

Meta-Learning on Spectral Images of Electroencephalogram of Schizophrenics

Maritza Tynes^{1,2}, Mahboobeh Parsapoor (Mah Parsa)^{1,3,4,5}

¹WinAIML, ² University of London, ³ University of Toronto, ⁴ Vector Institute for Artificial Intelligence, ⁵ McGill University
mtynes@acm.org, mahparsa@cs.toronto.edu

Abstract

Schizophrenia is a complex psychiatric disorder involving changes in thought patterns, perception, mood, and behavior. The diagnosis of schizophrenia is challenging and requires that patients show two or more positive symptoms (e.g., disorganized speech) for at least one month. Delays in identifying this debilitating disorder can impede a patient’s ability to receive much needed treatment. Advances in neuroimaging and machine learning algorithms can facilitate the diagnosis of schizophrenia and help clinicians to provide an accurate diagnosis of the disease.

This paper presents a methodology for analyzing spectral images of Electroencephalography collected from patients with schizophrenia using convolutional neural networks. It also explains how we have developed accurate classifiers employing Model-Agnostic Meta-Learning (MAML) and prototypical networks. Such classifiers have the capacity to distinguish people with schizophrenia from healthy controls based on their brain activity.

Introduction

Schizophrenia (SZ) is a chronic mental disorder characterized by impairments in cognitive function, including executive function, sensory perception, and affective function. Patients typically present with disorganized thinking, auditory hallucinations, emotional dysfunction, and social withdrawal (Patel et al. 2014). The diagnosis of schizophrenia is challenging for psychiatrists. The differential diagnosis must exclude other disorders which can elicit similar symptoms, such as post-traumatic stress disorder, subtypes of major depressive disorder, obsessive-compulsive disorder, and schizoaffective disorder. It must also take into account possible substance abuse and co-morbid disorders (Association et al. 2013; Patel et al. 2014)

Thus the process of diagnosis often involves using different assessment tools (Balda et al. 2015), including neuroimaging techniques. Various studies have supported the use of EEG to detect abnormal brain activity associated with schizophrenia (Seiler et al. 2019; Shatte, Hutchinson, and Teague 2019; Li et al. 2008; Karson et al. 1988; Chu et al. 2017).

Machine learning algorithms, including random forest (RF), support vector machines (SVMs), and deep learning algorithms such as convolutional neural networks (CNN) have received substantial attention in the ML community for their ability to distinguish between EEG data from patients with SZ and data from healthy controls (Chu et al. 2017; Hiesh et al. 2013; Li et al. 2008; Karson et al. 1988; Craik, He, and Contreras-Vidal 2019). However, while deep learning algorithms have been successful in efficiently analyzing neuroimaging data, they cannot be trained correctly using the small sample size typically available to clinicians. To address this issue, using meta-learning on neuroimaging data was introduced (Duan et al. 2020; Wang et al. 2019).

In this paper, we present a methodology for using meta-learning on spectral images extracted from EEG. To the best of our knowledge, it is the first time that meta-learning has been utilized to improve the classification accuracy of deep learning models when using neuroimaging data from healthy controls and patients with schizophrenia. After this introduction, we describe the EEG datasets used in our experiments and illustrate our methodology to employ Model-Agnostic Meta-Learning (MAML) and prototypical networks with CNN to classify the EEG data. We show the preliminary results obtained using our methodology and highlight our main contributions, and possible directions for future work.

Datasets

We have utilized two publicly available raw EEG datasets (see Table 1 for detailed information about the two datasets). Both datasets utilize the 10-20 electrode (Homan, Herman, and Purdy 1987) placement system, Figure 1, with 16 overlapping channels. The first dataset (hereafter referred to as **Dataset-1**) was obtained from ($N=14$) patients with paranoid SZ and ($N=14$) controls (HC) (Olejarczyk and Jernajczyk 2017). The second dataset, which is referred to as **Dataset-2**, consists of EEG readings from healthy adolescents (HC) and adolescents exhibiting traits of SZ. The subtype of schizophrenia for Dataset-2 has not been specified. The original data contains 39 samples for the HC group and 45 samples for the SZ group (Gorbachevskaya and Borisov. 2015). To match the size of the HC group ($N=39$), we have randomly selected $N=39$ of the SZ subjects.

