Search for a heavy resonance decaying to a pair of Higgs bosons in 4tau final state (boosted)

Mentee: Kirstin Poppen (Gustavus Adolphus College)
Mentor: Abdollah Mohammadi (University of Wisconsin- Madison)

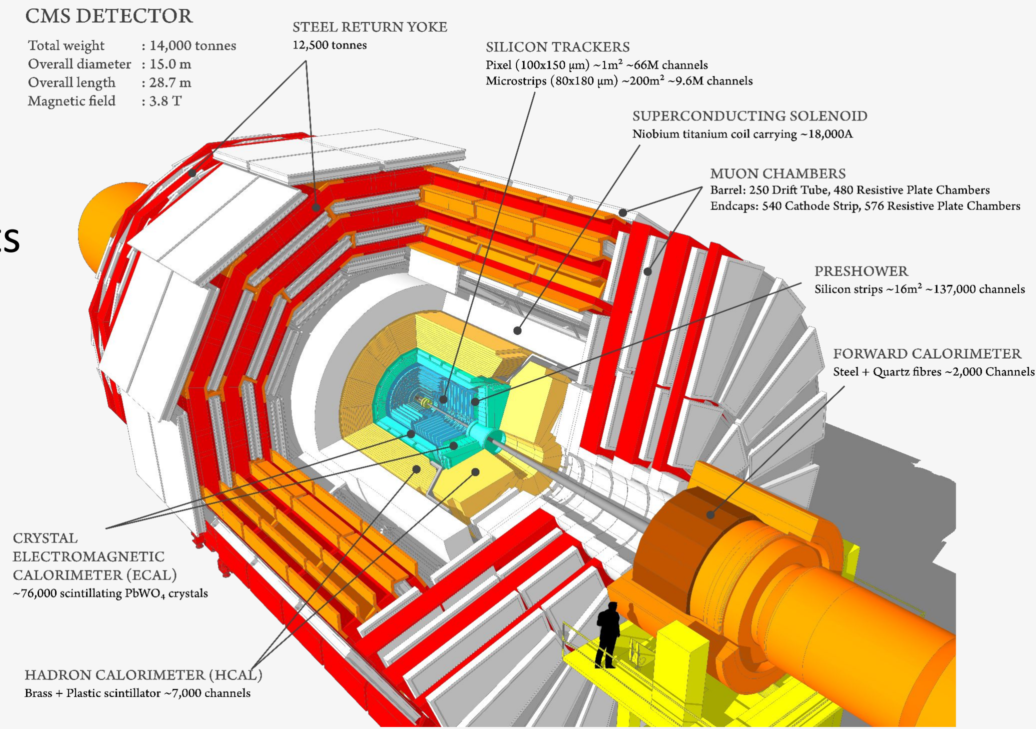
08/10/2023

Overview of My Work

1. Design an analysis to narrow down events consistent with a heavy resonance decaying a pair of boosted Higgs boson in 4tau final state.
2. Developed a binary neural network to better separate signal from major backgrounds
3. Ran the BNN on signal and backgrounds, evaluated how well the model performed
4. Introduced two high purity and low purity categories to better constrain the SM backgrounds
5. Ran the limit from Higgs Combine tools produce expected limit on cross section of the Signal

CMS Detector

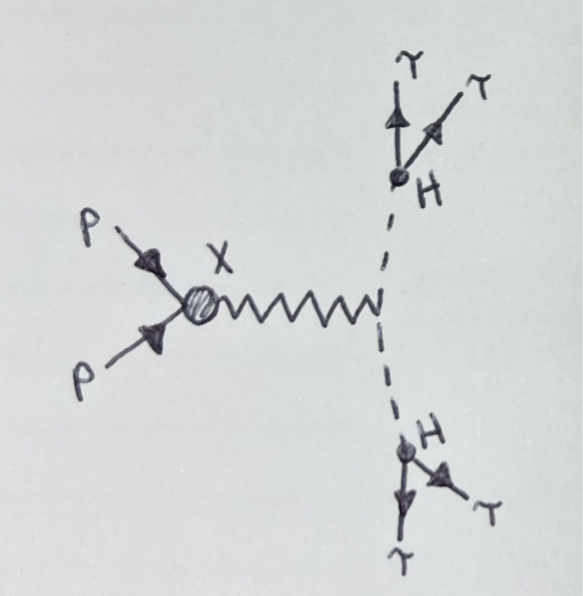
- Multipurpose detector
- 15 meters high, 21 meters long
- Collisions every 25 ns
- Trigger reduces the rate to abt 1 kHz in order to see potentially important events
- Records charged particles through the silicon tracker
- Measures energy of electrons and photons with the ECAL and measures energy of hadrons with the HCAL
- 3.8 T magnetic solenoid, bends charged particles to measure charge and momentum of a particle
- Designed to detect muons in outermost layer



