Deep learning topology–preserving EEG–based images for autism detection in infants*

Cosmin Stamate[1,2][0000−0003−0386−7331], George D. Magoulas[1][0000−0003−1884−0772], Michael S.C. Thomas[2][0000−0002−8231−6011], and the BASIS Team

[1] Birkbeck Knowledge Lab, Birkbeck College, University of London, UK
{cosmin, gmagoulas}@dcs.bbk.ac.uk
[2] Developmental Neurocognition Lab, Birkbeck College, University of London, UK
m.thomas@bbk.ac.uk

Abstract. Developing digital biomarkers that would enable reliable detection of autism–ASD early in life is challenging because of the variability in the presentation of the autistic disorder and the need for simple measurements that could be implemented routinely during checkups. Electroencephalography, widely known as EEG, is an electrophysiological monitoring method that has been explored as a potential clinical tool for monitoring atypical brain function. EEG measurements were collected from 101 infants, beginning at 12 to 15 months of age and continuing until 36 months of age. In contrast to previous work in the literature that analyses EEG signals, our approach considers EEG–as–an–image, using an appropriate signal transformation that preserves the spatial location of the EEG signals to create RGB images. It employs Residual neural networks to detect atypical brain function. Prediction of the clinical diagnostic outcome of ASD or not ASD at 36 months was accurate from as early as 12 months of age. This shows that using end-to-end deep learning is a viable way of extracting useful digital biomarkers from EEG measurements for predicting autism in infants.

Keywords: Autism spectrum disorder (ASD) · Deep Learning · EEG · Residual Networks · Early Stopping.

1 Introduction

Autistic spectrum disorder (ASD) refers to a range of conditions characterised by a reduced level of social interaction, impaired communication and language, as well as a narrow scope of repetitive interests and activities. Symptoms develop within the first five years of life and often persist into adolescence and adulthood. Intellectual functioning is extremely variable, ranging from profound impairment to superior abilities. One in every 160 children in the world is diagnosed with ASD [3]. This estimate reflects an average prevalence, it varies

* Project affiliated with The British Autism Study of Infant Siblings (BASIS) Research Network (www.basisnetwork.org).

greatly across studies, and there are some studies that report much higher figures [15]. High–risk infant siblings studies have demonstrated that the defining behavioural features of ASD become apparent during the first and second years of life [9]. As ASD is behaviourally and not biologically defined, the diagnosis in the first years of life, when intervention should be most effective remains challenging [12,13]. In particular, milder forms of ASD are especially difficult to detect while young because of diagnostic uncertainty that often surrounds the diagnosis of neurodevelopmental disorders [9]. However, even in that case, developing useful biomarkers for detecting the occurrence of emerging ASD symptoms might be helpful for implementing early interventions tailored to the individual's needs or to trigger more sensitive behavioural assessments [11].

During an EEG, electrical signals from the brain are recorded through a circular array of small metal discs, called electrodes, placed on the scalp. This activity is generated by brain cells, even when we are asleep, and appears on an EEG recording as waveforms. This has proven to be a low cost, reliable method for assessing various neurodegenerative diseases. Its non–invasive and low cost manner makes it a very good candidate for the detection of ASD in infants.

This paper proposes an end–to–end data driven approach to finding reliable and robust digital biomarkers in EEG signals of infants (12 to 15 months old) at high familial risk of developing ASD. Most previous investigations have focused on uni–dimensional biomarkers [4,5], which, albeit successful in predicting a higher chance of developing autism symptoms, fall short of achieving very high predictability or helping to identify clinical subgroups. A data transformation procedure inspired from [2] is applied to translate the fast Fourier transform (FFT) of raw EEG signal windows into images, whilst preserving the spatial location of the signal on the scalp. The result is a sequence of topology–preserving multi–spectral images, as opposed to standard EEG analysis techniques that ignore such spatial information. Once such an EEG "movie" is obtained, Deep Residual Networks with convolutional layers [6] are used to learn robust representations from the sequence of images, or frames.

The next section describes the EEG data collection process. Section 3 introduces the proposed framework. Experimental results are presented in Section 4. The paper ends with concluding remarks and future work.

