Attentive Adversarial Network for Large-Scale Sleep Staging

Samaneh Nasiri SNASIRI@EMORY.EDU
Department of Biomedical Informatics
Emory University School of Medicine
Atlanta, GA, USA

Gari D. Clifford GARI@GATECH.EDU
Department of Biomedical Informatics
Emory University School of Medicine
Department of Biomedical Engineering
Georgia Institute of Technology
Atlanta, GA, USA

Abstract

Current approaches to developing a generalized automated sleep staging method rely on constructing a large labeled training and test corpora by leveraging electroencephalograms (EEGs) from different individuals. However, data in the training set may exhibit changes in the EEG pattern that are very different from the data in the test set due to inherent inter-subject variability, heterogeneity of acquisition hardware, different montage choices and different recording environments. Training an algorithm on such data without accounting for this diversity can lead to underperformance. In order to solve this issue, different methods are investigated for learning an invariant representation across all individuals in datasets. However, all parts of the corpora are not equally transferable. Therefore, forcefully aligning the nontransferable data may lead to a negative impact on the overall performance. Inspired by how clinicians manually label sleep stages, this paper proposes a method based on adversarial training along with attention mechanisms to extract transferable information across individuals from different datasets and pay attention to more important or relevant channels and transferable parts of data, simultaneously. Using two large public EEG databases - 994 patient EEGs (6,561 hours of data) from the Physionet 2018 Challenge (P18C) database and 5,793 patients (42,560 hours) EEGs from Sleep Heart Health Study (SHHS) - we demonstrate that adversarially learning a network with attention mechanism, significantly boosts performance compared to state-of-the-art deep learning approaches in the cross-dataset scenario. By considering the SHHS as the training set, the proposed method improves, on average, precision from 0.72 to 0.84, sensitivity from 0.74 to 0.85, and Cohen's Kappa coefficient from 0.64 to 0.80 for the P18C database.

1. Introduction

A third of the US population experiences less than the recommended amount of sleep, which is linked to many chronic diseases and conditions, such as type 2 diabetes, heart disease, obesity, and depression (con (2015)). As sleep pathologies are increasingly recognized as crucial factors in many illnesses, both as effects and causes, and the improved availability of low-cost sleep monitoring devices continues to accelerate the field, the volume of data continues to expand. The need for automated sleep staging and diagnostics is, therefore, more acute, particularly in low resource regions of the world. The ground truth for sleep staging remains the multi-lead electroencephalogram (EEG) and the standard rules for sleep staging are still focused on 30-sec windows of data (or 'epochs') and manual labeling by a sleep expert into five stages: Wake (W), Rapid Eye Movement (REM), Non-REM 1 (N1), Non-REM 2 (N2) and Non-REM 3 (N3) (Berry et al. (2012b)). In addition to the time and cost involved in manual sleep staging, the significant inter-expert variability remains an issue (Younes and Hanly (2016)). However, the lack of a sizeable public database with heterogeneous populations has limited the development of verifiable algorithms that generalize well across the population. Due to the characteristics and complexities of EEG signals, accurate interpretation of them by human experts requires several years of training. Therefore, developing an accurate classifier with high generalizability on other datasets is a challenging task in this area. Due to the non-stationary nature of the EEG signal (Kaplan et al. (2005)), the changes in statistical characteristics of the signal with time, a classifier that is trained on a temporally-limited amount of data from an individual may poorly generalize on EEG data recorded at a different time on the same subject. Another issue with the low generalizability issue in EEG data is related to high inherent inter-subject variability in how an EEG manifests, limiting the usefulness of EEG applications. This phenomenon arises due to physiological differences (e.g., skull shape) between individuals, and neural activity does not propagate similarly in different subjects. In particular, cortical folding, tissue conductivity, and brain tissue shapes are different across people (Gayraud et al. (2017)). Moreover, electrode sensor montages (the points at which the electrodes are attached, and the reference points) may differ, and different manufacturers' acquisition hardware may filter the EEG differently. Finally, when electrodes are applied, small differences in the locations on the skull may exist, reflecting the EEG technicians' variety of training or even attentiveness on a given day. All these factors lead to significant variabilities in EEG signals.

In this paper, a multi-adversarial neural network with an attention mechanism is proposed to tackle these challenges to develop a generalized model for automated EEG sleep staging.

Technical Significance: The proposed method is the first work to combine multi-adversarial networks with attention mechanisms for sleep staging with two large datasets. The proposed method can operate in an unsupervised manner to highlight the critical channels contributing to the class estimate and pay attention to the more transferable part of EEG patterns across subjects, contributing more to the classification task. Using two large EEG databases, 994 patient EEGs from the PhysioNet 2018 Challenge database ($\approx$ 6,561 hours of data) and 5,793 patients ($\approx$ 42,560 hours) EEGs from Sleep Heart Health Study (SHHS), we demonstrate that adversarially learning a network with an attention mechanism significantly boosts performance compared to state-of-the-art deep learning approaches in the cross-dataset scenario. The proposed method improves, on average, precision from 0.72 to 0.84, sensitivity from 0.74 to 0.85, and a Cohen's Kappa coefficient from 0.64 to 0.80 for the PhysioNet 2018 Challenge database.

Clinical Relevance: Automated sleep staging from the EEG has been previously proposed to incorporate a particular carefully engineered feature extraction part and calibrating the data for each subject or dataset. These methods are time-consuming and costly and do not generalize well to other datasets or even subjects. In general, previous studies have analyzed data from fewer than 100 individuals, and most of them are on homogeneous and/or non-public datasets. Disregarding heterogeneity between individuals and insufficient sample size leads to essential limitations of clinical usage in real-life problems. Therefore, the proposed method attempts to solve this problem by training a network on a large dataset and testing it on another large dataset by learning transferable features, only paying attention to the essential part of data. The proposed method finds the important channels in the dataset, which can provide explanatory information for the clinician.

2. Related Work

As mentioned before, to build a data set large enough to make health-AI models work, studies often combine data from multiple hospitals. Therefore, the condition or device used to capture the data can vary from hospital to hospital and even department to department. Electrode mismatching, inherent inter-subject variability and the non-stationary nature of EEG signals, lead to different joint distribution, $P(X, Y)$ between different recording, where X and Y are feature and label space, respectively. Moreover, the generalizability of a model that is trained on a dataset is low when it is going to be tested on another dataset acquired in different environments with different acquisition hardware. Class imbalances between hospitals can then be associated with hardware differences (such as filter cut-offs). While this can be mitigated through careful inspection of the data, electrode placement differences, and patient physiological differences are harder to identify and mitigate. Therefore, the transferability of the trained model for an application on unseen subjects is degraded. The reason behind this problem is that the primary assumption in machine learning techniques is that training and test data should be drawn from the same distribution, an assumption that does not necessarily hold in large biomedical datasets. In other words, data from two hospitals that are recorded with different devices and set-ups, but for the same task, can not necessarily be leveraged directly in a machine learning approach. The main question raised is that of how to boost performance in the real-life application of EEG through the development of a generalized model across a large population. This issue could be interpreted as how one can diminish spatial and temporal shifts across individuals from different hospitals or recording environments to handle these different variabilities.

As noted, the spatial shift in data can be caused by the variation of sensors' location on the brain in different datasets or mismatching of electrodes in one dataset. This issue can be partly solved by finding an invariant representation across data-sets (Biswal et al. (2018)). In the literature, it has been shown that Symmetric Positive Definite (SPD) matrices provide a strong ability to representations the brain signals (Congedo et al. (2017); Barachant et al. (2010)). The covariance matrix is a typical example of SPD matrices, which has been employed in several studies (Saifutdinova et al. (2019); Rodrigues et al. (2019); Li et al. (2012)). These studies showed that using second-order statistics of multi-channel signals reduce inter-subject and intra-subject variabilities between EEG signals. The spatial covariance matrix can well separate useful information about brain functional connectivity structure (Barachant et al. (2010)) and create a feature space that is comparable across subjects. Moreover, it has been shown that SPD matrices have excellent robustness to the considerable variability of real-world environmental conditions such as instrument noise (Congedo et al. (2017)).