https://cms.cern/detector

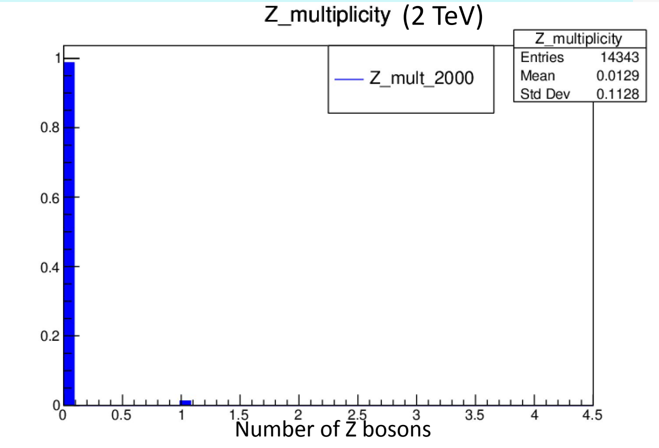
Physics Motivation/Background

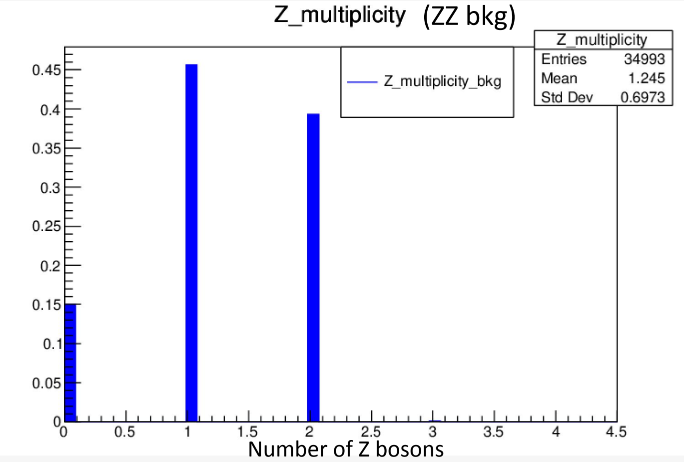
- Graviton/Radion (~TeV scale) are hypothetical particles that arise many Beyond standard models and might decay to SM Higgs bosons
- Higgs decay to SM particles
- Each Tau pair should decay from Higgs
- Two of the main backgrounds are ZZ to 4l and ZZ to 2l2q together with ttbar
- Higgs boson has an approximate mass of 125 GeV
- Z boson has an approximate mass of 90 GeV
- Previous research only looked for resonant particle with mass up to 1 TeV in 4 tau final state (HIG-21-002)
- This research will look in the range of 1-3 TeV



Event selection, Pairs formation

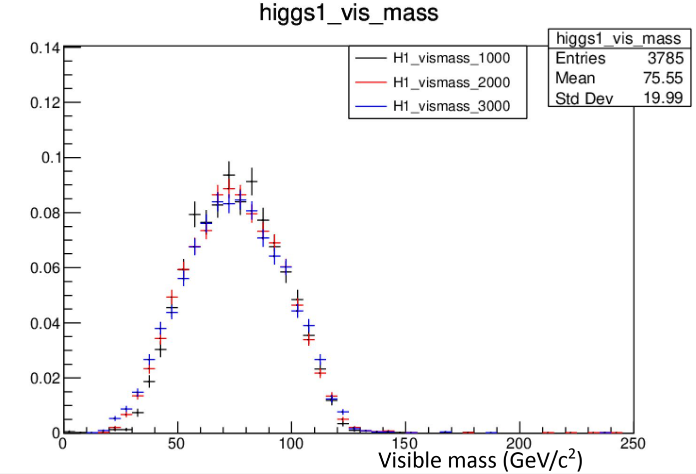
- Cuts were made in order to only select events that match the following criteria:
  - Event has 4 or more boosted Tau
  - Tau passes very loose isolation >0.5 (isolation ensures that the particles we want are correctly selected, and other particles i.e. jets are not)
  - Tau passes |eta| <2.3 and PT > 30 GeV cuts
- Checked the Z multiplicity, by counting number of Z to ee and Z to muon-muon pairs
  - Events with 1 or more Z bosons were cut
- When creating Tau pairs, matched Tau based on ΔR between two Tau

