Deep Learning based Tagger for Boosted $WW^{(*)}$ semi-leptonic ($l\nu$qq) Decays at ATLAS

Dewen Zhong
University of Illinois Urbana-Champaign

ILLINOIS

APS/DPF meeting July 2021

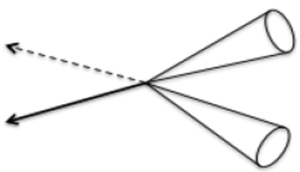
ATLAS EXPERIMENT

Motivation and Approach

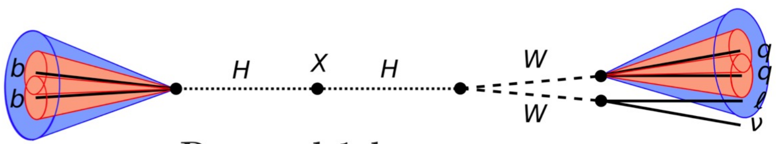
Motivation

- Di-Higgs production (*hh*) is a sensitive probe of new physics (e.g. a new boson *X*→ *hh*)
- When h → $WW^{(*)}$ is produced with high p_T (e.g. when m(X) ≫ m(h))
    - Hadronic W decay can form a large-R jet
    - Traditional methods are difficult to identify *h* by overlapping detector info from final states

Resolved



Boosted



Boosted 1-lepton

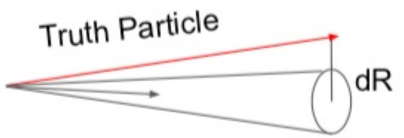
Approach

- Use deep learning for boosted WW : CNN, DNN, Hybrid-NN
- Semi-leptonic ($l\nu qq$) final state for boosted WW tagging ("3-prong" tagging)
    - One prong required to be constant with a lepton
- In this talk, we present our preliminary studies for the electron channel
- Note: Our approach is not specific to scalar boson masses with m = 125 GeV

Simulated data and jet features

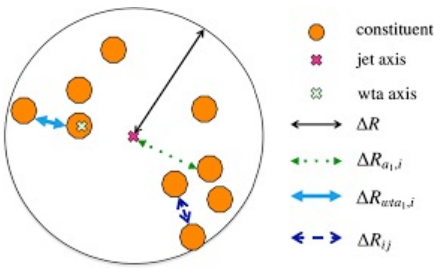
Data sources

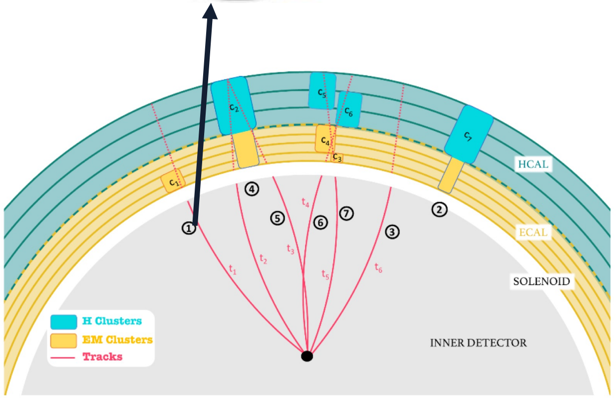
- Signal (29K events)
  - Single jet triggers
  - Leading Antikt jet with p_T > 150 GeV and $|\eta| < 2.5$
  - $\mathrm{X} \rightarrow hh \rightarrow wwb\bar{\boldsymbol{b}} \rightarrow lv\mathrm{qq}b\bar{\boldsymbol{b}}$
    for $\mathrm{m(X)} \in [1, 1.5, 2\ TeV]$
  - Signal jet selection: ΔR (truth jet) < 0.75



- Background (1.3M events)
  - $W' \rightarrow WZ$ m(W') from [0, 5.5TeV]
  - $Z' \rightarrow t\bar{t}$ m(Z') from [0, 5.5TeV]
  - Dijets Jet p_T from [60GeV,5.3TeV]

