Comparison of Neural Networks performance for $H \to ZZ \to 4l$ analysis using the TMVA ROOT package

Sergey A. Uzunyan,

Department of Physics, Northern Illinois University, DeKalb, IL 60115, USA

October 24, 2018

Abstract

Neural Networks (NNs) are an important part of modern high energy physics analyses. Among a variety of applications that are available for construction and evaluation of complex NNs, the especially interested those that can work with Deep Neural Networks (DNNs). However, the training of DNN discriminants is slow on traditional CPUs while use of graphics processors does significantly speed up the DNN analysis. The software packages for DNN construction that support GPUs are the $TensorFlow$ [1] via $Keras$ [2] interface and the Toolkit for Multivariate Data Analysis with ROOT (TMVA) package [3]. As the ROOT TMVA also allows the fast comparison between different types of Neural Networks we choose it to analyse the perfomance of popular networks .

1 Introduction

We used a framework based on the TMVA ROOT application package for training and evaluation of different neural networks using events selected for $H \to ZZ \to 4l$ analysis. We compared the performance of a DNN, Multy Layer Perception (MLP), and a Boosted Decision Tree (BDT) networks. The topology of the DNN network was taken from the [4] while we used the configurations of MLP and BDT networks as defined in the ROOT TMVA package.

Two sets of test were performed - first, we studied the discrimination power of selected networks for the inclusive Higgs signal for different sets of input variables, and second we evaluated the performance of TMVA networks for the Higgs vector-boson fusion (VBF) signal subset.

2 Software and hardware configuration

The presented results obtained with the package compiled with ROOT 6.10 and CUDA development framework 8.0. For the DNN training the Nvidia 1080 TI GPU was used.

3 Training of TMVA networks

For training of TMVA networks we used data, signal and standard model backgrounds Monte Carlo (MC) event samples shown in Tables 1 and 2, Section 10 of the analysis paper, [6]. For convinience all input signal, data, and backround samples, the trained MVA discriminats and classification weights of data events along with the TMVA service histograms were combined into the single output ROOT file. We then found the optimal cuts on trained discriminats using the product of the selection efficiency $\epsilon = \frac{S_{selected}}{S_{total}}$ and purity $\pi = \frac{S_{selected}}{(S+B)_{selected}}$ as a criterion for the network discrimination performance and compared the results for all tested networks and sets of input variables. The input event compositions in an inclusive Higgs signal and in the VBF channel are shown in Figure 1. The detailed description of the training framework package can be found at [5].

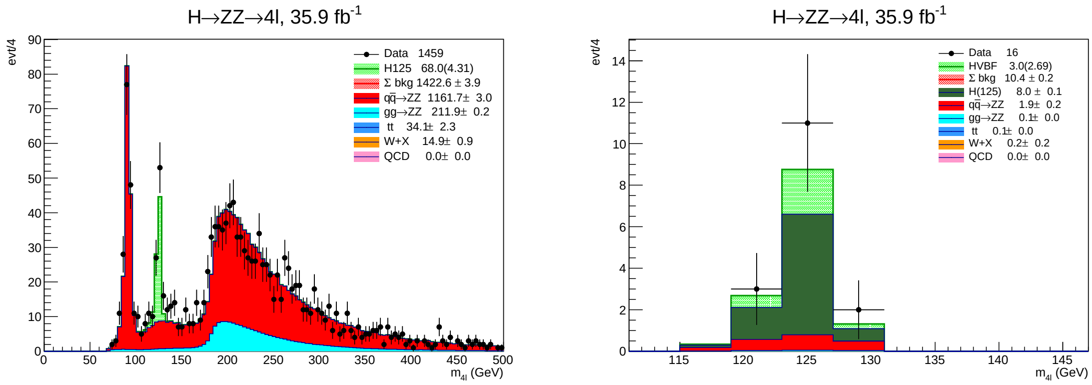


Figure 1: The four lepton invariant mass distributions for signal and SM backgrounds after selections of (a) inclusive Higgs signal , (b) VBF Higgs signal :

