Search for heavy resonances in the $HH \rightarrow \gamma\gamma b\bar{b}$ channel

Aleksei Chernov, supervised by James Robinson

September 1^{st}, 2017

Abstract

Monte-Carlo simulations of unknown heavy particle $m_X \geq 260Gev$ undergoing double-Higgs decay: $X \rightarrow HH$, as well as SM double Higgs production were analysed. Two techniques were used - simple cuts on various variables, and Boosted Decision Trees (BDT) to separate signal from background. It was demonstrated that BDTs give much better results as far as signal efficiency/background rejection is concerned. This analysis was later applied to a sample of real data taken from LHC collisions.

Contents

channel	BR	decay width [MeV]
$H \to b\bar{b}$	58.2%	2.38
$H \to WW$	21.4%	0.874
$H \to gg$	8.2%	0.335
$H \to \tau\bar{\tau}$	6.3%	0.256
$H \to c\bar{c}$	2.9%	0.118
$H \to ZZ$	2.6%	0.107
$H \to \gamma\gamma$	0.23%	$9.28 \cdot 10^{-3}$
other	$\leq 1\%$	

Table 1: Branching ratios for Higgs boson of the mass $m_H = 125\,GeV$ [4].

1 Introduction

In 2012, ATLAS and CMS experiments at LHC, CERN announced discovery of a particle consistent with Standard Model Higgs boson [1]. It has a mass around $m_H = 125\,GeV$, with intrinsic width of a few 10 MeV, which is not accessible at LHC - the visible width of Higgs resonance is due to limitations imposed by detector resolution. It's the only scalar ($J = 0$) particle observed up to date. [2] However, not all properties of a newest observed particle had been identified - in particular, Higgs self-coupling λ_{HHH} is of interest, considering that cubic and quartic terms play significant role in Higgs mechanism of electroweak symmetry breaking.

1.1 Single Higgs production and observation

Four dominant modes for Higgs production at the energies accessible to the LHC are:

- Gluon fusion $gg \to H$
- Vector boson fusion $qq' \to qq'H$
- Higgs-strahlung $qq' \to HW$
- Associated top production $qq' \to t\bar{t}H$, $gg \to t\bar{t}H$

Higgs couples to heavy particles preferentially; in $\approx 60\%$ of cases, it decays into a pair of b-quarks (top is not accessible kinematically). Other decay channels include $WW*$ and $ZZ*$, where one of the bosons is produced off-shell due to kinematics, $H \to l^- l^+$ where τ lepton dominates, since fermionic coupling to Higgs is proportional to the mass. Initial discovery of Higgs by ATLAS and CMS collaborations in 2012 was made in a rare $H \to \gamma\gamma$ channel. As Higgs doesn't possess an electric charge, it cannot directly couple to photons - even at leading order this has to proceed via top quark loop. Fig. 1 shows single H boson branching ratios as a function of the mass. One immediately can notice that $\gamma\gamma$ BR is very small compared to $b\bar{b}$, or WW. However, it's very clean. Table 1 shows values of H branching ratios for $m_H = 125\,GeV$.

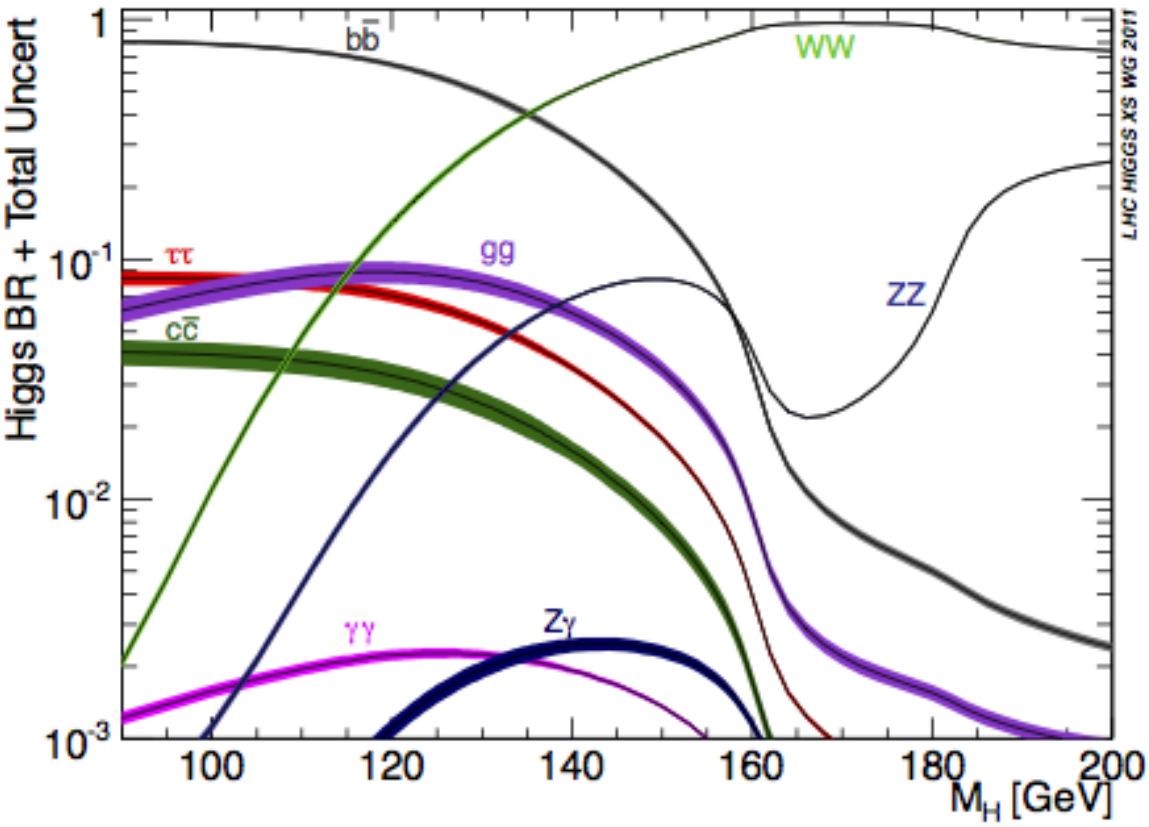


Figure 1: Higgs SM cross-sections, from ATLAS website

$\sqrt{s},\ TeV$	$\sigma_{gg\to HH},\ fb$	$\sigma_{qq'\to qq'HH},\ fb$	$\sigma_{qq'\to WHH},\ fb$	$\sigma_{qq\to ZHH},\ fb$	$\sigma_{qq\,gg\to HH\bar{t}t},\ fb$
8	8.16	0.49	0.21	0.14	0.21
14	33.89	2.01	0.57	0.42	1.02
$\sqrt{s},\ TeV$	$\sigma_{gg\to H},\ pb$	$\sigma_{qq'\to qq'H},\ fb$	$\sigma_{qq'\to WH},\ fb$	$\sigma_{qq\to ZH},\ fb$	$\sigma_{qq\,gg\to H\bar{t}t},\ fb$
8	21.42	702.50	420.70	1650	133.0
13	48.57	1373	883.70	3925	507.1

Table 2: Comparison between H and HH production cross-sections in the Standard Model [3].