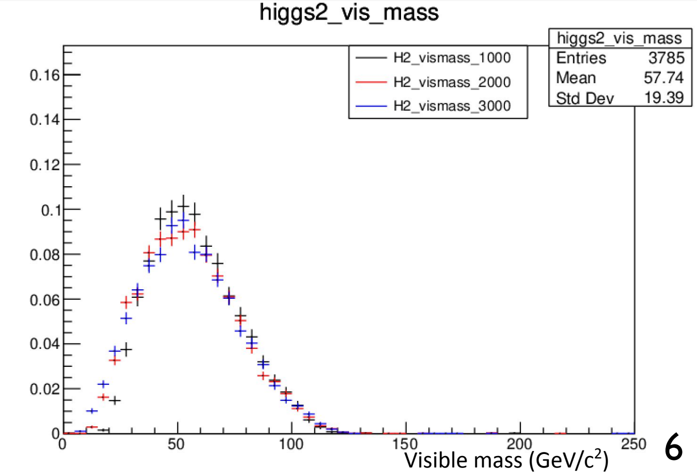




Understanding Features in Analysis

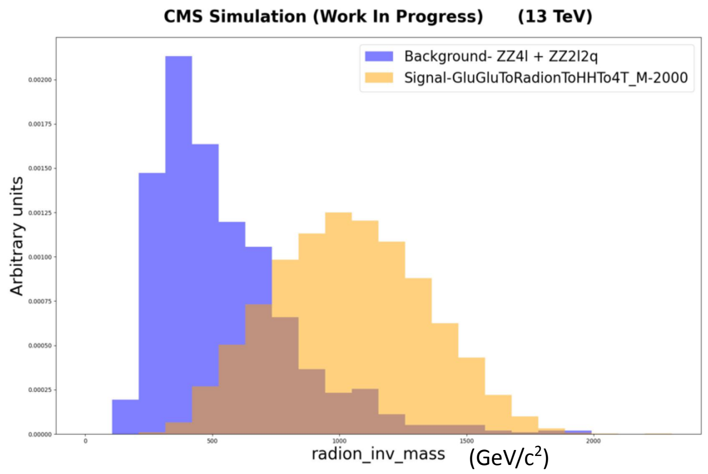
- First compared signal samples of different masses
  - 1 TeV, 2 TeV, 3TeV
- For example, created plots of the visible mass of each Higgs (mass of the Tau pair that make up each Higgs)
- The visible mass is less than 125 GeV because some mass is lost due to the existence of neutrinos
- Also created histogram plots to compare features from signal and ZZ background samples

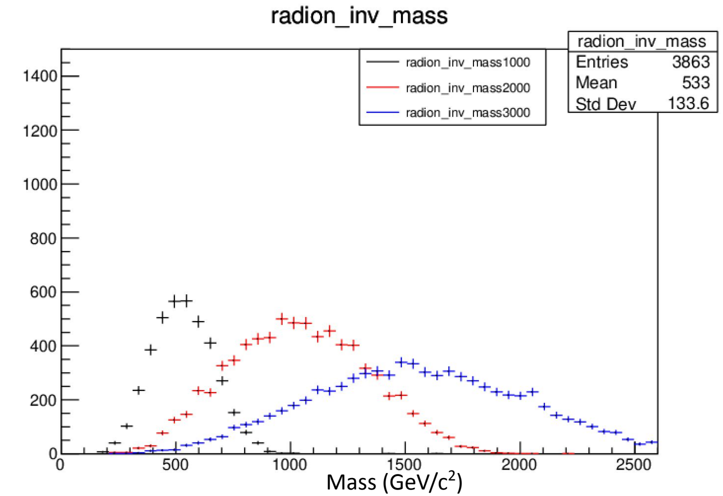




Observable Feature for BNN

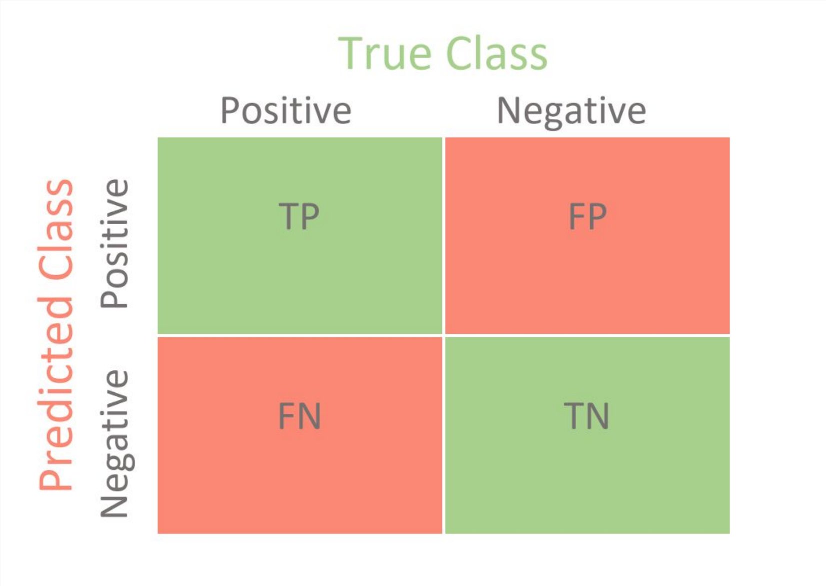
- Radion invariant mass is the mass of the two Higgs added together (4 tau mass)
- Radion invariant mass is excluded from the list of feature for the BNN
- The ZZ background peaks around 300-400 GeV while the 2 TeV signal peaks around 1000 GeV
- The invariant mass of the Radion in the 1, 2 and 3 TeV samples are ~500, 1000 and 1500 GeV/c^2
    - Half of the mass goes to other particles (neutrinos)