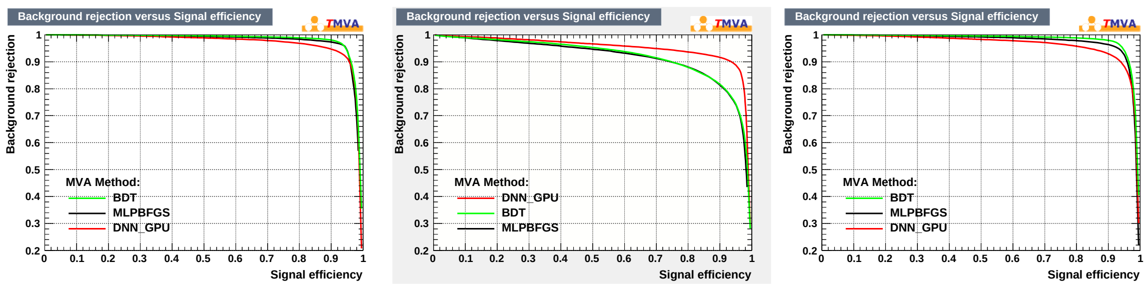
4 NN performance comparison (Higgs Signal)

For the comparison tests we defined three sets of input variables:

- the High Level Input set (HLI), that included invariant masses M_{Z1} , M_{Z2} of Z bosons candidates , and invariant masses, pseudorapidity and polar angle of M_{4l} , η_{4l} and phi_{4l} of Higgs candidates
- the Low Level Input set (LLI) - momentum and angles of leptons, jets and missing energy
- the Combined Input set (CI), LLI+HLI

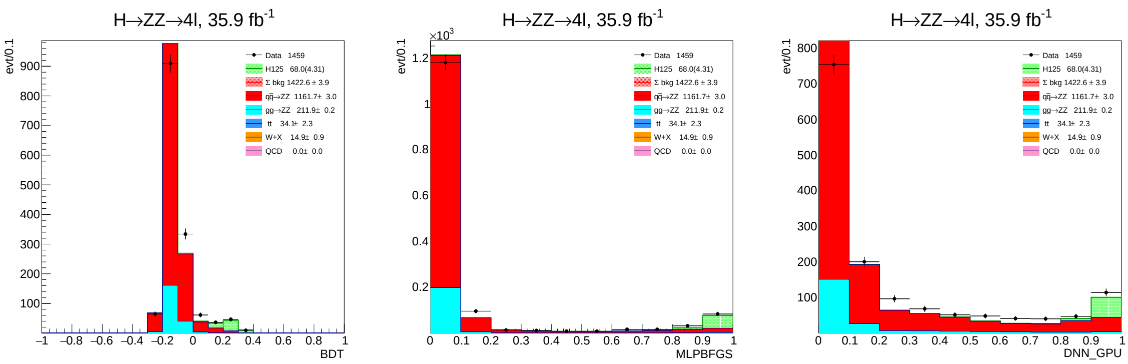
We trained The background rejection versus signal efficiency curves, or "ROC curves" for each trained discriminant are shown in Figure 2. Figures 3-11 shows the NN discriminants distributions for each set and for each network, distributions of $\epsilon \times \pi$ as a function of applied cuts on NN discriminats and distributions of four lepton invariant mass after optimal cuts. Calculations of area under ROC curves, $max(\epsilon \times \pi)$ for each network and input set are summarized in Tabe 2.

The BDT shows the best performance in terms of $max(\epsilon \times \pi)$ for both HLI and combined set of input variables, the DNN network works significantly better for the low level set of input varibles. BDT training is also significantly faster on the same hardware.



(a) (b) (c)

Figure 2: ROC curves comparison for different MVA methods obtained for : a) LLI set; b) HLI set; c) combined set of LLI and HLI variables.