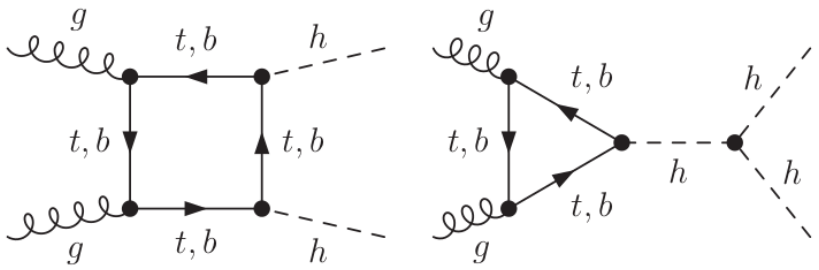


Figure 2: Production of Higgs pair via gluon fusion at the leading order. Box and triangle diagrams interfere destructively, resulting in a very low cross-section.

1.2 Di-Higgs production and observation

Just like single Higgs production, di-Higgs is dominated by a gluon fusion which is at least an order of magnitude bigger than VBF, Di-Higgstrahlung or 2-Higgs production in association with top quarks. However, cross-section for 2 Higgs production is significantly lower, with numbers of events around 50 per $3000 fb^{-1}$ integrated luminosity [3].

Table 2 shows comparison between single Higgs and double Higgs production within the Standard Model. As one can see, there is a factor of $10^{-3} - 10^{-4}$ between the two. HH is very rare - one event per several thousands Higgs produced; if there is any New Physics significantly affecting HH production, such as existance of heavy ($m_X \geq 260\ GeV$) resonances coupled to Higgs and thus undergoing $X \to HH$, it should be observed as deviation from predicted very low number of events.

We assume that, after a pair of Higgs bosons is produced, each boson decays according to measured branching ratios independently. Thus, requiring both Higgses to undertake $H \to \gamma\gamma$ decay would suppress an already rare process by an additional factor of $(2.27 \cdot 10^{-3})^2 \approx 5.15 \cdot 10^{-6}$. If one wishes to use methods developed in $H \to \gamma\gamma$ searches, but avoid suppressing signal too much, it is necessary to consider another boson decaying with a high BR. The highest BR, as Table 1 and Fig.1 show, is $H \to b\bar{b}$. However, that introduces a considerable background of single H production in association with 2 b-jets, or 2 jets wrongly tagged as b-jets.

2 Analysis

2.1 Initial Data

In this project, I worked with Monte-Carlo (MC) simulated data sets for SM background, SM di-Higgs decays, and templates for massive resonances at 260, 275, 300, 325, 350, 400, 450, 500, 750 and 1000 GeV. In the last part of the project, I applied analysis we developed to a sample of real data from LHC Run 2. In all samples an initial cut was applied to jets: $p^T \geq 20\,GeV$ to get rid of soft QCD background.

2.2 Event Selection

Events in the initial ntuples have the following variables:

- event weight - used in MC samples; set to 1 for all events in real data
- jet_n - number of jets in the event
- photon_n - number of photons in the event
- $m_{\gamma\gamma}$ - reconstructed invariant mass of two photons with highest transverse momenta p^T
- E_{jet} - energy of each jet
- η_{jet} - pseudorapidity of each jet
- ϕ_{jet} - ϕ angular component for each jet
- p^T_{jet} - transverse momenta of jets
- Jet_from_Higgs - Monete-Carlo only; boolean flag that identifies whether jet is final state of a Higgs boson or not
- jet_MV2c10 - variable related to b-tagging
- E_γ - energy of each photon
- η_γ - pseudorapidity of each photon
- ϕ_γ - angular component of each photon
- p^T_γ - transverse momenta of photons

Each sample came with the following cuts applied:

- $jet\_n \geq 2$
- $photon\_n \geq 2$

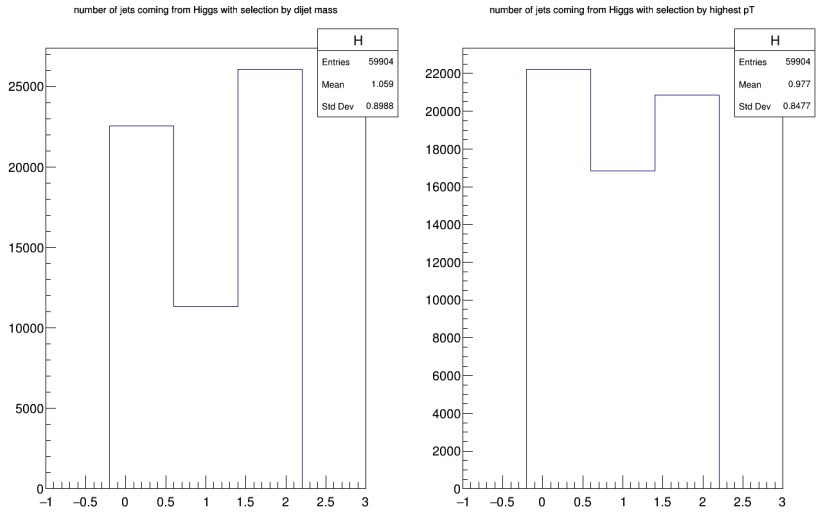


Figure 3: Comparison between selection by highest p^T and by reconstructed mass for $m_X = 300\,GeV$.

- $p^T_{jet} \geq 20\,GeV$ for all jets

Events that do not pass those cuts have zero probability to be signal. Then, for events that have three or more jets, we need to select a pair to be Higgs candidate. Two options are possible:
- Select two jets with the highest tranverse momenta p^T
- Reconstruct invariant mass for each of the jets; select one that is closest to $m_H = 125\,GeV$. Both of the approaches were tested, using "Jet_from_Higgs" flag for MC simulated data. Figs. 3-7 show the results. The left bin is number of events with 0 selected jets from Higgs decay, the middle bin is number of events with 1 selected jet from Higgs and the other not, and the right bin is amount of events with both jets correctly identified as coming from Higgs.
As one can see, for all samples selection by invariant mass produces better results than by highest p^T, and therefore this is the one subsequently used in analysis. For photons, we select two with the highest p^T as Higgs' final state candidates.
We expect to see signal (if there is any) as a clear peak in $m_{\gamma\gamma}$ - reconstructed mass of two photons, similar to single Higgs production. As Fig.8 shows, we're looking for a very weak signal on top of large background; our goal is to create a classification algorithm that will improve signal significance, defined as:

$$Significance = \frac{S}{\sqrt{B}}$$

while retaining high signal efficiency. We started with cut based analysis on our input variables. Figure 9 shows plots of signal efficiency and signal significance (as defined above) against the low cut on highest p^T (of 2 selected jets) for lightest $m_X = 260\,GeV$

