Improving a search for heavy neutral Higgs bosons in the $t\bar{t}Z$ final state at CMS using parameterized neural networks

DPG Frühjahrstagung 2025, Göttingen

Bianca Weidner, Matteo Bonanomi, Lukas Ebeling, Yannick Fischer, Johannes Haller, Daniel Hundhausen, Matthias Schröder

Two Higgs Doublet Models (2HDM)

- Second Higgs doublet → 5 different Higgs bosons (h , H, A, $H^{\pm}$)
- Free parameters
  - Masses of the Higgs bosons (m_h, m_H, m_A, $m_{H^{\pm}}$)
  - Ratio of the vacuum expectation values: $\tan(\beta) = \frac{v_1}{v_2}$
  - Mixing angle α between the CP-even Higgs bosons
- "Alignment limit": $\cos(\beta - \alpha) \to 0$
  - Beyond the standard model h boson couples like the standard model Higgs boson

h H

Neutral CP-even

A

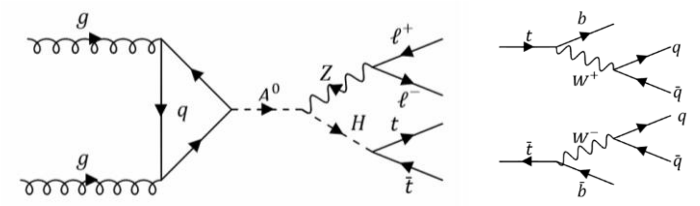
Neutral CP-odd

H^{+} H^{-}

Charged

Overview

- Search targets $A \rightarrow ZH$ with decay of $H \rightarrow t\bar{t}$ (see previous talk by Yannick Fischer)
- Focus on hadronic decay of $t\bar{t}$
- CMS result with 138/fb of Run 2 data at 13 TeV [1]
- Now: analysis of Run 3 data
- Here: study of neural network to separate signal of heavy Higgs boson events from background



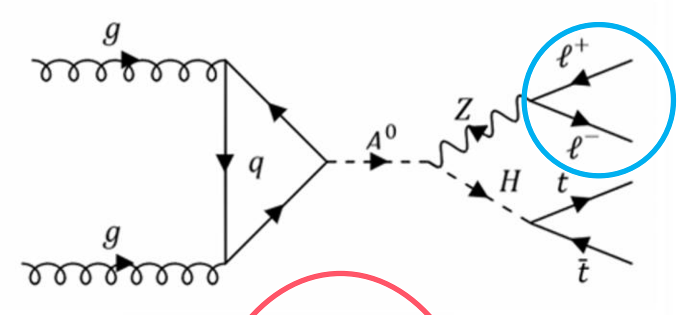
Limit plot Yannick

[1] [arXiv: 2412.00570 (subm. to PLB)]

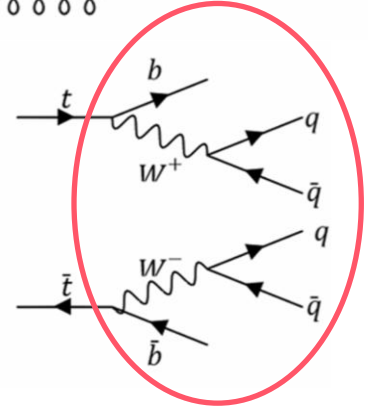
Setup for the analysis

Monte Carlo simulated events
(Run 2 – 2017)

$$\frac{\Gamma_{A/H}}{m_{A/H}} = 0.03$$



Exactly two leptons
(ee / $\mu\mu$)
$|m_{ll} - m_Z| \leq 5$ GeV



5 jets,
$\geq$ 1 b tagged jet

Mass range:
$330 \text{ GeV} \leq m_H \leq 850 \text{ GeV}$
$430 \text{ GeV} \leq m_A \leq 950 \text{ GeV}$
$\triangle m \geq 100 \text{ GeV}$

Background processes:
Drell-Yan, $t\bar{t}$, $t\bar{t}Z$

Characteristics of the neural networks (NN)

Network structure

→ 5 hidden dense layers
→ Leaky ReLU, Sigmoid activation function
→ Vs weight initializer