Binary Neural Networks

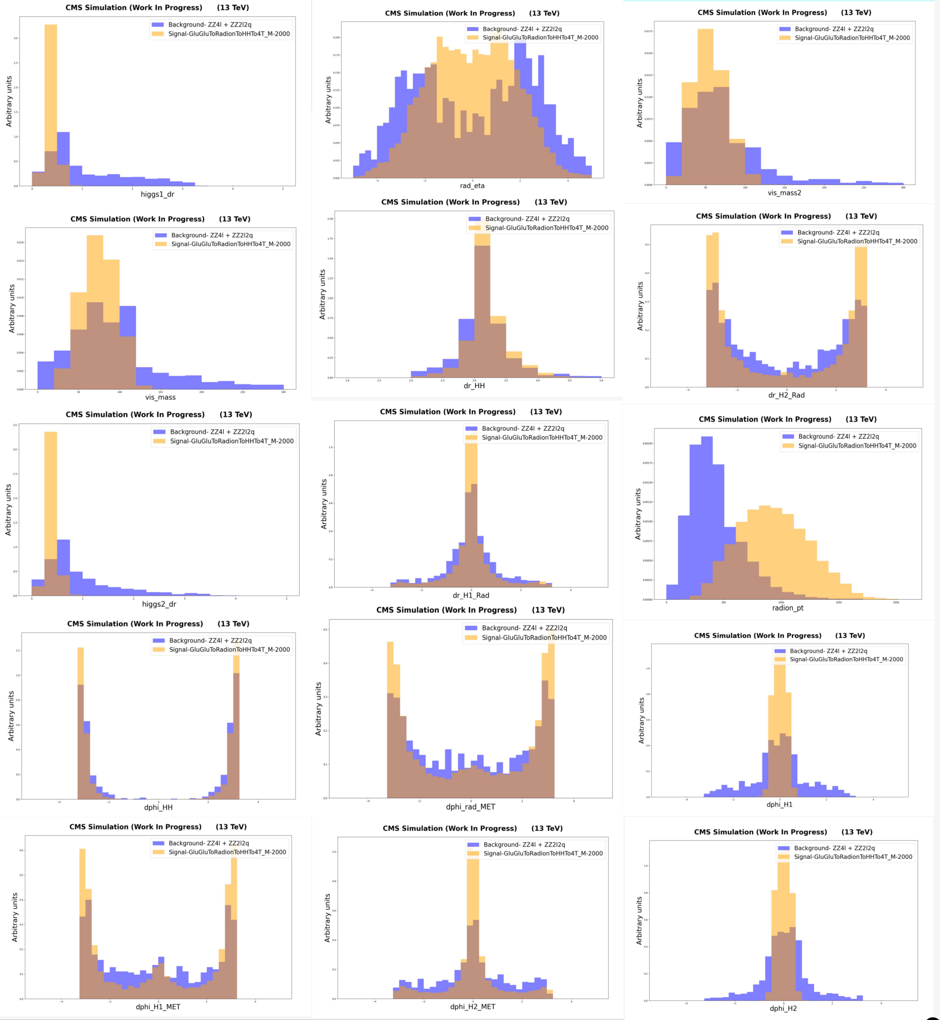
- Binary neural network assigns signal a probability close to 1, while close to 0 for backgrounds
- Training script: Takes the output root files from the analysis. Trains on signal and background.
- Evaluation script: Takes a sample as an input. Based on the training script, classifies the sample as signal-like or background-like and evaluates how well the training model performed
- BNN used in this analysis has 2 hidden layers
- Produces a confusion matrix to visualize the accuracy of the model