2 EEG Data Collection in Infants

The EEG data used in this work were collected in the first phase of the British Autism Study of Infant Siblings–BASIS (`www.basisnetwork.org`). This is a large scale collaborative project aiming to understand the biological origin of ASD by following up longitudinally cohorts of infants at elevated family risk for this disorder. Infants participating in the study had an older sibling with ASD (high risk group; HR) or without ASD (low risk group; LR). Families, enrolled from various regions when their babies were younger than 5 months of age, were

invited to attend multiple research visits (6–9, 12–15, 24 and 36 months) until their children reach 3 years of age or beyond [4,5].

At the time of enrolment, none of the participating infants had been diagnosed with any major medical or developmental condition. All infants belonging to the HR group had an older sibling with a community clinical diagnosis of an autism spectrum disorder [4]. Infants in the LR group were recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. Inclusion criteria included full–term birth (with one exception), normal birth weight, and lack of any ASD within first–degree family members, as confirmed through parent interview about family medical history [3,5]. Children were only given a formal diagnosis of ASD at 36 months.

Continuous EEG was sampled while infants were sitting on their parents' laps at a distance of 60cm from a 40×29cm monitor watching three video stimuli, each lasting for 30–40sec: (1) woman singing nursery rhymes or playing peek-a-boo ('soc' social video); (2) brightly coloured toys moving and producing sounds ('toy' nonsocial video); (3) the same sounding toys manipulated by a human hand ('hands' nonsocial video). Three triplets of video stimuli were presented in random order within the triplet (1–2–3, 2–3–1, etc.), but constant across the triplets for each participant. This resulted in nine 30–40sec EEG segments. Infants' behaviour during EEG session was recorded with a video camera.

EEG signals are inherently noisy because of EEG's sensitivity to both external (electrical lines, ambient noise) and internal influences (facial muscles, eyes, heartbeats). In an attempt to eliminate the inherent noise and capture the true distribution of EEG signals, the same EEG experiment can be repeated multiple times and the results can be averaged across trials. However, obtaining EEG data from infants presents a whole new range of challenges on top of those discussed above, such as problems with spontaneous crying, laughing, not engaging with the experiments, etc, and in general it is difficult to get infants to do several trials of the same experiment as they get tired and lose interest quickly. All these factors contribute to losing much of the time dependency of the data, since unwanted behaviour in infants takes place at different periods of time, and indicate that after data processing and behaviour analysis the clean signal available for modelling is quite scarce.

EEG was recorded using a 128-electrode HydroCel Geodesic Sensor Net (EGI, Eugene, OR) with respect to the vertex and sampled at 500Hz. Twelve ridge electrodes most often contaminated by artefacts were excluded from analysis resulting in a 116-electrode layout. Data preprocessing and analysis was performed using FieldTrip (`http://fieldtrip.fcdonders.nl`) as well as various Python libraries. The behavioural coding results were synchronised with EEG, and the periods when the baby was not looking at the screen, performed gross body, head, or arm movements, or cried, as well as the periods of interference, were excluded from analysis. The raw EEG was also visually inspected for artefacts.

One hundred and one infants from BASIS took part in the study (53 HR, and 48 LR) at the 12 to 15–month time–point. Data from 13 infants were excluded because of technical failures (6 HR and 7 LR). Two infants fell asleep during the

EEG session (1 HR, 1 LR) and one participant in the HR group missed diagnostic assessment at the 36–month visit. From the rest of the subjects sufficient artefact–free EEG data were obtained in 39 LR and 39 HR infants.

Each uninterrupted data period was segmented into 1–sec (500 samples each) segments with 50% overlap starting from the beginning of a clean EEG segment. The end part of the period shorter than one EEG epoch was not analysed. Fast Fourier transform (FFT) was computed for each 1–sec segment after removal of the mean (baseline correction) and application of the Hanning window.