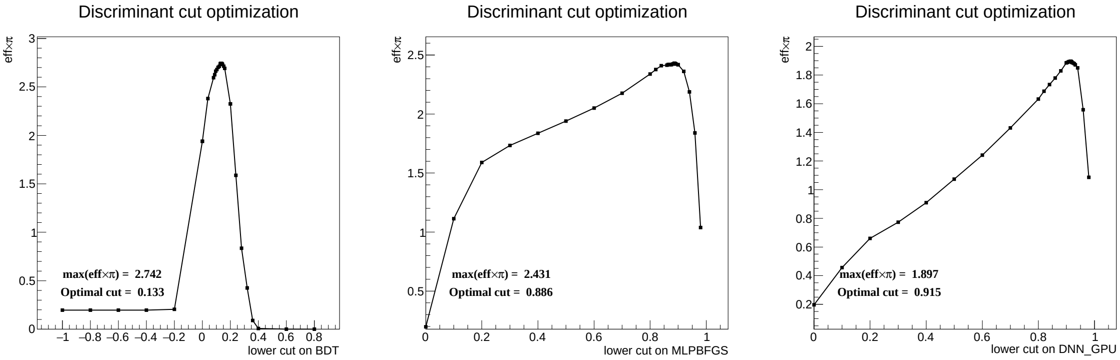


(a) (b) (c)

Figure 3: Distributions of MVA discriminants for the HLI set of input variables: a) BDT; b) MLP; c) DNN.

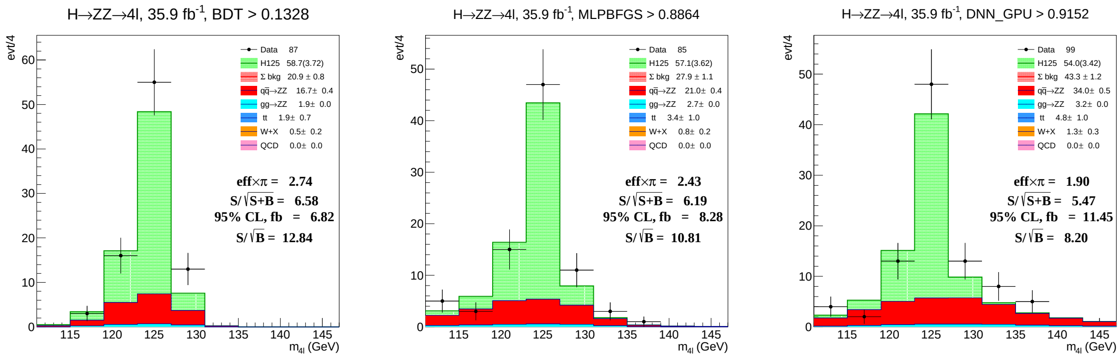
5 NN performance comparison (Higgs VBF Signal)

We applied the same MVA configurations (BDT, MLP and DNN) for the Higgs VBF signal selections. We now compared the significance of the optimal selection obtained using MVA trained discriminants with the significance of selections applied in the VBF analysis via independently trained DNN network based on KERAS interface. The Higgs VBF selections were the same in both cases as well as the input set of analysis variables, except that we added the VBF Matrix Element Likelihood Analysis (MELA, [8]) discriminant to the TMVA set. Figure 12 shows the ROC curves comparison for different MVA methods, distributions of MVA discriminants are shown in Figure 13, and the four leptons invariant mass distributions after optimal cuts on MVA discriminants are shown in Figure 14 . For comparison, Figure 15 shows the four leptons invariant mass distribution obtained in the CMS VBF analysis, $eff \times purity$=1.12. The maximum of $eff \times purity$, 1.14, is observed using the MVA BDT method .



(a) (b) (c)

Figure 4: Optimization of cut on MVA discriminants via Max(efficiency $\times$ purity) : for the HLI set of input variables: a) BDT; b) MLP; c) DNN.



(a) (b) (c)

Figure 5: Distributions of invariant mass of four leptons after optimal cuts on MVA discriminants : for the LLI set of input variables: a) BDT; b) MLP; c) DNN.

