEG problem, with an increase of best classification accuracy from 62.37% to 93.82% (+29.95). A very slight increase was also observed in reverse, when EEG network weights were used as the initial distribution for the EEG problem, with best accuracy rising from 84.76% to 85.12% (+0.36). In terms of starting accuracy, that is, accuracy of classification with no training at all, a success of knowledge transfer also occured; EEG classification increased from 54.7% to 84.95% (+30.25), and thus even prior to any training the network outperformed the network initially trained on EEG data. Likewise, the EMG classification prior to any training at epoch 0 increased from 62.84% to 65.22% (+2.38). It was observed that learning had ceased prior to epoch 30 being reached.

The epoch zero results are particularly interesting since transfer learning has occurred between two completely different domains, from EMG gesture classification to EEG mental state recognition. This shows that knowledge transfer is possible even without training being required.

TABLE 3: Comparison of the CNN Training Processes of EMG and EEG with random weight distribution compared to weight transfer learning between EMG and EEG

Experiment	**Training Accuracy (%)**		
	Epoch 0	***Final Epoch***	***Best Epoch***
EMG	52.4	88	**88.55**
EEG	72.5	95.3	96.24
Transfer Learning (EMG to EEG)	**82.39 (+9.89)**	96.4	**97.18 (+0.94)**
Transfer Learning (EEG to EMG)	**58.18 (+5.78)**	84.24	85.18 (-3.37)

B. EXPERIMENT 2: CNN TRANSFER LEARNING

Figure 9 shows the learning processes for the four networks. It was observed that learning was still occurring at epoch 30 (unlike in the MLPs in Experiment 1), and due to this, learning time was increased to 100 epochs. Table 3 shows the outcome of the experiments. Some transfer learning successes were achieved, with higher starts in TL experiments,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2979074, IEEE Access

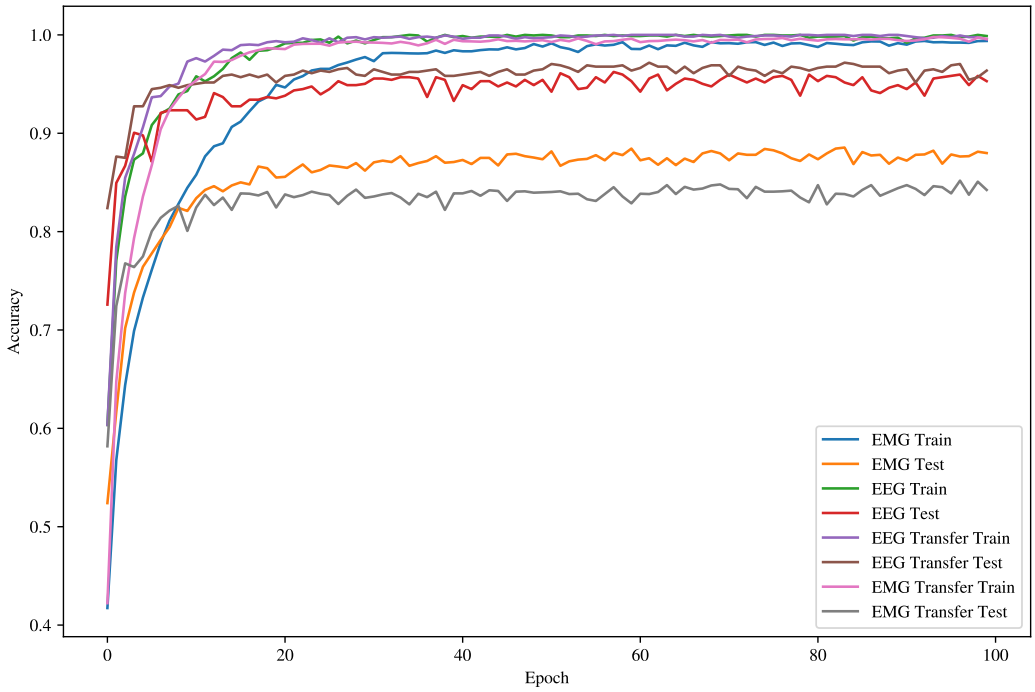

FIGURE 9: Test and Training Accuracies of EMG, EEG, and transfer between EMG and EEG with a Convolutional Neural Network, over 100 epochs. As with the previous figure, *'EEG Transfer' denotes EMG to EEG and likewise for 'EMG Transfer'*.