https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826

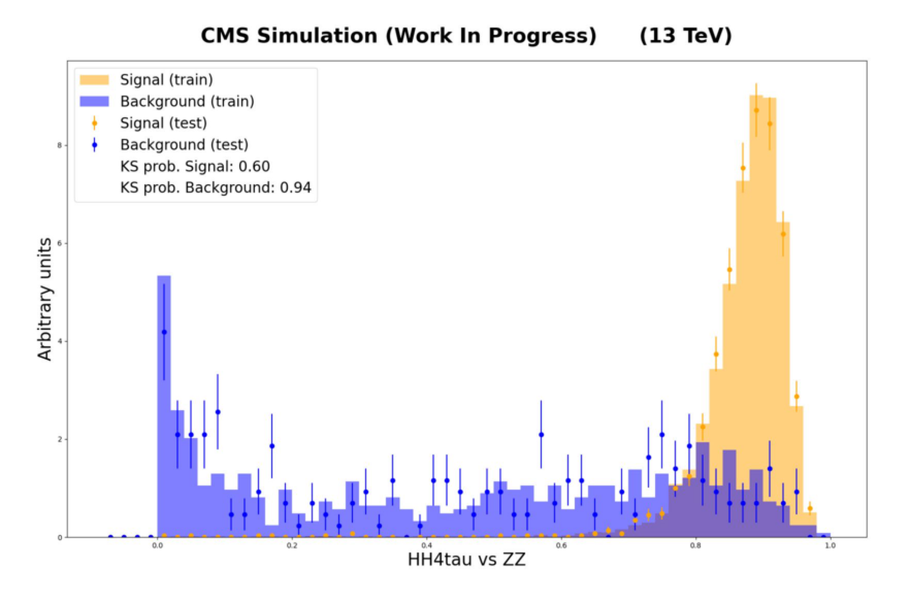
Features Used in BNN

- Radion PT
- Visible mass of each Higgs
- Radion eta
- ΔR between the two Higgs
- ΔR between each Higgs and radion
- Δφ between the two Higgs
- Δφ between radion and MET
- Δφ between each Higgs and MET
- Δφ of each Tau pair in each Higgs
- Δφ between each Higgs and MET



Neural Network Results

- Trained on 2 TeV signal and ZZ background
- The features from the previous slide were used as Machine Learning inputs
- Test loss: 0.2682
- Test accuracy: 0.9344
- Produced a KS binary classification plot
  - Above 0.7 is high purity (signal)
  - Below 0.7 is low purity (background)



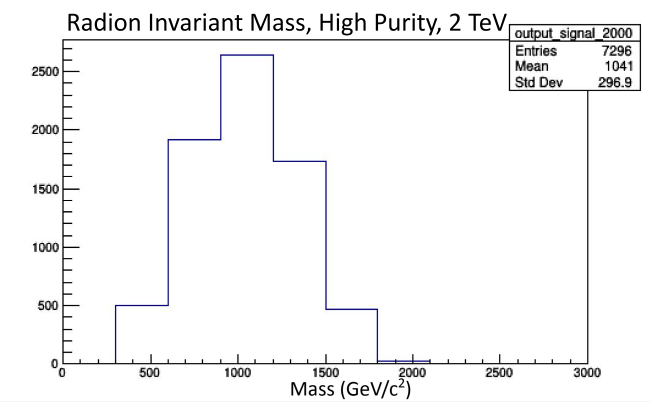
Evaluation Results

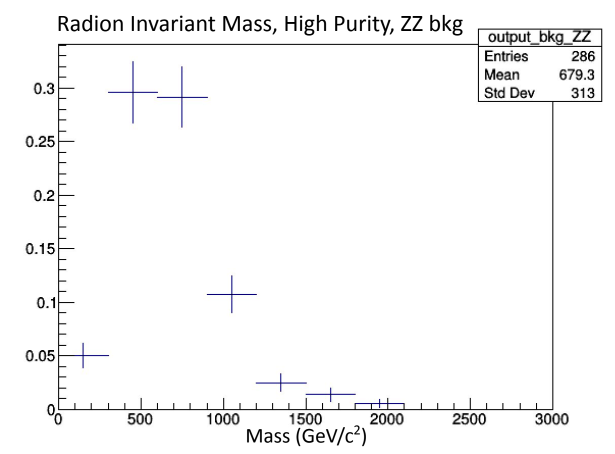
- After training on 2 TeV and ZZ, evaluated all samples and classified them as signal or background:
  - Signal: GluGluToRadionToHHTo4T (1, 2, 3 TeV samples)
  - DYJetsToLL
  - ZZ2l2q, ZZ4l
  - WJetsToLNu_HT
  - VV2l2nu
  - TTToSemiLeptonic, TTToHadronic, TTTo2L2Nu
  - T-tchan
  - Tbar-tW
  - WZ1l1nu2q, WZ1l3nu, WZ2l2q WZ3l1nu
  - QCD_HT (excluded, not enough events to evaluate)
  - T-tW (excluded, not enough events to evaluate)
  - Tbar-tchan (excluded, not enough events to evaluate)
- Evaluation results included a TTree with branches for the observable (4 tau mass) and the NN output

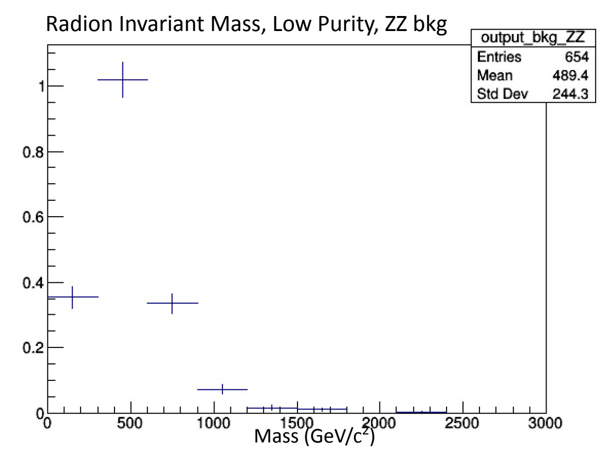
Samples from : (SkimBoostedHH4t/2018/4t/v2_Hadd/)