3 Framework of Proposed Approach

Cortical activity related to memory operations is present primarily in three frequency bands: Theta (4-7Hz), Alpha (8-13Hz) and Beta (13-30Hz) [8]. These channels were used in this work, and a noise removal process was used since EEG signals from infants are particularly noisy, as mentioned above. However, removing noisy EEG windows may result in losing part of the time dependency between samples. Our approach considers windows of an EEG signal independent of each other and only uses the order of the samples to represent the temporal aspect of the EEG signal. This was considered reasonable since the behavioural noise happens at different times in the experiment for each infant. An overview of the EEG dataset is presented in Table 1 according to the type of audio/video stimulus detailed previously. Data shown, [8], are split according to the diagnosis participants received at three years of age: Low Risk-no ASD (LR), High Risk initially but no ASD diagnosis at age 3 (HR-TD), Atypical development at age 3 but still no ASD (ATY), and High Risk ASD diagnosed with ASD at age 3 (ASD). The FFTs of the 1–sec EEG windows were further transformed into RGB images as explained below.

	LR		HR-TD		ATY		ASD	
	#S	#W	#S	#W	#S	#W	#S	#W
soc	39	2755	16	1248	11	853	12	1137
hand	39	3061	16	813	11	867	12	1092
toy	39	1865	16	572	11	441	12	660

Table 1. Overview of the EEG data. #S denotes the number of subjects with that diagnosis at 36 months; #W is the number of non-noisy EEG windows, i.e. selected windows sampled at 500Hz from all recordings of each participant, excepting noise.

3.1 Transforming EEG signals into topology-preserving images

Traditionally the raw values of electrodes or their spectral measures are used to represent EEG feature vectors in EEG data analysis. However, this approach does not take into account the spatial properties of the data– each electrode being located above a region of the brain and having neighbouring regions.

Following in the footsteps of [2], we transformed the EEG signals into an image that preserves the spatial structure of the data, and we used multiple colour channels to represent the spectral dimension, as shown in Figure 1.

To ensure that the relative distance between electrodes (spatial aspect) is preserved within each EEG window, we used the Azimuthal Equidistant Projection (AEP) also known as Polar Projection, borrowed from mapping applications [14]. At either pole, the Equator, or some intermediate point, the AEP is shaped into a plane that is typically tangential to the globe at that point. Distances between the projected middle and every other point are retained in the AEP. Similarly, the shape of the cap worn on a human's head can be approximated by a sphere (in our case), and the same method may be used to compute the projection of electrodes positions on a 2D surface that is tangential to the top point of the head using the same method. The relative distances of all pairs of electrodes would not be accurately maintained since the distances between the points on the map are only preserved with respect to a single point (the middle point). We obtained 2D predicted electrode positions using AEP applied to 3D electrode locations as in [2]. The spatial distribution of behaviours over the cortex was represented by the image's width and height. For interpolating the distributed power measurements over the scalp and estimating the values in–between the electrodes over a 224×224 mesh, we used the Clough–Tocher scheme [1]. This process was repeated for each of the three frequency bands of interest, yielding three topographical activity maps for each. After that, the three spatial maps were combined to produce a three–channel view. A graphical representation of the entire pipeline is shown in Figure 1. As described in the following section, this three–channel EEG image was fed into a deep network architecture.

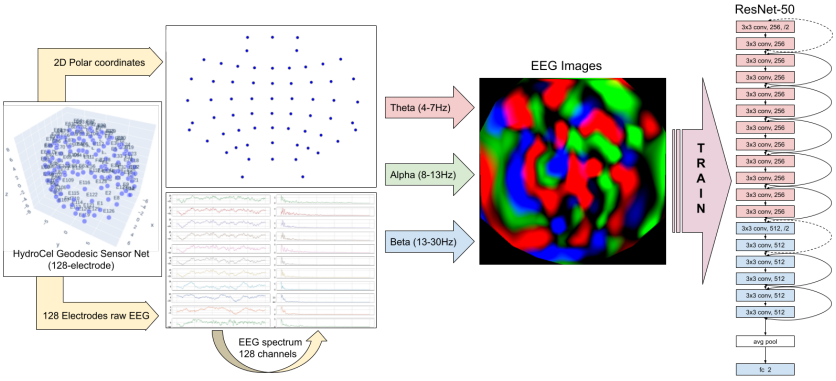

Fig. 1. EEG processing and training pipeline.