Jet constituents





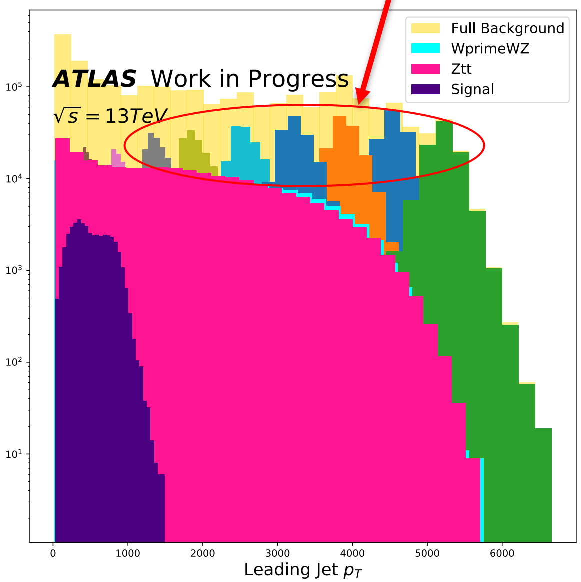
Jet features

- jet four vectors and jet constituents (shown in right)

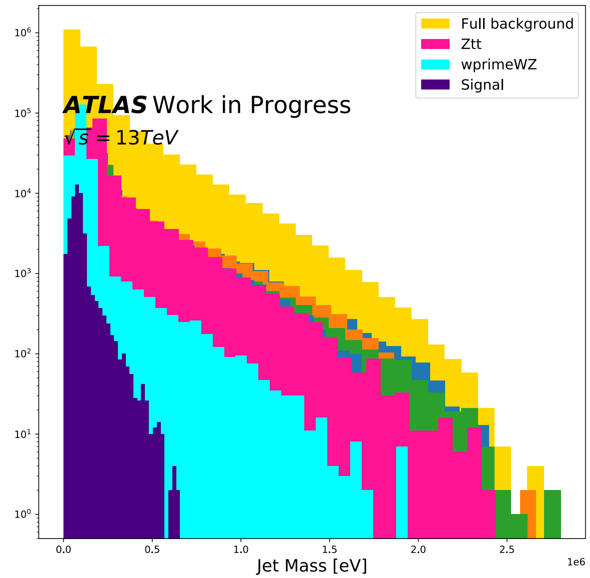
Jet four vectors	Jet Constituents(cluster moments)
p_T, η, ϕ, E	*FIRST_PHI*, *FIRST_ETA*, etc. (total 18)

Kinematic Distributions of Jet Features (examples)

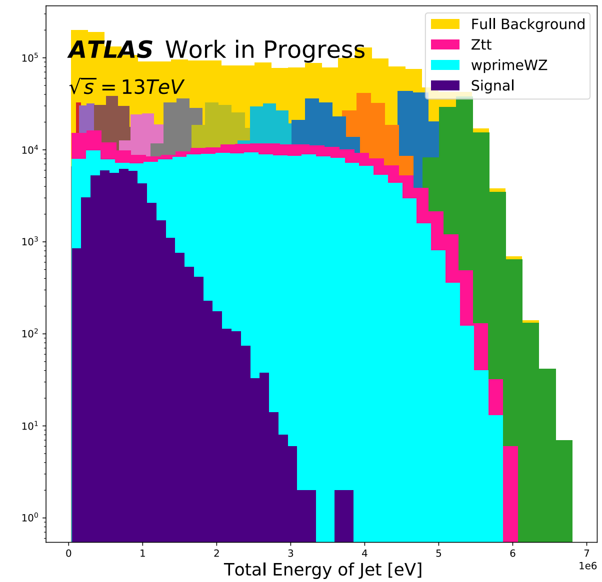
Dijet spectrum



Leading Jet p_T



Jet Mass [eV]



Total Energy of Jet [eV]

Several machine learning models were investigated (next slides)

