

Enhanced Error Decoding from Error-Related Potentials using Convolutional Neural Networks

Juan M. Mayor Torres¹, Tessa Clarkson², Evgeny A. Stepanov³, Christian C. Luhmann²,
Matthew D. Lerner², and Giuseppe Riccardi³

Abstract—Error-related potentials are considered an important neuro-correlate for monitoring human intentionality in decision-making, human-human, or human-machine interaction scenarios. Multiple methods have been proposed in order to improve the recognition of human intentions. Moreover, current brain-computer interfaces are limited in the identification of human errors by manual tuning of parameters (e.g. feature/channel selection), thus selecting fronto-central channels as discriminative features within-subject. In this paper, we propose the inclusion of error-related potential activity as a generalized two-dimensional feature set and a Convolutional Neural Network for classification of EEG-based human error detection. We evaluate this pipeline using the BNCI2020 – Monitoring Error-Related Potential dataset obtaining a maximum error detection accuracy of 79.8% in a within-session 10-fold cross-validation modality, and outperforming current state of the art.

I. INTRODUCTION

Error-related potentials are electrophysiological event-related potentials (ERPs) that occur after an individual realizes an error has been made. They are generated in the anterior cingulate cortex (ACC) [1] and are important electrophysiological biomarkers for studying human intentionality [2], [3]. The *error-related potential* (ErrP) is a group of EEG stimulus-locked waveforms associated with the elicitation of self/others erroneous trials. ErrPs occur around 50-100ms after the commission or realization of an error has been made, though they vary slightly in latency and amplitude.

One particular example of ErrP is the ERN [4] measured during fast Go-Nogo and Flanker tasks and is a neural response associated with the commission of an error by the individual. Alternatively, the FRN is measured during interactive tasks such as reward-seeking [3] and decision-making [5]. In general an ErrP is elicited after self/other erroneous trial has been perceived by the participant [1], [6].

However, despite the substantial amount of studies associating ErrP with erroneous trial processing, Brain Computer Interfaces (BCI) have not included generalized enough pipelines that can outperform the current ErrP-based error decoding pipelines [7], [4], and overcome the incidence of ErrP inter-trial and inter-subject variability [8]. Most pipelines perform an off-line pre-selection of the most discriminative channel or a set of channels to obtain better error decoding accuracy. This process implies a feature imbalance within-subject and poor generalization between subjects. Thus, ErrP-based classification of erroneous trials is a high demanded problem in neuroscience [1], [6], requiring better processing of ErrP single trials.

In this paper, we propose a novel pipeline for error/correct trials decoding using ErrP activity as a 2D feature set representation for training a deep Convolutional Neural Network (ConvNet). This pipeline does not require the selection of different set of fronto-central channels per subject but successfully outperforms current state-of-the-art methods for error decoding. We evaluate this new pipeline on the *Monitoring Error-Related Potential* dataset and using the same feature representation per trial in a within-session 10-fold cross-validation per subject modality.

II. METHODOLOGY

Figure 1 shows the complete error decoding pipeline with its corresponding subtasks: (1) Artifact Removal, (2) Whitening, (3) Cropping, and (4) A 2 Convolution-Pool Stages ConvNet. In the next subsections, we will describe the pipeline sub-processes in detail.

A. BNCI2020: Monitoring Error-Related Potential dataset

This pipeline was evaluated using the Monitoring Error-Related Potential dataset described in [1] as a part of the BNCI2020¹ initiative. This dataset is composed of a group of EEG stimulus-locked recordings elicited by a moving cursor (green square) and a randomly positioned target screen (red square). Six healthy participants attempted to guess the position of a target controlled by an artificial agent. A trial was considered erroneous if the agent’s cursor was not in the same position as the random target. Trials on which the agent’s cursor reached the target position were considered correct. After correct trials the target position randomly changes positions at least 3 slots away from the last moving cursor position.

¹J. M. Mayor Torres is with University of Trento, Italy, Signals and Interactive Systems Lab, 38123, and Stony Brook University, NY, Social Competence and Treatment Lab, NY,11733,USA (juan.mayortorres@unitn.it, juan.mayortorres@stonybrook.edu)

²Tessa Clarkson (tessa.clarkson@stonybrook.edu), Christian C. Luhmann (christian.luhmann@stonybrook.edu) and Matthew D. Lerner (matthew.lerner@stonybrook.edu) are with Stony Brook University, Department of Psychology, Social Competence and Treatment Lab (SCTL), NY, 11733, USA.