Other studies (Li et al. (2019); Ma et al. (2019); Tang and Zhang (2020)) tackled this challenge using domain adaptation techniques to increase generalization of a model that is trained on EEG data and tested on unseen subjects in Brain-Computer Interface (BCI), Motor Imagery (MI), and emotion recognition tasks. In the literature, it has been shown that domain adaption, which can be considered a particular case of transfer learning, solves the dataset bias of domain shifts, which is common in biomedical applications. The key technique of domain adaption is to diminish the discrepancy between these two distributions using the Maximum Mean Discrepancy (MMD) metric (Long et al. (2015)). Previous studies have employed domain adaptation in biomedical time-series data, bridging the training and test datasets from different individuals by learning subject-invariant representations or estimating instance importance using labeled training samples and unlabeled test samples (Ma et al. (2019); Li et al. (2019); Jayaram et al. (2016)). Shin et al (Shin and Oh (2012)) and Lee et al (Lee and Choi (2009)) used a group EEG analysis using non-negative matrix factorization (NMF) to seek common features for handling intra- and inter-subject variations.

Other methods to increase the generalization ability of a model involve transfer learning - finding subsets of past subjects to initialize a classifier for training on a new subject (Zanini et al. (2017)). Bolagh et al. (Bolagh et al. (2016, 2017)) proposed subject-selection and subject clustering to select relevant individuals based on the similarity between the EEG pattern of different individuals. Raza et al. (Raza and Samothrakis (2019)) proposed bagging methods to handle mismatching between training and test distributions. Chai et al. (Chai et al. (2017)) proposed an adaptive subspace feature matching (ASFM) to match both the marginal and conditional distributions between EEG data from different sessions/subjects. All of these studies tried to develop a method for reducing inter-subject variability by removing the irrelevant subjects in the training set and enabling efficient knowledge transfer from previous subjects to a new unseen patient.

Recently, multiple authors have focused on developing an automated sleep scoring based on applying deep learning (DL) methods (Biswal et al. (2018); Malafeev et al. (2018); Tagluk et al. (2010); Perslev et al. (2019)). Due to the nature of EEGs, which consist of spatial and temporal information, most convolutional and recurrent processing methodologies are suitable for EEG processing. Biswal et al. Biswal et al. (2018) proposed to use a combination of deep recurrent and convolutional neural networks to classify sleep stages as well as sleep abnormalities events. Spectrograms from EEG channels were fed to the CNN module as input, and then the CNN output was fed into a bidirectional recurrent neural network. Zhang et al. (Zhang et al. (2019)) also used the same approach for assessing the generalization capability of their model by testing their model on two different datasets. These methods have gained attention these days since they simplify processing pipelines through end-to-end learning, removing the need for domain-specific knowledge for feature engineering. DL methods are clearly appealing, but they present some dangers and ignore the nature of the EEG, and how it is acquired has limited the impact of DL in this domain. Although DL architectures have been very successful in processing complex data such as images, text, and audio signals (Liu et al. (2017); Hershey et al. (2017)), the generalization and interpretation of a DL method across different patients are still the main challenges for using DL in most clinical applications. DL architectures are hard to “trust” due to their complexity and non-linearity, further reducing their real-life application in a clinical setting.

Recently, the use of generative adversarial networks (GANs) (Goodfellow et al. (2014)) to handle temporal and spatial shifts has received more attention (Tzeng et al. (2017); Sankaranarayanan et al. (2018); Liu et al. (2019)). In fact, similar to GANs, a two-player minimax game is constructed, in which the feature extractor is trying to confuse the domain discriminator via adversarial training (Ganin et al. (2016)). These networks try to align the representations extracted from all EEG channels across all subjects. It is evident that some parts of the brain are more involved in a given task (or are more active during a given state); thus all channels are not equally transferable. Moreover, some parts of the EEG pattern are significantly dissimilar across subjects. Those patterns might be related to the patient's specific health history, which could affect EEG patterns. Therefore, forcefully aligned the irrelevant channels, and EEG patterns may have a large impact on overall performance. An attention mechanism (Vaswani et al. (2017)) is an effective method to focus on essential regions of data, with numerous successes in deep learning tasks such as classification, segmentation, and detection.

In this work, we use the Sleep Heart Health Study (SHHS) database (a private database available on request from the study investigators) to develop an attention mechanism to highlight relevant channels and transferable part of EEG pattern across datasets. The attention mechanisms explore the parts of data that are more similar across different subjects and contribute more to the classification task. More specifically, a multi-adversarial neural network with an attention mechanism is proposed to tackle the challenges detailed above in creating a generalized model for EEG processing. Finally, the algorithm's performance is assessed on the largest open-access EEG database – the Physionet 2018 Challenge (P18C) database (Ghassemi et al. (2018); Goldberger et al. (2000)). All data used in the study are de-identified, and therefore an ethics/institutional review board waiver was provided for this research.

3. Method

In this paper, we focus on the cross-dataset scenario. Two datasets from different hospitals, with different individuals, acquisition hardware, and environments, are leveraged to develop a generalized sleep staging algorithm. The goal is to develop a network on a training dataset (labeled domain $\mathcal{D}_{tr}$), which generalizes well on a test dataset (unlabeled domain $\mathcal{D}_{te}$), where the distributions of training and test sets are different ($P_{tr}(\mathbf{x}_{tr}) \neq P_{te}(\mathbf{x}_{te})$). Note that due to the variations mentioned earlier in biomedical signals, specifically in EEGs, training and test sets follow different probability distributions. The key technical challenge is raised based on the discrepancy between these two distributions.

In the following, we describe the proposed method based on a multi-adversarial neural network with an attention mechanism for the cross-dataset sleep staging task. At first, a high-level overview of the adversarial domain adaptation method, which is proposed in (Ganin et al. (2016)), is given, and then the proposed method for automatically classify sleep stages with attention mechanisms is presented.

3.1. Adversarial Learning for Domain Adaptation

An intuitive idea for developing a network with high generalizability is to minimize the distance between two distributions. Ganin et al. (Ganin et al. (2016)) developed a method to learn a robust representation across two domains and used an approach similar to GAN (Goodfellow et al. (2014)) for the domain adaptation problem, where adaptation behavior is achieved via adversarial training. To match the distribution of training and test set, a domain discriminator connected to the feature extractor via a gradient reversal layer that multiplies the gradient by a certain negative constant during the backpropagation-based training (Ganin et al. (2016)).

The domain adversarial network has three components; a feature extractor ($G_f(\cdot, \theta_f)$), a label classifier ($G_y(\cdot, \theta_y)$), and a domain discriminator ($G_d(\cdot, \theta_d)$). The feature extractor is a neural network that learns an invariant representation across training and test sets by finding a robust transformation. The label classifier is a neural network trained on the training set (labeled domain D_{tr}). Finally, the domain discriminator is a neural network that predicts whether the feature is coming from the training or test set. Here a two-player minimax game is constructed, in which the feature extractor is trying to confuse the domain discriminator via adversarial training. The key idea of domain-adversarial training is to use a Gradient Reversal Layer (GRL), placed between feature extractor and domain discriminator. The GRL acts like an identity function during forwarding propagation and multiplies the gradient by a certain negative constant during the backpropagation, leading to the opposite of gradient descent. Figure (1) shows the framework of the adversarial domain adaption model proposed in (Ganin et al. (2016)).