For simplicity and speed we have merged all non-ASD classes (HR-TD, ATY and TD) into one class named NON-ASD. By creating the NON-ASD class and trying to separate it from the ASD, the typical ASD detection problem based on

EEG signal windows is transformed into a binary image classification problem. The new class distribution, for the three stimuli is 81% NON-ASD and 19% ASD, as shown in Table 2.

	NON-ASD		ASD	
	#S	#I	#S	#I
soc	66	4856	12	1137
hand	66	4741	12	1092
toy	66	2878	12	660

Table 2. Overview of the EEG images based on the number of non-noisy epochs kept from each subject. #S denotes the number of subjects with that diagnosis at 36 months and #I is the number of total images (transformed from non-noisy EEG windows).

3.2 Deep learning considerations

Convolutional Neural Networks (CNN) can detect and extract local informative patterns and excel when we have a grid or volume–like input and the features in this input space are locally correlated. They were considered a good fit for the binary classification problem considered in this paper. However, they suffer, as do other deep neural network models, from the vanishing gradient problem caused by the gradient of the error w.r.t. weights travelling through so many layers of representation and gradually becoming too small for learning.

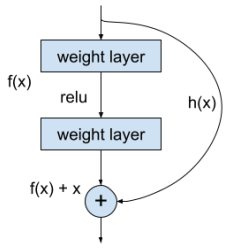

Fig. 2. Shortcut connection in a ResNet, where $h(x) = x$.

Residual Neural Networks–ResNets [6] alleviate the vanishing gradient situation. The ResNet architecture is similar to traditional multi–layer CNN but it includes the so-called shortcut or skip connections. ResNet is based on the assumption that by attempting to estimate residual functions that were not adequately captured by blocks of CNN layers, one can approximate the desired mapping (see Figure 2).

To this end, the ResNet architecture is using an identity mapping as a layer:

$$y = f(x, W_i) + x, \tag{1}$$

where the input layer is x, the output layer is y, the weights are W, and the residual mapping to be learned is $f(x, W)$. The fundamental principle is that by stacking non-linear layers, identity mapping will correct the degradation of the gradient. One might imagine an extreme case where, if identity mapping is the ideal mapping, y, this model would better capture reality by simply setting $f(x, W)$ to 0, while directly learning such a mapping without the use of the identity mapping would be more complicated [6]. The shortcut/skip connections approach is used to train the model based on this identity mapping, which facilitates the propagation of signals in both forward and backward passes [7].

When ResNets were first introduced [6], three distinct models were released to demonstrate that the effect of vanishing gradients can be alleviated at various levels of depth, and that increasing the model depth is not always necessary. In this instance, preliminary experiments with a 50–layer ResNet (ResNet–50), and 101–layer and 152–layer ResNets showed that all three architectures, when tested perform similarly, thus we opted for the ResNet–50 architecture (over 23 million weights), since it has only 3.8 billion FLOPs, as opposed to 11.3 billion FLOPs for the 152–layer counterpart.

In order to speed up training, transfer learning [16] was adopted. The types of convolutional kernels that the CNNs learn when trained on images, especially in the layers closest to the input, are very similar to edge detectors, which have been traditionally hand crafted by computer vision experts and used in machine vision systems. The benefits of transfer learning were also confirmed in our preliminary experiments, where by leveraging existing pre–trained models, on well established datasets, we were able to dramatically reduce training time.

To make full use of the ResNet–50 architecture, which was trained on ImageNet, the EEG images were created with the same width and height as the ImageNet images, which this ResNet model (from PyTorch model zoo) had been trained on. The PyTorch [10] machine learning library was used to implement and train the neural architectures mostly because of its ease of use and rich library of pre–trained models developed by different researchers.

Hence, each EEG image has a size of $224 \times 224 \times 3$, and the original 1000 classes ResNet–50 output layer changed to a layer with two outputs. The categorical–cross entropy was used as loss function, as in the ImageNet problem. Lastly, in the experiments, presented in the following section, the class imbalance issue was addressed by the Weighted Random Sampler (WRS) provided by the PyTorch library [10], which gives specific weights to every EEG image (belonging to a class) in the dataset thereby producing a balanced mini-batch when sampling. Moreover, the WRS with replacement was used, i.e. sampling the same data points multiple times to create a balanced mini-batch, and thus, a balanced training set of EEG images overall.

