Abstract Submitted
for the APR21 Meeting of
The American Physical Society

Explainable AI for ML jet taggers using expert variables and layerwise relevance propagation[1] MARGARET MORRIS, GARVITA AGARWAL, LAUREN HAY, BENJAMIN MANNIX[2], CHRISTINE MCLEAN, IA IASHVILI, ULRICH SCHUBERT[3], SALVATORE RAPPOCCIO, University at Buffalo, State University of New York, UNIVERSITY AT BUFFALO TEAM — A framework is presented to extract and understand decision-making information from a deep neural network classifier of jet substructure tagging techniques. There are two methods studied. The first is using expert variables that augment the inputs ("expert-augmented" variables, or XAUGs). These XAUGs are concatenated to the classifier steps immediately before the final decision. The second is layerwise relevance propagation (LRP).The results show that XAUG variables can be used to interpret classifier behavior, increase discrimination ability when combined with low-level features, and in some cases capture the behavior of the classifier completely. The LRP technique can be used to find relevant information the network is using, and when combined with the XAUG variables, can be used to rank features, allowing one to find a reduced set of features that capture part of the network performance. These XAUGs can also be added to low-level networks as a guide to improve performance.

[1]This work was supported under NSF Grants PHY-1806573, PHY-1719690 and PHY-1652066. Computations were performed at the Center for Computational Research at the University at Buffalo.
[2]Now at University of Oregon
[3]Now at Google

Salvatore Rappoccio
State Univ of NY - Buffalo

Date submitted: 05 Jan 2021 Electronic form version 1.4