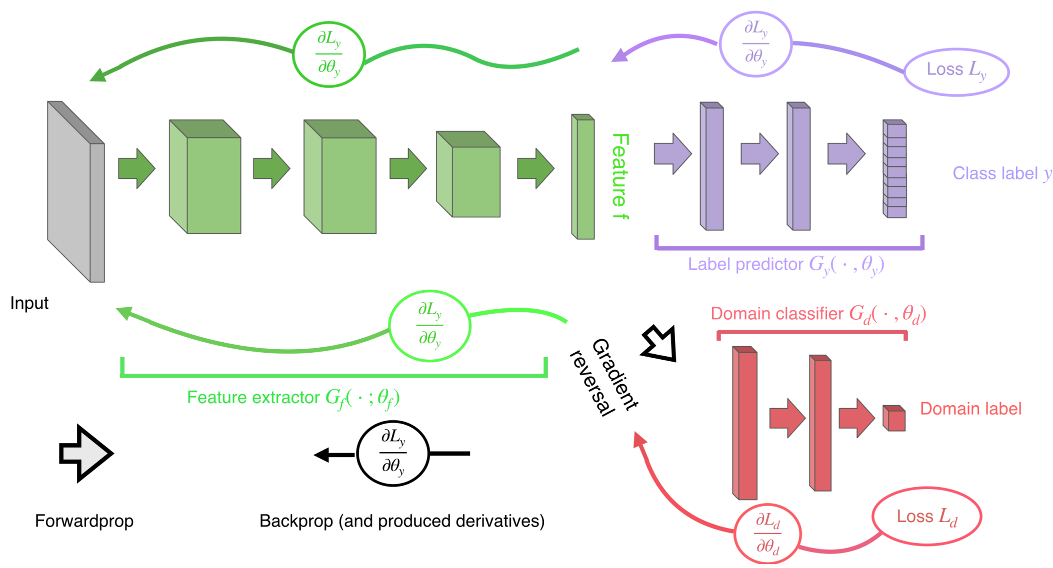


Figure 1: Adversarial Domain Adaption (Ganin et al. (2016)): The domain adversarial network has three components; a feature extractor ($G_f(\cdot, \theta_f)$), a label classifier ($G_y(\cdot, \theta_y)$), and a domain discriminator ($G_d(\cdot, \theta_d)$). This network consists of two losses, the classification loss which is minimized on training features, and the domain confusion loss which is minimized for all features from the training and test sets (while maximizing the domain confusion loss for the feature extraction). The network contains a gradient reversal layer to match the feature distributions from the training and test sets.

The optimization framework of the domain adversarial network proposed in (Ganin et al. (2016)) can be written as follows:

$$C(\theta_f, \theta_y, \theta_d) = \frac{1}{n_{tr}} \sum_{x_i \in \mathcal{D}_{tr}} L_y(G_y(G_f(\mathbf{x}_i)), y_i) - \frac{\lambda}{n} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr} \cup \mathcal{D}_{te}} L_d(G_d(G_f(\mathbf{x}_i)), d_i) \tag{1}$$

where $n = n_{tr} + n_{te}$, n_{tr} and n_{te} are number of sample in training and test sets, respectively. λ is a hyper-parameter that trades-off the domain discriminator loss L_d with the classification loss L_y corresponding to the training classifier G_y.

3.2. Attentive Adversarial Network

As mentioned earlier, this type of domain adversarial network, which proposed in (Ganin et al. (2016)), tries to align the representations extracted from all EEG channels across subjects. Since it is obvious that some parts of the brain are more involved in a given task, all channels are not equally transferable, which is called local attention. Moreover, some parts of the EEG pattern are significantly dissimilar across subjects due to subjective variations in the population, which is called global attention. Inspired by how clinicians manually label sleep stages, two attention mechanisms, local and global attention, along with adversarial training is used to extract transferable information across individuals from different datasets and pay attention to more important or relevant channels and transferable parts of data, simultaneously.

To define the local attention mechanism that matches the training and test sets over the important electrodes, and highlight the important channels in a given task, we split the discriminator G_d in Eq (1) into K channel-wise discriminators G_k^d; $k = 1, 2, \ldots, K$. Therefore, applying this to all K discriminators G_k^d (shown in Figure (2) in blue), $k = 1, 2, \ldots, K$ yields:

$$\mathcal{L}_{ch} = \frac{1}{Kn} \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr} \cup \mathcal{D}_{te}} L_d(G_d^k(\mathbf{f}_i^k), d_i) \tag{2}$$

where $\mathbf{f}_i^k = (G_f^k(\mathbf{x}_i))$ is the feature representation in channel k, d_i is the domain label of point $\mathbf{x}_i$, L_d is the cross-entropy loss. The output of each channel-wise discriminator G_d^k, defined as:

$$\hat{d}_i^k = G_d^k(\mathbf{f_i^k}) \tag{3}$$

where $\hat{d}_i^k$ is the probability of the channel k in windows i belonging to the training set. When this probability approaches 1, the k^{th} channel belongs to the training set. Conversely, as the probability approaches 0, the k^{th} channel belongs to the test set.

The goal of the local attention mechanism is to increase the weighting for those channels that are transferable across training and test set. Thus, a larger local attention value should be generated over transferable channels across the population. The entropy function is an uncertainty measure, defined as $H(p) = -\sum_j p_j log(p_j)$, and creates the transferability criterion. The local weight for each channel is then given by