Fig. 3. Example of training and testing accuracy on the balanced dataset with 80/20 split. The markers indicate the early stopping (ES) locations of the three ResNets (soc, hand, toy).

4 Experiments and Results

As mentioned above, a standard ResNet–50 architecture, pre-trained on ImageNet, showed in preliminary experiments higher performance than other architectures tested, whilst it required a lot less training (mainly fine-tuning) compared to a ResNet–50 trained from scratch. The standard Adam optimizer provided by PyTorch [10], with a learning rate of $1e^{-3}$ and mini-batch size of 100, was used for training, and a separate ResNet–50 was trained for each video stimulus (soc, hand, toy).

In this paper, two types of experiments are reported: (i) split each participant's samples into 80% for training and 20% for testing; (ii) leave–one–participant out (i.e. all one–second EEG windows which belong to that participant and have been transformed into images are left out), train on data from the rest of the participants, and evaluate performance on all windows of the participant that was left out, and repeat the procedure for each participant.

All experiments were run with and without Early Stopping (ES), to alleviate over-fitting, which was most evident in the leave–one–out experiments. Early stopping requires a validation set, which includes images that are not used for training. A small balanced validation set is built for each ResNet–50 trained on a video stimulus by removing 12 ASD images and 12 NON-ASD images from the training set of that stimulus (see Table 2). Training starts on the remaining EEG images using the validation set's predictions as the criterion for early stopping. This criterion is monitored during the training process. As training progresses, fluctuations in the prediction accuracy of the validation set occur. When the

	Training (%)			Testing (%)		
ID	NO-ES	ES	**Change**	NO-ES	ES	**Change**
soc	100.0	100.0	**0.0**	80.0	82.0	**+2.0**
hand	99.0	100.0	**-1.0**	82.0	81.0	**-1.0**
toy	100.0	100.0	**0.0**	78.0	79.0	**+1.0**

Table 3. Average accuracy per video stimulus (calculated over all participants) for ImageNet pre–trained ResNets–50, trained with and without early stopping (ES/NO-ES) on each video stimulus, and tested using the 80/20 split.

networks begins to over–fit the data, the validation accuracy starts degrading. If that happens for a specified consecutive number of epochs, the training is stopped, and the parameters of the model with the best validation accuracy are returned and used on the test set.

The results for the 80/20 split can be seen in Table 3, with and without ES. The table shows that ES produces only marginally improved performance when training on EEG images of the toy and social datasets. As shown in Figure 3, the approach is resilient to over-fitting when a 80/20 split is used.

However, over–fitting becomes significant when leave–one–participant out is used, especially when participants of the minority class should be classified (i.e. predict ASD at 3 years of age). Table 4 presents average precision, recall, F1–score and accuracy across all participants with ES and without ES (NO–ES), revealing the beneficial role of ES in the more realistic leave–one-participant scenario. Looking closer the performance of ResNets when tested with ASD participants, results in Table 5 and Table 6 reveal how the networks massively miss-classify ASD as NON-ASD when training is done without early stopping (NO-ES). Introducing early stopping significantly improves performance, particularly on the minority class, as shown in the same tables. It is worth noticing that since all the EEG data of one participant are left out each time and this is a binary problem, a test score over 50% can be considered a good prediction since the model classifies more than half the EEG images of that participant to the ASD class.

Figure 4 illustrates training and testing accuracies per epoch for four random ASD participants. It shows that the network learns the training set quite rapidly but the predictions on the test images of the participant that was left out fluctuate a lot, resulting in a rapid over–fit as the network focuses on predicting the majority class. The use of the simple early stopping technique adopted appears to help, and in Figure 4 one can see where the adopted early stopping technique terminates the training process (indicated by the four ES markers in the figure) within the specified maximum of 100 epochs.