TABLE 4: Best CNN accuracy observed for ResNet50, Baseline (Non-Transfer) Learning, and Transfer Learning

Best CNN Accuracy Observed (%)					
ResNet50		***Baseline***		***Transfer Learning***	
EEG	***EMG***	***EEG***	***EMG***	***EEG***	***EMG***
92.34	74.92	96.24	**88.55**	**97.18**	85.18

of +9.89% and +5.78% for EEG and EMG, respectively. The best classification accuracy of EEG was improved by 0.94% whereas this was not the case for EMG, which actually decreased by 3.37%. Thus, the CNN transfer learning approach is only successful in the case of EMG to EEG but not vice-versa.

It is important to note that previously, the One Rule Random Forest approach [60] gained 87.16% accuracy and the image representation and CNN approach [69] gained 89.38% accuracy on EEG data. Our network is competitive at 82.39% accuracy on the same dataset with no training whatsoever, using simply the weights from the EMG network. Similarly, it is also important to note that the final accuracy of 97.18% substantially outperforms these previous approaches.

1) Comparison to ResNet50

For comparison of transfer quality, the ResNet 50 CNN architecture is used. Table 4 shows that the ResNet50 achieves weaker results for both problems. The ResNet50 architecture was observed to stop improving after 35 and 39 epochs for EEG and EMG respectively, similarly to the behaviour of our architecture shown in Figure 9.

VII. FUTURE WORK AND CONCLUSION

This study demonstrated that cross-domain transfer learning is possible between the domains of electroencephalography and electromyography, between the electrical signals produced by the frontal lobes of the brain and forearm muscles.To the best of our knowledge, cross-domain transfer learning with EMG to EEG and vice-versa has not been explored elsewhere. Firstly, models were trained via a train/test split and with some hard-coded hyperparameters such as activation function and gradient descent optimisation algorithm; future work should explore the effects and implications of differing hyperparameter sets.

Limited selection of network topologies was performed through a single multi-objective evolutionary search. With the possibility of a local minimum being encountered and stagnation occurring, further executions of the search should be performed in a subsequent study in order to explore the problem space and thus reduce the chance of stagnation.

Scalarisation was considered equal between the two datasets, though the EMG dataset was more diverse and much larger than the EEG dataset and thus alternative scalars with preference to either dataset should also be benchmarked. Future work could also involve the possibilities of cross-domain transfer learning in multiple biological signal domains, such as including other areas of the muscular system and brain, and additionally, other domains such as electrocardiography. The potential for transfer learning between these domains should be applied in Human-machine interaction in the future, since the application of a framework as described here shows not only the advantage of improved accuracy of classification, but additionally, the derivation of a less computationally expensive process compared to learning from scratch.

In this study, dimensionality reduction or feature selection was not performed. Although previous studies have shown the effectiveness of removing certain features in favour of others (both for increased classification ability and reduction of complexity), it was important that the neural networks were identical in structure to one another. In doing this, transfer of weights between networks was directly comparable to the original network to be transferred. In future, a strategy of synchronised feature selection could be explored in order to match network inputs. This could be done in a relatively simple manner, by ranking a defined number of attributes (by Information Gain, Symmetrical Uncertainty, etc.) and choosing those which both algorithms would choose individually. In a far more complex experiment, the selection of attributes could be presented as a combinatorial optimisation problem. Analogously to the famous knapsack problem, transfer learning could provide a fitness function to the combination of selected attributes similar to that in Equation 4. A set of attributes useful to both problems could then be derived to present a synchronised set of inputs for a final transfer learning benchmark. These suggestions could be benchmarked in an 'optimisation extension' to this study. Additionally, dimensional-reshape techniques other than the traditional 2D image reshape considered in this study could also be performed.