6 Summary

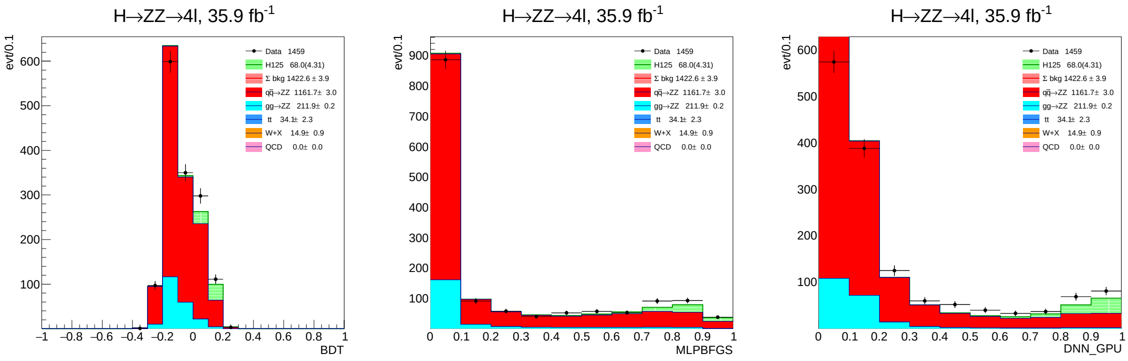
Performance of BDT, MLP and DNN networks evaluated with ROOT TMVA package using $H \to ZZ \to 4l(e, mu)$ datasets (2016). Results show that DNN outperforms BDT and MLP with low-level inputs, however the BDT works better using the most complete set of the analysis variables. Application of tested TMVA discriminants to the selection of Higgs VBF signal shows similar discrimination power.

References

[1] "An open source machine learning framework for everyone", https://www.tensorflow.org.

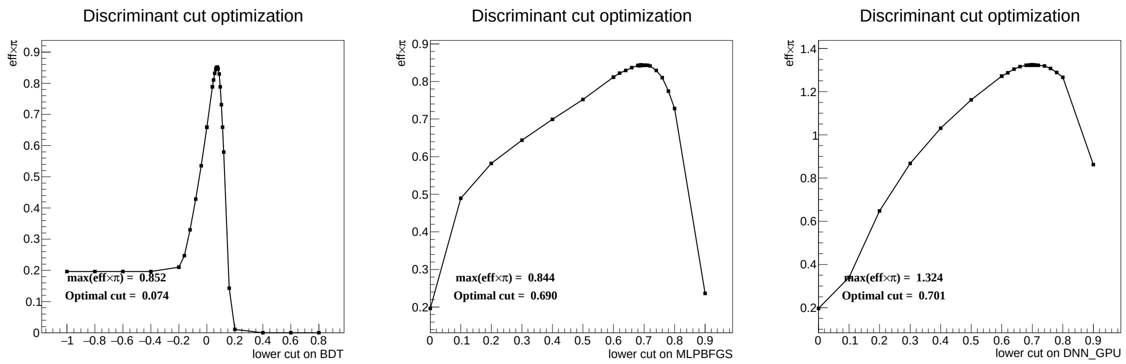
[2] "Keras: The Python Deep Learning library", https://keras.io/.

[3] "The Toolkit for Multivariate Data Analysis with ROOT (TMVA)", https://root.cern.ch/tmva.



(a) (b) (c)

Figure 6: Distributions of MVA discriminants for the LLI set of input variables: a) BDT; b) MLP; c) DNN.



(a) (b) (c)

Figure 7: Optimization of cut on MVA discriminants via Max(efficiency $\times$ purity) : for the LLI set of input variables: a) BDT; b) MLP; c) DNN.

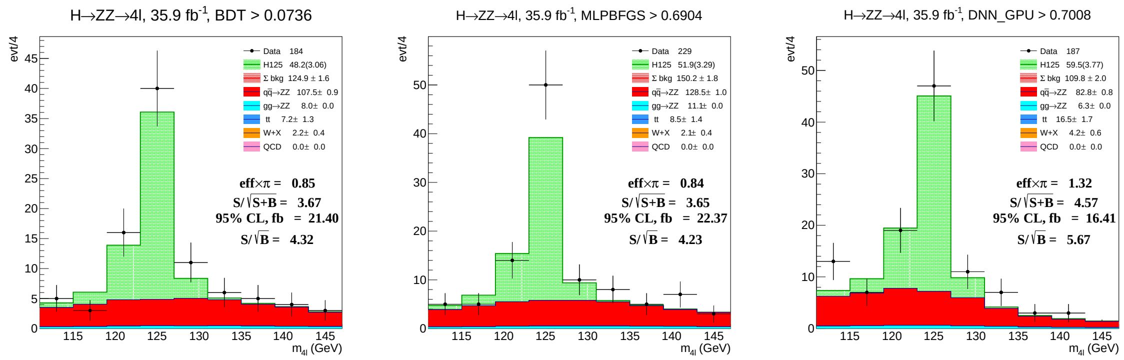
[4] P. Baldi, P. Sadowski and D. Whiteson, Nature Commun. 5, 4308 (2014), doi:10.1038/ncomms5308 [arXiv:1402.4735 [hep-ph]].