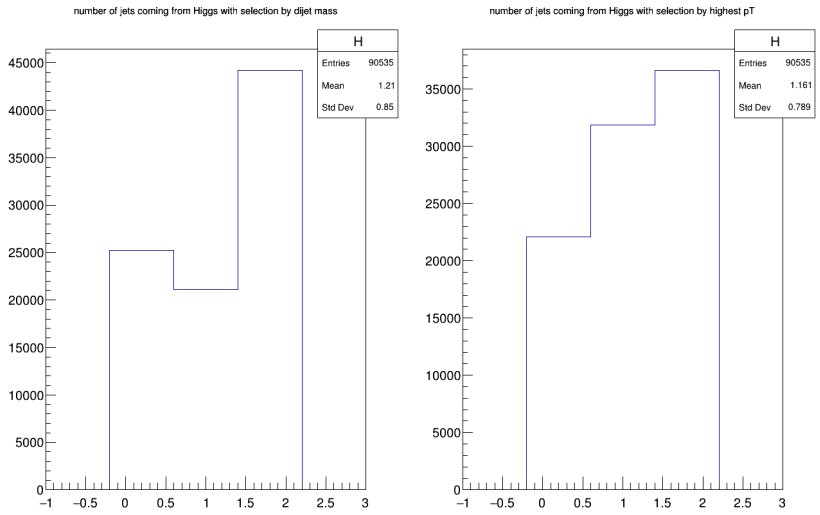


Figure 4: Comparison between p^T and invariant mass selection for $m_X = 500\,GeV$ template

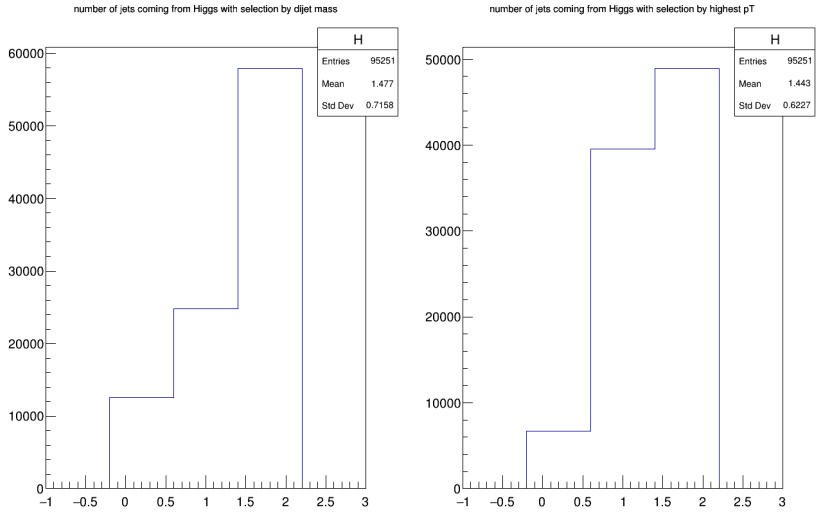


Figure 5: Comparison between p^T and invariant mass selection for $m_X = 750\,GeV$

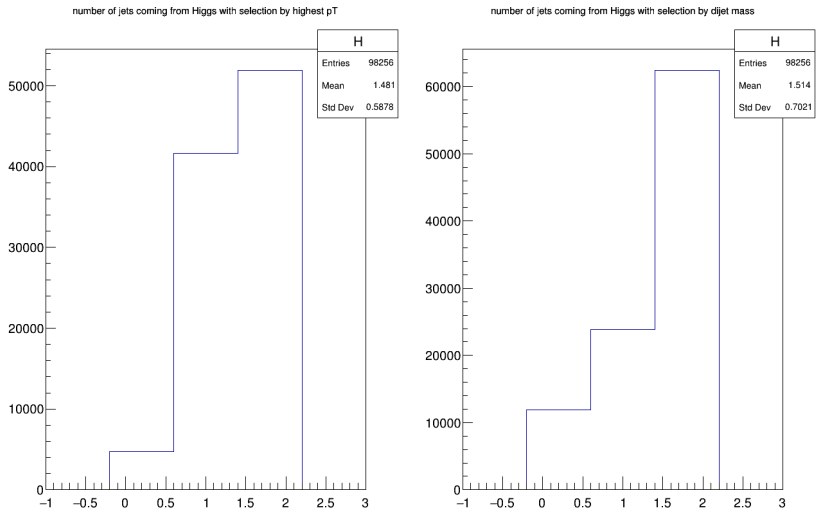


Figure 6: Comparison between p^T and invariant mass selection for $m_X = 1000\,GeV$

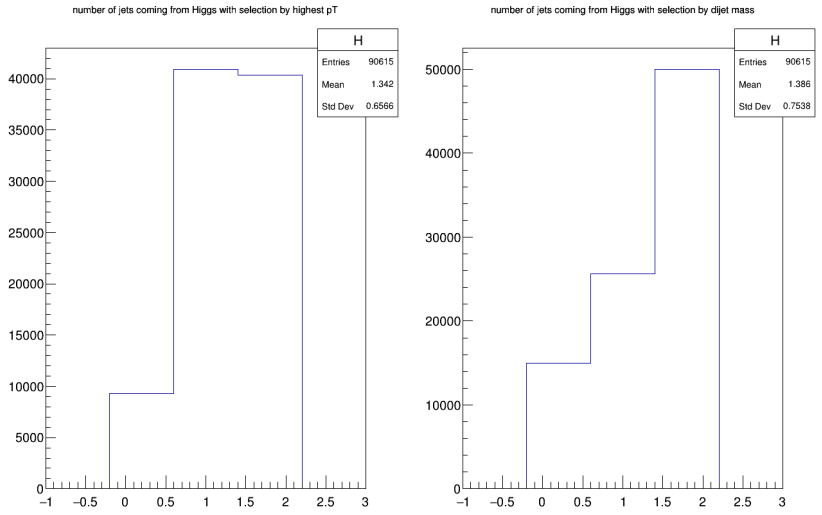


Figure 7: Comparison between p^T and invariant mass selection for Standard Model HH sample.

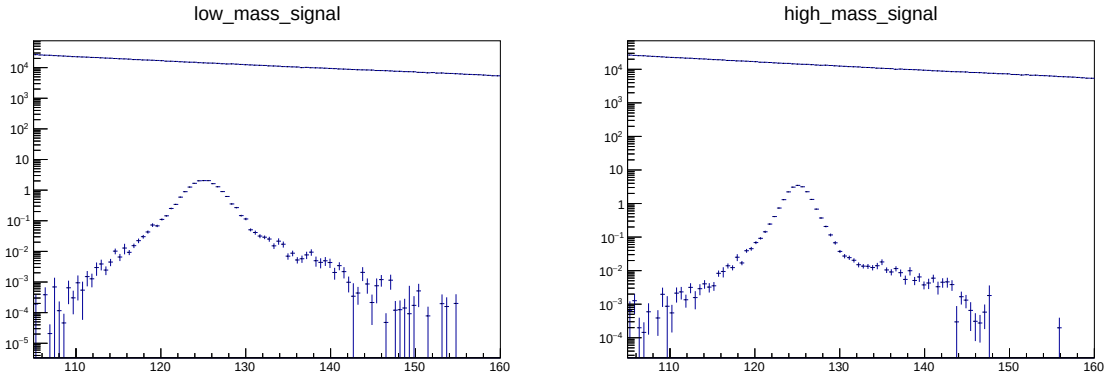


Figure 8: Spectrum of $m_{\gamma\gamma}$. Continious line at the top is SM background; peak around $m_H = 125$ is from decay of $m_X = 300$ heavy resonance template on the left and from $m_X = 1000$ on the right.

sample. Fig.10 shows the same plot for 2^{nd} jet, and Fig.11 shows the plot of reconstructed dijet mass (that is not among initial variables, but can be reconstructed by reading E, ϕ, η, p^T of the selected jets and combining their 4-vectors). Blue dashed line is drawn at 0.95 signal efficiency; signal significance is normalized to its' maximal value, that is in that case of the order of magnitude 10^{-3}–10^{-4}.