³Evgeny A. Stepanov (evgeny.stepanov@unitn.it) and Giuseppe Riccardi (giuseppe.riccardi@unitn.it) are with University of Trento, Italy, Department of Information Engineering and Computer Science (DISI), Signals and Interactive Systems Lab (SISLab), 38123

¹<http://bnci-horizon2020.eu/database/data-sets>

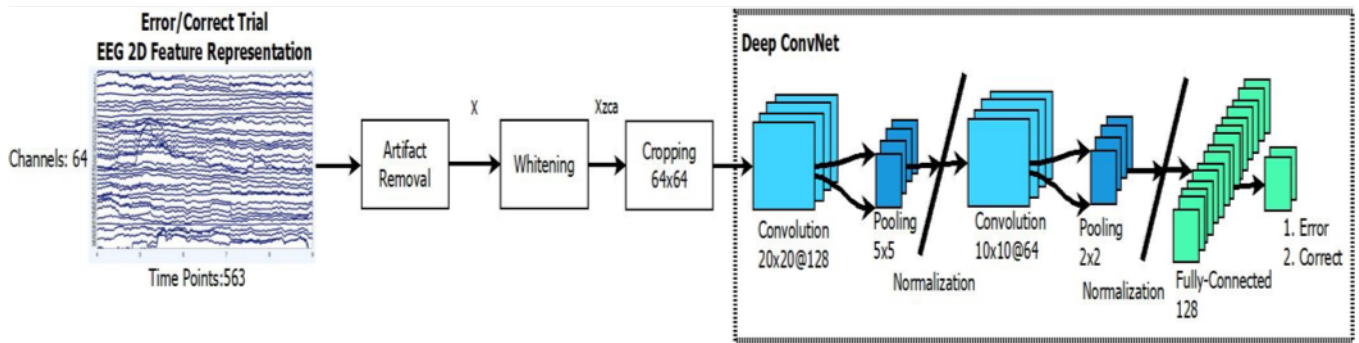


Fig. 1. Error decoding pipeline using ErrP neural activity and a ConvNet classifier. This pipeline is composed of (1) The construction of the channels \times time-points 2D ErrP feature representation, (2) the preprocessing pipeline composed by sub-tasks such as Artifact Removal, ZCA Whitening, and Cropping, and (3) The ConvNet classifier with 2 Convolutional - maximum Pooling stages. A final *argmax* layer is used for computing the Error and Correct class posteriors.

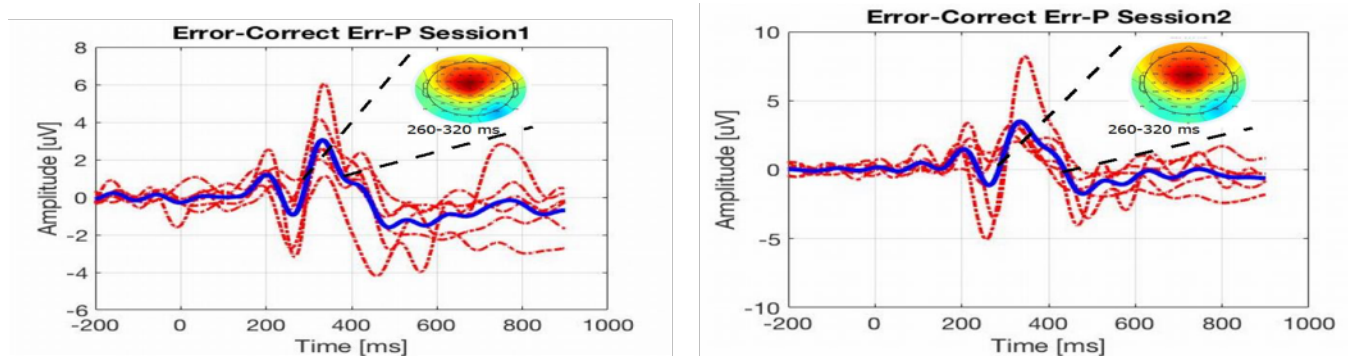


Fig. 2. Figure 2.A and 2.B show the complete set of grand-average, and subjects average for session 1 and session 2 data respectively. ErrP FCz grand-average average neural activity are shown in blue and the average for each subject in red. The topo-plot added on the top shows the average amplitude between 260 and 320 ms.