Table 1: Summary of the EEG datasets

Specification	Dataset-1	Dataset-2
Number of healthy controls	14	39
Number of schizophrenic patients	14	45*
Schizophrenia subtype	Paranoid	Not determined
Minimum sample length (seconds)***	740	60
Maximum sample length (seconds)	2170	60
Median sample length (seconds)	925	60
Sampling rate (Hz)	250	128
Montage	10-20	10-20
Number of channels	19	16
Labelled electrodes	Fp1**, Fp2**, F7, F3, Fz**, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2	F7, F3, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2
Reference electrode	Between Fz and Cz	N/A

*39 SZ subjects were randomly chosen to match the number of HC subjects. ** Channels labeled Fp1, Fp2 and Fz were used in Dataset 1 but not in Dataset 2 ***All samples were limited to the length of the shortest sample.

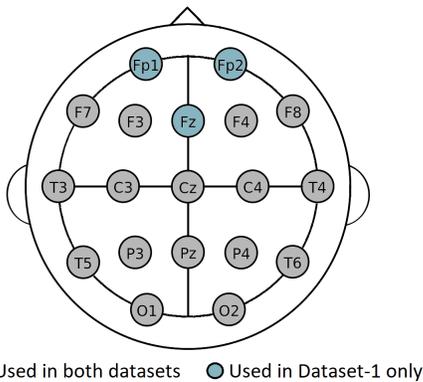


Figure 1: Electrode placements of two datasets. It describes electrode placements to collect two datasets. For both data collection, the 10-20 system was used. The 16 nodes colored with gray used in both datasets; While the nodes colored with blue used in Dataset-1.

Data Cleaning and Pre-processing

We have removed the EEG samples of five participants in Dataset-1 using an outlier detection method based on tree classifiers. Since Dataset-2 presented a more homogeneous distribution, we have not removed any raw EEG samples from this dataset. We have not excluded any channels in the data cleaning procedure. As previously mentioned, both datasets utilize the standard 10-20 placement system. Only the Fp1, Fp2, and Fz channels are present in the first dataset and not the second (See Figure 1).

For Dataset-1, all data was cropped to the length of the shortest subject sample (740 seconds). A padding of 60 seconds was also applied to exclude artifacts present from subjects moving during the beginning and end of the data acquisition process. The result is a set of 500-second samples; one for each participant. In Dataset-2, no cropping was used and the full sample (60 seconds) was processed for all participants.

We have divided raw EEG into time windows with substantial overlap. This method is similar to (Firpi, Goodman, and Echauz 2005) and (Cosmo et al. 2015), where a "sliding

window" is used to sub-sample the data by taking a sample, applying an offset/overlap, and taking another sample (See figure 2, part 1)). Such a procedure allows us to produce a higher number of total samples for each class, and places the focus on differences in fine features. We hypothesize that using shorter samples can also encourage the classifiers to ignore sub-samples with aberrant data. For Dataset-1, the time window used was 20 seconds, while for Dataset-2, a time window of 5 seconds was used. The overlap rate was set at 70 percent in both cases to provide more substantial and representative samples. From this point, the raw data was converted into spectrograms while maintaining all channels.

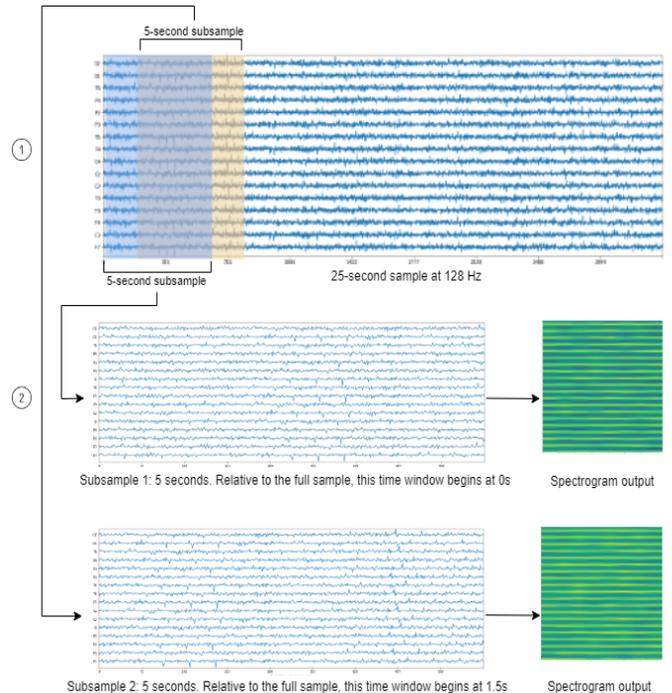


Figure 2: Data preprocessing (1)Input is subsampled using 5-second time windows and a 70% overlap rate. (2) Subsamples are converted into spectrograms while retaining all channels.