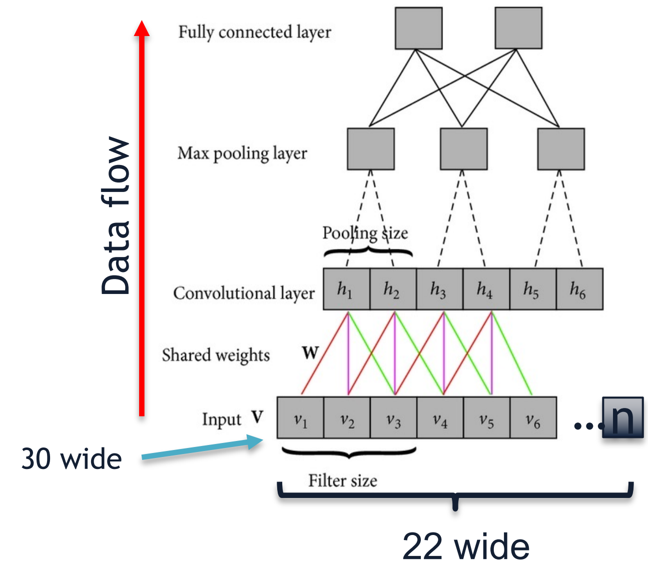
1-Dimensional Convolutional Neural Network (CNN)

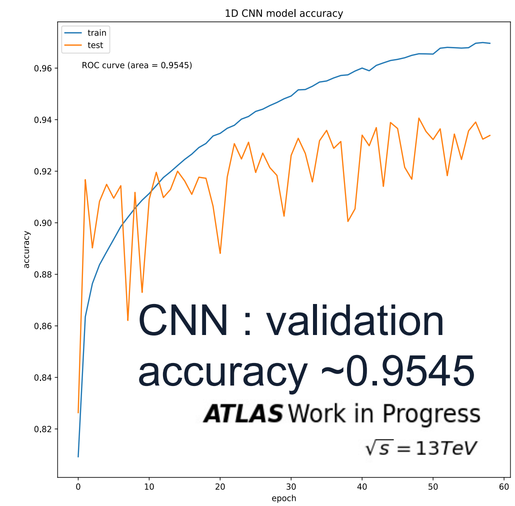
CNN

- Advantages
  - Convolutional NN could withdraw features from matrix data types, like images and arrays
  - Form raw jet constituents as 1D matrix dataset
  - learn features from each jet constituents

- Architecture Overview

Input Vector (V)	22 features in row 30 fix data lens (30*22) array
Convolutional Layer	Compute the output of neurons, 3 layers
Full Connected Layer	Prediction for 2 categories (signal, background)



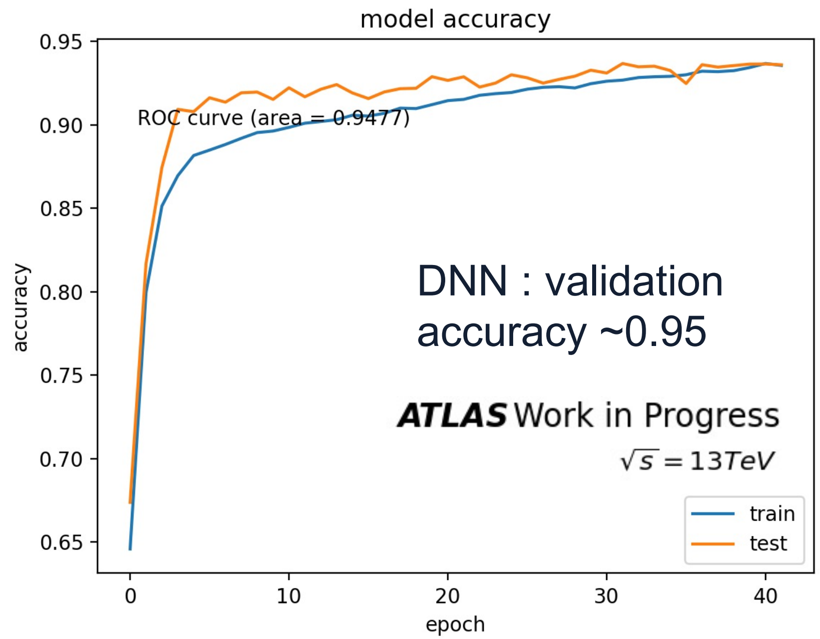


Deep Neural Network (DNN)

DNN:

- Advantages
    - a stacked neural networks (several layers)
    - The set of neural features will be used to uniquely identify a particular output (signal or background)
- Structure Overview