[5] S. Uzunyan, https://twiki.cern.ch/twiki/bin/view/Main/HzzTmva

[6] “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at sqrt(s)=13 TeV”, CMS Collaboration (Sirunyan, Albert M et al.) JHEP 1711 (2017) 047 arXiv:1706.09936 [hep-ex] CMS-HIG-16-041, CERN-EP-2017-123.

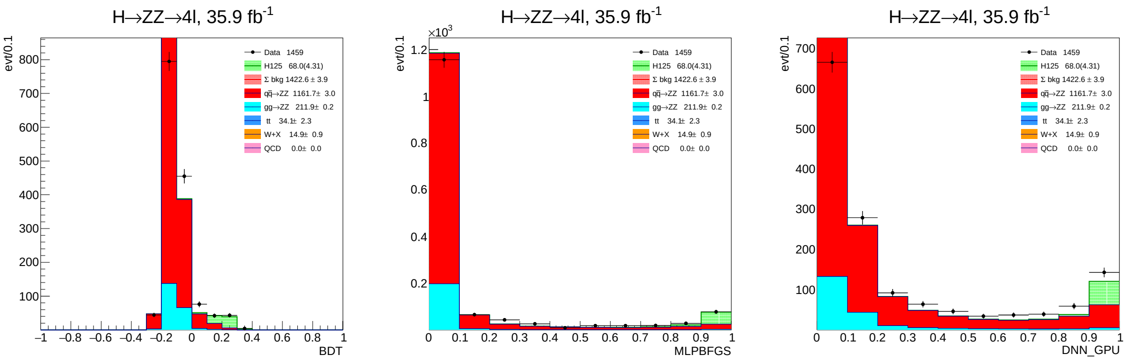
[7] M. M. Almeida *et. al.* “Measurement of Higgs Production Cross Section via Vector Boson Fusion in H-¿ZZ-¿4l final state at 13 TeV using Deep Neural Networks”, CMS Analysis Note AN-18-120.

[8] “The Matrix Element Method within CMS”, https://indico.cern.ch/event/258092/contributions/1588574/attachments/454217/629628/ACA



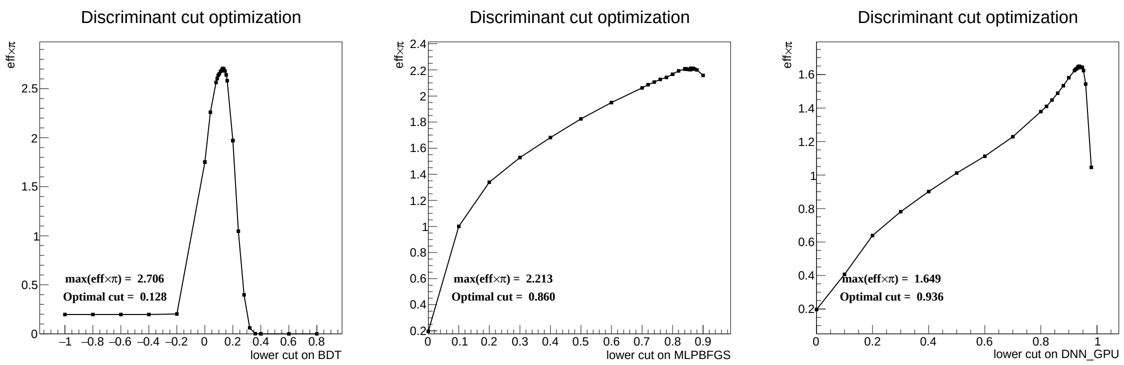
(a) **(b)** **(c)**

Figure 8: Distributions of invariant mass of four leptons after optimal cuts on MVA discriminants : for the LLI set of input variables: a) BDT; b) MLP; c) DNN.