For experiments where a single dataset has been used, training, validation and testing sets have been randomly selected with a static seed. The ratios for the split have been set at 0.6, 0.2 and 0.2, respectively. Throughout our experiments, we have labelled data from healthy controls (HC) as 0 and data from schizophrenic subjects as 1.

Methodology

We utilize a simple CNN as our benchmark and propose two methodologies that combine meta-learning algorithms with CNN to analyze image extracted from two EEG datasets that were described in the previous section. The first meta-learning methodology uses MAML with CNN. The second methodology combines CNN with prototypical networks.

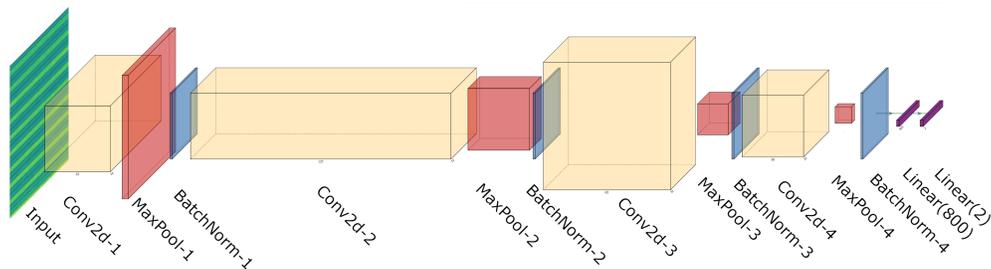


Figure 3: Benchmark CNN architecture

CNN

To first analyze the use of a common deep learning model with our data, we have implemented a simple CNN whose architecture is summarized in Table 2 and Figure 3. It consists of 4 convolutional layers, with 32 channels and a kernel size of 3. Each layer is followed by a ReLU layer, a max-pooling layer with a kernel size of 2, and a batch normalization layer of 32 features. This is followed by two fully connected layers which use linear activation. Training has been implemented over 400 epochs. We have utilized an Adam optimizer with a learning rate of $1e-3$ and a binary cross-entropy loss function.

Table 2: Description of the CNN architecture

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 32, 254, 254]	896
BatchNorm2d-2	[-1, 32, 127, 127]	64
Conv2d-3	[-1, 32, 125, 125]	9,248
BatchNorm2d-4	[-1, 32, 62, 62]	64
Conv2d-5	[-1, 32, 60, 60]	9,248
BatchNorm2d-6	[-1, 32, 30, 30]	64
Conv2d-7	[-1, 32, 28, 28]	9,248
BatchNorm2d-8	[-1, 32, 14, 14]	64
Linear-9	[-1, 800]	5,018,400
Linear-10	[-1, 2]	1,602

Total params: 5,048,898

Trainable params: 5,048,898

As another experiment, we have used the same CNN architecture and added a pre-training step using the miniImageNet dataset, and a fine-tuning step where one of EEG datasets is used to further train the model and the other EEG dataset is used for testing the model (i.e., where Dataset-1 is tested, Dataset-2 is used for fine-tuning). Pre-training was run over 200 epochs, followed by 200 epochs of fine-tuning. This was constructed as a point of comparison for a subset of the MAML experiments, which employs a similar use of the datasets.

MAML

Finn et al., introduced MAML as an optimization algorithm which is compatible with any class of machine learning model capable of learning through gradient descent (e.g., regression, reinforcement learning and classification) (Finn,

Abbeel, and Levine 2017). MAML is a kind of few-shot learning algorithm which mirrors human learning by utilizing representations of one dataset (training step) to improve generalizations about a second dataset (fine-tuning step). It uses a few or only one gradient step to build the representations. Data from the second distribution is normally limited and the model attempts to find parameters which are optimal for the specific task (Finn, Abbeel, and Levine 2017).