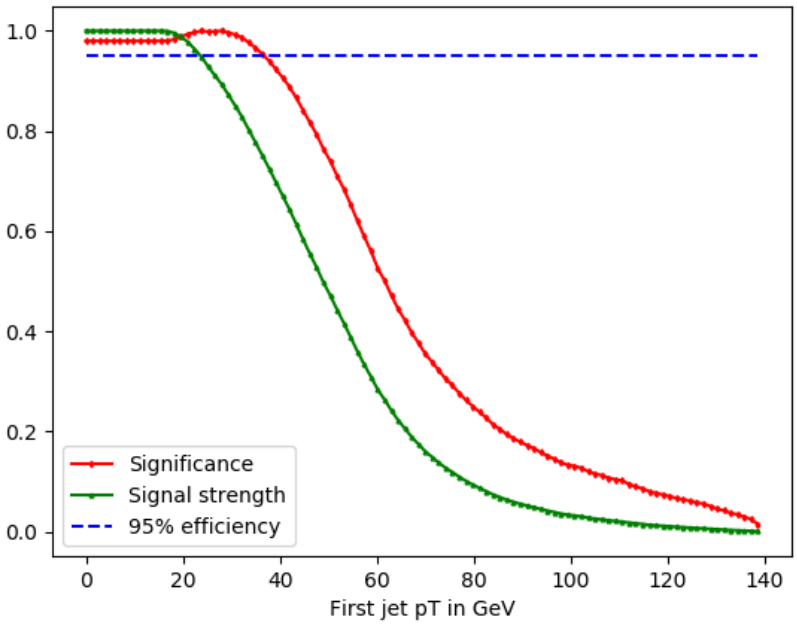


Figure 9: Signal significance and signal efficiency as functions of low cut on p_1^T

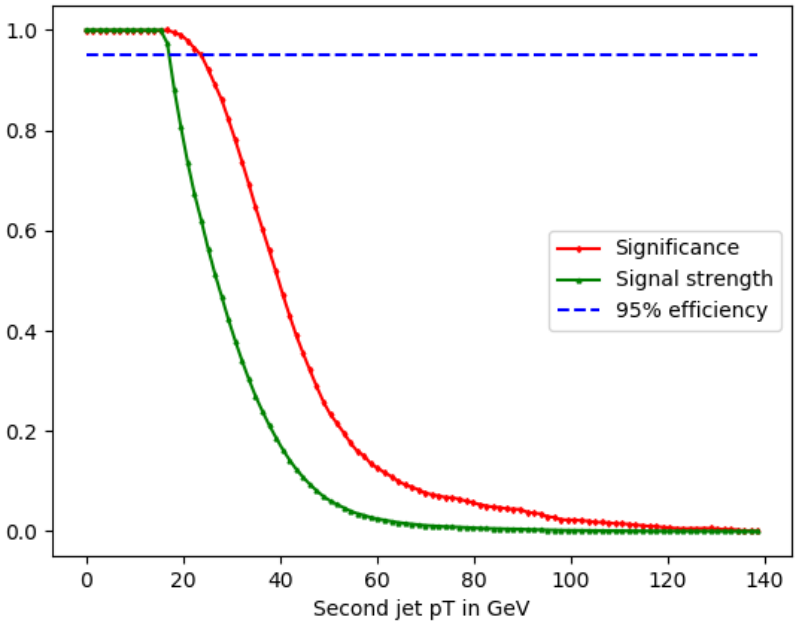


Figure 10: Signal significance and signal efficiency as functions of low cut on p_2^T

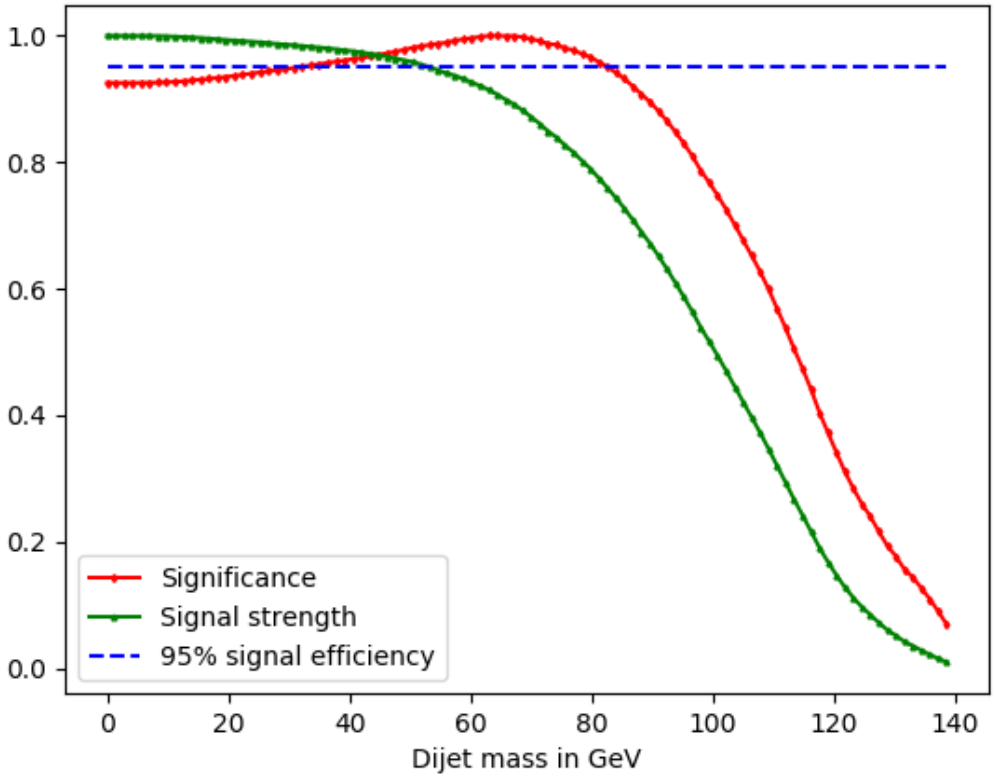


Figure 11: Signal significance and signal efficiency as function of low cut on dijet mass

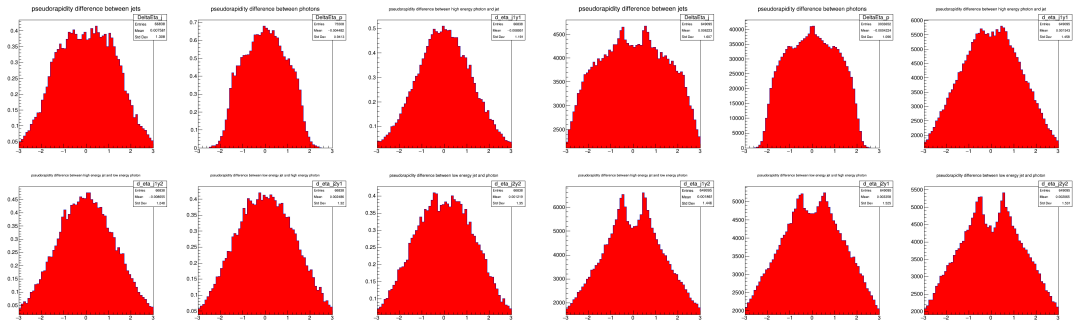


Figure 12: On the left-hand side - $m_X = 400$ signal in some of the 83 variables we constructed; on the right-hand side is the background in the same variables.