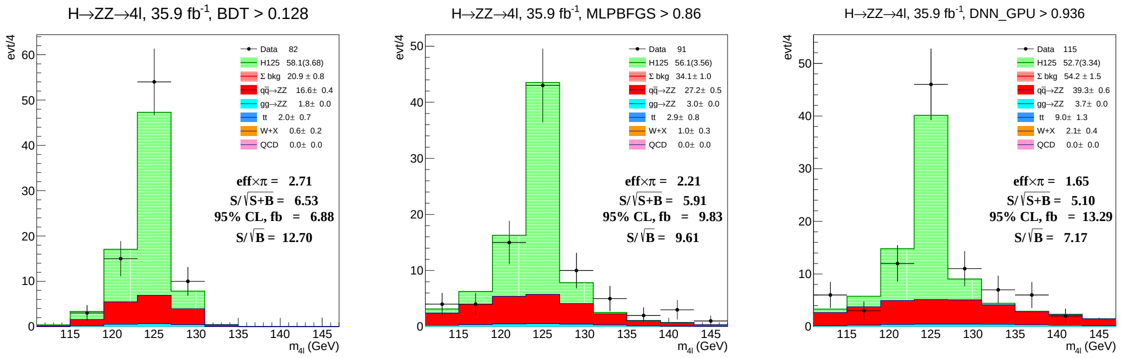
(a) **(b)** **(c)**

Figure 9: Distributions of MVA discriminants for the combined (LLI+HLI) set of input variables: a) BDT; b) MLP; c) DNN.



(a) **(b)** **(c)**

Figure 10: Optimization of cut on MVA discriminants via Max(efficiency $\times$ purity) : for the combined (LLI+HLI) set of input variables: a) BDT; b) MLP; c) DNN.



(a) (b) (c)

Figure 11: Distributions of invariant mass of four leptons after optimal cuts on MVA discriminants : for the combined (LLI+ HLI) set of input variables: a) BDT; b) MLP; c) DNN.

Table 1: Comparison of MVA methods for different sets of training variables.

Training set	HLI			LLI			HLI+LLI		
MVA	DNN	MLP	BDT	DNN	MLP	BDT	DNN	MLP	BDT
ROC AUC	0.969	0.974	0.980	0.950	.919	0.913	0.963	0.973	0.981
$\epsilon \times \pi$	1.90	2.43	**2.74**	**1.32**	0.85	0.84	1.65	2.21	**2.71**
Training time, sec (346620 events)	1250	3080	**127**	6660	26100	**374**	4280	44100	**492**