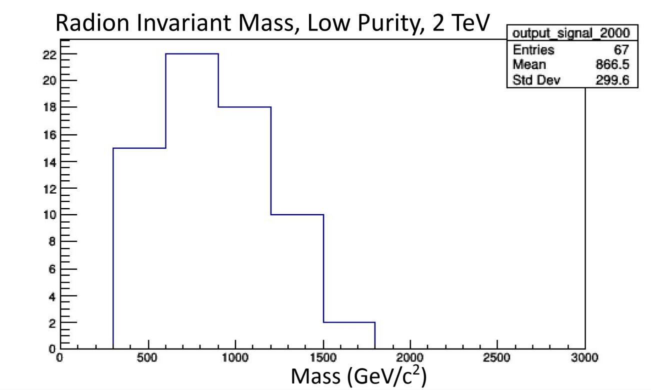
Output of BNN

- From the output of the BNN classification, created two categories
  - Low purity (dominated by background)
  - High purity (dominated by signal)



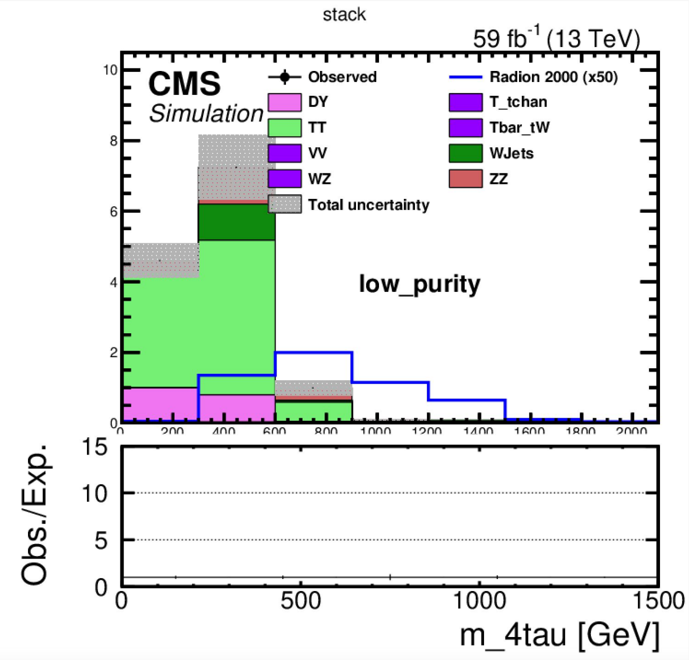


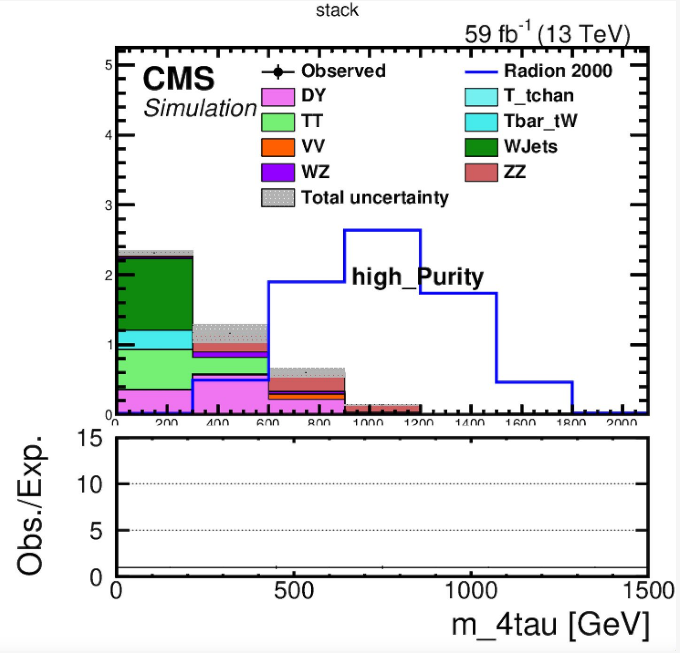




Final Observables

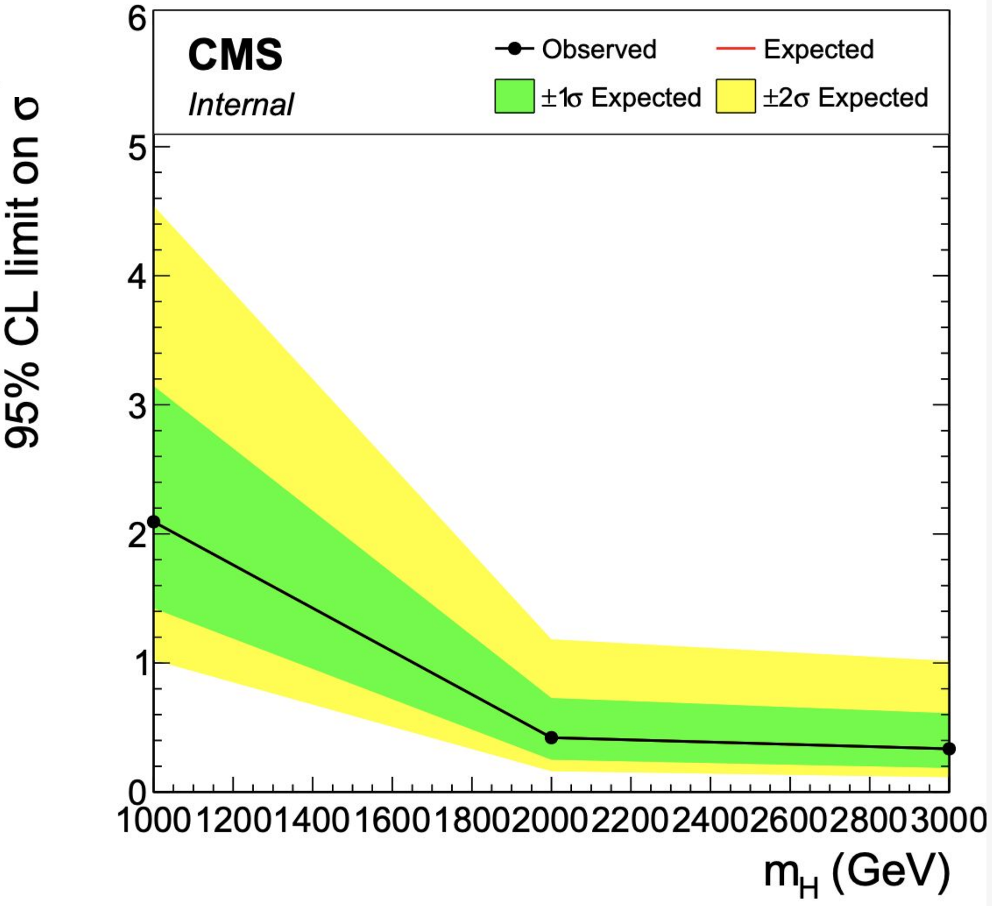
- Created plots for each of the two categories (high and low impurity), stacking the 4tau mass of background and signal





Limit Plot

- Expected limit on cross-section time branching ratio for the Signal
- Goal of a limit plot is to see if signal exists in data (which is mostly background)
- If there's no excess in data, signal is excluded with a certain cross section
- Performed a blinded fit in this plot
- Not yet looking at data
- This is the expected limit, not the observed limit