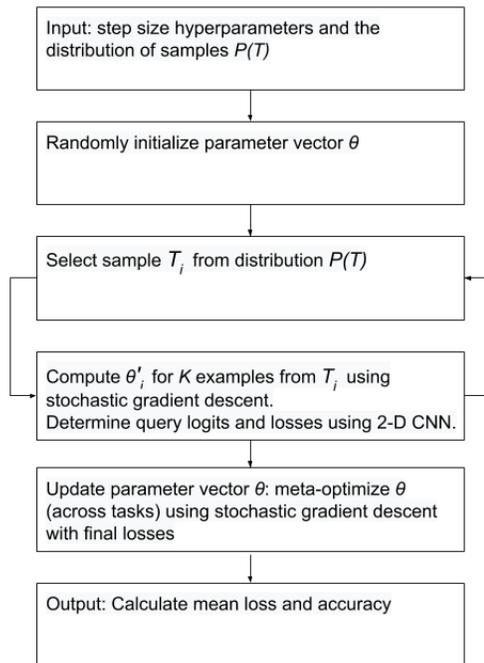


Figure 4: Flowchart of MAML training: forward sequence Steps for computing theta parameters, logits and losses during forward function during training.

We have considered an implementation of MAML using a CNN with cross entropy loss. The CNN used is similar in structure to the CNN architecture described in Table 2 . With each EEG dataset, we execute two experiments using MAML. For both experiments, the model is first pre-trained with two classes from the miniImageNet dataset.

For the simple implementations of MAML with CNN, the fine-tuning step is completed with the spectrograms of

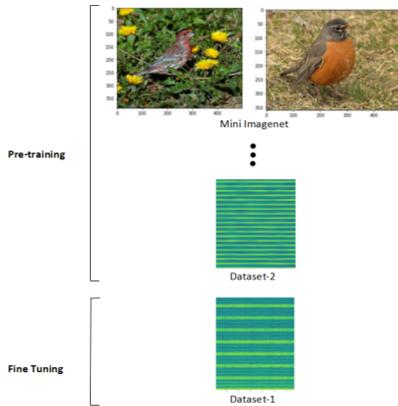


Figure 5: MAML input for Dataset-1 with further training. Samples of input for Dataset-1 experiments. Two classes from miniImageNet are used for pre-training. In the further training experiment, training continues with input from Dataset-2. Fine tuning is done with Dataset-1.

Dataset-1 and Dataset-2 separately.

To further improve results, a second training phase was added for our second MAML experiment. This was done to evaluate the use of further training with a similar but qualitatively distinct distribution. This phase of training would normally be referred to as fine-tuning. However, we refer to it as further training to avoid confusion with the step in the MAML pipeline which evaluates the query set.

In this further training phase, for each of the EEG datasets (i.e., Dataset-1, Dataset-2), we use the other dataset to provide the model with data from a similar distribution. For example, for Dataset-1, training is first done with miniImageNet, then further training is done with Dataset-2. In this case, Dataset-1 is used for fine-tuning, which is performed every 500 steps. See Figure 4 for a visual explanation.

Prototypical Networks

Prototypical networks are another class of few-shot learning algorithms. They have been proposed to address the issues of limited data and resources within most clinical settings (Snell, Swersky, and Zemel 2017). In this approach, each class has a support set (i.e., randomly selected portion of the sample input, which is split into episodes), and this is used to calculate the mean embedding of that class’s labeled subset. The support set is used to represent the prototype of that class. With each training step, the distance between classes is determined using stochastic gradient descent (SGD) which minimizes the loss function. Then a query set, which is randomly selected embedded input, is chosen from the remaining sample. Next, a set of Euclidean distances is calculated to define the distances between the query set and prototypes of various classes. Finally, the query set’s class is determined to be the class of the prototype with the smallest relative Euclidean distance. (See Figure 6 for an illustration of the mean class embeddings (prototype), and the Euclidean distance between a query embedding and a prototype.)