Input data vector	Pure Jet Constituents (p_T, η, ϕ,..) in sequence ($p_{T1}, \eta_1, \phi_1, \ldots, p_{T2}, \eta_2, \phi_2, \ldots$)
3 hidden layers	1000 nodes,1000 nodes, 448 nodes
Data split	80% for training,10% test, 10% validation



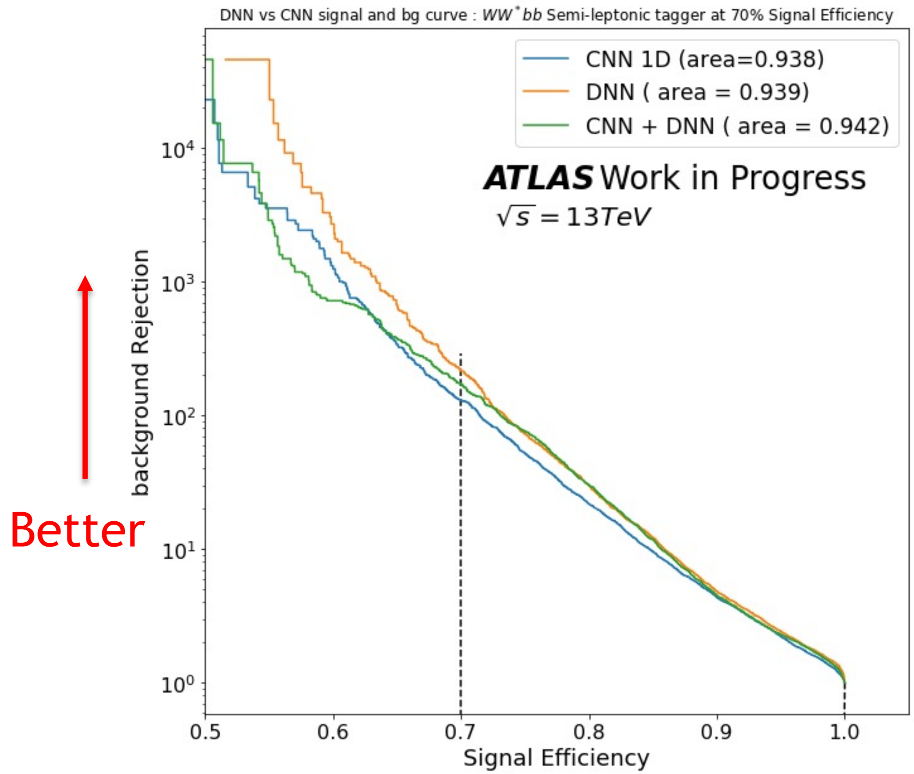
Compared with CNN

- DNN training converges quick
- higher validation accuracy

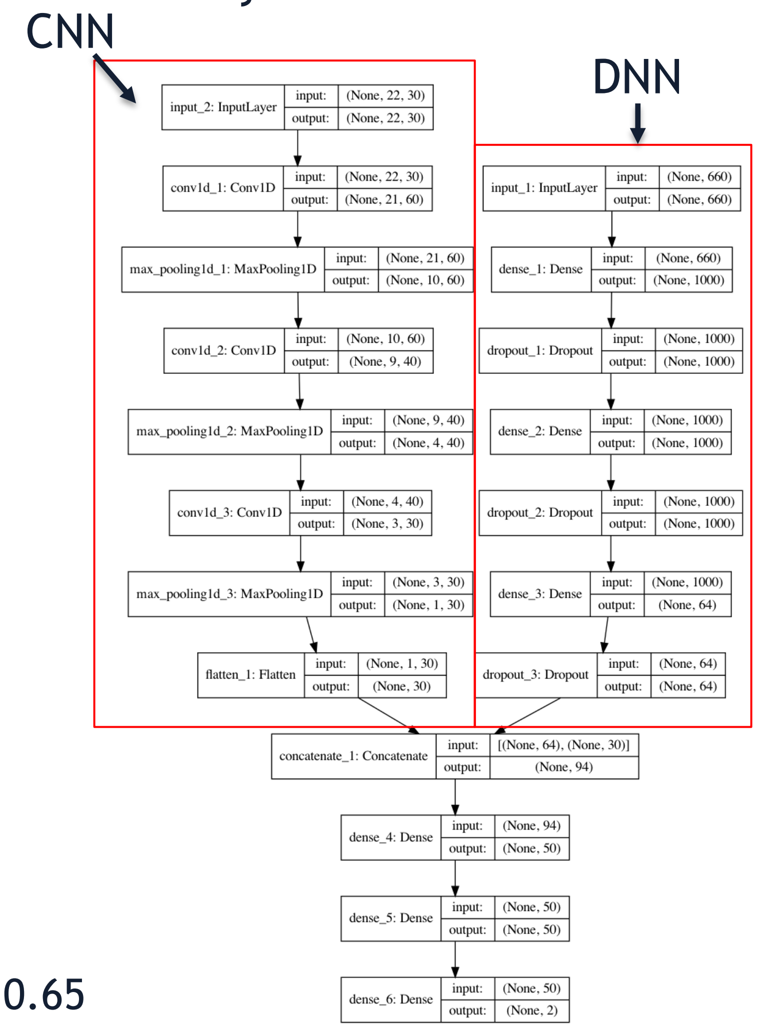
Hybrid Neural Network

Hybrid-NN

- Combined CNN and DNN
