

Multi-Site Mild Traumatic Brain Injury Classification with Machine Learning and Harmonization*

Biozid Bostami¹, Flor A. Espinoza¹, Harm J. van der Horn², Joukje van der Naalt²,
Vince D. Calhoun¹, Member, IEEE and Victor M. Vergara¹, Member, IEEE

Abstract—Traumatic brain injury (TBI) can drastically affect an individual's cognition, physical, emotional wellbeing, and behavior. Even patients with mild TBI (mTBI) may suffer from a variety of long-lasting symptoms, which motivates researchers to find better biomarkers. Machine learning algorithms have shown promising results in detecting mTBI from resting-state functional network connectivity (rsFNC) data. However, data collected at multiple sites introduces additional noise called site-effects, resulting in erroneous conclusions. Site errors are controlled through a process called harmonization, but its use in classifying neuroimaging data has been addressed lightly. With the ongoing need to improve mTBI detection, this study shows that harmonization should be integrated into the machine learning process when working with multi-site neuroimaging datasets.

I. INTRODUCTION

Traumatic brain injury (TBI) can have profound adverse effects on an individual's neurocognitive functions. Even patients with mild traumatic brain injury (mTBI) may suffer from various symptoms, including dizziness, fatigue, anxiety, depression, lack of concentration, vertigo, irritability, and impulsiveness [1]. Different technologies have been applied to examine the effects of mTBI on the human brain. Used neuroimaging techniques include diffusion magnetic resonance imaging (dMRI), tomography and structural MRI [2, 3]. Later, resting-state functional network connectivity (rsFNC) was found to be an optimal biomarker for mild TBI [4]. However, when data is collected from different cohorts at multiple sites, additional non-biological variability is added as noise to the combined dataset, commonly known as site effects [5]. These sites' effects are due to differences in scanners, acquisition methods, etc. The additional noise may lead analysis to erroneous conclusions. To reduce undesired site effects from the combined dataset, researchers need to perform harmonization. For neuroimaging, a harmonization algorithm known as 'ComBat' has been successfully applied in several studies [5, 6, 7, 8, 9]. This study explored the effects of harmonization on machine learning classification algorithms for datasets gathered from different site sources. This work shows how integrating harmonization and machine learning classifiers can significantly improve mTBI detection.

II. METHODS

This study includes fMRI cohort data from two countries USA, North America, and the Netherlands, Europe (EU). At the USA New Mexico (NM) site, first fMRI data was collected, getting the approved consent of all participants following the declaration of Helsinki and the institutional guidelines at the University of New Mexico. Second fMRI data received from the Netherlands (EU), following the local medical ethics committee of the University Medical Center Groningen (UMCG) guidelines, and written informed consent was obtained from all participants.

A. Subjects

New Mexico cohort data contained 96 participants (48 mTBI patients and 48 healthy controls) with an age range of 27.3 ± 9.0 . European cohort data included 74 participants (54 mTBI patients and 20 healthy controls) with an age range of 19-64.

B. Imaging protocols

New Mexico cohort data was collected using a 3 Tesla Siemens Trio Scanner. TR (Repetition Time) = 2000 ms; TE (Time of Echo) = 29 ms; flip angle = 75°; FOV (Field of View) = 240 mm; matrix size = 64 x 64. European cohort data was collected using a 3.0 T Philips Integra MRI scanner. TR (Repetition Time) = 2000 ms; TE (Time of Echo) = 20 ms; flip angle = 8°; FOV (Field of View) = 224 x 224 x 136.5 mm.

C. fMRI Pre-processing

fMRI data was first transformed into Montreal Neurological Institute standard space using Statistical Parametric Mapping (SPM; <http://www.fil.ion.ucl.ac.uk/spm>) [10]. AFNI software v17.1.03 was used for de-spiking. Then time courses were converted to orthogonal to 1) linear, quadratic, and cubic trends, 2) 6 realignment parameters, 3) derivatives of realignment parameters. We used group independent component analysis (GICA) [11] using the Group ICA fMRI Toolbox (GIFT; <http://trendscenter.org/software/gift/>) [12] on NM cohort data and collected a set of functionally independent components.

*Research supported by NSF 2112455 and NIH R01DA040487.

¹Biozid Bostami, ¹Flor A. Espinoza, ¹Vince D. Calhoun and ¹Victor M. Vergara with Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TRENDS), Georgia State University, Georgia Institute of Technology, and Emory University, Atlanta, GA, USA.