$$w_i^k = 1 - H(G_d^k(\mathbf{f}_i^k)) = 1 – H(\hat{d}_i^k). \tag{4}$$

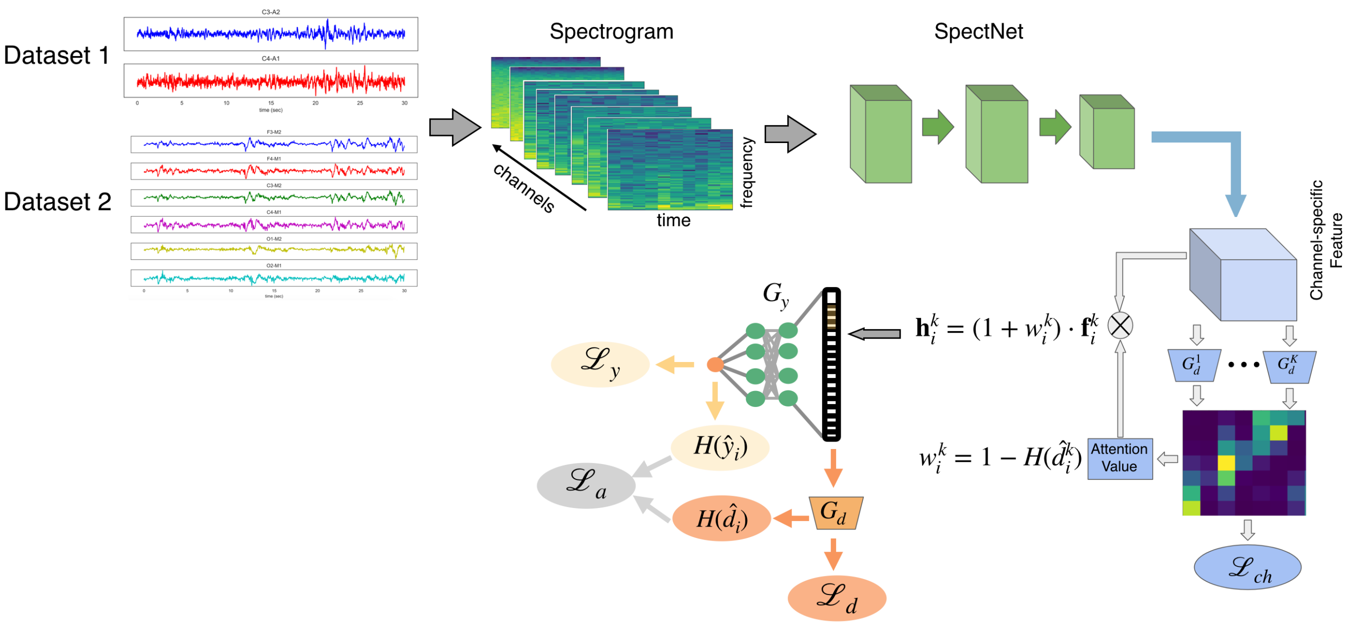


Figure 2: Framework of proposed method: After extracting the spectrograms of all EEG channels, they are fed to the feature extractor, 3-layer convolutional layers, which is called SpectNet here. A multi-adversarial network (blue) is developed for highlighting important channels across datasets attention, and a adversarial network (orange) is used to boost the certainty of output for transferable windows in the feature space across the population. Reproduced from (Nasiri (2020)) “CC by 4.0”.

In order to mitigate the detrimental effect on the network from the selection of incorrect attention values, a residual connection is added in the framework to provide robust optimization. The final channel feature representation $\mathbf{h}_i$ generated from the attended channel features is then expressed as:

$$\mathbf{h}_i = \sum_{k=1}^{K} (1 + w_i^k) \cdot \mathbf{f}_i^k \tag{5}$$

Using the local attention mechanism helps the network to focus on the transferable and important channels based on a given task. Due to varying quality of electrode contact impedance, varying skull shapes, different hardware acquisition systems, or other variations across the population and datasets, the domain discriminator may not be able to find any channels to align. However, some parts of the data might still be transferable by applying an appropriate projection into a new space. Therefore, it is necessary to add a global attention mechanism to the extracted feature, $\mathbf{h}_i$, before the classifier G_y, can assist in the transfer of information. The global discriminator is trained with the following objective function:

$$\mathcal{L}_g = \frac{1}{n} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr} \cup \mathcal{D}_{te}} L_d(G_d(\mathbf{h}_i, d_i)) \tag{6}$$

where d_i is the label of feature $\mathbf{x}_i$ coming from the training or test set and L_d is the cross-entropy loss of the global domain discriminator G_d.

Finally, the minimum entropy regularization is utilized to refine the classifier adaptation. The entropy minimization principle encourages the low-density separation between classes by minimizing the entropy of class-conditional distributions on the test set, which is useful for refining the classifier adaptation (Wang et al. (2019)). Minimizing entropy increases the confidence of the classifier predictions. However, not all windows (features extracted from 30-sec epochs of EEG) in the training set are transferable. For example, windows that are significantly dissimilar in the feature space across the whole population from training and test sets are much less likely to carry useful information. Therefore, forcing minimizing the entropy of these windows hurts the overall performance. Increasing the confidence of these windows in the classifier predictions will confuse the classifier, which is harmful. Therefore, to generate an attention value for each window's entropy loss, the output of the global discriminator $\hat{d}_i = G_d(\mathbf{h}_i)$ is used, by aiming to enhance the certainty of those windows that are more similar across training and test sets. The global attention value for each window is defined as:

$$\mathcal{G} = 1 + H(\hat{d}_i) \tag{7}$$

where the larger global attentions correspond to transferable windows.

Thus, the attentive entropy with the global attention value is defined as follows:

$$\mathcal{L}_a = -\frac{1}{n} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr} \cup \mathcal{D}_{te}} \sum_{j=1}^{c} (1 + H(\hat{d}_i)) \cdot \mathbf{p}_{i,j} \cdot \log(\mathbf{p}_{i,j}) \tag{8}$$

where c is the number of classes, and $\mathbf{p}_{i,j}$ is the probability of predicting that point $\mathbf{x}_i$ is in class j. By minimizing the attentive entropy penalty, the predictions of transferable windows will become certain, and thus improve the classifier's performance.

Using the local and global attention mechanisms along with adversarial learning alleviates negative impacts from forcefully aligning the nontransferable part of data, thus enhances the generalizability of the network. The local attention module focuses on important channels in a given task, while the global attention module focuses on more transferable parts of signals across the population. Like other networks, a appropriate classification loss function is used to lead the classifier to generate correct predictions. Since the labeled data is just available in the training set, classification loss is evaluated on the training set as follows:

$$L_y = \frac{1}{n_{tr}} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr}} L_y(G_y(\mathbf{h}_i), y_i) \tag{9}$$

where L_y is the cross-entropy loss function, and G_y is the training classifier.

Therefore, the end-to-end optimization framework can be express as follows:

$$\begin{aligned} C(\theta_f, \theta_y, \theta_d, \theta_d^k|_{k=1}^K) =& \frac{1}{n_{tr}} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr}} L_y(G_y(G_f(\mathbf{x}_i)), y_i) \\ &+ \frac{\gamma}{n} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr} \cup \mathcal{D}_{te}} \sum_{j=1}^{c} (1 + H(\hat{d}_i)) \cdot \mathbf{p}_{i,j} \cdot \log(\mathbf{p}_{i,j}) \\ &- \frac{\lambda}{n} \left[\sum_{\mathbf{x}_i \in \mathcal{D}_{tr} \cup \mathcal{D}_{te}} \mathcal{L}_d(G_d(\mathbf{h}_i, d_i))] + \frac{1}{K} \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in \mathcal{D}_{tr} \cup \mathcal{D}_{te}} \mathcal{L}_d(G_d^k((G_f(\mathbf{x}_i)))) \right] \end{aligned} \tag{10}$$

Where λ_s, and γ, chosen via grid search, are hyper-parameter that trade-off between domain discriminator and attentive entropy loss, respectively. The network parameters can be learned end-to-end by a minimax optimization procedure as follows:

$$\begin{aligned} (\hat{\theta}_f, \hat{\theta}_y) &= \underset{\theta_f, \theta_y}{\operatorname{argmin}}\, C(\theta_f, \theta_y, \theta_d, \theta_d^k|_{k=1}^K) \\ (\hat{\theta}_d, \hat{\theta}_d^i, \cdots, \hat{\theta}_d^K) &= \underset{\theta_d, \theta_d^i, \cdots, \theta_d^K}{\operatorname{argmax}}\, C(\theta_f, \theta_y, \theta_d, \theta_d^k|_{k=1}^K) \end{aligned} \tag{11}$$

4. Experimental Set-Up

4.1. Data

Sleep Heart Health Study: The SHHS database consists of two rounds of polysomnographic recordings (SHHS-1 and SHHS-2) sampled at 125 Hz in a sleep center environment. Following (Duggal et al. (2020)), we use only the first round (SHHS-1) containing polysomnographic records from participants included 52.9% women and 47.1% men, over two channels (C4-A1 and C3-A2). Recordings were manually classified into one of six classes (W, REM, N1, N2, N3, and N4). As suggested in (Berry et al. (2012a)), we merge N3 and N4 stages into a single N3 stage. Table 1 shows number of sleep stages per class.

Physionet 2018 Challenge: The P18C database includes PSG data from 1,985 subjects included 65% male and 35% women, which were monitored at the MGH sleep laboratory for the diagnosis of sleep disorders. The data were partitioned two-part: public dataset (n = 994) and hidden dataset (n = 989). The sleep stage labels for 994 of the recordings were made available for the public dataset, where includes Wake, REM N1, N2, and N3 stages. It includes multiple physiological signals that are all sampled at 200 Hz and were manually scored by certified sleep technicians at MGH sleep laboratory according to the AASM guidelines into 30 second 'epochs'. In this work, we use the EEG channels, which include 'F3-M2', 'F4-M1', 'C3-M2', 'C4-M1', 'O1-M2', and 'O2-M1' channels.