- Obtain CNN and DNN model's advantages



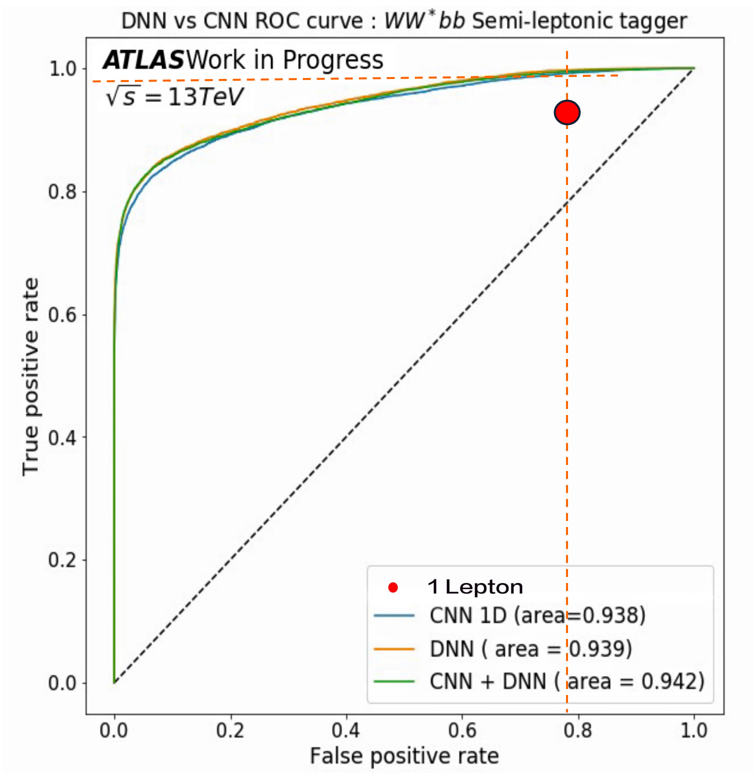
Hybrid-NN structure



Hybrid NN performs better when signal Efficiency > 0.65

ML models vs. single lepton selection (electron channel)

- Single lepton cut method for boosted $WW^{(*)}$ jet
  - Pt > 25 GeV
  - $|\eta|$ < 2.47
  - $d_0\, sig < 5$
  - $|\Delta z_0 sin\theta|$ < 0.5 mm
  - Isolated lepton in the W jet
  - loose likelihood-based Electron ID