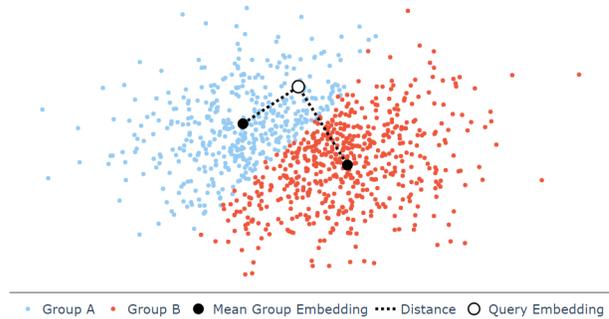


Figure 6: Prototypical networks with two classes. Prototypes (black dots) for each class are calculated as the mean vector of the embeddings of the support set (colored dots). The class of the query embedding (white dot with gray circle) is determined by finding the nearest prototype using Euclidean distance. This diagram was created using sample data only.

We have used prototypical networks along with a simple CNN to analyze the spectral images. The CNN consists of a stack of convolutional segments, each identical except for the input and output dimensions. Each block is comprised of a Conv2d layer with a kernel size of 3 and padding of 1 unit. This is followed by a batch normalization layer (with the number of features set as the Conv2d output size), a ReLU (Rectified Linear Unit) layer and a max-pooling layer with a size of 2. No pre-training is performed.

Results

To evaluate the performance of our suggested methodology, we have considered the following datasets 1) Dataset-1 and Dataset-2; 2) the miniImageNet dataset (Vinyals et al. 2016). The dataset consists of 100 unique object classes, each with 600 84 x 84 pixel samples per class. For its usage in our experiments, we have manually selected two classes to act as training input. These classes have been chosen for their visual similarity. This has been done to encourage the models to focus on utilizing finer details to distinguish between the classes. No pre-processing or cleaning was performed on this data.; As another experiment, we have used the same CNN architecture and added a pre-training step using the miniImageNet dataset, and a fine-tuning step where one of EEG datasets is used to further train the model and the other EEG dataset is used for testing the model (i.e., where Dataset-1 is tested, Dataset-2 is used for fine-tuning). Pre-training was run over 200 epochs, followed by 200 epochs of fine-tuning. This was constructed as a point of comparison for a subset of the MAML experiments, which employs a similar use of the datasets. We have employed a simple CNN model (see Figure 3) on both Dataset-1 and Dataset-2 (see Table 3) and also combined it with MAML and have conducted fine-tuning/further training-see the rows named "MAML+CNN" in the Table 3. To do so, we have used the CNN model and added a pre-training step using the miniImageNet dataset, and a fine-tuning step where one of EEG datasets is used to further train the model and the other EEG dataset is used for testing the model (i.e., where Dataset-1

is tested, Dataset-2 is used for fine-tuning). Pre-training was run over 200 epochs, followed by 200 epochs of fine-tuning. This was constructed as a point of comparison for a subset of the MAML experiments, which employs a similar use of the datasets. We have also employed prototypical networks with the CNN model for both datasets (see Table 3). Our results showed that the CNN classifier failed to accurately classify the first EEG dataset (i.e., the CNN classifier produced relatively low test accuracy (0.6646)). One possible explanation for this result is that underfitting occurred with Dataset-1. For the second dataset, it produced perfect accuracy for the validation and test sets. Using CNN with prototypical networks produced high validation and test accuracy, yielding 0.9768 and 0.8587 test accuracy for the Dataset-1 and Dataset-2, respectively.

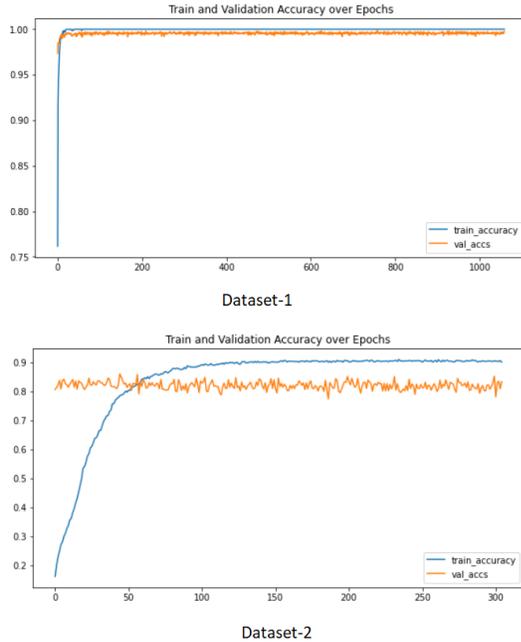


Figure 7: Prototypical networks: training and validation accuracy

For Dataset-1, prototypical networks resulted in higher test accuracy than MAML (0.9768 compared to 0.9342). The reverse was true for Dataset-2, where prototypical networks had an accuracy of 0.8587, compared to 0.9489 for MAML. Among the MAML experiments, tests including further training produced substantially higher accuracy for both datasets. This improvement can be seen in the graphs of Figure 8 which show training and validation accuracy by epoch. Validation in this case represents the fine tuning step using the query set. There are substantial improvements immediately when the more similar dataset is used.

