AIPhysics Final Report: Machine Learning on Higgs Production

Hsin Yu Wu
Student Number: 106022231
Physics Department, National Tsing Hua University.
(Dated: May 22, 2022)

The Higgs Boson was discovered at the LHC in 2012, so the study of it has begun. Higgs Bosons produced the gluon-gluon fusion, which hid physics beyond the Standard Model. Therefore, if we want to do research on the BSM, we need to seperate this mode from other modes. In this project, we try to use machine learning to classify different modes. We train DNN model and Decision Tree model by the high level feature of Higgs.

I. INTRODUCTION

A. Higgs production

The Higgs Boson was discovered at the LHC in 2012, and the study has just begun. They are produce from 4 different modes, as figures show.(Fig.1)

Since it was discovered, we have to find the technique to distinguish them. When we research the paper, we see many ways to do so. They use two stream CNN and boosted decision tree. [1][2][3]

A major motivation for the study of boosted H $\rightarrow$ bb final states is that it allows us to study the structure of the gg $\rightarrow$ H process, which is at high pT. However, H $\rightarrow$ bb has other processes, including VBF(vector boson fusion), VH(production associated with a vector boson) and ttH(production associated with a top-quark pair).

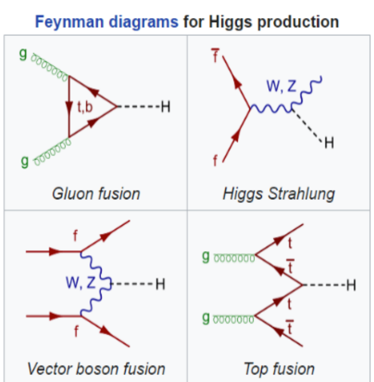

FIG. 1. 4 modes of higgs production. [4]

ggF means the two gluons combine to form a loop of virtual quarks. In the VH part, If a fermion collides with an anti-fermion, they can merge to form a Higgs boson. VBF is that when two fermions collide and the two exchange a virtual W or Z boson, and it will emit a Higgs boson. ttH involves two colliding gluons, and they decay into a heavy quark–antiquark pair. A quark and antiquark from each pair can form a Higgs. particle.[4]

If gg $\rightarrow$ H could be clearly separated from the other Higgs production modes, the sensitivity to BSM would be enhanced. However, these four higgs production channel are all higgs to two b quarks. Thus, we can't distinguish them by having a look at bb, so we put focus on non-Higgs jet, which are circled in Fig.2. Using the data like momentum, angle of particle relative to beam axis, or using image, we can classify different mode.

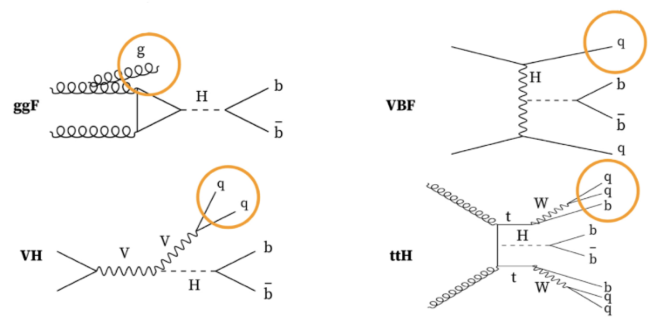

FIG. 2. 4 decay modes of higgs production. [1]

B. Features

We have three major groups of data which are Low Level Features, High Level Features and jet image. The low level feature is primitive physical quantity, which contains the momentum, the direction of particle motion (the angle with the incident particle beam). And then, with some calculation on the low level feature, we can get high level features. The third part of data is leading non-Higgs jet images. Images show the eta and phi of jets.

In order to distinguish four different modes, We take focus on High level features to train the model. we use high level features below:

1. M_j means the invariant mass of the leading jet.
2. M_{jj} means the invariant mass of two leading jets.
3. η_j means the the pseudorapidity of the leading jet, which is the angle with the incident particle beam. The equation is: $\eta = -ln(tan(\frac{\theta}{2}))$

4. $|\Delta\eta_{jj}|$ means the absolute η difference between the leading and subleading jets.