Signal Efficiency = $\frac{N(TP)}{N\,(TP)+N(FN)}$

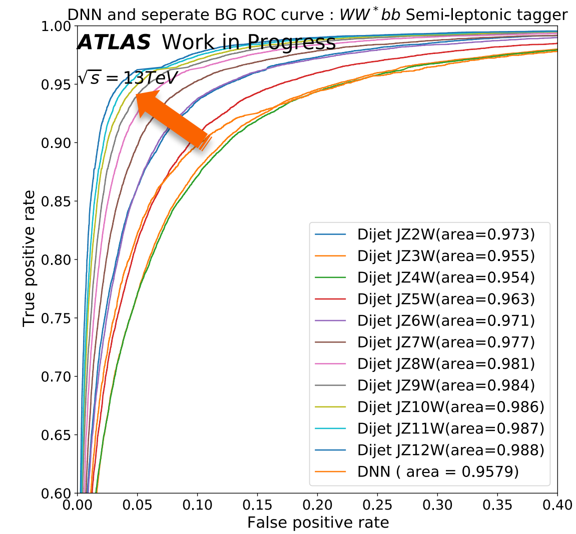
TP : event pass single lepton cut with signal label
FN : event not pass single lepton cut with signal label

Single lepton cut Signal Eff ~ 89.65%
ML-based selection eff ~ 98%

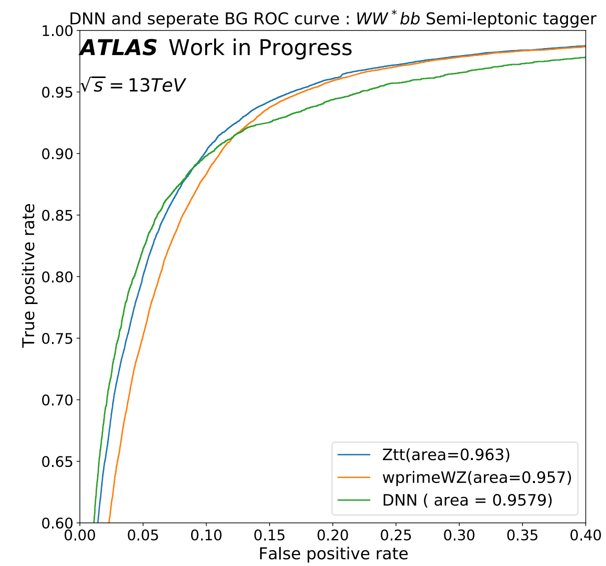
ML based boosted $WW^{(*)}$ jet tagging signal efficiency is higher than single lepton cut method at same background rejection rate

Single background efficiency study for DNN model

Single background + Full signal samples → Validation set → NN model → Prediction



- ➢ Strategy: using single background as validation set's background
- ➢ Mixing single background and full signal as a validation dataset
- ➢ Testing Dijet, $Z' \to t\bar{t}$ and $W' \to WZ$ separately.