Table 1: Number of subjects and samples per class for each dataset

Dataset	# Subjects	# Wake	# N1	# N2	# N3	# REM
SHHS	5,792	1,690,997	217,535	2,397,062	739,230	817,330
P18C 2018	994	145,558	135,409	372,257	101,678	113,872

Fig (3) illustrates the electrode position on the scalp (looking from top down, with the nose at the top of the diagram). Note that the green electrodes (C3 and C4) are common to both databases and were used in this study.

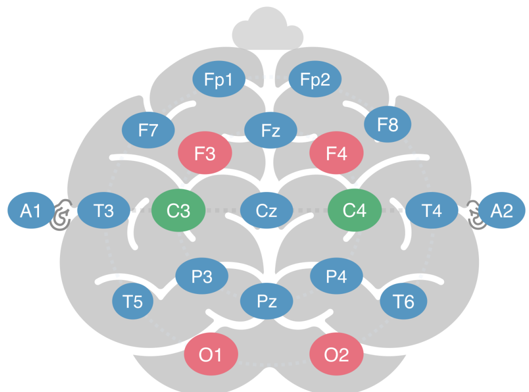


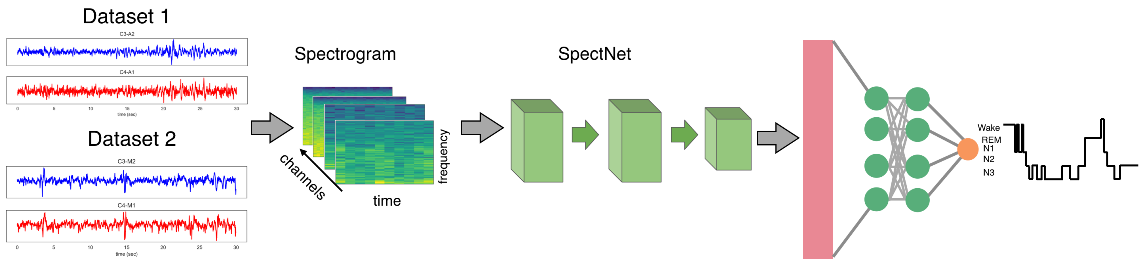
Figure 3: 10-20 EEG Placement: Red electrodes were used in the P18C database and blue electrodes were used in the SHHS database. Green electrodes (C3 and C4) are common to both databases. Reproduced from (Nasiri (2020)) “CC by 4.0”.

4.2. Preprocessing

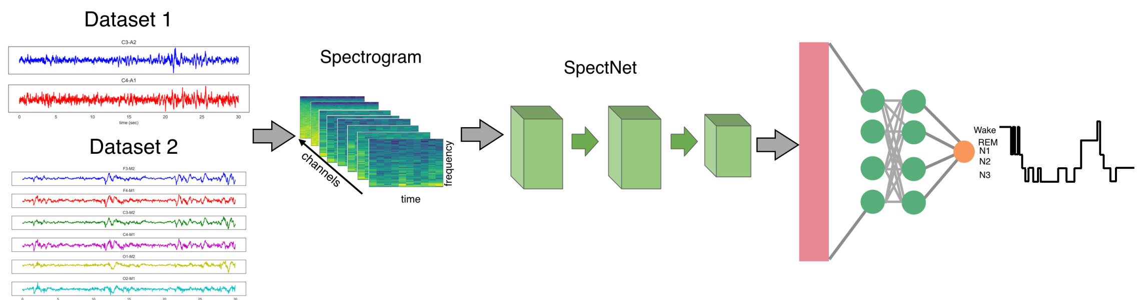
Before presenting the signal to the network, preprocessing is performed to reduce the negative effects of signal artifacts. Two filters were applied to the EEG channels: a notch filter to remove 60 Hz power line interference, and a band-pass filter to allow a frequency range of 0.5-180 Hz through. Normalization of EEG amplitude is then carried out as the last step to minimize the difference in EEG amplitudes using min-max normalization across different subjects. After the preprocessing steps, spectrograms are generated for each EEG channel to transform data to the time-frequency domain. Each 30-second epoch is transformed into log-power spectra via a short-time Fourier transform (STFT) with a window size of two seconds and a 50 % overlap, followed by logarithmic scaling. A Hamming window and 256-point Fast Fourier Transform (FFT) are used on each epoch. This results in an image $\mathbf{S} \in \mathbb{R}^{F \times T}$ where $F = 129$ (the number of frequency bins), and $T = 29$ (the number of spectral columns).

4.3. Network Implementation

For extracting features for the adversarial neural network, we use the same architecture of Biswal et al. (Biswal et al. (2018)). It includes a 3-layer of 1-D CNN (kernel size = 3), which was applied to each EEG channel, followed by batch normalization (BatchNorm), rectified linear (ReLU) units, and max pooling units, we called it as SpectNet here. A cross-entropy loss function is used as a domain discriminator $\mathcal{L}_d$ and classification $\mathcal{L}_y$. We apply back-propagation to train the classifier layer and all domain discriminators. Mini-batch stochastic gradient descent (SGD) is employed with the momentum of 0.95 using the



(a) Common Channels (C3 and C4) with 3-layer Deep Neural Network



(b) All available channels with 3-layer Deep Neural Network DNN

Figure 4: To assess the efficacy of the proposed attentive adversarial network and relevant prior work, we developed two baseline systems for performance comparison. a) Extract the spectrogram from two common channels and feed to 3-layer convolutional layers followed by a softmax as the classifier. b) Extract the spectrograms from all available channels, feed to 3-layers convolutional layer, and utilize the softmax as a classifier for 5-stage sleep staging. Reproduced from (Nasiri (2020)) “CC by 4.0”.

learning rate and progressive training strategies as in (Ganin et al. (2016)) to learn the weights of a deep neural network. To address the class imbalance, we balance each batch for positive and negative examples, which leads to oversampling the positive class. The proposed methods were implemented with PyTorch 0.4 and Python3. See Figure 4.

4.4. Evaluation Metrics

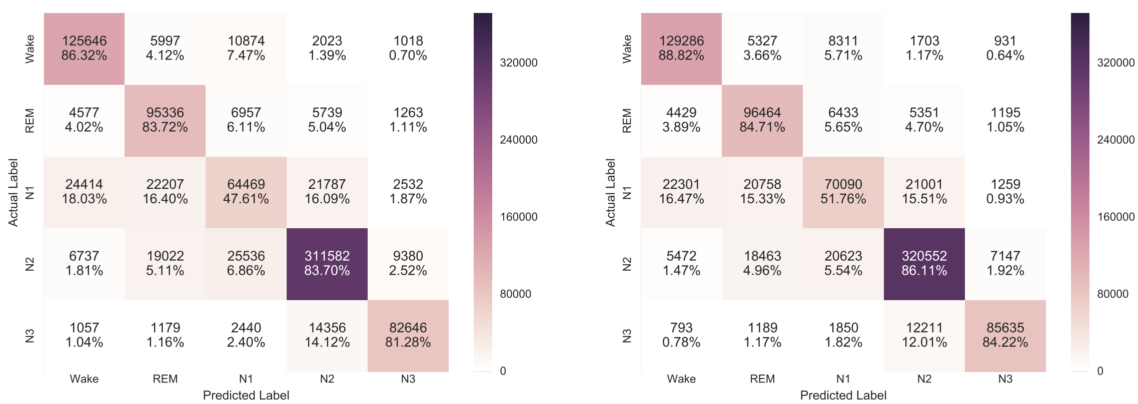
To evaluate the proposed approach performance and assess how adversarial domain adaptation network helps to develop a model with high generalizability, we initially conduct simple experiments. Similar to the literature on sleep stage assessment, to evaluate model performance, accuracy, specificity, sensitivity, and F1-score per class are reported. The other primary metric that we have used for performance evaluation of our proposed method is Cohen’s Kappa coefficient (κ). This metric measures the agreement between the labels obtained by the algorithm and the ground truth annotations. Due to a large number of patients, the SHHS database and the P18C database are considered as the training (labeled/source) and test (unlabeled/target) sets, respectively.