(e-mail: bbostami@student.gsu.edu, fespinozahidalgo@gsu.edu, vcalhoun@gsu.edu and vvergara@gsu.edu).

²Harm J. van der Horn, ²Joukje van der Naalt are with the Department of Neurology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

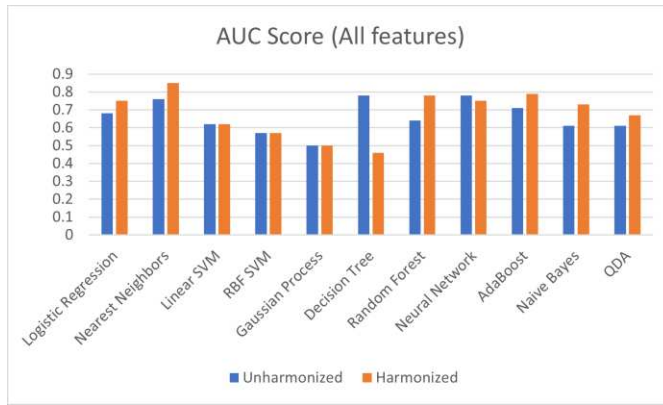


Figure 1: AUC score comparison of different classifier for unharmonized and harmonized datasets considering all features.

Also used group information guided ICA (GIGICA) [13], available through the GIFT software, algorithm to match the 70 components in the EU cohort dataset. Artificial components were discarded, and 48 noise-free components were selected as resting-state networks (rsFN) for further study.

D. ComBat and Machine Learning Algorithms

This study combined harmonization using the ComBat method with a machine learning workflow. ComBat was initially proposed in genomics to reduce the batch effects [14]. Later, several studies showed that ComBat could harmonize different neuroimaging modalities [5, 6, 7, 8, 9]. Another advantage of selecting ComBat is that this algorithm has a decentralized version to harmonize data in a decentralized fashion [15]. Decentralized algorithms do not pool data from the original source, also do not create additional copies of the same dataset which reduce the use of computational resources. Another advantage of using decentralized algorithms is that ensure privacy of the original data. ComBat is a widely used algorithm for harmonization in the field of neuroimaging. For machine learning algorithms, 'scikit-learn' python library was used [16]. This study explored the most common classical machine learning classifiers to predict healthy controls and mTBI patients. The classifiers which were selected include logistic regression, nearest neighbor, gaussian process, support vector machine (SVM), decision tree, random forest, neural network, naïve Bayes, adaboost and quadratic discriminant analysis (QDA).

III. ANALYSIS AND RESULTS

In this experiment, fMRI data was collected from two site sources. Scan data were preprocessed to obtain rsFNC values. Next, form a large dataset with sites' rsFNC data. Then, harmonized the combined dataset using the ComBat algorithm. To minimize any potential confounding influence of age and gender, linear regression was used to regressed out age and sex from rsFNC data. These residuals were further used as rsFNC data for the machine learning classifiers. The dataset was prepared for machine learning classifiers by splitting it into training and testing datasets (80:20). The training dataset was used to train the classifiers and the test data to evaluate their performances.

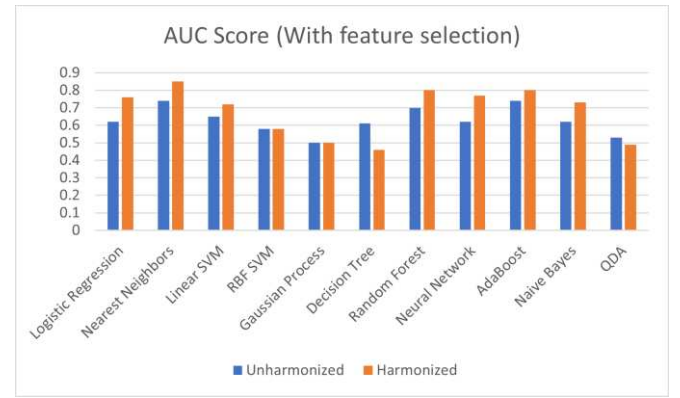


Figure 2: AUC score comparison of different classifier for unharmonized and harmonized datasets considering feature selection.