This procedure was repeated in ten 3-min blocks, yielding 64 2,000ms trials each. Each participant completed the task twice, weeks apart ($M = 38.97$ weeks, $SD = 41.09$) to examine different error probabilities. For each session the agent error probability P_{err} was programmed in two modalities $P_{err} = 0.2$ and $P_{err} = 0.4$ for balancing the number of erroneous occurrences per subject, but only 4 out of six subjects took session 2 for $P_{err} = 0.4$. In our experiments we used data from $P_{err} = 0.2$ given the unbalanced amount errors and correct trials included in this subset. Grouping the data for all subjects the first session yielded 665 errors and 2,442 correct trials, and the second session yielded 657 errors and 2,673 correct trials. EEG data was recorded at 512 Hz using a 64 channels Biosemi ActiveTwo device.

Figures 2.A and 2.B show the Error-Correct FCz average waveform for session 1 and 2 respectively. Averaged topoplots are also added in Figures 2.A and 2.B showing the corresponding Error-Correct average neural activity between 260 and 320 ms.

A significant difference between error and correct is found subtracting between Error/Correct individual average waveforms. This difference is evident at fronto-central region, especially FCz ($F(1,11) = 7.78, p < 0.001$) and Cz ($F(1,11) = 6.55, p < 0.001$). However, this difference

is not observed using a RM-ANOVA between sessions – $P_{err} = 0.2$ ($F(1,23) = 0.122, p = 0.353$). With a clear confirmation of a statistical difference between error and correct neural activity we design this novel ConvNet based pipeline for EEG single-trial error decoding using ErrP features, in order to discriminate systematically this Error-Correct difference.

B. Pre-Processing

The ConvNet pipeline and all the baseline classifiers (Section II-C) received an epoched EEG training subset after the Artifact Removal process. Whitening and Cropping are only used as a pre-emphasis for the ConvNet classifier.

1) *Artifact Removal*: This process is composed of the execution of two EEGlab plugins in a sequence. First, we use the Koethe’s *clean_raw* function from PREP pipeline [9] using Artifact Subspace Removal (ASR) for bad channel removal. After bad channels are removed we used interpolation for infer the new neural activity from the 4 nearest channels. Second, we use ADJUST [10] for removing eye blinking and movement artifacts using spatio-temporal high order statistics. ADJUST uses Independent-Components Analysis (ICA) to evaluate all the possible artifactual independent-components (IC) that constitute each *ErrP* trial. The non-artifactual IC selection is implemented based on high-order spatio-temporal features.

2) *Whitening*: The cleansed feature set is normalized using a ZCA whitening process. Following [11], the ZCA processes each Error/Correct trial per batch, implementing (1) a subtraction between the EEG trial representation and its corresponding average and (2) a Principal Component Analysis (PCA) Singular Value Decomposition SVD task described in Equation 1.

$$X_{ZCA} = \lambda Tr \left(\frac{1}{\sqrt{Tr(S) + \varepsilon}} \right) \Sigma X \lambda^T \quad (1)$$

ErrP trial X is normalized and transformed into the ZCA 2D feature representation denoted as X_{ZCA} using the covariance matrix, the rotation matrix λ extracted from the initial SVD, and the eigenvector matrix S .

3) *Cropping*: In this step, for each ConvNet training iteration and for each fold, we randomly crop a 64x64 size feature set varying the center position across the training iterations. Previous implementations [2], [11] assert that this process reduces the probability of identifying a false training local minima.

C. Baseline Classifiers: Gaussian, SVM, DBN, Majority Voting

We propose a group of baseline classifiers to assess our ConvNet-based pipeline. The classifiers we used include Gaussian [1], a linear Support Vector Machine (SVM) [4], and a 2-layer Deep Belief Network (DBN). These classifiers represent generative alternatives to establish a fair comparison with robust discriminative classifiers introduced here, such as ConvNets.

We use the same parameters described in [1] for the Gaussian classifier such as learning rates of 0.001 and 0.0001 for computing the covariance per class. From [4] we set the linear kernel SVM with a regularization parameter $C = 1$. We train a 2 hidden-layer DBN composed by 20 units in the first hidden layer and 100 in the second. All hidden units' activation functions are *sigmoid* except for *softmax* output layer's units. The training method is composed of a Contrastive Divergence (CD) with 100 iterations pre-training and complemented with a subsequent 2,500 iterations fine-tuning training [12], [13]. Pre-training learning rate is set in 0.01, and fine-tuning learning rate is set in 0.15 with a weight decay factor of 0.0001 per iteration.

The majority voting strategy is simple rule of *argmax* majority voting learner dividing the decision probabilities evenly across Gaussian, SVM and DBN classifiers.