- 2 fb limit is set on the cross section for 1 TeV Radion. The limit on 2 TeV and 3 TeV are about 0.3 fb at 95% confidence level



Next Steps

- After this, next steps would be to:
    - NN output of 1 TeV signal sample doesn’t look quite right, need to figure out what’s wrong with that
    - Have not applied any trigger yet
    - Have only relied on MC for background estimation so far
    - Need to run on data
    - Consider systematic uncertainties

What I’ve Learned This Summer

- How to use the terminal
- How to use Git and Github
- How to perform data analysis
  - C++ and python
- Constructing neural networks
  - How to combine data analysis with a neural network
- How the CMS detector works
- The standard model and current research in this field
- What grad school and physics career paths look like
  - How to create resumes/personal statements
  - Networking skills

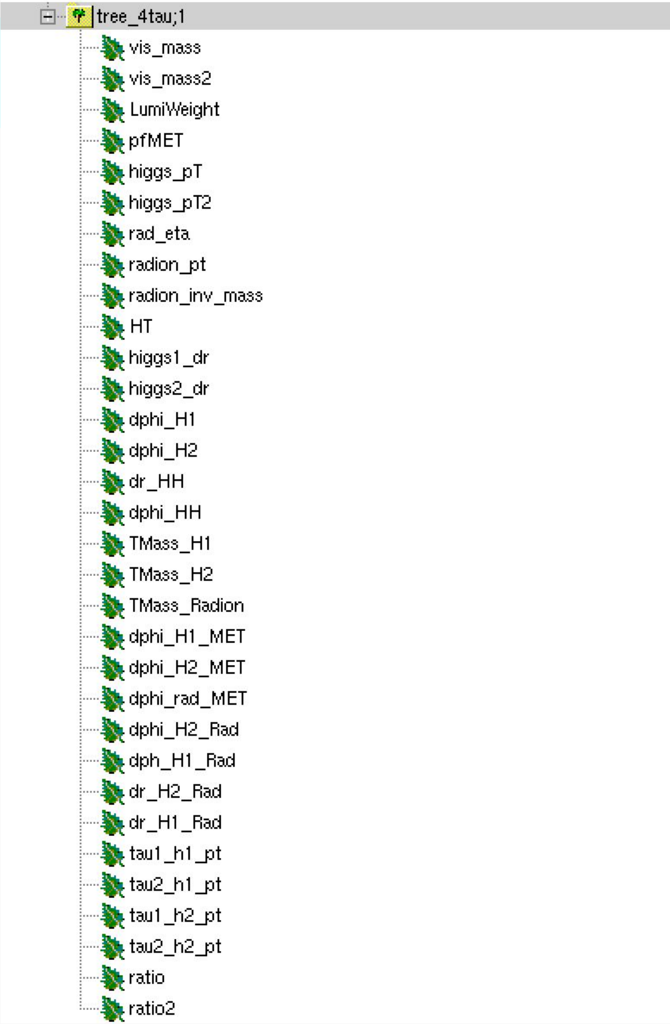
Acknowledgements:

Thank you for this summer research opportunity!!

Backup Slides

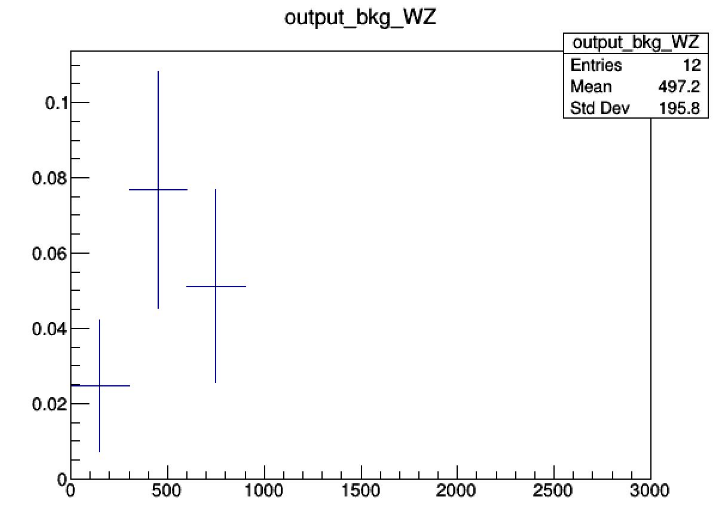
Analysis and TTrees

- Wrote code (C++) for analysis, then produced a TTree using ROOT
- A TTree is a columnar dataset, in which any C++ type can be stored
- Consists of a list of columns or *branches*
- The analysis filtered events and created branches of desirable features
- The TTree was then used as input for the neural network
- The neural network learns from the characteristics stored in the branches

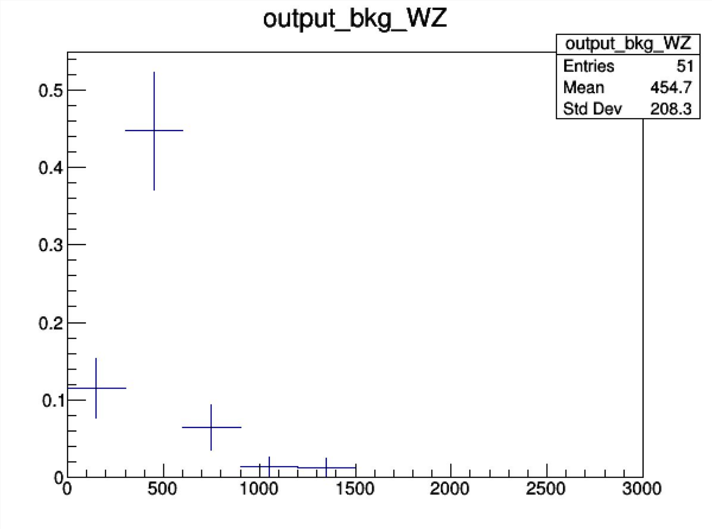


Output of BNN Cont.

- More examples of what background looks like when put into high and low purity categories:



High Purity



Low Purity

Number of Events

- Number of events in each sample after selection cuts
  - 1 TeV Signal⟶ 3706
  - 2 TeV Signal⟶ 7363
  - 3 TeV Signal⟶ 7325
  - DYJetsToLL⟶ 88
  - ZZ2l2q, ZZ4l⟶ 940
  - WJetsToLNu_HT⟶ 63
  - VV2l2nu⟶ 1
  - TTToSemiLeptonic, TTToHadronic, TTTo2L2Nu⟶ 109
  - T-tchan⟶ 1
  - Tbar-tW⟶ 1
  - WZ1l1nu2q, WZ1l3nu, WZ2l2q WZ3l1nu⟶ 63
  - QCD_HT⟶ 0
  - T-tW⟶ 0
  - Tbar-tchan⟶ 0