5. $girth = \sum_{i\in J}^{N} \frac{P_T^i r_i}{P_T^J}$

6. CIJS: $\phi_c = \frac{1}{N}\sum_{j=1}^{2}\sum_{i\in J}^{N} \frac{P_{T,i}^{j}(0<r_i^j<0.1)}{P_T^j}$

7. SIJS: $\phi_s = \frac{1}{N}\sum_{j=1}^{2}\sum_{i\in J}^{N} \frac{P_{T,i}^{j}(0<r_i^j<0.2)}{P_T^j}$

II. METHOD

A. DNN

In DNN model part, we use all high level features to train the model. We just extend example code in class and add the number of hidden layer. The first hidden layer consists of 256 nodes and decrease in follow up layers. Activation function of six hidden layer is ReLU. Activation function of output layer is softmax because it is great for multi-classification. More clearly description about our architecture is below:

1. structure: 6 hidden layer with 256, 128, 64, 32, 16, 8 nodes
2. activation function: ReLU
3. activation function of output layer: softmax
4. optimizer: adam
5. loss function: cross entropy
6. batchsize: 512
7. epoch: 50

B. Decision tree

When we searched the paper, we saw that they used boosted decision tree, so we tried to use decision tree method, which taught in the class. Decision Trees is a supervised learning method, which used for classification and regression. It create a model that uses the features of data to learn simple rule to make the decision.

By using whole high level features, we trained the model with different level and observed the result.

III. RESULT

A. Feature exploration

We have three kinds of data: low level features, high level features, and jet image. We think that low level features can't distinguish each modes, so we use high level features to train the model. Fig.3 shows the histogram of high level features. It has more obvious characteristic for each modes, such as pt, CIJS, etc.

The label of the data is one hot encoding. There is four label: ggF, VBF, VH, ttH. By this, we can train the model easily. The data has been already divided into three group: testing, training, and validation data. However, when we train the model, we think this is a little strange. Thus, we remix the whole data and divide them again. In this way, we can also change any ratio of data.

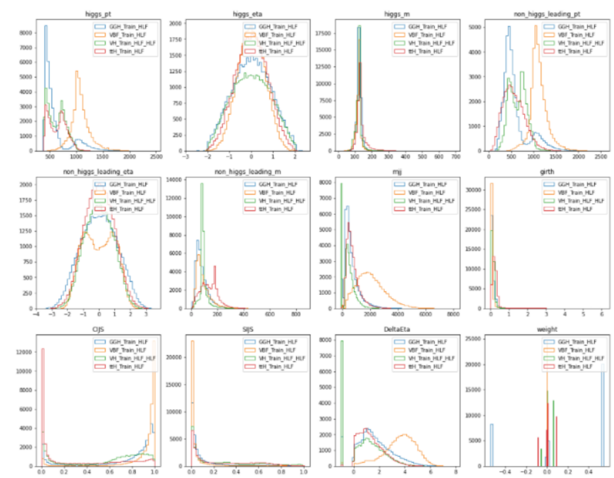

FIG. 3. Histogram of High level feature.

B. DNN

We made the ROC curve (Fig.4) and confusion matrix. It is useful to present the effectiveness of the methods using the ROC curves. Our AUC value is 0.92. See Fig.5, the horizontal axis is real label and vertical axis is prediction. We can see VBF, VH, ttH have better results while GGH only get 0.672. Besides, our total test accuracy is 0.78. Since we don't so much about the model, we just try different function and get this the best result. it still has to be improved.

C. Decision tree

We tried different level of decision tree and different ways to seperate the data. We remixed the test data and train data and divide them in a ratio of 3:7. We also check the label is seperated equally to testing and training data.

The accuracy of 7 level is 0.989. We find that we can get very high value just in 5 to 7 level. Fig.6 is tree model and Fig.7 is its confusion matrix. It just has some mistake on distinguishing VH as ttH. However, most of the predictions are correct.

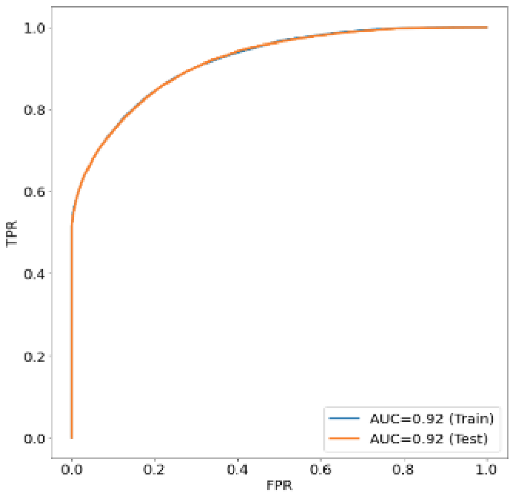

FIG. 4. ROC curve of DNN model.

