Heart rate variability enhances the EEG-based machine learning prediction of Internet gaming disorder

Young Eun Park[1], Jieun Seo[1], Sunkyung Lee[1], Donghwan Lee[1], So Young Yoo[2], Jung-Seok Choi[3]

[1]Department of Statistics, Ewha Womans University, Republic of Korea
[2]Department of Psychiatry, SMG-SNU Boramae Medical Center, Republic of Korea
[3]Department of Psychiatry, Samsung Medical Center, Republic of Korea

Abstract

Electroencephalography (EEG) identifies amplitude changes in the evoked response potential (ERP) among patients with internet game disorder (IGD) or Alcohol use disorder (AUD), so it can be a biomarker for predicting IGD or AUD. Recently, with the high-dimensional EEG features, various machine learning methods have been applied and they achieved satisfactory prediction performance. Meanwhile, heart rate variability (HRV) is also being proved that it is associated with disorders like IGD which involves disrupted regulatory function. In this study, we propose the use of HRV as well as EEG for classifying the IGD and AUD patients from normal subjects. We compare various machine learning methods such as XGBoost, Random forest, Elastic Net, and so on. Because the combined predictors are high-dimensional, we first employ two types of iterative sure independence screening (ISIS) to select relevant features. Even though the number of predictors is dramatically reduced, we showed the machine learning results with ISIS are better than the results with whole predictors in terms of both prediction and interpretability. We investigated the utility of HRV on overall performance and found that the integration of EEG and HRV outperforms the single modality method especially in predicting IGD.

Data and Methods

Data

- Patient samples: Collected from SMG-SNU Boramae Medical Center
- Label: AUD (Alcohol Use Disorder; n = 48), IGD (Internet Gaming Disorder; n = 49), and HC (Healthy Control; n = 59)
- Features: EEG (p = 1330) + HRV (p = 12) + demographic variables (age and sex)

Iterative Sure Independence Screening (ISIS) (Saladana and Feng, 2018)

Iterative Sure Independence Screening effectively reduces high-dimensionality, especially in case of n < p.

- **Step 1.** For the first iteration, compute the marginal regression of each predictor and form an index set $\hat{A}_1$ by choosing top-ranked features of size d. Then, maximize the penalized likelihood to get estimators $\hat{\beta}_{\hat{A}_1}$ and leave the non-zero coefficient features to get a feature set $\widehat{M}_1$.(SIS procedure).

- **Step 2.** For the second iteration, compute the conditional marginal regression of each predictor that is not in $\widehat{M}_1$ and form an index set $\hat{A}_2$ by ranking the conditional features.

- **Step 3.** With the set of features $\widehat{M}_1 \cup \hat{A}_2$, maximize the penalized likelihood to get estimators $\hat{\beta}_{\widehat{M}_1 \cup \hat{A}_2}$ and then get updated estimate $\widehat{M}_2$.

- **Step 4.** Iterate Step 2 and 3 until we get an index set $\widehat{M}_l$ with size of d or $\widehat{M}_l = \widehat{M}_j$ for some $j < iter$ or $iter = iter_{max}$.

Learning pipeline of machine learning classifier

- 5-fold cross-validation
- Feature selection
  - Use vanilla ISIS for 2-class classification model and SIS for 3-class classification model. Since there is no R package supporting SIS for multiple classification, we implemented 3-class SIS by coding. Moreover, AIC is used instead of coefficients as ranking criteria in step 1.
  - Select features from the whole dataset using ISIS/SIS (Type-1)
  - Select features from each modality with ISIS/SIS and combine the selected features (Type-2)
  - Without feature selection (Whole)
- Log-scaling on HRV features
- Standard scaling on entire features
- Machine learning methods - Random Forest (RF), Gradient Boosting (GBM), XGBoost (XGB), Elastic Net (EN)
- Two combinations of features are used
  - EEG + age, sex
  - EEG + HRV + age, sex

Result 1: Number of Selected Features (min - max)

	Whole	(a) HC vs AUD	(b) HC vs IGD	(c) HC vs AUD vs IGD
EEG	1332	6 - 22	2 - 24	10 - 18
EEG, HRV	1344	4 - 22 / 6 - 23	9 - 18 / 5 - 25	10 - 17 / 15 - 22

Table 1. The number of selected features (Type-1 / Type-2): (a) HC vs. AUD, (b) HC vs. IGD, (c) HC vs. AUD vs. IGD.