D. Deep 2-conv/pool stages ConvNet: FCz and Cz channels, and 64 channels

The ConvNet classifier illustrated in detail on Figure 1 is initially set with random distributed weights for all the Conv-Pooling stages. The first ConvNet stage is composed of one Convolutional layer with 128 filters 64x64 each, and a subsequent 5x5 max pooling layer. The corresponding features obtained after pooling are sub-sampled with a factor of 2. In the second stage, a similar scheme was set using a reduced number of convolutional filters (64), each with a

TABLE I
AVERAGE ACCURACY FOR ERROR AND CORRECT TRIALS
CLASSIFICATION. PER CLASSIFIER ACCURACIES ARE REPORTED FOR
SESSIONS 1 AND 2 RESPECTIVELY.

Classifier	Error	Correct
Session 1: $P_{err} = 0.2$		
<i>Gaussian</i>	0.619	0.771
<i>SVM</i>	0.621	0.781
<i>DBN:20-100</i>	0.631	0.786
<i>Majority Voting</i>	0.627	0.792
<i>ConvNet: FCz, Cz</i>	0.721	0.813
<i>ConvNet: 64 ch</i>	0.798	0.844
Session 2: $P_{err} = 0.2$		
<i>Gaussian</i>	0.616	0.762
<i>SVM</i>	0.583	0.779
<i>DBN:20-100</i>	0.616	0.766
<i>Majority Voting</i>	0.625	0.784
<i>ConvNet: FCz, Cz</i>	0.678	0.804
<i>ConvNet: 64 ch</i>	0.751	0.838

10x10 kernel size. The subsequent maximum pooling layer is 2x2 with another striding factor of 2.

Following [2], we use exponential linear units (ELU) as activation functions in all the stages except for the fully-connected layer in which we use a step function. The 64x64 cropped feature sets are assigned randomly into a mini-batch of size 120, thus changing the cropping location across the training iterations. We use a learning rate of 0.0001 and train each subject for both $P_{err} = 0.2$ sessions with 1,200 iterations using a weight decay factor of 0.002.

The ConvNet uses features from (1) a 2x563 image wrapping up the ErrP neural activity from FCz and Cz channels, and (2) 64x563 representation grouping the features for all the 64 channels per trial. For the two channel case, we change the height of the convolutional filters from 20 to 2 in the first stage, and from 10 to 1 in the second stage. Therefore, the pooling layers are 1x5 and 1x1 respectively; the sub-sampling is applied only for the first stage. A final two-unit *argmax* layer is used to compute the Error and Correct classes' posterior probabilities.

III. RESULTS

For evaluating the ConvNet pipeline we use two different sets of features: (1) A 2x563 image wrapping up the features from FCz and Cz channels, and (2) 64x563 representation grouping the features for all the 64 channels per trial. Results of the classification experiments are presented in Table I and Figures 3.A and 3.B. From Table I we can observe that ConvNet models outperform the other classification algorithms – Gaussian, SVM, DBN, and their majority voting ensemble. Among the two ConvNet models, the 64-channel model yields superior performances for both correct and error classes across the six subjects. Figures 3.A and 3.B, on the other hand, present the error trial classification accuracy, using error/correct data from session 1 and session 2 as a test-set. Standard deviations are reported in the legends.

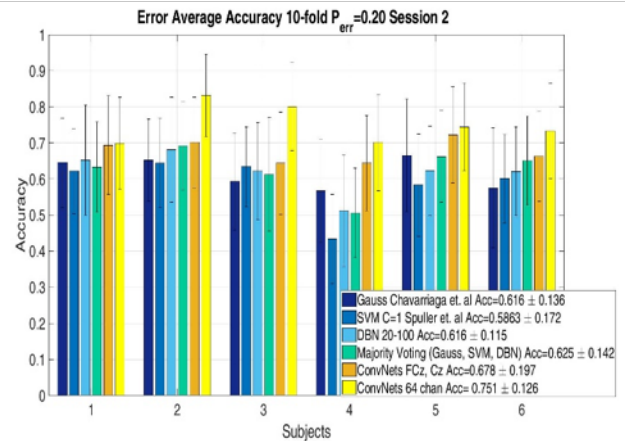
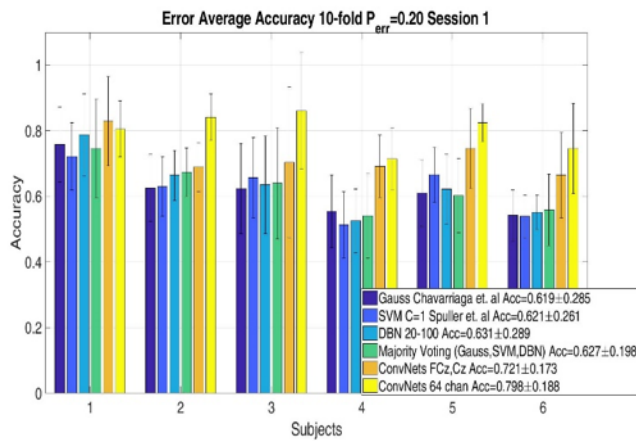