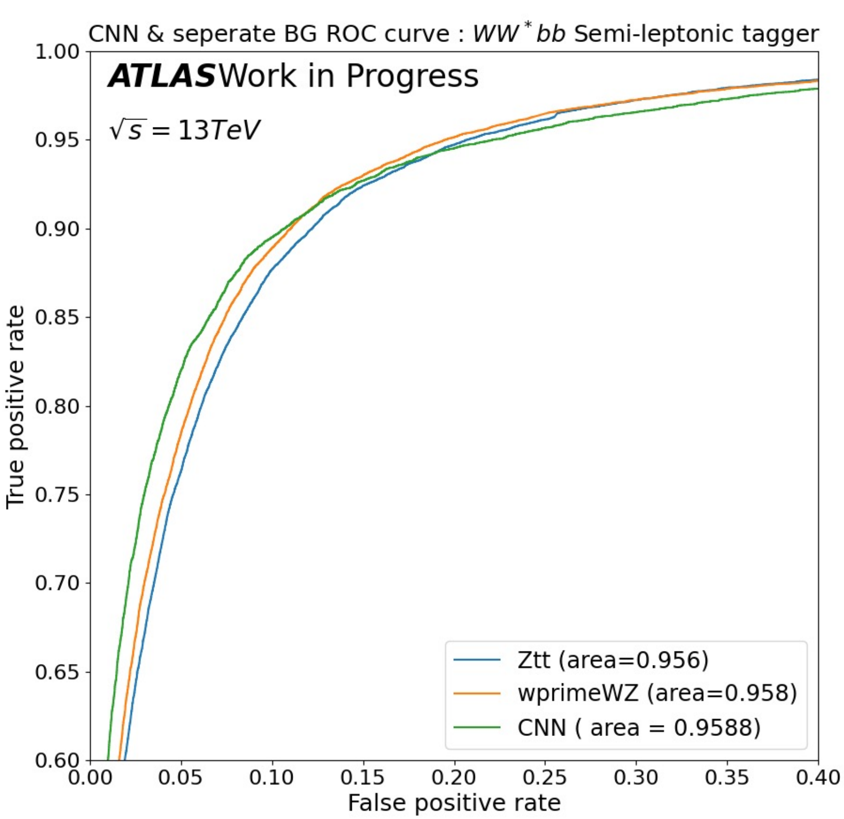


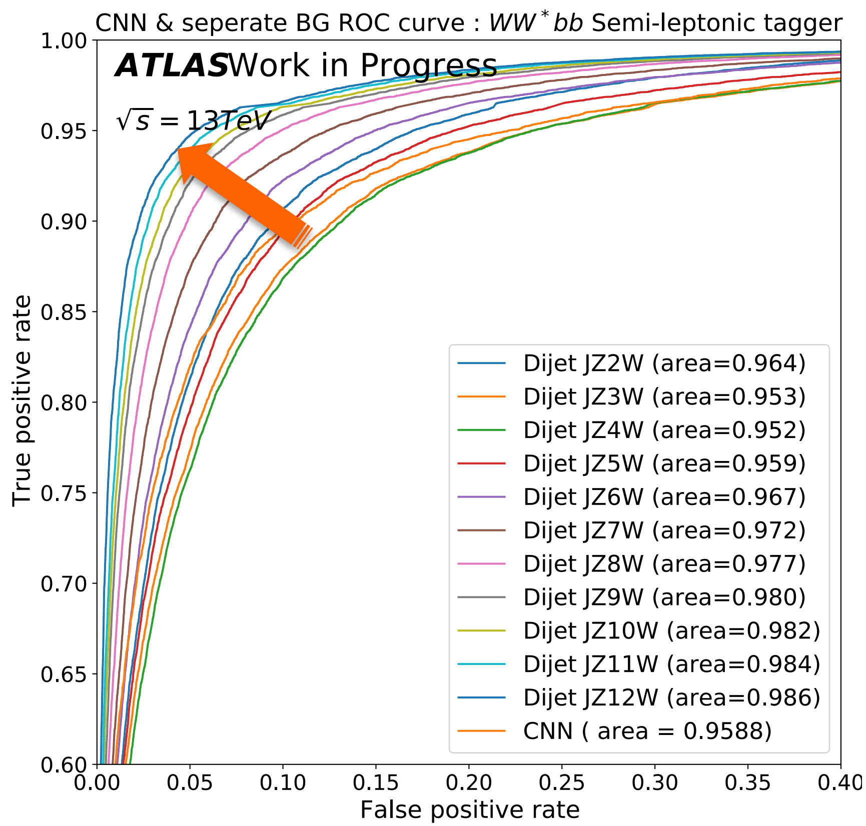
Preliminary results:

- High leading p_T Dijet samples' performances are better than DNN model
- $Z' \to t\bar{t}$ and $W' \to WZ$ samples' performances are similar as DNN model

Single background efficiency study for CNN model

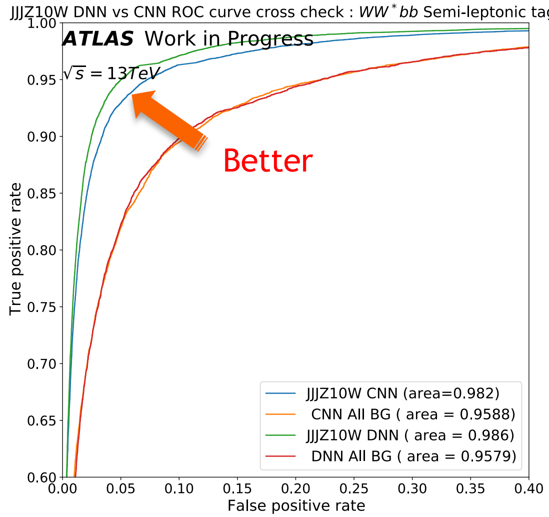
- Similar trend seemed with the Dijet background samples
  - High leading p_T Dijet samples performances are better in CNN model
  - $Z' \rightarrow t\bar{t}$ and $W' \rightarrow WZ$ samples performances are similar as CNN model