The final comparison experiment was designed specifically to assess whether the cross-domain transfer of knowledge has occurred as opposed to simply CNN transfer learning. Success in cross-domain transfer of knowledge was shown compared to the ResNet50 model. In the future, further CNN architectures could be explored. For example, InceptionV3 and VGG16 models outperform the ResNet50 model in the state-of-the-art, but their minimum input dimensions are above those of the data in this experiment. With larger feature selection bounds for the CNN input, in order to generate larger images (which are also to be benchmarked in a related experiment prior to discern the best feature extraction processes), other models such as these could be implemented and provide further evidence for cross-domain transfer of knowledge.

To conclude, we argue that, through initial weight distribution, cross-domain transfer learning between two biological signal domains is possible and, in some cases, to a great positive effect. Identical mathematical features were extracted from the waves to provide a stationary description fit for classification, and transfer between features was also noted. Initial abilities pre-training were higher than random weight distribution, the learning curves and final classification abilities for both domains were also better, indicating that useful knowledge had been shared between both domains during the transfer learning process. The exploration of the possibility of transfer of knowledge from/to other biological signal domains such as ECG is also an exciting topic for future study.

ACKNOWLEDGEMENT

This work was partially supported by "UK ACADEMIES - RESEARCH MOBILITY" PI 02/2018 in Partnership with CONFAP-Brazil and Fundação Araucária Pr-Brazil (protocol 50715.538.33202.06072018), through the project: *"Stepping-stones to Transhumanism: Merging EMG and EEG signals to control a low-cost robotic hand"* awarded to Diego R. Faria (Aston University, UK) and Eduardo P. Ribeiro (Federal University of Parana, Brazil).

REFERENCES

[1] S. Shi, Q. Wang, P. Xu, and X. Chu, "Benchmarking state-of-the-art deep learning software tools," in 2016 7th International Conference on Cloud Computing and Big Data (CCBD), pp. 99–104, IEEE, 2016.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[3] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-vgg16 cnn model for big data places image recognition," in 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), pp. 169–175, IEEE, 2018.

[4] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in European conference on computer vision, pp. 694–711, Springer, 2016.

[5] M. A. Oskoei and H. Hu, "Myoelectric control systems—a survey," Biomedical signal processing and control, vol. 2, no. 4, pp. 275–294, 2007.

[6] D. P. Subha, P. K. Joseph, R. Acharya, and C. M. Lim, "EEG signal analysis: a survey," Journal of medical systems, vol. 34, no. 2, pp. 195–212, 2010.

[7] K. Tatarian, M. S. Couceiro, E. P. Ribeiro, and D. R. Faria, "Stepping-stones to transhumanism: An EMG-controlled low-cost prosthetic hand for academia," in 2018 International Conference on Intelligent Systems (IS), pp. 807–812, IEEE, 2018.

[8] C. Kast, B. Rosenauer, H. Meissner, W. Aramphianlert, M. Krenn, C. Hofer, O. C. Aszmann, and W. Mayr, "Development of a modular bionic prototype arm prosthesis integrating a closed-loop control system," in World Congress on Medical Physics and Biomedical Engineering 2018, pp. 751–753, Springer, 2019.

[9] J. Edwards, "Prosthetics' signal processing connection: Sophisticated prosthetic controls allow amputees to engage more fully in everyday life [special reports]," IEEE Signal Processing Magazine, vol. 36, no. 4, pp. 10–172, 2019.

[10] F. Michaud, P. Boissy, D. Labonte, H. Corriveau, A. Grant, M. Lauria, R. Cloutier, M.-A. Roux, D. Iannuzzi, and M.-P. Royer, "Telepresence robot for home care assistance.," in AAAI spring symposium: multidisciplinary collaboration for socially assistive robotics, pp. 50–55, California, USA, 2007.

[11] J. Broekens, M. Heerink, H. Rosendal, et al., "Assistive social robots in elderly care: a review," Gerontechnology, vol. 8, no. 2, pp. 94–103, 2009.

[12] J. Scholtz, M. Theofanos, and B. Antonishek, "Development of a test bed for evaluating human-robot performance for explosive ordnance disposal robots," in Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, pp. 10–17, ACM, 2006.