To assess the efficacy of the proposed attentive adversarial network, in addition to relevant prior work, we developed two baseline systems for performance comparison.

- Extract spectrograms of common EEG channels, C3 and C4, and use a 3-layer SpectNet, where it is followed by a fully connected neural network and softmax to classify sleep stages.
- Repeat the above experiment with all EEG channels.

5. Results

Figure (5) provides the confusion matrix for sleep staging, which shows the SpectNet agreement with expert scores. Sleep experts score each 30 second EEG epoch as wake, REM, non-REM stage 1, 2, or 3. Table (2) presents the performance of each class achieved by using a simple three-layer, SpectNet, with two common channels and all channels of P18C dataset, respectively. Based on this experiment, using all available channels of P18C database boosts the performance by 3% on average. It seems that when the algorithm exploits all channels, the N1 class can be better distinguished than with fewer channels. The N1 stage is often confused for wake and N2, and it is considered a transition period from being awake to falling asleep. Colten et al. (Altevogt et al. (2006)) defined the N1 stage as “active sleep”, which means N1 may also occur between other stages of sleep, such as between N3 and REM. Therefore, it is often confused with many other stages, as we can see in confusion matrices in Figure (5).



(a) Using common channels (C3 and C4)

(b) Using all channels

Figure 5: Confusion matrix for sleep staging for using all data from P18C dataset, showing SpectNet agreement with expert scores. The SpectNet is trained on SHHS dataset and tested on two common channels and all channels. Reproduced from (Nasiri (2020)) “CC by 4.0”.

Table 2: Per-class performance achieved using **two common channels** (C3 and C4) and **all available** channels by a 3-layer CNN network and a softmax layer

Sleep Stage	two common channels					all channels					# Samples
	Precision	Sensitivity	F1-Score	Kappa	Acc	Precision	Sensitivity	F1-Score	Kappa	Acc	
Wake	0.77	0.86	0.81	0.76	0.92	0.79	0.88	0.83	0.79	0.93	145558
REM	0.66	0.83	0.74	0.68	0.91	0.67	0.84	0.75	0.70	0.91	113872
N1	0.58	0.47	0.52	0.43	0.85	0.65	0.51	0.57	0.50	0.87	135409
N2	0.87	0.83	0.85	0.73	0.86	0.88	0.86	0.87	0.76	0.88	372257
N3	0.85	0.81	0.83	0.80	0.95	0.89	0.84	0.86	0.84	0.96	101678
avg	0.75	0.76	0.75	0.68	0.90	0.78	0.79	0.78	0.72	0.91	

Using all channels from two datasets and using adversarial domain adaptation (ADA) (Ganin et al. (2016)), with the SHHS database as the training set and the P18C database as the test set, as shown in Figure (6) we see an improved performance with all metrics. The performance on the test set is presented in Figure (7). Table (3) presents the performance of each class achieved with this method. One can conclude that adversarially learning transferable features across subjects boosts the performance of N1 class significantly. Finally, the performance of multi-stage classification using multi-adversarial neural network with attention mechanism is reported in Figure (7) and per-class performance is given in Table (3). In terms of evaluation metrics, which are mostly used in the sleep staging task, the proposed method outperforms the state-of-art algorithms on the 2018 P18C database. For instance, Perslev et al. (Perslev et al. (2019)) obtained 0.77 F1-score under 5-fold cross-validation on the same dataset. The average accuracy of proposed method is 0.94 on the unseen (P18C) database, which is significantly higher than other state-of-the-art methods (Biswal et al. (2018)). Our proposed method significantly beats their results after using adversarial training with an attention mechanism.

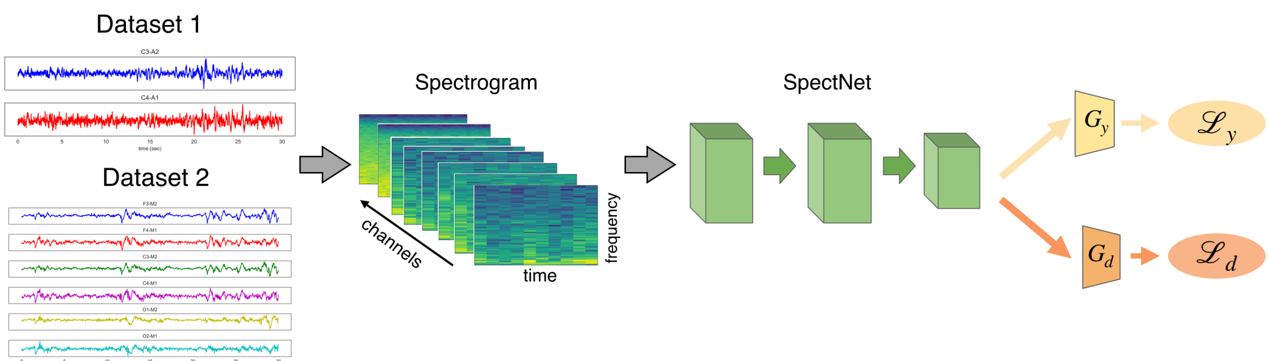
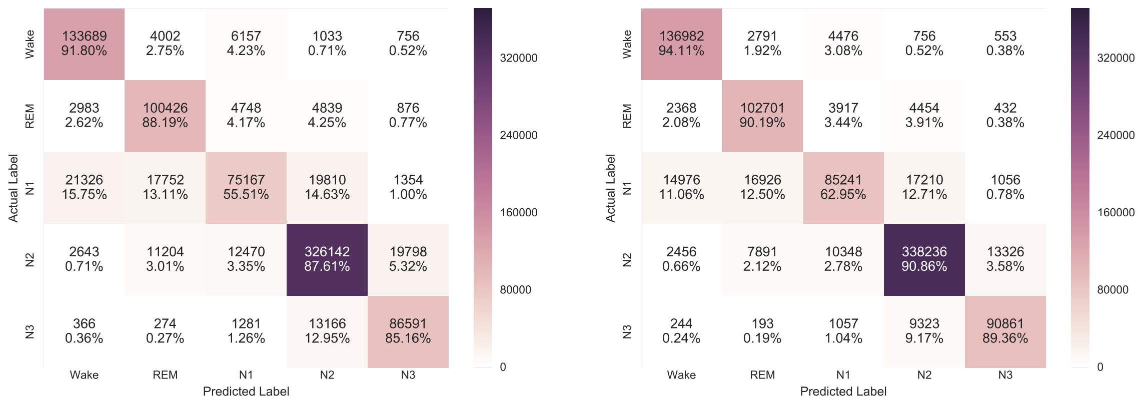


Figure 6: All channels from two datasets are used, and the spectrograms of EEG channels are extracted. The spectrograms are fed to the 3-layers convolutional network (SpectNet), followed by the domain discriminator and the classifier predictor. These three networks adversarially train (Ganin et al. (2016). Note that the SHHS database and the P18C database are used as the training and test sets, respectively. Reproduced from (Nasiri (2020)) “CC by 4.0”.