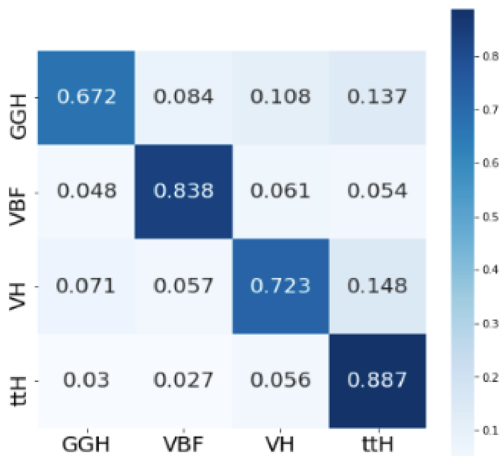

FIG. 5. Confusion matrix of DNN model.

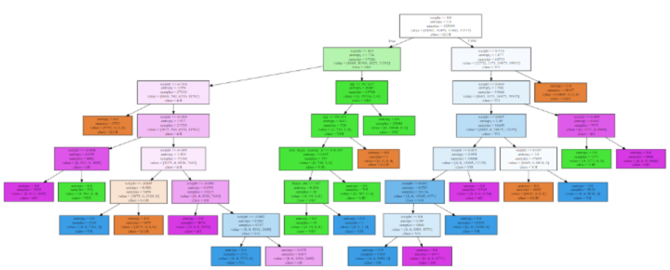

FIG. 6. Decision tree for 7 level, test:train=3:7

Second, we try to divide the testing data and the training data equally. This means the number of training data is decrease. We want to see if the number of training data is small, how the result will be. The answer is that if the layer of tree is still 7 layers, the accuracy will be still above 0.9. Also, we can see Fig.8, if we just use 3 layers, the accuracy is still 0.73. Fig.9 is the confusion matrix of 3 layers. This model see the others as ttH, which circled

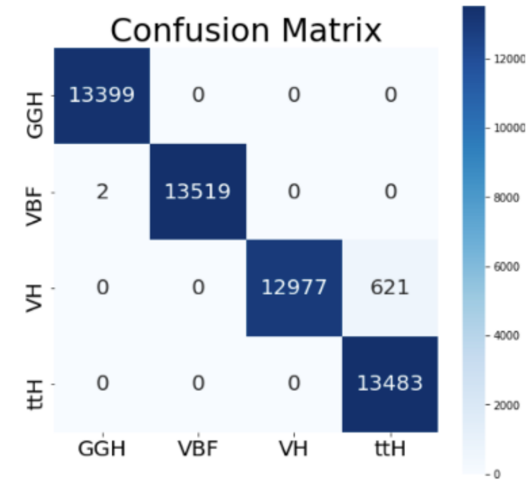

FIG. 7. Confusion matrix of decision tree(7 level) model, test:train=3:7

in red line. To observe the architecture of the tree, we find that the "weight" feature is the most important. It appears in the first level, so it means that weight is the first consideration to classify the data.