[13] E. Welburn, T. Wright, C. Marsh, S. Lim, A. Gupta, B. Crowther, and S. Watson, "A mixed reality approach to robotic inspection of remote environments,"

[14] V. F. Annese, M. Crepaldi, D. Demarchi, and D. De Venuto, "A digital processor architecture for combined EEG/EMG falling risk prediction," in Proceedings of the 2016 Conference on Design, Automation & Test in Europe, pp. 714–719, EDA Consortium, 2016.

[15] A. Heydari, A. V. Nargol, A. P. Jones, A. R. Humphrey, and C. G. Greenough, "EMG analysis of lumbar paraspinal muscles as a predictor of the risk of low-back pain," European Spine Journal, vol. 19, no. 7, pp. 1145–1152, 2010.

[16] D. De Venuto, V. Annese, M. de Tommaso, E. Vecchio, and A. S. Vincentelli, "Combining EEG and EMG signals in a wireless system for preventing fall in neurodegenerative diseases," in Ambient assisted living, pp. 317–327, Springer, 2015.

[17] P. Werbos, "Beyond regression:" new tools for prediction and analysis in the behavioral sciences," Ph. D. dissertation, Harvard University, 1974.

[18] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv preprint arXiv:1609.04747, 2016.

[19] S. Haykin, Neural networks: a comprehensive foundation. Prentice Hall PTR, 1994.

[20] K. P. Murphy, Machine learning: a probabilistic perspective. MIT press, 2012.

[21] S. Kullback and R. A. Leibler, "On information and sufficiency," The annals of mathematical statistics, vol. 22, no. 1, pp. 79–86, 1951.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[23] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop, coursera: Neural networks for machine learning," University of Toronto, Technical Report, 2012.

[24] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in International conference on machine learning, pp. 1139–1147, 2013.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.

[26] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[27] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on deap dataset.," in Twenty-Ninth IAAI Conference, 2017.

[28] D. Purves, G. Augustine, D. Fitzpatrick, W. Hall, A. LaMantia, J. McNamara, and S. Williams, Neuroscience. Sinauer Associates, 2004.

[29] E. Niedermeyer and F. L. da Silva, Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins, 2005.

[30] R. Mercelis, "Practical approach to electroencephalography," Spinal Cord, vol. 48, pp. 840 EP –, Nov 2010. Book Review.

[31] K. G. Jordan, "Emergency EEG and continuous EEG monitoring in acute ischemic stroke," Journal of Clinical Neurophysiology, vol. 21, no. 5, pp. 341–352, 2004.

[32] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang, C. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang, "Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback," in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 5549–5552, IEEE, 2010.

[33] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time–frequency analysis," IEEE transactions on information technology in biomedicine, vol. 13, no. 5, pp. 703–710, 2009.

[34] A. Aarabi, R. Grebe, and F. Wallois, "A multistage knowledge-based system for EEG seizure detection in newborn infants," Clinical Neurophysiology, vol. 118, no. 12, pp. 2781–2797, 2007.

[35] G. Kamen and D. Gabriel, Essentials of Electromyography. Human Kinetics 10%.

[36] A. Chowdhury, R. Ramadas, and S. Karmakar, "Muscle computer interface: a review," in ICoRD'13, pp. 411–421, Springer, 2013.

[37] K. Nymoen, M. R. Haugen, and A. R. Jensenius, "Mumyo–evaluating and exploring the myo armband for musical interaction," 2015.

[38] J. G. Abreu, J. M. Teixeira, L. S. Figueiredo, and V. Teichrieb, "Evaluating sign language recognition using the myo armband," in 2016 XVIII Symposium on Virtual and Augmented Reality (SVR), pp. 64–70, IEEE, 2016.

[39] M. E. Benalcázar, A. G. Jaramillo, A. Zea, A. Páez, V. H. Andaluz, et al., "Hand gesture recognition using machine learning and the myo armband," in 2017 25th European Signal Processing Conference (EUSIPCO), pp. 1040–1044, IEEE, 2017.

[40] L. Torrey and J. Shavlik, "Transfer learning," in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, pp. 242–264, IGI Global, 2010.

[41] U. Kamath, J. Liu, and J. Whitaker, "Transfer learning: Domain adaptation," in Deep Learning for NLP and Speech Recognition, pp. 495–535, Springer, 2019.