As one can see, any cut on p_2^T beyond the initial one at 20GeV decreases significance. For p_1^T, moving cut from 20 GeV to 25 GeV results in small ($\approx 7\%$) gain in significance while retaining 95% of the signal. For dijet mass, discarding events with $m_{jj} \leq 50\,GeV$ also improves significance while retaining most of the signal. Those values are different for higher mass samples, but we need to be as generous as possible to avoid losing any possible signal. To summarize, following simple cuts were placed during analysis:

- $n_{jets} \geq 2$
- $n_\gamma \geq 2$
- $p_1^T \geq 25\,GeV$
- $p_2^T \geq 20\,GeV$
- $m_{jj} \geq 50\,GeV$

2.3 Multivariate techniques

In the next part of this project, we turned to Multivariate Analysis (MVA) in order to obtain better separation between signal and background events. First of all, we took initial data (as described in 2.1) and created 83 different variables by combining 4-vectors and taking their various components. Figs. 12-14 shows how $m_X = 400GeV$ looks in some of those variables compared to background (right-hand side) - if we have different shapes for signal and background events, we can hope to distinguish between them.
We used TMVA[5] toolkit in order to implement our multivariate analysis. Boosted Decision Trees (BDT) are fairly simple and robust algorithm, that makes it a popular choice for data analysis in HEP. We also tried to implement a Multilayered Perceptron classifier, but weren't successful with it.

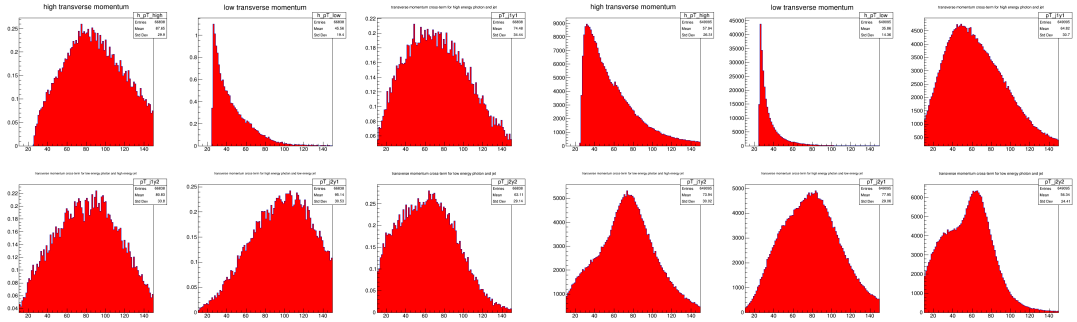


Figure 13: LHS - $m_X = 400GeV$ sample; RHS - SM background

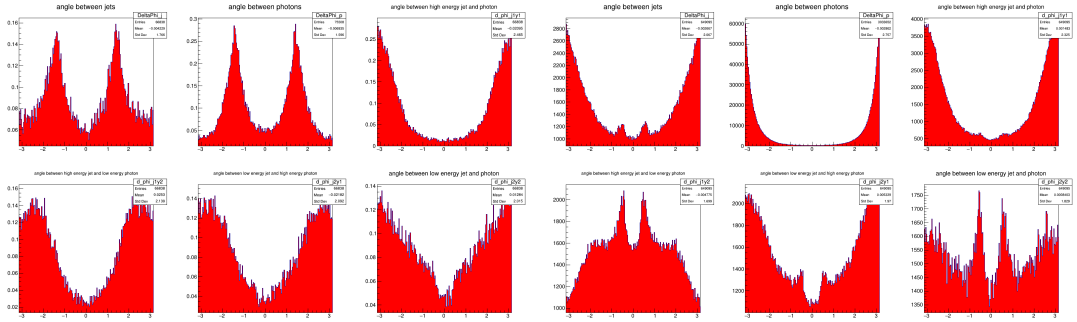


Figure 14: LHS - $m_X = 400GeV$ sample, RHS - SM background

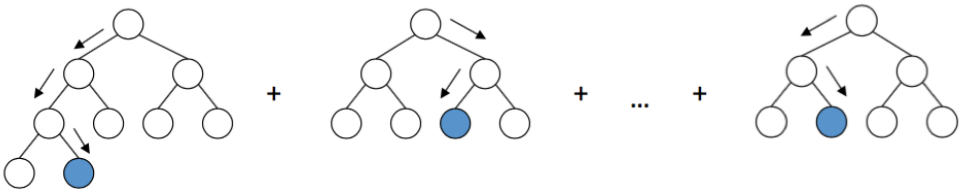


Figure 15: Schematic of a boosted decision tree, courtesy A. Rogozhnikov on github

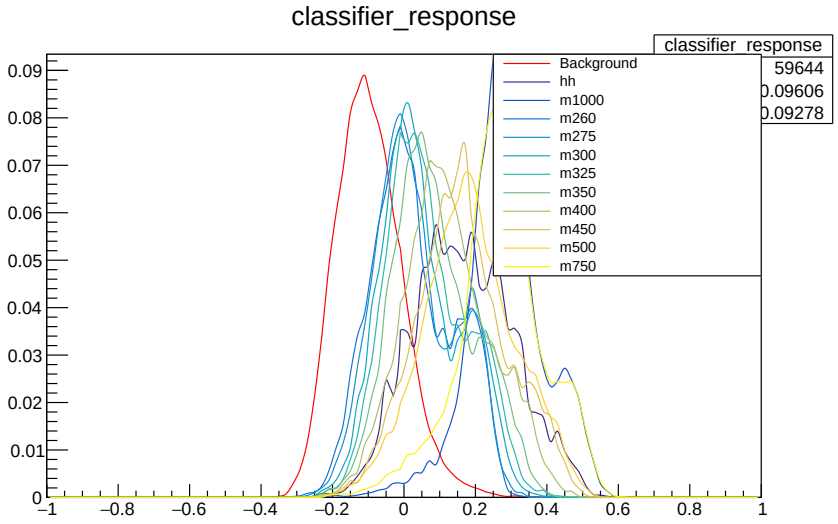


Figure 16: Classifier response for a single BDT trained on all samples simultaneously

Fig. 15 demonstrates a very simple schematic of a BDT. In our analysis, we used $N = 800$ decision trees with $MaxDepth = 3$, i.e no more than three forks on each individual tree. Each individual decision tree attempts to classify whether a given event is signal-like or background-like, resulting in a score of -1 or +1. In the end, scores from all trees are added and normalized to N, therefore an output of a BDT classifier is a real number in range from -1 to 1.