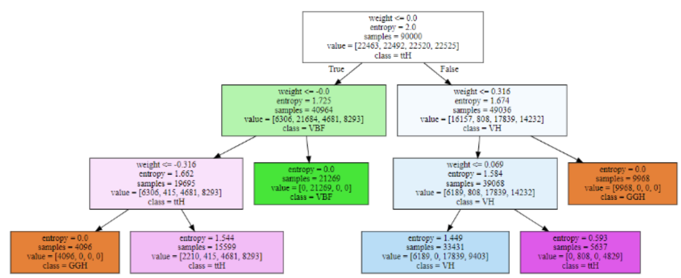

FIG. 8. decision tree(3 level) model, test:train=1:1

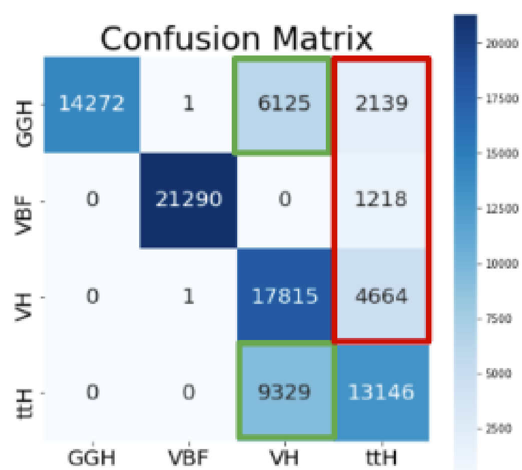

FIG. 9. Confusion matrix of decision tree(3 level) model, test:train=1:1

If we take away the weight, the accuracy will decrease. Fig.10 shows the 3 level tree without feature "weight", and Fig.11 is the confusion matrix of this model. The accuracy is only 0.61. Thus, we can find the importance of weight. We can also observe which features can also reduce the entropy effectively. We can find 'Delta Eta' , 'non higgs leading pt' , 'girth' , 'non higgs leading m' are selected first to distinguish the data. This result is also what we predict. For taking a look at the properties of non leading higgs jet, we can classify different mode in effect.

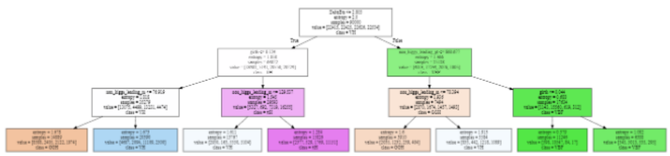

FIG. 10. decision tree(3 level) model, without weight.

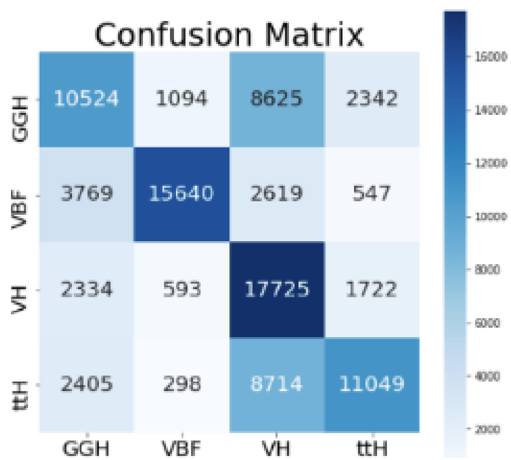

FIG. 11. Confusion matrix of decision tree(3 level) model, without weight.

IV. DISCUSSION

In our result, our DNN model get 0.78 accuracy, and the best accuracy of decision tree is 0.989. We are surprised in our result that we get better accuracy in Decision Tree model. When we use 7 level decision tree, the prediction is very close to the label.

In DNN part, we try to use more layers to train the model because we think when we use more features, we have to increase the number of layers. We use full connection layers, and try different activation function and loss function. The result gets better when we use adam as optimizer. We also try to change the activation function of output layer, and softmax and sigmoid get similar result. We finally decide to use softmax because it is for multi-classification.

In Decision Tree part, we use different ratio of data and different level of tree. By doing this, we can check whether the data ratio will have an effect on the result or not, and also check the situation of overfitting.

Therefore, it is more easy for us to find the best model of decision tree. Maybe that's why we can't get better result of DNN because we don't know how to find the best model and the best parameters.

V. CONCLUSION

In this report, we use two model: DNN and decision tree to do classification on Higgs production. We use high level features to train the model. The accuracy of decision tree is 0.989 and the DNN is 0.78. Both of them can roughly distinguish four different mode.

However, we still have some data that we didn't use. We think jet image can use CNN to do image classification, and low level features may use DNN. Neuron network is more complicated that it may find the connection between each feature. Thus, it still has many interesting things let us to attempt.

[1] Y.-L. Chung, S.-C. Hsu, and B. Nachman, Disentangling boosted higgs boson production modes with machine learning, arXiv preprint arXiv:2009.05930 (2020).

[2] K. Becker, F. Caola, A. Massironi, B. Mistlberger, P. Monni, X. Chen, S. Frixione, T. Gehrmann, N. Glover, K. Hamilton, *et al.*, Precise predictions for boosted higgs production, arXiv preprint arXiv:2005.07762 (2020).

[3] J. Lin, M. Freytsis, I. Moult, and B. Nachman, Boosting h→ bb with machine learning, Journal of High Energy Physics **2018**, 1 (2018).

[4] Wikipedia contributors, Higgs boson — Wikipedia, the free encyclopedia (2021), [Online; accessed 30-June-2021].