Next, for each classifier, a tuned model was built by performing grid-search 10-fold cross-validation providing a set of hyperparameters and the training set as the input data. The model with the best area under the curve (AUC) average test score was selected as the classifier's tuned model. Feature selection is achieved by extracting the random forest feature importance values. The lower-dimensional features, referred to as selected features, are obtained by keeping the features with non-zero discriminative power. Then, the classifiers' tuned models were trained considering two input data cases: higher-dimensional data (all features) and lower-dimensional

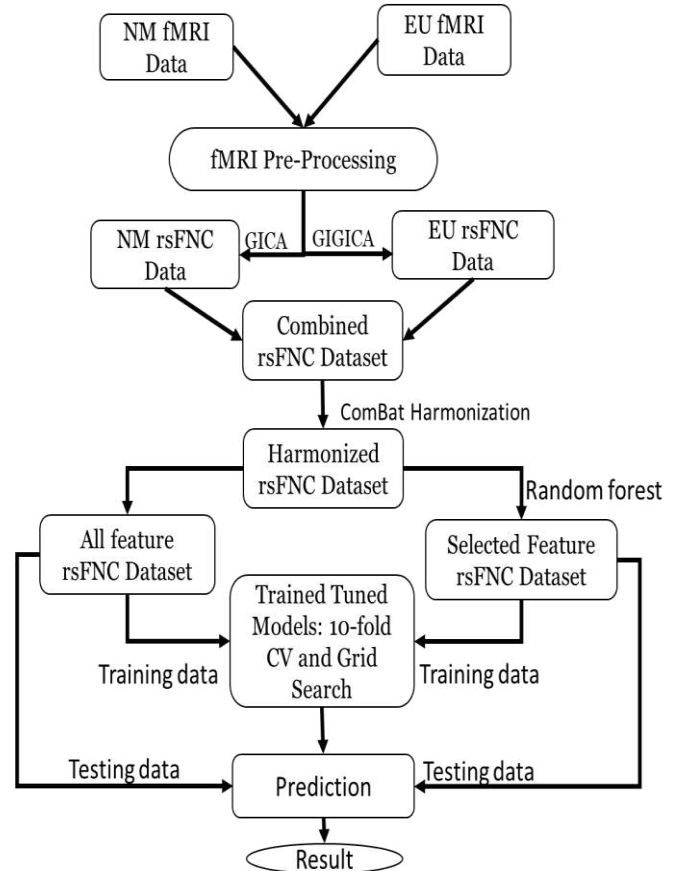


Figure 3: Flow chart of the experiment including Harmonization and Machine Learning

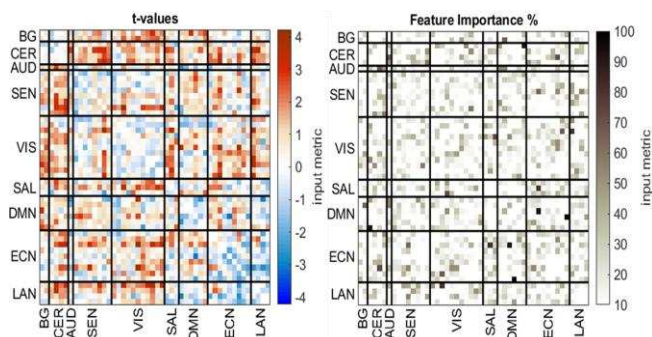


Figure 4: t-values of group-difference (mTBI-HC) (left) and Feature Importance (right) for the harmonized dataset.

data (selected features). Tuned models' performance was evaluated with the test data using the AUC metric.

The model having the best average AUC score was selected as the classifier's tuned model. Finally, tested the tuned models with the test dataset and report the AUC scores. Also performed another experiment without applying the feature selection step in parallel and collected the results.

Similarly, repeated the analysis with unharmonized data and collected the results. Finally, plot the AUC scores of different classifiers for visualization and comparison. Fig. 1 shows the AUC scores of different machine learning classifiers when trained on a harmonized and unharmonized dataset. Results show that we got better predictive scores than the unharmonized dataset when using harmonization. Fig. 2 shows feature selection's AUC scores of different machine learning classifiers. We observed a similar AUC score increase in machine learning classifiers when harmonization was applied. Fig. 1 and Fig. 2 showed that with or without the feature selection, the highest AUC score of 0.85 was achieved for the nearest neighbor classifier for harmonized dataset compared to the highest of 0.76 on the unharmonized dataset. The performance of the algorithms on harmonized dataset changes very slightly on few cases. Fig. 3 showed the overall experimental setup with a flow chart.

For further analysis on the effect of harmonization we calculate t-values of the group difference between mTBI and healthy controls. Fig. 4 and Fig. 5 showed the t-values and feature importance (%). First, calculated the correlation between the t-values of group difference (mTBI-HC) and feature importance; found that the p-value was 0.0256 for the harmonized dataset. However, in the unharmonized dataset, the p-value was 0.7825. Since harmonization removed additional noise due to site effects, the true correlation between the features and group differences was only found in the harmonized results.