(a) Without any attention mechanisms

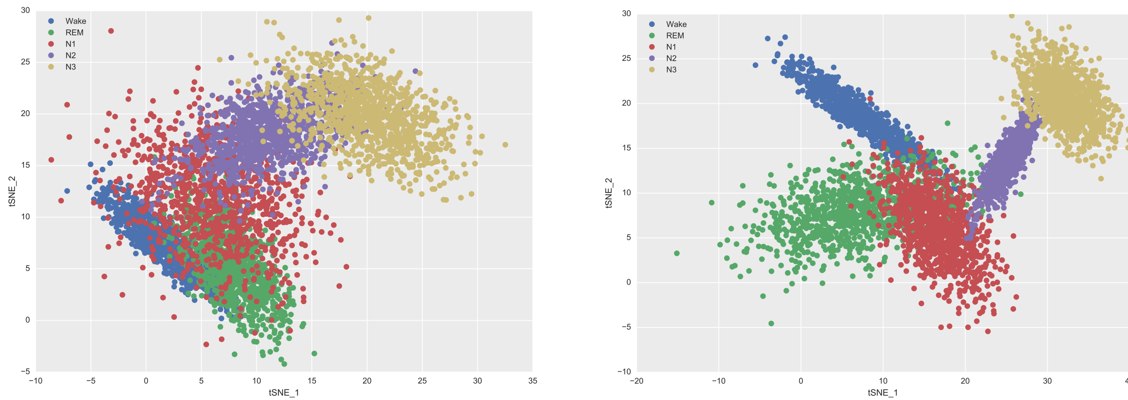
(b) With two attention mechanisms

Figure 7: Confusion Matrix using all channels with domain adversarial network, using SHHS database as training set and P18C database as test set without attention mechanism (Figure (6)) and with attention mechanisms (Figure (2)). Reproduced from (Nasiri (2020)) "CC by 4.0".

Table 3: Per-class performance achieved using adversarial domain adaptation network with/without attention mechanisms with all available

Sleep Stage	**without attention mechanism**					**with attention mechanism**					# Samples
	Precision	Sensitivity	F1-Score	Kappa	Acc	Precision	Sensitivity	F1-Score	Kappa	Acc	
Wake	0.83	0.91	0.87	0.83	0.94	0.87	0.94	0.90	0.88	0.96	145558
REM	0.75	0.88	0.81	0.77	0.93	0.78	0.90	0.84	0.81	0.95	113872
N1	0.75	0.55	0.63	0.57	0.89	0.81	0.62	0.70	0.66	0.91	135409
N2	0.89	0.87	0.88	0.78	0.98	0.91	0.90	0.91	0.83	0.91	372257
N3	0.80	0.85	0.82	0.79	0.95	0.85	0.89	0.87	0.85	0.96	101678
avg	0.81	0.81	0.80	0.75	0.92	0.84	0.85	0.84	0.80	0.94	

Feature Visualization: Figure (8) illustrates the discriminability of learned features in the deep neural network without/with adversarial training for scoring sleep stages, using t-distributed Stochastic Neighbor Embedding (Maaten and Hinton (2008)). The figure visualizes the network activations of the last hidden layer of DNN for each segment from 1000 samples for each class; 500 samples from the SHHS database and 500 samples from Physionet Challenge 2018 database. Figure (8(a)) shows the representations generated by a conventional deep neural network, where classes are less easily distinguished; there is significant confusion between N1 and wake, and between REM and N2. However, an adversarial neural network with an attention mechanism learns features with high transferability and discriminability. (see Figure (8(b).) As mentioned earlier, this figure shows that the N1 stage may not be like any stage, and it is considered as a transition stage between other stages (Wake, REM and N2).



(a) Without domain adapatation (b) With adversial domain adapatation

Figure 8: t-SNE visualization of the last hidden layer representations in the feature extraction network without/with adversarial training. Colored points represent the different stages, showing how the algorithm discriminate classes. Wake (blue), REM (green), N1 (red), N2 (purple) and N3 (flax). Reproduced from (Nasiri (2020)) “CC by 4.0”.

Attention Mechanism: To investigate the key channels (sensors) on the scalp, we show the attention weights across channels for a randomly selected sample from Physionet 2018 Challenge database. Figure 9 illustrates this - the hotter the color, the larger the attention value. It can be seen that the network pays more attention to features extracted from channel C4 rather than channel C3. Moreover, it seems that the C4 channel is a more transferable channel across databases and subjects. These results intuitively show which channel can be used for a wearable devices to capture sleep stages.

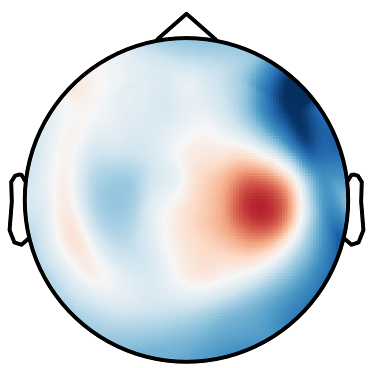


Figure 9: Attention visualization of of the last convolutional layer over sensors on the brain by illustrating the corresponding attentions weights over the channels. It can be seen that the network pays more attention to features extracted from channel C4 rather than channel C3. Moreover, it seems that the C4 channel is a more transferable channel across databases and subjects. Reproduced from (Nasiri (2020)) “CC by 4.0”.

6. Conclusion

In this work, adversarial training with an attention mechanism was proposed for the sleep staging task across two large and heterogeneous datasets. In the cross-dataset classification task, inherent inter-subject variability, hardware acquisition heterogeneity, and recording environment differences lead to different probability distributions between individuals, and hence poor generalization across subjects/dataset. Potentially, individuals with different biomedical demographics and phenotypes would provide enough diversity in the dataset, in which a conventional network cannot be robust to such variabilities, although given the need to factor in differences in montages, electrode placement errors and hardware systems, the dataset would likely be prohibitively large. The proposed method uses a multi-adversarial network to attend to relevant channels across datasets and highlight the important part of a segment of signal, and extract transferable features across the dataset, which achieves state-of-the-art performance (without prior knowledge) on a large public dataset, the Physionet 2018 Challenge database. The proposed method identified the important channel (C4), which suggests single-channel sleep staging with acceptable performance is possible. The method developed in this work can be applied to other biomedical signals (e.g. the electrocardiogram (ECG), electromyogram (EMG) and photoplethysmogram (PPG)), where multiple datasets from different hospitals are recorded for the same task. The ultimate goal of the research presented here, however, is to solve real-world automate sleep stage classification problems. Therefore, in addition to integrating adversarial training with attention mechanism, there are two main directions we would like to pursue for future work: 1) to apply the method in the cross-modality scenario, where we combine different modality such as EEG, ECG, and PPG which are recorded simultaneously in sleep; 2) to extend this method to leverage a dataset with different labels, i.e., partial domain adaptation, where the label sets are not equal across the dataset. This is a is much more challenging, but closer to real-world scenarios.

Moreover, it should be noted that DL models do not generalize well to unseen data. Typically, they are fine-tuned later on new patients, requiring new labels that are costly and time-consuming, which reduces clinical applicability. This work presents a method which reduces the labeling cost, by leveraging readily-available labeled data from a different but related dataset.

For generalizable insights, we note that the approach described here can apply to other similar domains that suffer from the same issues, such as electrocardiography (ECG). Placement of ECG leads is as subject to the user as in the case of EEG, and the body type variance is even larger, particularly between genders. Moreover, there is the potential to generalize the approach presented here could generalize to very different modalities, where the population clusters into different phenotypes, such as in imaging, voice recordings, or gait analysis, for example.