Like other MVA algorithms, BDT needs to be trained, then tested - evaluated on a known datset that is not part of the training. Each of our simulated data samples was split into two, according to a very simple algorithm: read the input TTree event-by-event basis, do a virtual coin-flip using Python's numpy library random function, if a result is 0, assign this event to a training tree; if a result is 1, assign this event to a testing tree. For training, we fed all training samples into BDT simultaneously - since we want to be able to see all possible signals. Then, it was applied to each of the testing samples simultaneously; results are presented at Fig.16. As one can see, the separation between signal and background is not very good. Figure 17 shows so-called ROC (Receiver

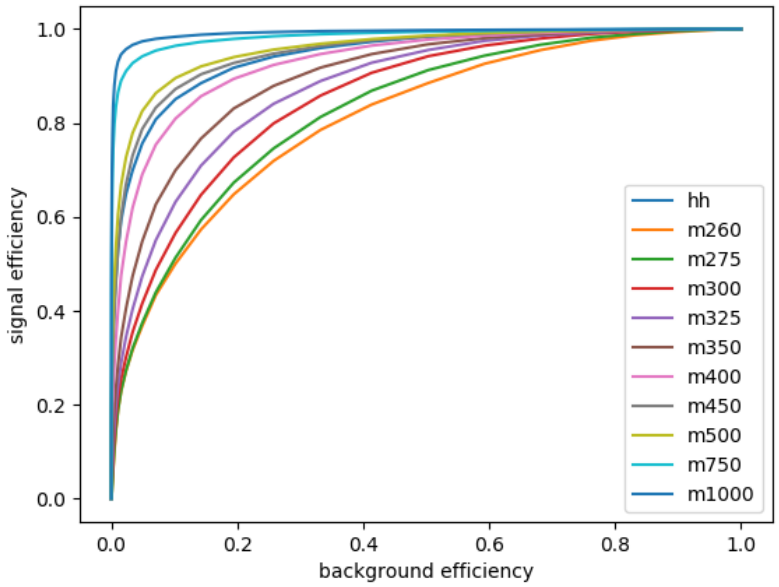


Figure 17: ROC for a single classifier trained on all training events

Operator Curves) for each of the samples - a common metric used to evaluate how good is a given classification. In order to improve our resolution, we decided to split our data further into two categories. One included Standard Model HH sample and heavy ($m_X \geq 450$) resonances, another included light ($260 \leq m_X \leq 400$) resonances. Two separate BDTs were created; one for each cathegory. Each of the BDTs was then tested on corresponding testing samples, no cross-category tests were performed (i.e we have not measured how well classifier for light resonances separates heavy resonances and SM HH from background, and vice versa). Figure 18 shows classifier response for each of those; a considerable improvement over a single classifier.
By comparing Fig.19 to Fig.17, one can see that two classifiers result in better background rejection at high singal efficiency; especially in the problematic (light resonances) region where we could not achieve a good separation. Now, instead of placing simple one-dimensional cuts on the input variables, we can place a cut on the classifier: all events with classifier score greater than this value are considered signal-like and all events with classifier score lower than this value are considered background-like. Figs. 20-21 shows significance and signal efficiency depending on the value of classifier score cut; compared to Figs.9-11, it demonstrates a great improvement.

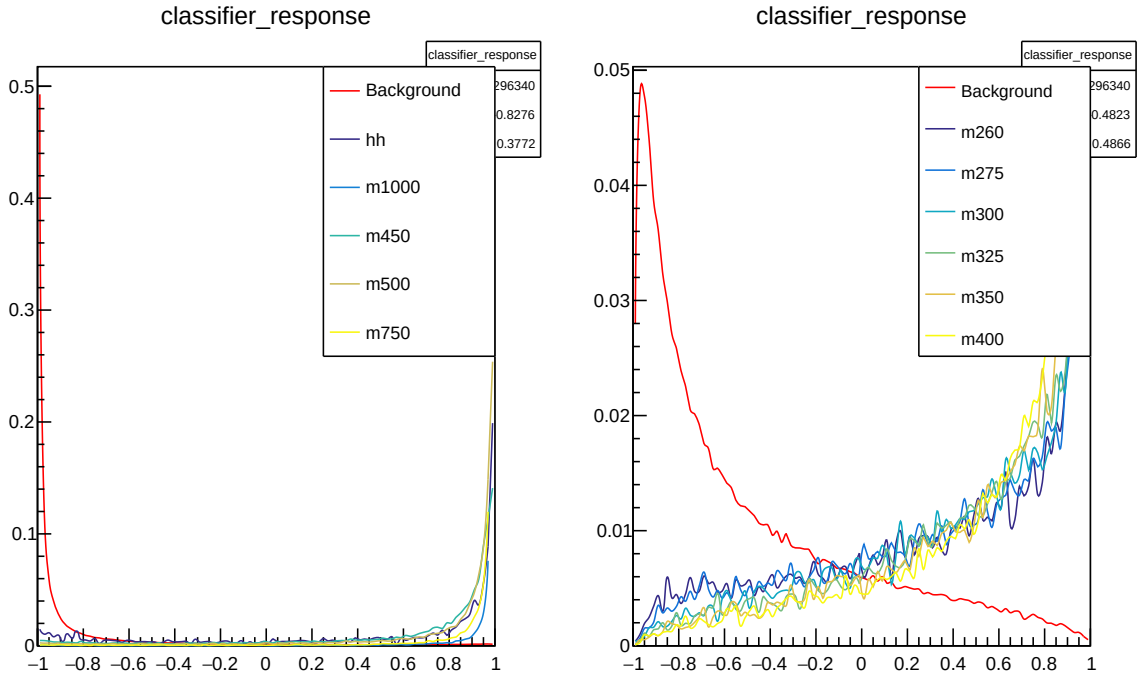


Figure 18: Two classifiers response for MC testing samples; LHS - high mass + SM classifier, RHS - low mass classifier