IV. DISCUSSION

When neuroimaging data is collected from various locations worldwide, each dataset introduces additional non-biological noise resulting from site effects. Harmonization aims to reduce the source dependency from the combined dataset.

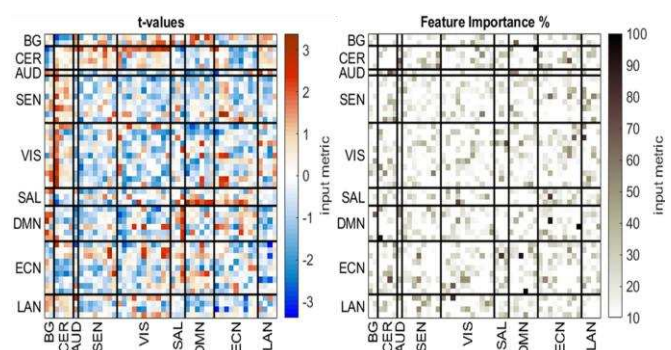


Figure 5: t-values of group-difference (mTBI-HC) (left) and Feature importance (right) for the unharmonized dataset.

This additional noise affects the machine learning classifiers and any other statistical analysis. For data collected at two sites that used different scanners, parameters and acquisition methods, machine learning classifiers performed relatively poorly in this study. After including data harmonization in the machine learning pipeline, we found that reducing site effects can improve machine learning classifiers performance. The New Mexico dataset was analyzed in a previous classification study where the authors showed that SVM has an AUC score of 0.85 [4]. Another study also used SVM to discriminate mTBI patients from healthy controls (HC) on a different dataset and found an AUC score of 0.72 [17]. However, when we combined the two-sites datasets, we found that the performance of the SVM classifier decreased to an AUC score of 0.62. When we harmonized the data and perform feature selection SVM algorithm reached an AUC score of 0.72. The predictive score decreased because the site effect heavily affected the combined dataset. After harmonization, we observed a high correlation between the t-values of group difference and feature importance. Moreover, the previous studies only considered single-source data collected by a single scanner and the same acquisition methods.

We emphasize that no comparisons were shown between other classifiers in previous studies. When data is combined with another dataset from a different source, SVM performs poorly when data remain unharmonized. A different study [18] compared different machine learning algorithms to predict mTBI patients from the EEG dataset. The authors showed that Nearest Neighbor has an accuracy of around 85% for a single source dataset. This study also found that nearest neighbor can be the higher-performing classifier for predicting mTBI patients. This study found that nearest neighbor classifier showed AUC score of 0.74 before harmonization and after harmonization the AUC score increased to 0.85. It was a 10% improvement. Since harmonization removed non-biological variances from the dataset resulted in a much clear group difference (mTBI and HC), it helped nearest neighbor classifier to perform more accurately.

In summary, to the best of our knowledge, no previous studies included data harmonization and machine learning classifiers to predict mTBI patients [4, 17, 18, 19, 20]. Our

study results showed that harmonization should be a part of a multi-site machine learning pipeline.

V. LIMITATIONS AND FUTURE DIRECTION

Among the limitations, this study presents results only considering fMRI data from two site sources. We explored the performance of one harmonization algorithm; however, our primary purpose was not to study harmonization algorithms. Instead, we aimed to examine harmonization effects on machine learning classification algorithms. Another limitation may be considering one metric (AUC) to compare classifiers' performance. However, other metrics also showed a similar performance increase for harmonized data. In future studies, we plan to explore the effects of harmonization with deep learning algorithms.

VI. CONCLUSION

Machine learning algorithms are highly dependent on the quality of the data. Various neuroimaging studies on mTBI have been done worldwide. Scattered data allows researchers to access data collected at different sites to perform better analysis. This study has shown that researchers need to consider harmonization as an integrated part of the machine learning workflow when working with multi-site datasets.