This shows that while MAML provides satisfactory performance when using an unrelated dataset for training, the results can be substantially improved by including data from a similar distribution. It should be noted that the loss values included for MAML are in reference to the query sets, which

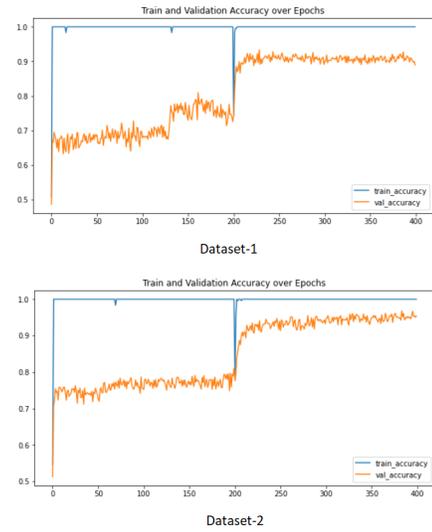


Figure 8: MAML with further training: training and validation (fine tuning) accuracy

are considered to be more relevant than support set loss in this case.

Table 3: Models' performance

Network Name	Val Loss	Val Acc	Test Acc	Test AUC	Test F1 Macro
Dataset-1					
CNN	1.6223	0.5000	0.6646	0.8128	0.6468
CNN + Fine Tuning	0.6348	0.5455	0.5021	0.5014	0.3579
Prototypical Networks	0.0112	0.9977	0.9768	0.9772	0.9768
MAML + CNN	0.5000	0.8945	0.8420	0.8426	0.8419
MAML + CNN with Further Training	0.2184	0.9342	0.9683	0.9342	0.9341
Dataset-2					
CNN	0.0000	1.0000	1.0000	1.0000	1.0000
CNN + Fine Tuning	0.5805	0.5014	0.6814	0.6892	0.6560
Prototypical Networks + CNN	0.5980	0.8333	0.8587	0.8580	0.8584
MAML + CNN	1.317	0.7749	0.8133	0.8122	0.8097
MAML + CNN with Further Training	0.1888	0.9541	0.9489	0.9489	0.9489

Discussion

The strategy of using deep learning algorithms such as CNN to analyze EEG data and detect mental disorders is an expanding area of interest for mental health professionals and machine learning researchers. One of our future research works will focus on detecting other mental disorders by employing deep learning, and meta learning with deep learning to analyze EEG datasets. Another possible direction for our future work involves the examination of CapsuleNet and "Emotion-inspired Deep Structure (EiDS)" (Parsapoor 2020) on the EEG datasets. We could also perform a more thorough analysis to determine the ideal CNN architecture for neuroimaging tasks with MAML and/or Prototypical Networks. Further, other methods of visualizing EEG can

be reviewed. While we have used spectral images in the experiments of this paper, we could also examine short-time Fourier transforms, and Stockwell transforms (Stockwell, Mansinha, and Lowe 1996) as visual representations of time-series data.

Conclusion

This paper shows the preliminary results of using MAML and prototypical networks with CNN to analyze spectral images of EEG from healthy individuals and patients with schizophrenia. Our results have verified that if prototypical networks or MAML are used with deep learning classifiers, we can classify neuroimaging data with higher accuracy than with deep learning alone. We have also shown that neuroimaging data from a different dataset or distribution can be used to improve the classification accuracy in deep learning and meta learning tasks.