Result 2: Classification Performance of MLs

(a) HC vs. AUD **(b) HC vs. IGD** **(c) HC vs. AUD vs. IGD**

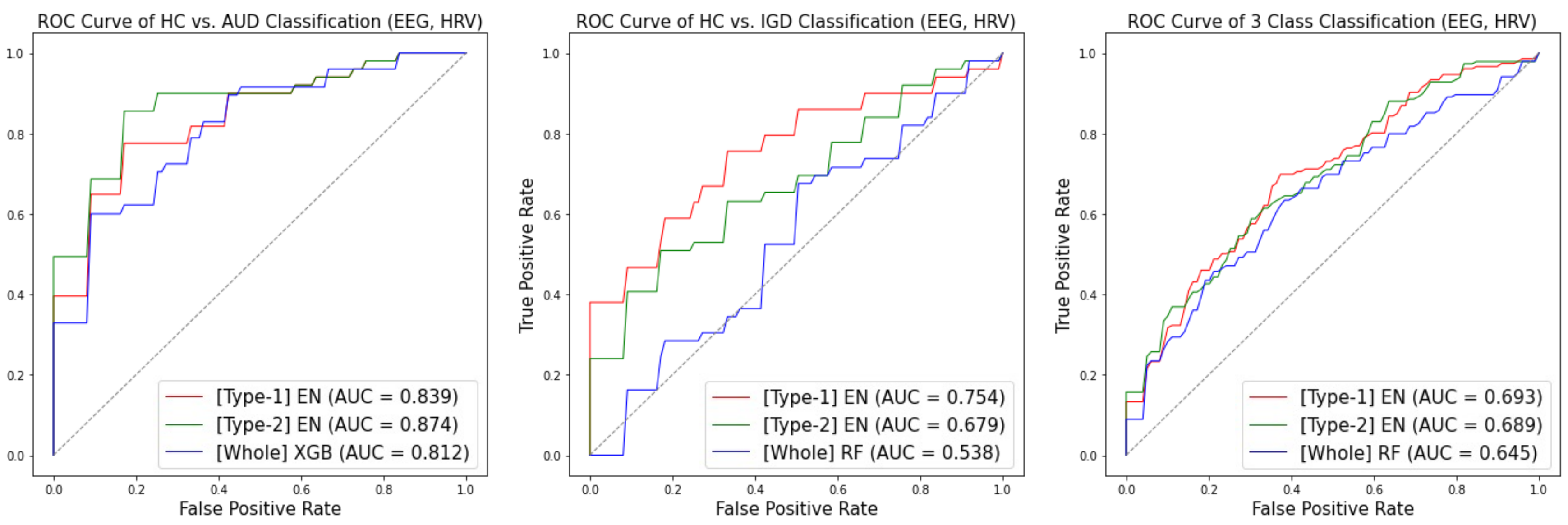

Figure 1. ROC curves of three feature selection methods(Type-1, Type-2, Whole) for each classification using EEG and HRV data. Overall, using ISIS/SIS improved AUC score.

Result 3: The Utility of HRV Features in EN (Best method)

(a) HC vs. AUD

Dataset	AUC score	Feature Selection Method
EEG only	0.8670	Type-1 ISIS
EEG+ **HRV**	**0.8740**	Type-2 ISIS

(b) HC vs. IGD

Dataset	AUC score	Feature Selection Method
EEG only	0.6174	Type-1 ISIS
EEG + **HRV**	**0.7545**	Type-1 ISIS

(c) HC vs. AUD vs IGD (3 classes classification)

Dataset	AUC score	Feature Selection Method
EEG only	0.6875	Type-1 SIS
EEG + **HRV**	**0.6929**	Type-1 SIS

Table 2. The best AUC scores for each classification problem.

