Asynchronous Detection of Erroneous Behaviors in Human-Robot Interaction with EEG: A Comparative Analysis of Machine Learning Models.

Zhezheng Ren[1,6,7], Xuzhe Xia[2,6,7], Yuzhi Tang[3,6,7], Bo Zhao[5,6], Dongsheng Xiao[4,6]*

[1]University of British Columbia, Department of Electrical and Computer Engineering, BC V6T 1Z3, Canada

[2]University of British Columbia, Department of Computer Science, BC V6T 1Z3, Canada

[3]University of Toronto, Department of Computer Science, 40 St George St, Toronto, ON M5S 2E4, Canada

[4]University of British Columbia, Department of Psychiatry, Kinsmen Laboratory of Neurological Research

[5]Bank of Montreal, Data & Analytics, Toronto, Canada

[6]NeuroPrior AI LTD, Neurointelligence research lab, Vancouver, Canada

[7]These authors contributed equally to this work.

*Correspondence to: Dongsheng Xiao, E-mail: michael.xiao@ubc.ca

Abstract:

We present a comparative analysis of two distinct machine-learning models designed to detect asynchronous errors during Human-Robot Interaction (HRI). The models under scrutiny are a customized ResNet model and an ensemble model, both trained and validated using EEG data. The ResNet model is a unique adaptation of the Residual Network, comprising a one-dimensional convolutional layer followed by batch normalization and ReLU activation. It also features a custom ResidualBlock structure and an adaptive average pooling layer, concluding with a fully connected linear layer for binary classification. The ensemble model, on the other hand, incorporates various machine learning models such as MLP, logistic regression, SVC, random forest, k-NN, XGBoost, LSTM, CNN, and a grid-search-optimized XGBoost, unified in a pipeline with feature extraction and transformation steps. A critical component of our research is the innovative probability map, which maintains a granularity of 0.1 seconds. This map forecasts the likelihood of forthcoming one-second intervals being classified as either Error (S96) or Non-error (non-S96). Our comparative analysis reveals significant variations in the performance of the two models, both of which exhibit promising results in detecting erroneous behaviors during HRI. We provide detailed validation results, including the accuracy, F1 score, and confusion matrix for each model. This study offers valuable insights into the potential of machine learning in enhancing HRI efficiency and accuracy, indicating promising directions for future research.

Methods

Our analysis employs two machine learning models: a customized ResNet and an ensemble model. We first create the training dataset by slicing out 1s trials of events (64 x 501) from the raw EEG sets. The ResNet model takes a 1s EEG trial (64 x 501) as input and predicts it as either Error (S96) or Non-error (non-S96). The ensemble model instead uses the same input and predict it as either Error (S96), Motor response (S80), or Non-error (non-S96).

Customized ResNet model: The ResNet model starts with a 1-D convolutional layer, followed by batch normalization and a ReLU activation. A custom ResidualBlock structure is defined, which consists of two convolutional layers with batch normalization and ReLU activation, as well as a skip connection path from the input to the output. The residual block is applied twice in sequence. After passing through the residual blocks, the output is passed to an adaptive average pooling layer, flattened, and finally fed to a fully connected linear layer for binary classification. For training, the data is first split into training and validation sets in a 7:1 ratio. The data is first standardized (Z-score normalization) per channel across the entire training data. The model with the best validation loss during training is saved. Its performance is measured on the validation data using the F1 score and its confusion matrix plotted. The model is additionally evaluated using 10-fold cross-validation for each subject. The average F1 score and average confusion matrix for each subject are recorded.

Ensemble model: An ensemble of machine learning models comprising MLP, logistic regression, SVC, random forest, k-NN, XGBoost, LSTM, CNN, and a grid-search-optimized XGBoost are bundled in a pipeline with feature extraction and transformation steps. XdawnCovariances are applied to the raw data, which is then transformed using TangentSpace. The classifiers are combined into a soft-voting classifier, leveraging the averaged predicted probabilities for final decision-making. This ensemble model is evaluated using 10-fold cross-validation with the SMOTE technique employed for each fold to oversample the minority class and address the class imbalance. Metrics, including accuracy, F1 score, and a confusion matrix, are calculated for each fold, visualized, and stored as text files, with the trained models saved as joblib files.

Probability Map: The granularity of the probability map maintains a resolution of 0.1 seconds, and it is designed to forecast the likelihood of one-second intervals being classified either as S96 or S80 at the onset of every 0.1-second period. The algorithm achieves this by predicting the probabilities of the forthcoming one-second interval receiving either the S96 or S80 label and subsequently summing them. Within each 0.1-second interval, the pre-trained model computes the likelihood of the ensuing one-second period being assigned the S96 or S80 label. This calculated probability is then allocated to the imminent ten data points, each symbolizing a fraction of one-tenth of a second. The algorithm perpetually revises the maximum probability for each point throughout this process. In the concluding phase, the algorithm pinpoints the six-time points that generate the highest probabilities.

Results

The two models displayed promising results in detecting erroneous behaviors during HRI. Table 1 shows the result of 10-fold validation for the Ensemble model (F1 score). Figure 1 presents the confusion matrix of the Ensemble model, Customized ResNet model, and probability map.

Table 1. The results of 10-fold validation for the Ensemble model and ResNet model (Accuracy)

Subject	AA56D	AC17D	AJ05D	AQ59D	AW59D	AY63D	BS34D	BY74D
Ensemble model (Mean±Std)	0.9123 ± 0.0197	0.9130 ± 0.0209	0.9146 ± 0.0212	0.9148 ± 0.0190	0.9137 ± 0.0203	0.9148 ± 0.0196	0.9144 ± 0.0205	0.9113 ± 0.0162
ResNet model (Mean±Std)	0.8584 ± 0.0821	0.9428 ± 0.0524	0.9425 ± 0.0418	0.9559 ± 0.0300	0.9391 ± 0.0573	0.8318 ± 0.1177	0.9193 ± 0.0632	0.8798 ± 0.1338

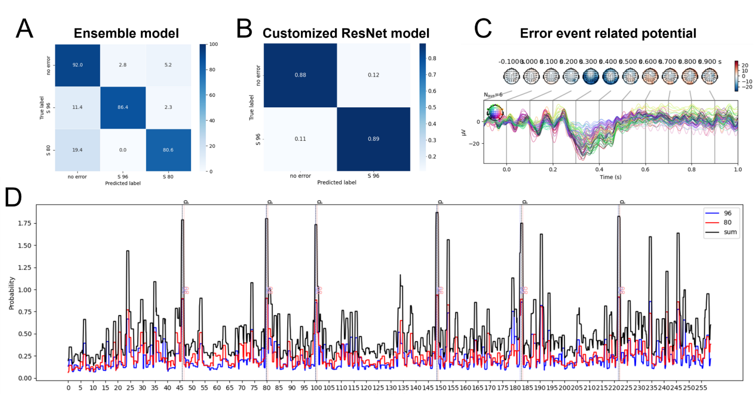

Figure 1. The result of the confusion matrix (A, B) using all data, example error event-related potential (C), and example error time point prediction using probability map (D).

Source code: https://github.com/NeuroPrior-AI/IntEr-HRI-Competition