References

- Association, A. P.; et al. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
- Balda, M.; Calvó, M.; Padilla, E.; Guerrero, G.; Molina, J.; Florenzano, N. V.; Kamis, D.; Escobar, J. I.; Cloninger, C. R.; and de Erausquin, G. 2015. Detection, assessment, and management of schizophrenia in an Andean population of South America: parkinsonism testing and transcranial ultrasound as preventive tools. *Focus* 13(4): 432–440.
- Chu, L.; Qiu, R. C.; Liu, H.; Ling, Z.; and Shi, X. 2017. Individual Recognition in Schizophrenia using Deep Learning Methods with Random Forest and Voting Classifiers: Insights from Resting State EEG Streams. *CoRR* abs/1707.03467.
- Cosmo, C.; Ferreira, C.; Miranda, J. G.; Rosário, R.; Baptista, A.; Montoya, P.; and De Sena, E. 2015. Spreading Effect of tDCS in Individuals with Attention-Deficit/Hyperactivity Disorder as Shown by Functional Cortical Networks: A Randomized, Double-Blind, Sham-Controlled Trial. *Frontiers in Psychiatry* 6. doi:10.3389/fpsyt.2015.00111.
- Craik, A.; He, Y.; and Contreras-Vidal, J. L. 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering* 16(3): 031001.
- Duan, T.; Chauhan, M.; Shaikh, M. A.; and Srihari, S. 2020. Ultra Efficient Transfer Learning with Meta Update for Cross Subject EEG Classification.
- Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. *arXiv preprint arXiv:1703.03400*.
- Firpi, H.; Goodman, E.; and Echaz, J. 2005. Epileptic seizure deflection by means of genetically programmed artificial features. 461–466. doi:10.1145/1068009.1068082.
- Gorbachevskaya, N. N.; and Borisov., S. V. 2015. EEG of healthy adolescents and adolescents with symptoms of schizophrenia. *EEG Database - Schizophrenia* URL http://brain.bio.msu.ru/eeg_schizophrenia.htm.
- Hiesh, M.; Andy, Y. L.; Shen, C.; Chen, W.; Lin, F.; Sung, H.; Lin, J.; Chiu, M.; and Lai, F. 2013. Classification of schizophrenia using Genetic Algorithm-Support Vector Machine (GA-SVM). In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6047–6050. ISSN 1094-687X. doi:10.1109/EMBC.2013.6610931.
- Homan, R. W.; Herman, J.; and Purdy, P. 1987. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology* 66(4): 376–382.
- Karson, C. N.; Coppola, R.; Daniel, D. G.; and Weinberger, D. R. 1988. Computerized EEG in schizophrenia. *Schizophrenia bulletin* 14(2): 193.
- Li, Y.; Tong, S.; Liu, D.; Gai, Y.; Wang, X.; Wang, J.; Qiu, Y.; and Zhu, Y. 2008. Abnormal EEG complexity in patients with schizophrenia and depression. *Clinical Neurophysiology* 119(6): 1232 – 1241. ISSN 1388-2457. doi:<https://doi.org/10.1016/j.clinph.2008.01.104>. URL <http://www.sciencedirect.com/science/article/pii/S1388245708000461>.
- Olejarczyk, E.; and Jernajczyk, W. 2017. EEG in schizophrenia. doi:10.18150/repod.0107441. URL <https://doi.org/10.18150/repod.0107441>.
- Parsapoor, M. 2020. Emotion-Inspired Deep Structure (EiDS) for EEG Time Series Forecasting.
- Patel, K. R.; Cherian, J.; Gohil, K.; and Atkinson, D. 2014. Schizophrenia: overview and treatment options. *Pharmacy and Therapeutics* 39(9): 638.
- Seiler, N.; Nguyen, T.; Yung, A.; and O’Donoghue, B. 2019. Terminology and Assessment Tools of Psychosis: A Systematic Narrative Review. *Psychiatry and Clinical Neurosciences*.
- Shatte, A. B.; Hutchinson, D. M.; and Teague, S. J. 2019. Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine* 49(9): 1426–1448.
- Snell, J.; Swersky, K.; and Zemel, R. 2017. Prototypical networks for few-shot learning. In *Advances in neural information processing systems*, 4077–4087.
- Stockwell, R.; Mansinha, L.; and Lowe, R. 1996. Localisation of the complex spectrum: the S transform. *Journal of Association of Exploration Geophysicists* 17(3): 99–114.
- Wang, Y.; Qiu, S.; Zhao, C.; Yang, W.; Li, J.; Ma, X.; and He, H. 2019. EEG-Based Emotion Recognition with Prototype-Based Data Representation. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 684–689. doi:10.1109/EMBC.2019.8857340.

Appendices

The EEG datasets and CNN codes are available in the WinAIML Github.