5 Conclusions

The proposed framework considered EEG–as–an–image by transforming power spectra of EEG signal windows into RGB images, and formulated the ASD

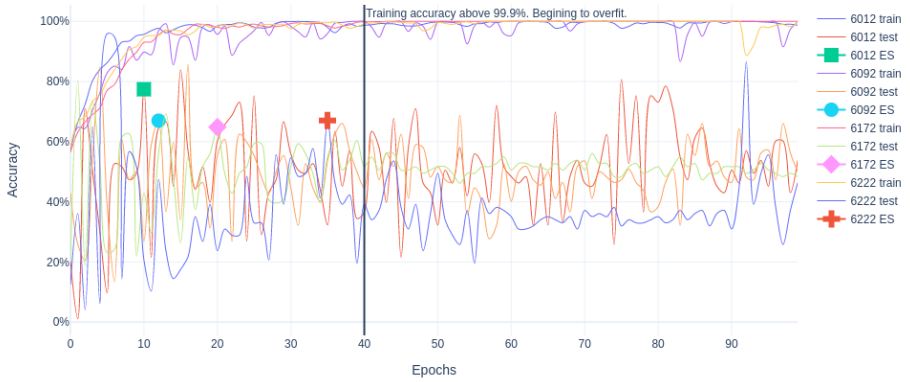

Fig. 4. Training and Testing accuracy (for the soc video stimulus) with leave–one–participant out cross-validation for four ASD participants. The markers indicate the early stopping (ES) locations of the soc stimulus ResNet.

NO-ES	Precision	Recall	F1-score
NON-ASD	0.87	1.00	0.93
ASD	1.00	0.17	0.29
Accuracy	0.87	0.87	0.87
Macro avg	0.93	0.58	0.61
Weighted avg	0.89	0.87	0.83

ES	Precision	Recall	F1-score
NON-ASD	1.0	1.0	1.0
ASD	1.0	1.0	1.0
Accuracy	1.0	1.0	1.0
Macro avg	1.0	1.0	1.0
Weighted avg	1.0	1.0	1.0

Table 4. Average precision, recall and F1 scores (calculated over all participants and video stimuli) for ImageNet pre–trained ResNets–50, trained with and without early stopping (ES/NO-ES), and tested using leave–one–participant out.

prediction problem in infants as binary classification. It employed transfer learning using ResNet–50, pre–trained on industry standard ImageNet data. Issues of class imbalance and over–fitting were alleviated with some simple techniques. Performance was evaluated using hold–out and leave–one–out methods and promising results were produced. This implies that there is latent structure in the EEG images, and therefore presumably the underlying brain processing, that can serve as a biomarker of ASD in infancy. Future work will attempt to treat class imbalance and improve generalisation using more advanced schemes. Also, the multi-class problem will be an objective of future work, together with an expansion of the sample of the ASD participants to reduce class imbalance.

Acknowledgement: The investigators wish to sincerely thank all BASIS families for their enormous contribution and commitment to the project. BASIS is supported by the BASIS funding consortium led by Autistica (`www.autistica.`

	Training (%)			Testing (%)		
ID	NO-ES	ES	**Change**	NO-ES	ES	**Change**
hand	100.00	92.04	**-7.96**	49.48	72.86	**+23.37**
soc	99.59	86.85	**-12.74**	40.54	73.76	**+33.22**
toy	99.80	87.24	**-12.55**	36.59	71.32	**+34.74**

Table 5. Average accuracy per video stimulus (calculated over all ASD participants) for ImageNet pre–trained ResNets–50, trained with and without early stopping (ES/NO-ES) on each video stimulus, and tested using leave–one–participant out.

org.uk), Autism Speaks (grant no: 1292/MJ) and by the UK Medical Research Council (grant no: G0701484). Support of the NVIDIA Corporation that donated the Tesla K40 GPUs used for this research is also gratefully acknowledged.

	Training (%)			Testing (%)		
ID	NO-ES	ES	**Change**	NO-ES	ES	**Change**
6012	99.57	75.85	**-23.72**	48.92	76.01	**+27.08**
6092	99.76	88.24	**-11.51**	41.11	72.18	**+31.07**
6122	100.00	90.88	**-9.12**	45.20	74.25	**+29.05**
6152	100.00	87.73	**-12.27**	39.05	73.74	**+34.69**
6162	100.00	88.74	**-11.26**	14.15	57.39	**+43.24**
6172	99.72	95.37	**-4.35**	48.52	73.38	**+24.86**
6222	99.89	94.53	**-5.36**	43.06	74.13	**+31.07**
6232	100.00	89.93	**-10.07**	38.41	60.29	**+21.88**
6262	99.72	90.54	**-9.18**	47.55	81.62	**+34.07**
6272	100.00	98.76	**-1.24**	62.38	74.52	**+12.14**
6292	99.00	92.51	**-6.49**	59.49	88.21	**+28.72**
6372	99.96	75.28	**-24.68**	27.22	67.11	**+39.90**

Table 6. Average accuracy of ResNets in correctly classifying images of each participant with ASD (across all three stimuli) trained with and without early stopping (ES/NO-ES), and testing by leaving all data of one participant out, as indicated by the ID.

References

1. Alfeld, P.: A trivariate clough—tocher scheme for tetrahedral data. Computer Aided Geometric Design **1**(2), 169–181 (1984)
2. Bashivan, P., Rish, I., Yeasin, M., Codella, N.: Learning representations from EEG with deep recurrent-convolutional neural networks. In: Bengio, Y., LeCun, Y. (eds.) 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings (2016), http://arxiv.org/abs/1511.06448
3. Elsabbagh, M., Divan, G., Koh, Y., Kim, Y., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C., Wang, C., Yasamy, M., Fombonne, E.: Global preva-

lence of autism and other pervasive developmental disorders. Autism Research **5**(3), 160–179 (Jun 2012). https://doi.org/10.1002/aur.239
4. Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., Pickles, A., Baron-Cohen, S., Bolton, P., Johnson, M.H.: Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. Current Biology **22**(4), 338–342 (2012). https://doi.org/10.1016/j.cub.2011.12.056, http://dx.doi.org/10.1016/j.cub.2011.12.056
5. Gliga, T., Bedford, R., Charman, T., Johnson, M.H.: Enhanced visual search in infancy predicts emerging autism symptoms. Current Biology **25**(13), 1727–1730 (2015). https://doi.org/10.1016/j.cub.2015.05.011, http://dx.doi.org/10.1016/j.cub.2015.05.011
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
7. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision. pp. 630–645. Springer (2016)
8. Orekhova, E.V., Elsabbagh, M., Jones, E.J., Dawson, G., Charman, T., Johnson, M.H.: EEG hyper-connectivity in high-risk infants is associated with later autism. Journal of Neurodevelopmental Disorders **6**(1), 40 (2014). https://doi.org/10.1186/1866-1955-6-40, http://dx.doi.org/10.1186/1866-1955-6-40
9. Ozonoff, S., Iosif, A.M., Baguio, F., Cook, I.C., Hill, M.M., Hutman, T., Rogers, S.J., Rozga, A., Sangha, S., Sigman, M., et al.: A prospective study of the emergence of early behavioral signs of autism. Journal of the American Academy of Child & Adolescent Psychiatry **49**(3), 256–266 (2010)
10. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
11. Pettersson, E., Anckarsäter, H., Gillberg, C., Lichtenstein, P.: Different neurodevelopmental symptoms have a common genetic etiology. Journal of Child Psychology and Psychiatry **54**(12), 1356–1365 (2013)
12. Reiersen, A.M.: Early identification of autism spectrum disorder: is diagnosis by age 3 a reasonable goal? Journal of the American Academy of Child and Adolescent Psychiatry **56**(4), 284–285 (2017)
13. Sheldrick, R.C., Garfinkel, D.: Is a positive developmental-behavioral screening score sufficient to justify referral? a review of evidence and theory. Academic pediatrics **17**(5), 464–470 (2017)
14. Snyder, J.P.: Map projections–A working manual, vol. 1395. US Government Printing Office (1987)
15. Supekar, K., Uddin, L.Q., Khouzam, A., Phillips, J., Gaillard, W.D., Kenworthy, L.E., Yerys, B.E., Vaidya, C.J., Menon, V.: Brain hyperconnectivity in children with autism and its links to social deficits. Cell reports **5**(3), 738–747 (2013)
16. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q.: A comprehensive survey on transfer learning. Proceedings of the IEEE **109**(1), 43–76 (2020)