[42] W. Van Der Aalst, "Data science in action," in Process Mining, pp. 3–23, Springer, 2016.

[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826, 2016.

[44] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[45] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345–1359, 2009.

[46] E. Hajiramezanali, S. Z. Dadaneh, A. Karbalayghareh, M. Zhou, and X. Qian, "Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data," in Advances in Neural Information Processing Systems, pp. 9115–9124, 2018.

[47] I. B. Arief-Ang, F. D. Salim, and M. Hamilton, "Da-hoc: semi-supervised domain adaptation for room occupancy prediction using co 2 sensor data," in Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments, p. 1, ACM, 2017.

[48] I. B. Arief-Ang, M. Hamilton, and F. D. Salim, "A scalable room occupancy prediction with transferable time series decomposition of co 2 sensor data," ACM Transactions on Sensor Networks (TOSN), vol. 14, no. 3-4, p. 21, 2018.

[49] C. B. Do and A. Y. Ng, "Transfer learning for text classification," in Advances in Neural Information Processing Systems, pp. 299–306, 2006.

[50] P. H. Calais Guerra, A. Veloso, W. Meira Jr, and V. Almeida, "From bias to opinion: a transfer-learning approach to real-time sentiment analysis," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 150–158, ACM, 2011.

[51] M. Sharma, M. P. Holmes, J. C. Santamaría, A. Irani, C. L. Isbell Jr, and A. Ram, "Transfer learning in real-time strategy games using hybrid cbr/rl.," in IJCAI, vol. 7, pp. 1041–1046, 2007.

[52] M. Thielscher, "General game playing in ai research and education," in Annual Conference on Artificial Intelligence, pp. 26–37, Springer, 2011.

[53] D. Wu, B. Lance, and V. Lawhern, "Transfer learning and active transfer learning for reducing calibration data in single-trial classification of visually-evoked potentials," in 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2801–2807, IEEE, 2014.

[54] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," IEEE Signal Processing Letters, vol. 16, no. 8, pp. 683–686, 2009.

[55] W. Tu and S. Sun, "A subject transfer framework for EEG classification," Neurocomputing, vol. 82, pp. 109–116, 2012.

[56] W.-L. Zheng, Y.-Q. Zhang, J.-Y. Zhu, and B.-L. Lu, "Transfer components between subjects for EEG-based emotion recognition," in 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 917–922, IEEE, 2015.

[57] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 4, pp. 760–771, 2019.

[58] C. Prahm, B. Paassen, A. Schulz, B. Hammer, and O. Aszmann, "Transfer learning for rapid re-calibration of a myoelectric prosthesis after electrode shift," in Converging clinical and engineering research on neurorehabilitation II, pp. 153–157, Springer, 2017.

[59] J. J. Bird, D. R. Faria, L. J. Manso, A. Ekart, and C. D. Buckingham, "A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction," Complexity, vol. 2019, 2019.

[60] J. J. Bird, L. J. Manso, E. P. Ribeiro, A. Ekárt, and D. R. Faria, "A study on mental state classification using EEG-based brain-machine interface," in 2018 International Conference on Intelligent Systems (IS), pp. 795–800, IEEE, 2018.
[61] Z. Xiong and J. Zhang, "Neural network model-based on-line re-optimisation control of fed-batch processes using a modified iterative dynamic programming algorithm," Chemical Engineering and Processing: Process Intensification, vol. 44, no. 4, pp. 477–484, 2005.
[62] H. Huttunen, F. S. Yancheshmeh, and K. Chen, "Car type recognition with deep neural networks," in 2016 IEEE Intelligent Vehicles Symposium (IV), pp. 1115–1120, IEEE, 2016.
[63] J. Zahavi and N. Levin, "Applying neural computing to target marketing," Journal of direct marketing, vol. 11, no. 1, pp. 5–22, 1997.
[64] R. F. Albrecht, C. R. Reeves, and N. C. Steele, Artificial neural nets and genetic algorithms: proceedings of the International conference in Innsbruck, Austria, 1993. Springer Science & Business Media, 2012.
[65] M. Suganuma, S. Shirakawa, and T. Nagao, "A genetic programming approach to designing convolutional neural network architectures," in Proceedings of the Genetic and Evolutionary Computation Conference, pp. 497–504, ACM, 2017.
[66] V. Maniezzo, "Genetic evolution of the topology and weight distribution of neural networks," IEEE Transactions on neural networks, vol. 5, no. 1, pp. 39–53, 1994.
[67] J. J. Bird, , E. Wanner, A. Ekart, and D. R. Faria, "Phoneme aware speech recognition through evolutionary optimisation," in The Genetic and Evolutionary Computation Conference, GECCO, 2019.
[68] J. J. Bird, A. Ekart, and D. R. Faria, "Evolutionary optimisation of fully connected artificial neural network topology," in SAI Computing Conference 2019, SAI, 2019.
[69] J. Ashford, J. J. Bird, F. Campelo, and D. R. Faria, "Classification of EEG signals based on image representation," in UK Workshop on Computational Intelligence, Springer, 2019.
[70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