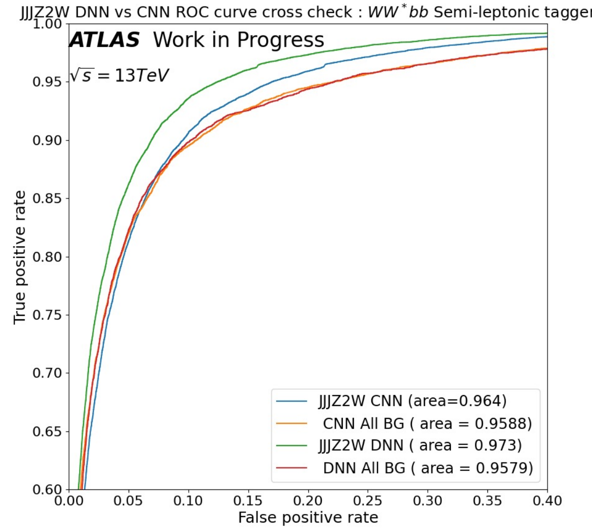


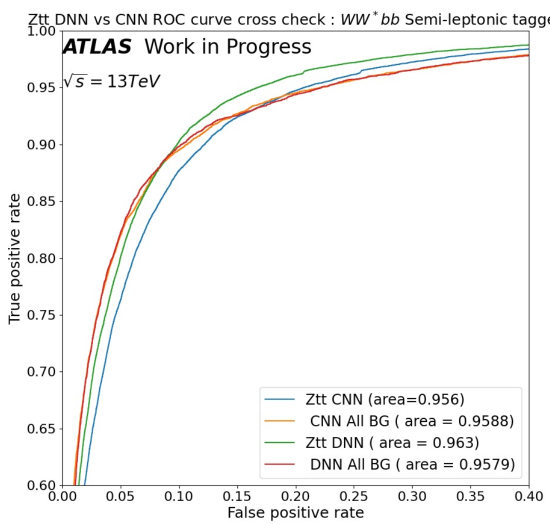


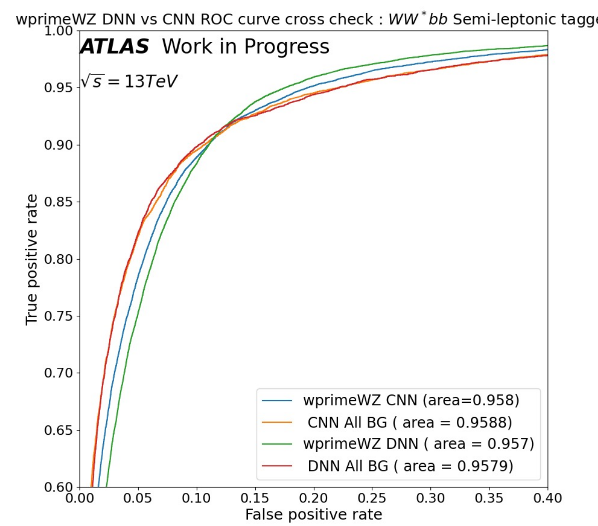
Cross comparison of NN models' ROC curves

- DNN's performance is better than CNN
- Dijet(single background) : high leading p_T samples performances are better than general NN models
- $Z' \rightarrow t\bar{t}$ and $W' \rightarrow WZ$: ROC curve crosses models at ~ 0.9 TPR









Summary and Conclusions

- We have developed a novel ML-based tagger for boosted WW decays in the semi-leptonic (electron) channel
- Preliminary (and promising!) result: This tagger shows a higher signal efficiency than a single lepton cut method for the same background rejection rate
- We plan to apply similar methods to the muon channel and continue other developments

Thanks !