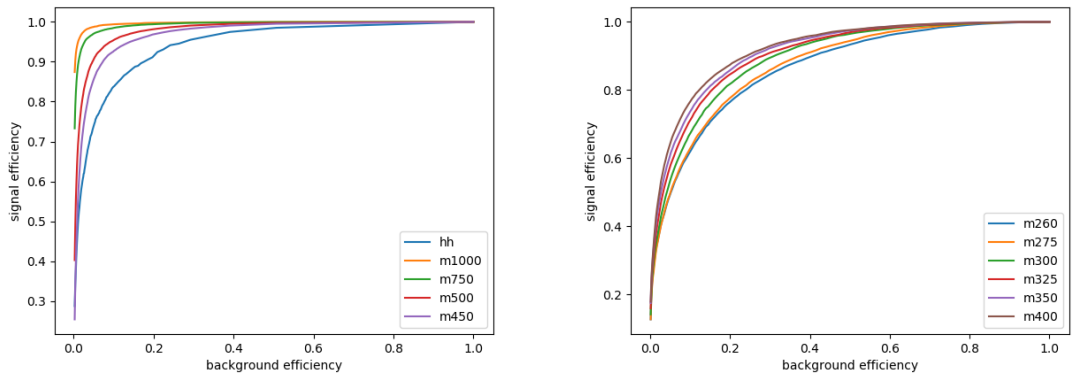
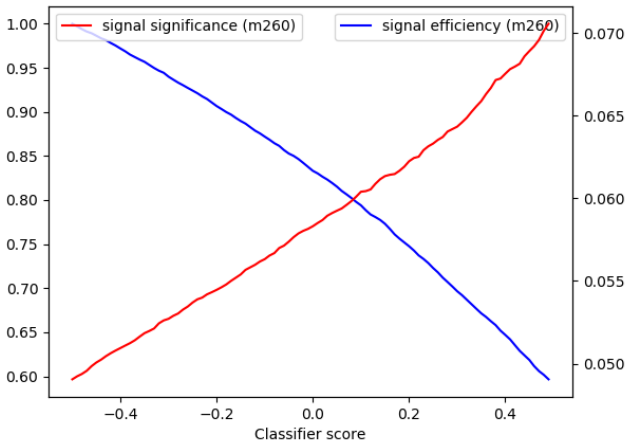
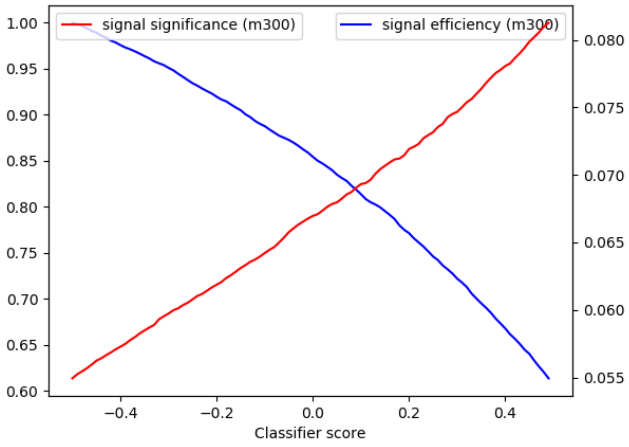


Figure 19: LHS - ROC for high mass classifier, RHS - ROC for low mass classifier





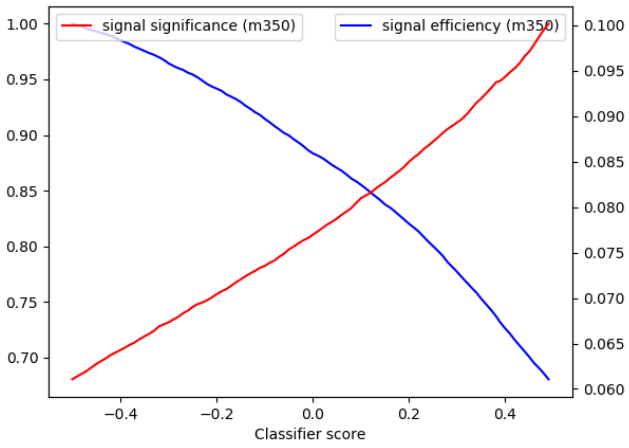
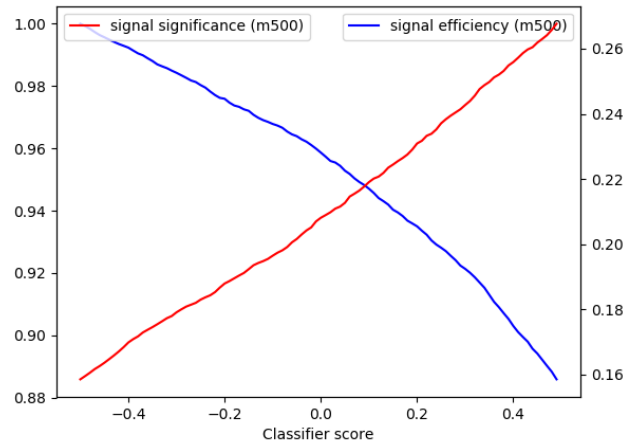
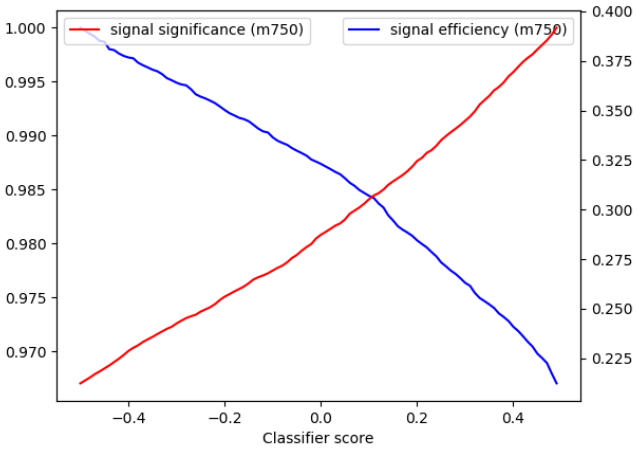