Fig. 3. Average accuracy for Erroneous trial classification in 10-fold within-session cross-validation modality. Figure 3.A shows the results for session 1 as test set, and Figure 3.B shows the results for session 2 as test set, both $P_{err} = 0.2$. These results are wrapped up per subject analyzing classifiers such as Gaussian classifier proposed in [1], the linear SVM proposed in [4], the DBN proposed in this study, a majority voting strategy grouping the decision-level of Gauss, SVM and DBN classifiers, and two ConvNet settings using features from FCz and Cz channels, and the complete set of 64 channels.

IV. CONCLUSION

ErrP is not only a group of early and late deflections statistically related to erroneous trial decoding, but is also considered an important set of features. Re-arranging ErrP features into a 2D representation for training a deep ConvNet architectures yields considerable results in error single trial decoding. The ZCA whitening and the random cropping are critical complements for the ConvNet training/testing and the ErrP single trial representation construction.

REFERENCES

- [1] R. Chavarriaga and J. d. R. Millan, "Learning from eeg error-related potentials in noninvasive brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, pp. 381–388, Aug 2010.
- [2] R. T. Schirmmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [3] T. E. Rosen and M. D. Lerner, "Error-related brain activity and anxiety symptoms in youth with autism spectrum disorder," *Autism Research*, vol. 11, no. 2, pp. 342–354, 2018.
- [4] M. Spüler, C. Niethammer, W. Rosenstiel, and M. Bogdan, "Classification of error-related potentials in eeg during continuous feedback," in *Proceedings of the 6th International Brain-Computer Interface Conference (BCI2014)(Graz)*, 2014, pp. 24–27.
- [5] A. Meyer, M. D. Lerner, A. D. L. Reyes, R. D. Laird, and G. Hajcak, "Considering erp difference scores as individual difference measures: Issues with subtraction and alternative approaches," *Psychophysiology*, vol. 54, no. 1, pp. 114–122. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.12664>
- [6] R. Chavarriaga, A. Sobolewski, and J. del R. Millán, "Errare machinale est: the use of error-related potentials in brain-machine interfaces," in *Front. Neurosci.*, 2014.
- [7] A. Rahimi, P. Kanerva, J. d. R. Milln, and J. M. Rabaey, "Hyperdimensional computing for noninvasive brain-computer interfaces: Blind and one-shot classification of eeg error-related potentials," *10th EAI Int. Conf. on Bio-inspired Information and Communications Technologies*, 2017.
- [8] M. Abu-Alqumsan, C. Kapeller, C. Hintermüller, C. Guger, and A. Peer, "Invariance and variability in interaction error-related potentials and their consequences for classification," *Journal of neural engineering*, vol. 14, no. 6, p. 066015, 2017.
- [9] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, "The prep pipeline: standardized preprocessing for large-scale eeg analysis," *Frontiers in Neuroinformatics*, vol. 9, p. 16, 2015. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fninf.2015.00016>
- [10] A. Mogron, J. Jovicich, L. Bruzzone, and M. Buiatti, "Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.
- [11] J. M. M. Torres and E. A. Stepanov, "Enhanced face/audio emotion recognition: Video and instance level classification using convnets and restricted boltzmann machines," in *1st International Workshop on Affective Computing and Emotion Recognition (ACER)*. Leipzig, Germany: ACM, August 2017, pp. 939–946.
- [12] J. M. M. Torres, E. A. Stepanov, and G. Ricciardi, "Eeg semantic decoding using deep neural networks," in *Rovereto Workshop on Concepts, Actions, and Objects (CAOS)*, Rovereto, Italy, May 2016.
- [13] A. Coates and A. Y. Ng, "Selecting receptive fields in deep networks," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2011, pp. 2528–2536. [Online]. Available: <http://papers.nips.cc/paper/4293-selecting-receptive-fields-in-deep-networks.pdf>