JORDAN J. BIRD achieved a first class Bachelor's Degree with Honours in Computer Science at Aston University in the United Kingdom, before continuing with PhD studies at the same institution in 2018 with an awarded scholarship. Garnered through a deep Scientific passion from an early age, his research interests exist largely within the field of Human-robot Interaction; these include, the Emergence of Artificial Intelligence, Intelligent Social Frameworks, Turing's Imitation Game, Deep Machine Learning, and Transfer Learning. Jordan is a founding member of the Aston Robotics, Vision and Intelligent Systems (ARVIS) laboratory at Aston University.

JHONATAN KOBYLARZ is an Electronic Engineering student at Universidade Federal do Paraná (UFPR), Brazil. His research interests include Deep Machine Learning towards Social Robotic Interaction, Bioengineering and Computer Vision.

DIEGO R. FARIA is a Senior Lecturer in Computer Science. He is with the School of Engineering and Applied Science, Aston University, Birmingham (UK). He is the coordinator and founder of the ARVIS Lab (Aston Robotics, Vision and Intelligent Systems Lab). Currently (2019-2022) he is the project coordinator of the EU CHIST-ERA InDex project (Robot In-hand Dexterous manipulation by extracting data from human manipulation of objects to improve robotic autonomy and dexterity) funded by EPSRC UK. Dr Faria is also PI and Co-I (2020-2022) of two projects with industry (KTP-Innovate UK scheme) related to perception and autonomous systems applied to autonomous vehicles, and NLP and image processing for multimedia retrieval. He received his Ph.D. degree in electrical and computer engineering from the University of Coimbra, Portugal, in 2014. He holds an M.Sc. degree in computer science from the Federal University of Parana, Brazil, in 2005. In 2001, he earned a bachelor's degree in informatics technology (data computing & information) and he has finished a computer science specialization in 2002 at the State University of Londrina, Brazil. From 2014 to 06/2016 Dr. D. Faria was a post-doctoral fellow at the Institute of Systems and Robotics, University of Coimbra where he collaborated on different projects funded by EU commission and the Portuguese government in areas of Robot Grasping and Dexterous Manipulation, Artificial Perception, Cognitive Robotics and Assisted Living. His research interests are: assisted living, perception systems, intelligent and autonomous systems, and cognitive robotics.

ANIKÓ EKÁRT is a reader in Computer Science at Aston University, where she is also Associate Dean for Postgraduate Studies for the School of Engineering and Applied Science. After obtaining a PhD from Eötvös Loránd University, Budapest, Hungary, she held research positions at the Computer and Automation Research Institute, Hungarian Academy of Sciences in Budapest, Hungary and a lectureship in Computer Science at University of Birmingham, UK. Her research is focused on theory and application of artificial intelligence techniques, and in particular genetic programming and evolutionary computation. Areas of application range from health to engineering, art and economics.

EDUARDO P. RIBEIRO is a Professor at Universidade Federal do Parana, Brazil, since 1997. He received Bachelor's degree in Electrical Engineering (1990), M.Sc. degree (1992) and Ph.D. degree (1996) from Pontifical Catholic University of Rio de Janeiro. He did research stage at Vanderbilt University, USA (1995) and at The University of British Columbia, Canada (2005). His research interests include machine learning, signal processing, instrumentation and communication.