Figure 20: Signal significance and efficiency curves for $m_X = 260, 300, 350$





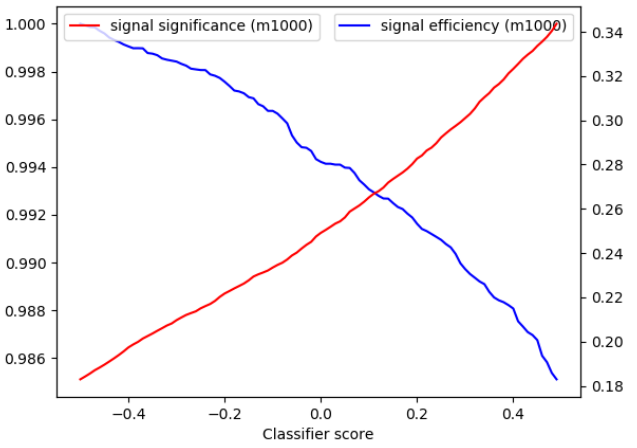


Figure 21: Signal significance and efficiency curves for $m_X = 500, 750, 1000$

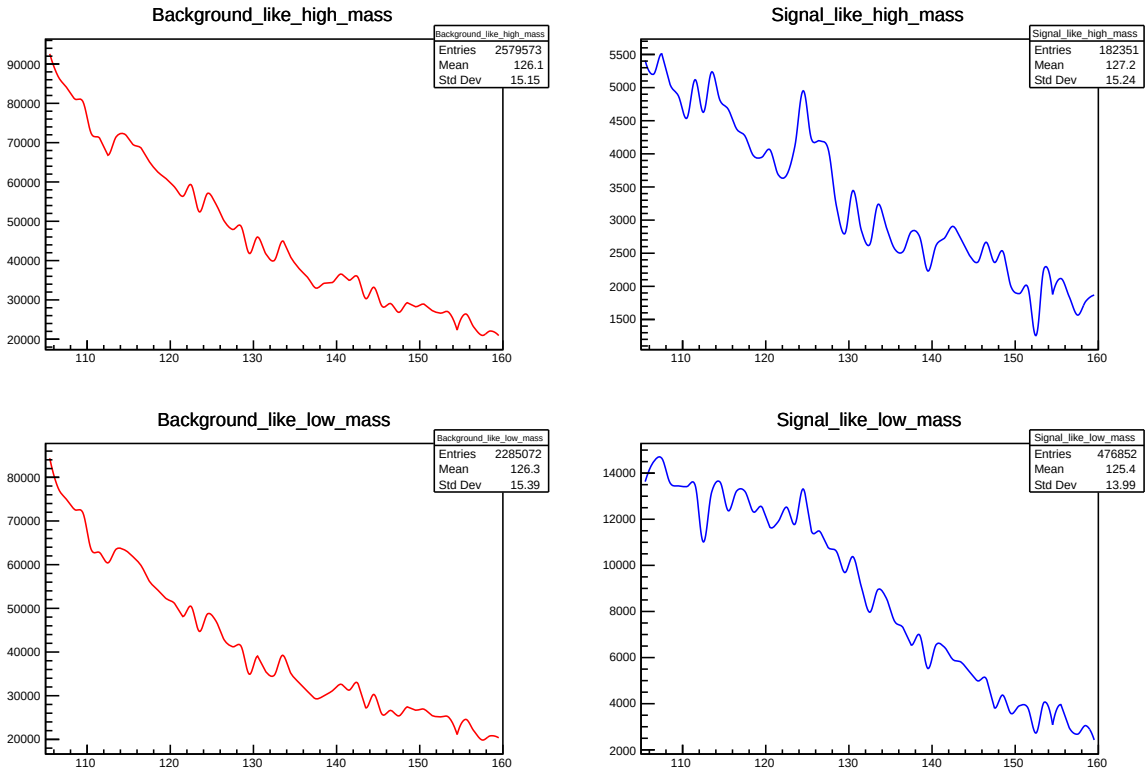


Figure 22: $m_{\gamma\gamma}$ of the real data sample after classification

3 Results

We applied classification algorithm described in Section 2 to a sample of real collisions data. Cut-off classifier score was set to be 0; i.e everything with positive classifier score was considered signal-like and everything with negative score was considered background-like. Fig.22 shows $m_{\gamma\gamma}$ plots after classification; background-like is in red and signal-like is in blue.
We modelled our SM background as an exponential fit to those functions, Fig.23 shows results after the fitting. Fits seems to be not so good for the low-mass signal-like events, but for the high-mass signal-like events, we do see a very clear peak around $m_H = 125\,GeV$.

3.1 Discussion

- Do we really see a signal? Hard to tell. We're considering $b\bar{b}\gamma\gamma$ final states; but they're not exclusive to the HH. A single Higgs produced in association with 2 jets - for example, in VBF, or $t\bar{t}H$ with tops decaying quickly into b-jets + W bosons, and possibly some other processes can end up with the same final state and pass our loose selection criteria. Our modelling of the background did not account for singe Higgs

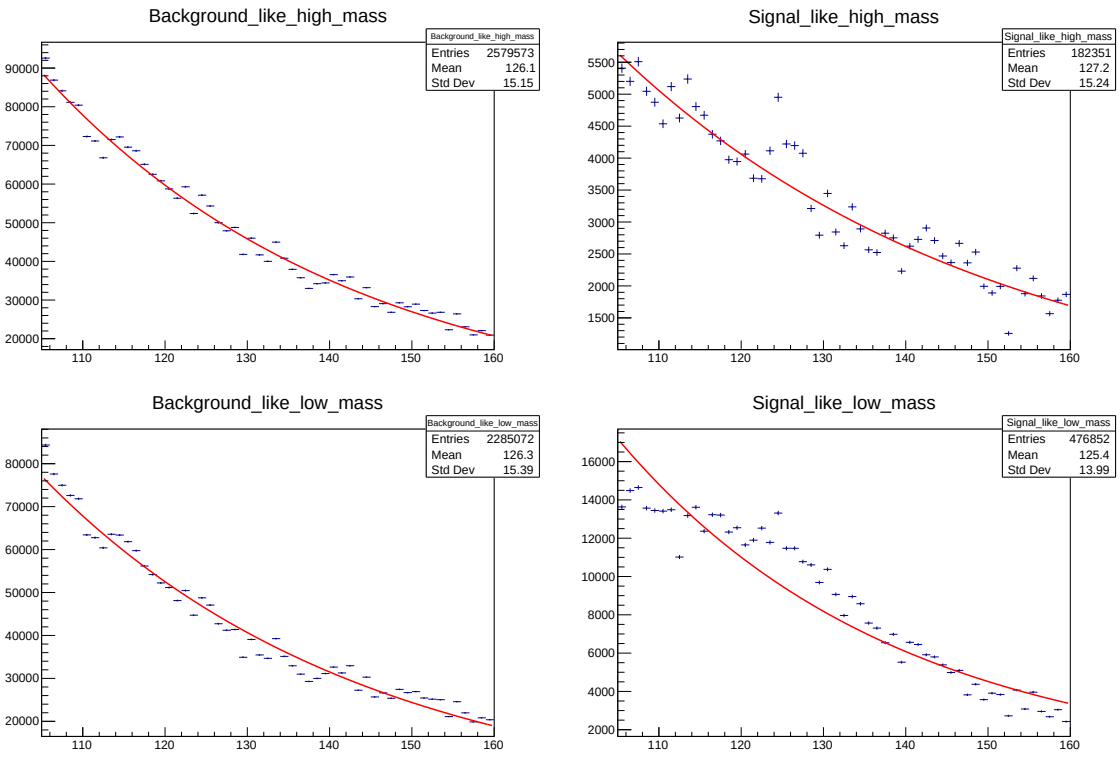


Figure 23: Exponential fits to $m_{\gamma\gamma}$

production; thus, it is irreducible in our current approach. A logical next step would be to deal with that: extend initial feature space to leptons, account for missing energy if there is any; come up with additional classifier focusing on jets to distinguish ones that are produced in Higgs' decay from other processes - which is not an easy task, that's why $H \to \gamma\gamma$ and $H \to W^+W^-$, as well as other states, are used in analysis despite lower branching ratios than $H \to b\bar{b}$.
In order to reach any conclusions and (possibly) place limits on $HH \to b\bar{b}\gamma\gamma$, one needs to evaluate this background from single Higgs production, using data from current Higgs studies and/or theoretical predictions for SM $H \to \gamma\gamma$ with at least 2 jets that can pass our criteria. Which is, unfortunately, beyond the scope of this summer project.

References

[1] The ATLAS Collaboration, Phys. Lett. B716 (2012) 1 and ATLAS-CONF-2012-162;
The CMS Collaboration, Phys. Lett. B716 (2012) 30 and CMS-PAS-HIG-12-045.

[2] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016) and 2017 update

[3] J.Baglio et al. KA-TP-44-2012 SFB/CPP-12-102 LPT-ORSAY-12-124 PSI-PR-12-10
The measurement of the Higgs self-coupling at the LHC: theoretical status

[4] D. de Florian, C. Grojean et al.
Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector
arXiv:1610.07922v2 [hep-ph] 15 May 2017

[5] https://root.cern.ch/tmva
The Toolkit for Multivariate Data Analysis with ROOT (TMVA)
TMVA User's Guide: http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf