Automated Artifact Detection for EEG Data Using a Convolutional Neural Network

Developmental Electrophysiology Laboratory
Yale Child Study Center

T. McAllister, A. Naples, J. McPartland

McPartland Lab, Yale Child Study Center, New Haven, CT

Yale Child Study Center
SINCE 1911

Background:

- Electroencephalography (EEG) is a temporally precise and inexpensive tool for studying brain activity in autism spectrum disorder (ASD)
- However, a weakness of EEG is contamination by movement and muscle activity
- It is necessary to identify and exclude artifactual muscle activity, but there is little consensus on methodology for its automated detection
- EEG is still checked for artifacts by hand, which is time intensive and error-prone
- In other domains, such as facial recognition and email spam detection, convolutional neural networks (CNNs) have been effective in automating classification tasks
- CNNs are complicated, layered, nonlinear, mathematical models with parameters that are roughly optimized to classify inputs
- CNNs show promise for automating artifactual data identification

Objectives:

1. Develop a CNN to classify contaminated EEG collected from infants at normal risk (NR) and high risk (HR) for ASD
2. Assess its performance against human experts
3. Assess its performance in classification between NR and HR infants to explore potential differences in artifact across groups

Methods:

- Data were collected in 118 EEG sessions with NR and HR infants
- Sessions involved 100 trials of looking at point light displays of biological and scrambled motion and were recorded at 500 Hz using EGI Net Amps 300 and 128-channel Hydrocel Geodesic Sensor Nets
- Data were split into event-related epochs and manually classified by a human expert as artifactual or normal
- Epochs were converted into simple 2D arrays of amplitude by time across 129 EEG channels, yielding 5834 artifactual and 5388 clean examples
- Epochs were clamped to the range [-100, 100] µV, then rescaled to [-1, 1]
- 3 datasets were combined, with a randomized ordering of epochs and differences in data augmentation
- Data augmentation creates more training data using existing epochs by slightly modifying them in ways that do not impact their classification
- Data augmentation options were: none, zero padding of 2-13ms before or after the epoch, and adding a .2µV noise signal
- Datasets with augmentation had double the number of epochs (N=22,044) as the dataset with no augmentation (N=11,022)
- Each dataset was split into a validation set (N=1000), a test set (N=1400), and a training set
- CNNs were built according to the design in *Figure 1* using the python library Keras, and the Tensorflow backend
- For results, the network was trained from a randomized initialization until it had seen all training data 12 times
- The CNN was trained and tested using a GeForce GTX 1080 Ti graphics card

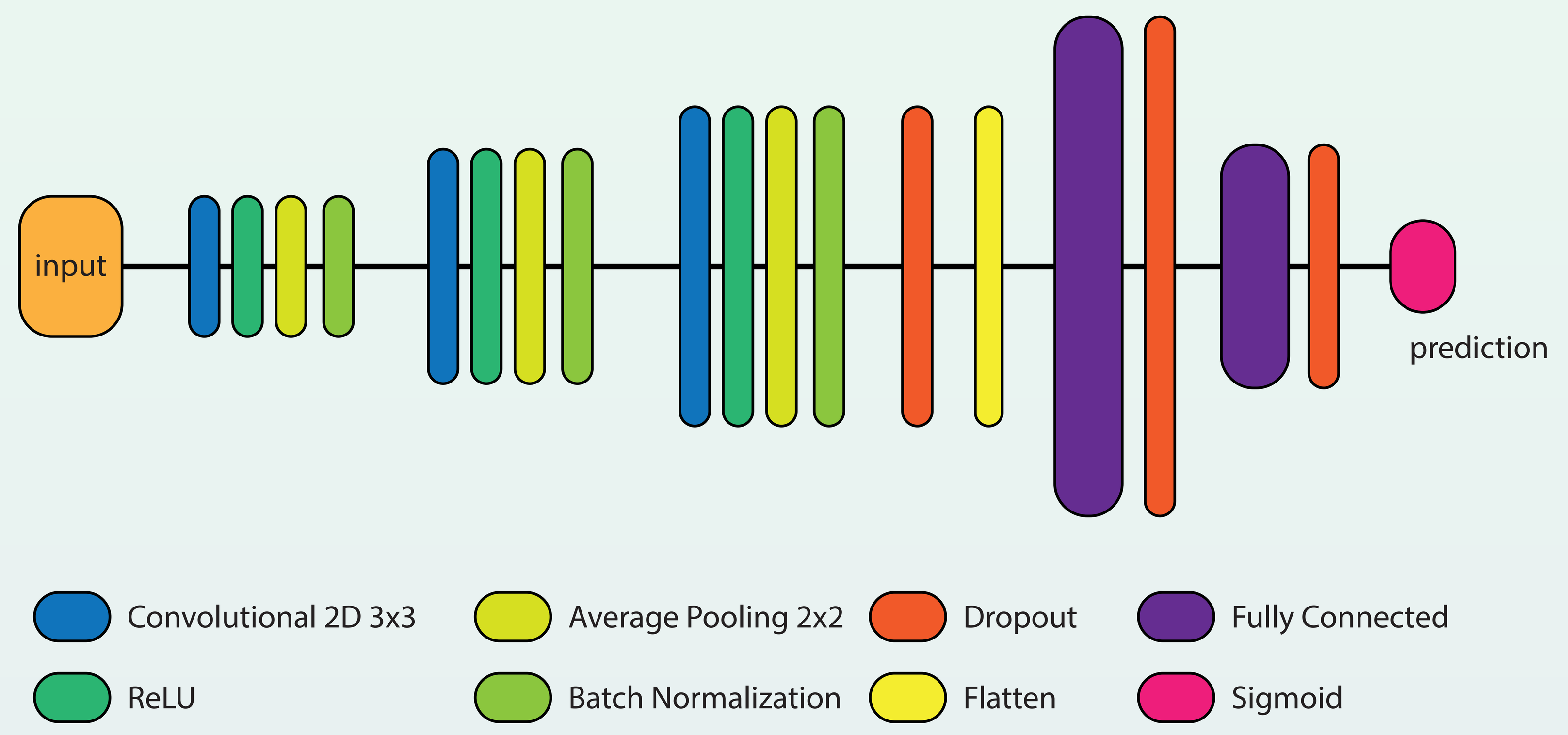

Figure 1 : Outline of Convolutional Neural Net architecture.
ReLU and sigmoid layers (color coded) introduce nonlinearity [1], batch normalization layers normalize the output of preceding layers to assist optimization [2], dropout layers reduce overfitting [3], flatten layers change the dimensions of output [4], convolutional 2D layers perform a mathematical convolution with a 3x3 kernel [4], and fully connected layers are a linear combination of the outputs of the preceding layer [4]

1. Glorot, X., Bordes, A., & Bengio, Y. (2011, June). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (pp. 315-323).
2. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
3. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929-1958.
4. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.

Base	Accuracy	False Negative	False Positive
Overall	79.86	5.86	14.29
High Risk	77.50	7.00	15.50
Normal Risk	81.62	5.00	13.38

Noise	Accuracy	False Negative	False Positive
Overall	79.93	9.36	10.71
High Risk	81.88	10.62	7.50
Normal Risk	77.33	7.67	15.00

Padding	Accuracy	False Negative	False Positive
Overall	85.07	6.93	8.00
High Risk	84.25	1.50	14.25
Normal Risk	85.40	9.10	5.50

Table 1 : Comparison of the performance of the network when trained with the base dataset (no augmentation), noise augmented dataset, and the padding augmented dataset. Values reflect percentage of the 1400 epochs in the test set. Metrics looking only at subsets of the test data by group are also shown.

Results:

- After training, the predictions of the CNN for the novel test set of epochs (N=1400) were compared with the labels created by a human-expert
- Modification of architectures (*Figure 1*) was able to increase accuracy from 75% (unoptimized models) to over 80% (final model: *Figure 1*)
- CNN changes included:
  - Adding preprocessing in the form of clipping epochs to a standard range of ±100µV, then normalizing them to a range of ±1
  - Increasing the network architecture from 14 layers to the current design with 24 layers
- Clean epochs incorrectly labeled as artifactual were considered false positive epochs, while overlooked artifacts were considered false negative epochs
- The performance of the network trained with the various datasets is summarized in *Table 1*
  - Augmentation with noise did not produce substantial improvements, but showed higher accuracy for the HR group
  - Augmentation by zero padding had the best results, achieving over 85% accuracy overall
  - Evaluating the test set of 1400 epochs as good or artifactual took roughly 1.5 seconds, excluding the time to load the data into memory
  - Editing these data by hand take approximately 182 hours

Conclusions:

- Our results show that CNNs can classify EEG artifact at rates approaching human expert performance
- Our CNN was able to classify epochs in a fraction of the time that a human evaluator would require
- The development process was able to improve performance significantly, and more complex architectures may bring further benefits
- Ongoing analyses of classification performance between groups could allow us to detect differential patterns of artifact by risk status
- Automated detection paired with a human reviewer via a visual check system could dramatically increase the accuracy and speed of the human while preserving input from human judgement

Acknowledgements:

NIH U19 MH108206 (McPartland), CTSA Grant Number UL1 RR024139 (McPartland), NIMH K23 MH086785 (McPartland)

McPartland Lab: https://mcp-lab.org • mcp.lab@yale.edu