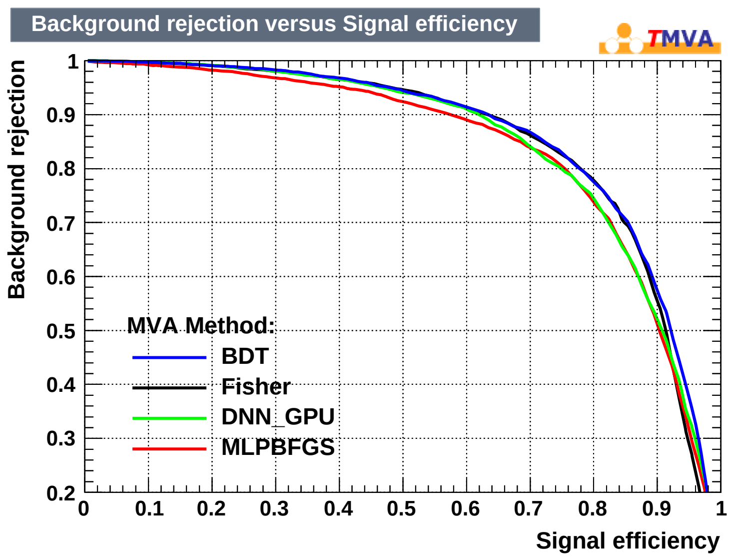
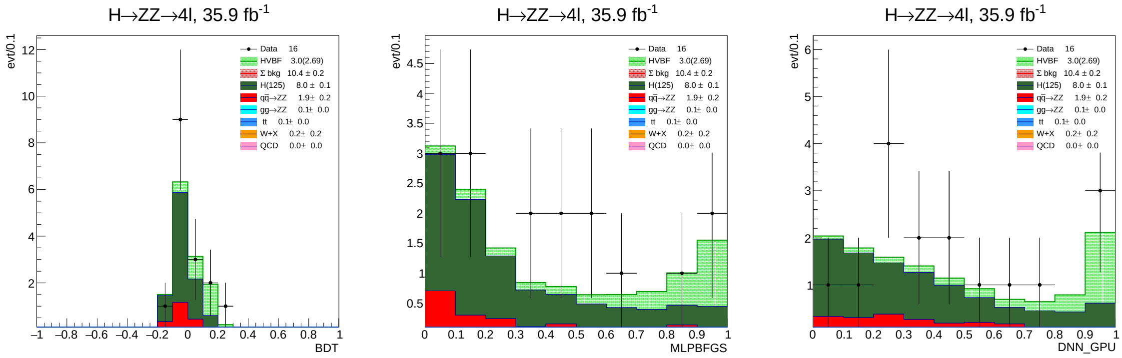
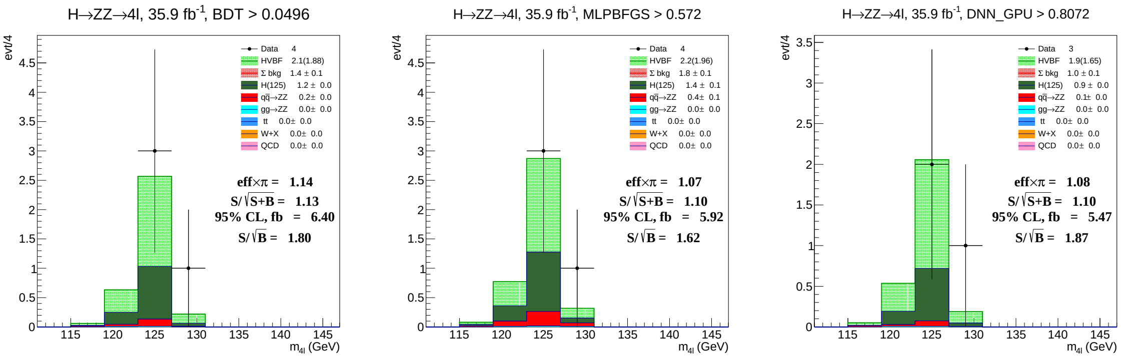


Figure 12: ROC curves comparison for different MVA methods for the Higgs VBF analysis.



(a) (b) (c)

Figure 13: Distributions of MVA discriminants for the Higgs VBF analysis a) BDT; b) MLP; c) DNN.



(a) (b) (c)

Figure 14: Distributions of the four leptons invariant mass after optimal cuts on MVA discriminants for the Higgs VBF analysis.

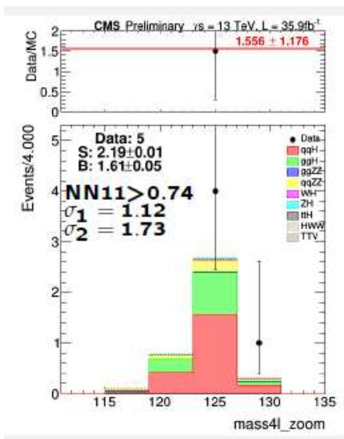


Figure 15: Distributions of the four leptons invariant mass after optimal cut on DNN discriminant (VBF analysis note)

Table 2: Comparison of MVA methods for Higgs VBF Signal selection

MVA	DNN	MLP	BDT
ROC AUC	0.848	0.838	0.860
$\epsilon \times \pi$	1.08	1.07	**1.14**
Training time, sec (18758 signal + 23980 background events)	179	2210	**26.4**