REFERENCES

- [1] DeKosky, S.T., Ikonomic, M.D., and Gandy, S. (2010). Traumatic brain injury: football, warfare, and long-term effects. *N. Engl. J. Med.* 363, 1293–1296.
- [2] Lipton, M.L., Kim, N., Park, Y.K., Hulkower, M.B., Gardin, T.M., Shifteh, K., Kim, M., Zimmerman, M.E., Lipton, R.B., and Branch, C.A. (2012). Robust detection of traumatic axonal injury in individual mild traumatic brain injury patients: intersubject variation, change over time and bidirectional changes in anisotropy. *Brain Imaging Behav.* 6, 329–342.
- [3] Huisman, T.A., Schwamm, L.H., Schaefer, P.W., Koroshetz, W.J., Shetty-Alva, N., Ozsunar, Y., Wu, O., and Sorensen, A.G. (2004). Diffusion tensor imaging as potential biomarker of white matter injury in diffuse axonal injury. *AJNR Am. J. Neuroradiol.* 25, 370–376.
- [4] Vergara, V. M., Mayer, A. R., Damaraju, E., Kiehl, K. A., & Calhoun, V. (2017). Detection of Mild Traumatic Brain Injury by Machine Learning Classification Using Resting State Functional Network Connectivity and Fractional Anisotropy. *Journal of neurotrauma*, 34(5), 1045–1053. <https://doi.org/10.1089/neu.2016.4526>
- [5] Jean-Philippe Fortin, Drew Parker, Birkan Tunc, Takanori Watanabe, Mark A Elliott, Kosha Ruparel, David R Roalf, Theodore D Satterthwaite, Ruben C Gur, Raquel E Gur, Robert T Schultz, Ragini Verma, Russell T Shinohara. Harmonization Of Multi-Site Diffusion Tensor Imaging Data. *NeuroImage*, 161, 149-170, 2017
- [6] Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fava, Patrick J. McGrath, Melvin McInnis, Mary L. Phillips, Madhukar H. Trivedi, Myrna M. Weissman, Russell T. Shinohara. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167, 104-120, 2018
- [7] B. Bostami, V. D. Calhoun, H. J. Van Der Horn and V. Vergara, "Harmonization of Multi-site Dynamic Functional Connectivity Network Data," 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE), 2021, pp. 1-4, doi: 10.1109/BIBE52308.2021.9635538.
- [8] M Yu, KA Linn, P A Cook et al., "Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data", *Hum Brain Mapp.*, vol. 39, pp. 4213-4227, 2018.
- [9] R. Pomponio, G. Erus, M. Habes, J. Doshi, D. Srinivasan, E. Mamourian, et al., "Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan", *Neuroimage*, pp. 116450, December 2019, ISSN 1095-9572.
- [10] Friston, KJ (2003). Statistical parametric mapping. In: *Neuroscience Databases*. Springer: New York, pps. 237–250.
- [11] Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151.
- [12] GIFT software (GIFT v4: <https://trendscenter.org/software/gift/>)
- [13] Y. Du, Y. Fan. Group information guided ICA for analysis of multi-subject fMRI data. 2011, 17th Annual Meeting of the Organization for Human Brain Mapping, Quebec City, Canada. Trainee Abstract Travel Awards, Interactive poster.
- [14] W. Evan Johnson and Cheng Li, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1):118-127, 2007.
- [15] B. Bostami, V. Vergara, V. Calhoun, F. Hillary (2021). "Networking Brain Networks: Federated Harmonization of Neuroimaging Data". The 10th International Conference on Complex Networks and Their Applications - Book of Abstracts. (p. 381).
- [16] Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.
- [17] Xiaoping Luo, Dezha Lin, Shengwei Xia, Dongyu Wang, Xinmang Weng, Wenming Huang, Hongda Ye, "Machine Learning Classification of Mild Traumatic Brain Injury Using Whole-Brain Functional Activity: A Radiomics Analysis", *Disease Markers*, vol. 2021, Article ID 3015238, 7 pages, 2021. <https://doi.org/10.1155/2021/3015238>
- [18] Vishwanath, M., Jafarlou, S., Shin, I., Lim, M. M., Dutt, N., Rahmani, A. M., & Cao, H. (2020). Investigation of machine learning approaches for traumatic brain injury classification via EEG assessment in mice. *Sensors*, 20(7), 2027.
- [19] Caracol G.R. et al. (2019) Prediction of Neurological Deterioration of Patients with Mild Traumatic Brain Injury Using Machine Learning. In: Nguyen H. (eds) *Statistics and Data Science. RSSDS 2019. Communications in Computer and Information Science*, vol 1150. Springer, Singapore. https://doi.org/10.1007/978-981-15-1960-4_14
- [20] McNerney, M.W., Hobday, T., Cole, B. et al. Objective Classification of mTBI Using Machine Learning on a Combination of Frontopolar Electroencephalography Measurements and Self-reported Symptoms. *Sports Med - Open* 5, 14 (2019). <https://doi.org/10.1186/s40798-019-0187-y>