Result 4: SHAP Values of the Relevant Features

(a) HC vs. AUD **(b) HC vs. IGD** **(c) HC vs. AUD vs. IGD**

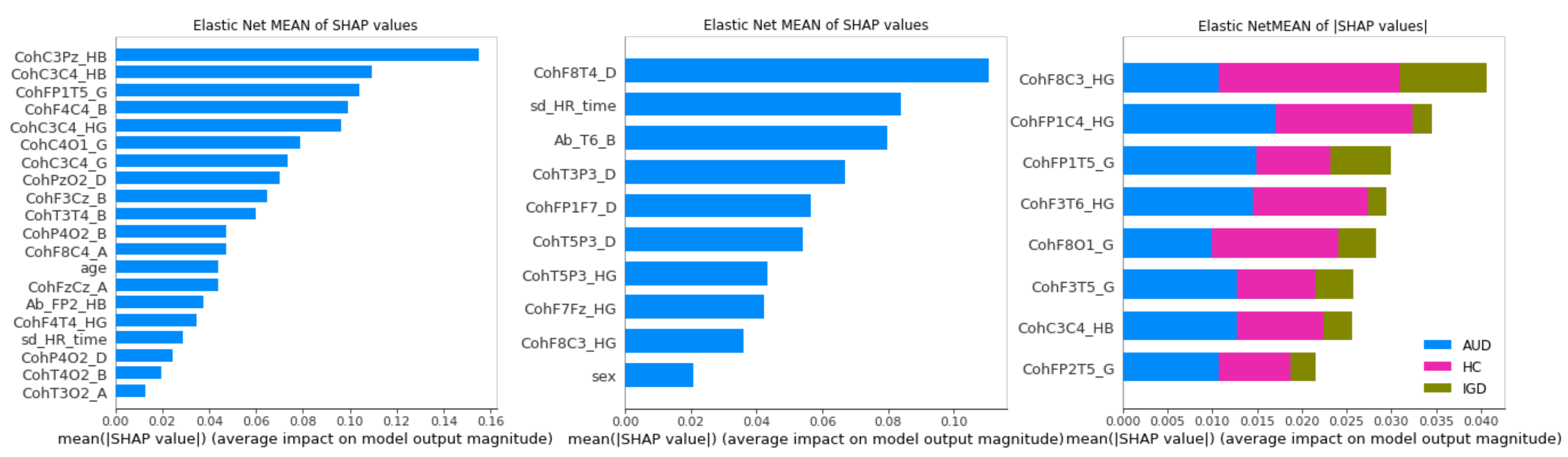

Figure 2. Important features' SHAP values (sorted by the mean of absolute value).

Overall, coherence features of EEG are top-ranked. Among HRV, sd_HR_time is second-ranked in HC vs IGD classification. Moreover, there are some overlapping features in each classification.

- Between (a) and (b): sd_HR_time
- Between (a) and (c) : CohC3C4_HB, CohFP1T5_G
- Between (b) and (c): CohF8C3_HG0

Conclusions

- Sure independence screening improved the prediction performance in general.
- By reducing the dimension while preserving individual features, both predictability and interpretability improved.
- AUC scores were high when using EEG and HRV data.
- When HRV features are integrated, the performance improved. Especially, when classifying HC and IGD, HRV feature seems influential.
- Elastic Net model outperforms among the MLs considered in all cases.

References

- Park SM, Lee JY, Choi AR, et al. (2020). Maladaptive neurovisceral interactions in patients with Internet gaming disorder: A study of heart rate variability and functional neural connect ivity using the graph theory approach. *Addiction Biology. 25*:e12805. https://doi.org/10.1111 /adb.12805
- Saldana DF, & Feng Y (2018). SIS: An R Package for Sure Independence Screening in Ultr ahigh-Dimensional Statistical Models. *Journal of Statistical Software*, *83*(2), 1–25. https://d oi.org/10.18637/jss.v083.i02