Acknowledgements

All images are adapted from Nasiri (2020) under a creative commons license “CC by 4.0”.

References

Recommended amount of sleep for a healthy adult: a joint consensus statement of the american academy of sleep medicine and sleep research society. *Sleep*, 38(6):843–844, 2015.

Bruce M Altevogt, Harvey R Colten, et al. *Sleep disorders and sleep deprivation: an unmet public health problem.* National Academies Press, 2006.

Alexandre Barachant, Stphane Bonnet, Marco Congedo, and Christian Jutten. Common spatial pattern revisited by riemannian geometry. In *2010 IEEE International Workshop on Multimedia Signal Processing*, pages 472–476. IEEE, 2010.

Richard B Berry, Rita Brooks, Charlene E Gamaldo, Susan M Harding, C Marcus, Bradley V Vaughn, et al. The aasm manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 176:2012, 2012a.

Richard B Berry, Rohit Budhiraja, Daniel J Gottlieb, David Gozal, Conrad Iber, Vishesh K Kapur, Carole L Marcus, Reena Mehra, Sairam Parthasarathy, Stuart F Quan, et al. Rules for scoring respiratory events in sleep: update of the 2007 aasm manual for the scoring of sleep and associated events. *Journal of clinical sleep medicine*, 8(05):597–619, 2012b.

Siddharth Biswal, Haoqi Sun, Balaji Goparaju, M Brandon Westover, Jimeng Sun, and Matt T Bianchi. Expert-level sleep scoring with deep neural networks. *Journal of the American Medical Informatics Association*, 25(12):1643–1650, 2018.

Samaneh Nasiri Ghosheh Bolagh, Mohammad Bagher Shamsollahi, Christian Jutten, and Marco Congedo. Unsupervised cross-subject BCI learning and classification using riemannian geometry. In *24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2016)*, 2016.

Samaneh Nasiri Ghosheh Bolagh, Gari Clifford, et al. Subject selection on a riemannian manifold for unsupervised cross-subject seizure detection. *arXiv preprint arXiv:1712.00465*, 2017.

Xin Chai, Qisong Wang, Yongping Zhao, Yongqiang Li, Dan Liu, Xin Liu, and Ou Bai. A fast, efficient domain adaptation technique for cross-domain electroencephalography (eeg)-based emotion recognition. *Sensors*, 17(5):1014, 2017.

Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174, 2017.

Rahul Duggal, Scott Freitas, Cao Xiao, Duen Horng Chau, and Jimeng Sun. Rest: Robust and efficient neural networks for sleep monitoring in the wild. *arXiv preprint arXiv:2001.11363*, 2020.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.

Nathalie TH Gayraud, Alain Rakotomamonjy, and Maureen Clerc. Optimal transport applied to transfer learning for p300 detection. 2017.

Mohammad M Ghassemi, Benjamin E Moody, Li-Wei H Lehman, Christopher Song, Qiao Li, Haoqi Sun, Roger G Mark, M Brandon Westover, and Gari D Clifford. You snooze, you win: Physionet/computing in cardiology challenge 2018. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4. IEEE, 2018.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Shawn Hershey, Sourish Chaudhuri, Daniel PW Ellis, Jort F Gemmeke, Aren Jansen, R Channing Moore, Manoj Plakal, Devin Platt, Rif A Saurous, Bryan Seybold, et al. Cnn architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)*, pages 131–135. IEEE, 2017.

Vinay Jayaram, Morteza Alamgir, Yasemin Altun, Bernhard Scholkopf, and Moritz Grosse-Wentrup. Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, 11(1):20–31, 2016.

Alexander Ya Kaplan, Andrew A Fingelkurts, Alexander A Fingelkurts, Sergei V Borisov, and Boris S Darkhovsky. Nonstationary nature of the brain activity as revealed by eeg/meg: methodological, practical and conceptual challenges. *Signal processing*, 85(11): 2190–2212, 2005.

Hyekyoung Lee and Seungjin Choi. Group nonnegative matrix factorization for eeg classification. In *Artificial Intelligence and Statistics*, pages 320–327, 2009.

Jinpeng Li, Shuang Qiu, Changde Du, Yixin Wang, and Huiguang He. Domain adaptation for eeg emotion recognition based on latent representation similarity. *IEEE Transactions on Cognitive and Developmental Systems*, 2019.

Yili Li, Kon Max Wong, and Hubert de Bruin. Electroencephalogram signals classification for sleep-state decision–a riemannian geometry approach. *IET signal processing*, 6(4): 288–299, 2012.

Hong Liu, Mingsheng Long, Jianmin Wang, and Michael Jordan. Transferable adversarial training: A general approach to adapting deep classifiers. In *International Conference on Machine Learning*, pages 4013–4022, 2019.

Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234: 11–26, 2017.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791*, 2015.

Bo-Qun Ma, He Li, Wei-Long Zheng, and Bao-Liang Lu. Reducing the subject variability of eeg signals with adversarial domain generalization. In *International Conference on Neural Information Processing*, pages 30–42. Springer, 2019.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Alexander Malafeev, Dmitry Laptev, Stefan Bauer, Ximena Omlin, Aleksandra Wierzbicka, Adam Wichniak, Wojciech Jernajczyk, Robert Riener, Joachim Buhmann, and Peter Achermann. Automatic human sleep stage scoring using deep neural networks. *Frontiers in neuroscience*, 12:781, 2018.

Samaneh Nasiri. *Generalizable Machine Learning Methods for Electrophysiology*. PhD thesis, Emory University, 2020.

Mathias Perslev, Michael Jensen, Sune Darkner, Poul Jørgen Jennum, and Christian Igel. U-time: A fully convolutional network for time series segmentation applied to sleep staging. In *Advances in Neural Information Processing Systems*, pages 4417–4428, 2019.

Haider Raza and Spyridon Samothrakis. Bagging adversarial neural networks for domain adaptation in non-stationary eeg. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2019.

Pedro Rodrigues, Marco Congedo, and Christian Jutten. A data imputation method for matrices in the symmetric positive definite manifold. 2019.

Elizaveta Saifutdinova, Marco Congedo, Daniela Dudysova, Lenka Lhotska, Jana Koprivova, and Vaclav Gerla. An unsupervised multichannel artifact detection method for sleep eeg based on riemannian geometry. *Sensors*, 19(3):602, 2019.

Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8503–8512, 2018.

Bonggun Shin and Alice Oh. Bayesian group nonnegative matrix factorization for eeg analysis. *arXiv preprint arXiv:1212.4347*, 2012.

M Emin Tagluk, Necmettin Sezgin, and Mehmet Akin. Estimation of sleep stages by an artificial neural network employing eeg, emg and eog. *Journal of medical systems*, 34(4): 717–725, 2010.

Xingliang Tang and Xianrui Zhang. Conditional adversarial domain adaptation neural network for motor imagery eeg decoding. *Entropy*, 22(1):96, 2020.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7167–7176, 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Ximei Wang, Liang Li, Weirui Ye, Mingsheng Long, and Jianmin Wang. Transferable attention for domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5345–5352, 2019.

Magdy Younes and Patrick J Hanly. Minimizing interrater variability in staging sleep by use of computer-derived features. *Journal of Clinical Sleep Medicine*, 12(10):1347–1356, 2016.

Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, and Yannick Berthoumieu. Transfer learning: a riemannian geometry framework with applications to brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 2017.

Linda Zhang, Daniel Fabbri, Raghu Upender, and David Kent. Automated sleep stage scoring of the sleep heart health study using deep neural networks. *Sleep*, 42(11):zsz159, 2019.