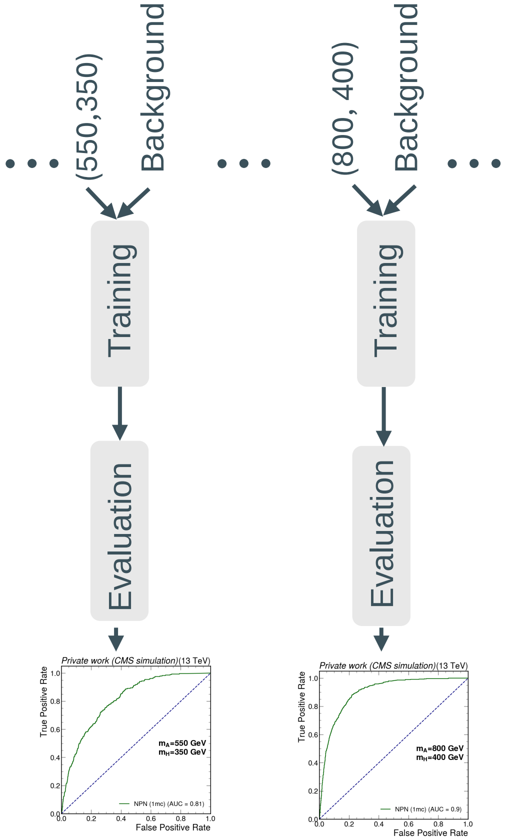
Tuned hyperparameters (Keras Hyperband Optimizer)

→ Number of nodes in the layers
→ Dropout rate
→ Learning rate

Input features

p_T, η, ϕ of the jets, leptons, A, Z, H, $\mathrm{M_A}$, $\mathrm{M_H}$, $\Delta\mathrm{M}$, b tagging scores, $\Delta\mathrm{R}$ leptons, $\Delta\mathrm{R}$ leading b jet and Z

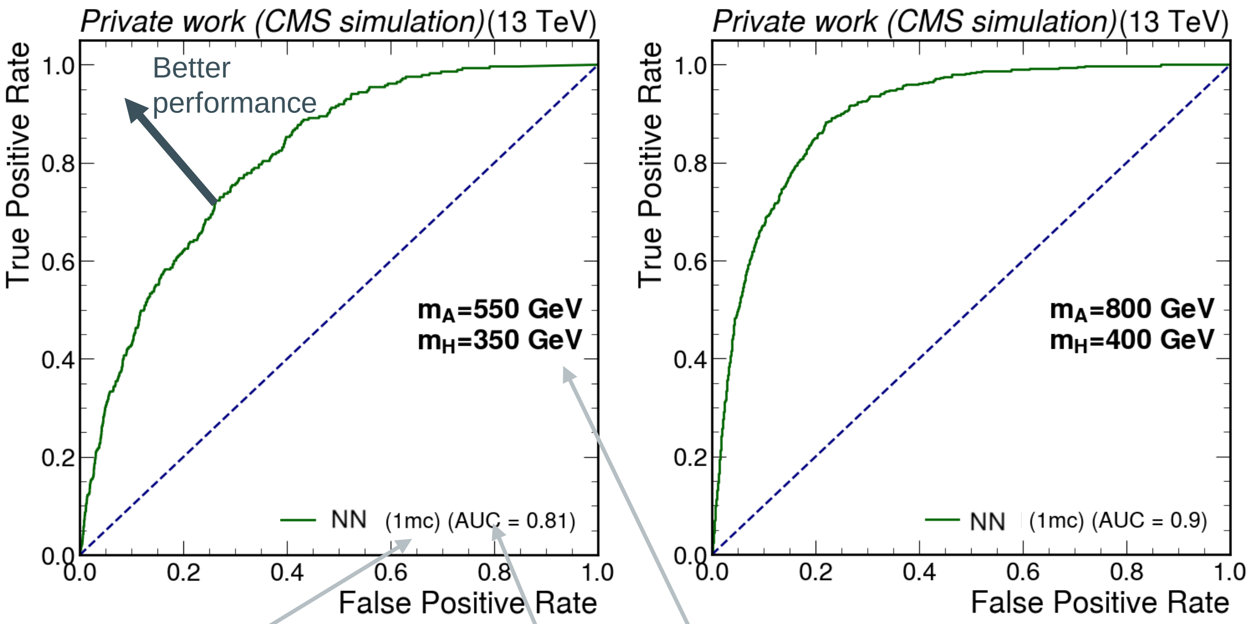
Training of individual NN



Input features: p_T, ϕ, η, M_A, M_H ….

(in total 79 different mass combinations considered)

ROC dedicated NN per mass combination



NN (1 mc): trained with 1 mass combination

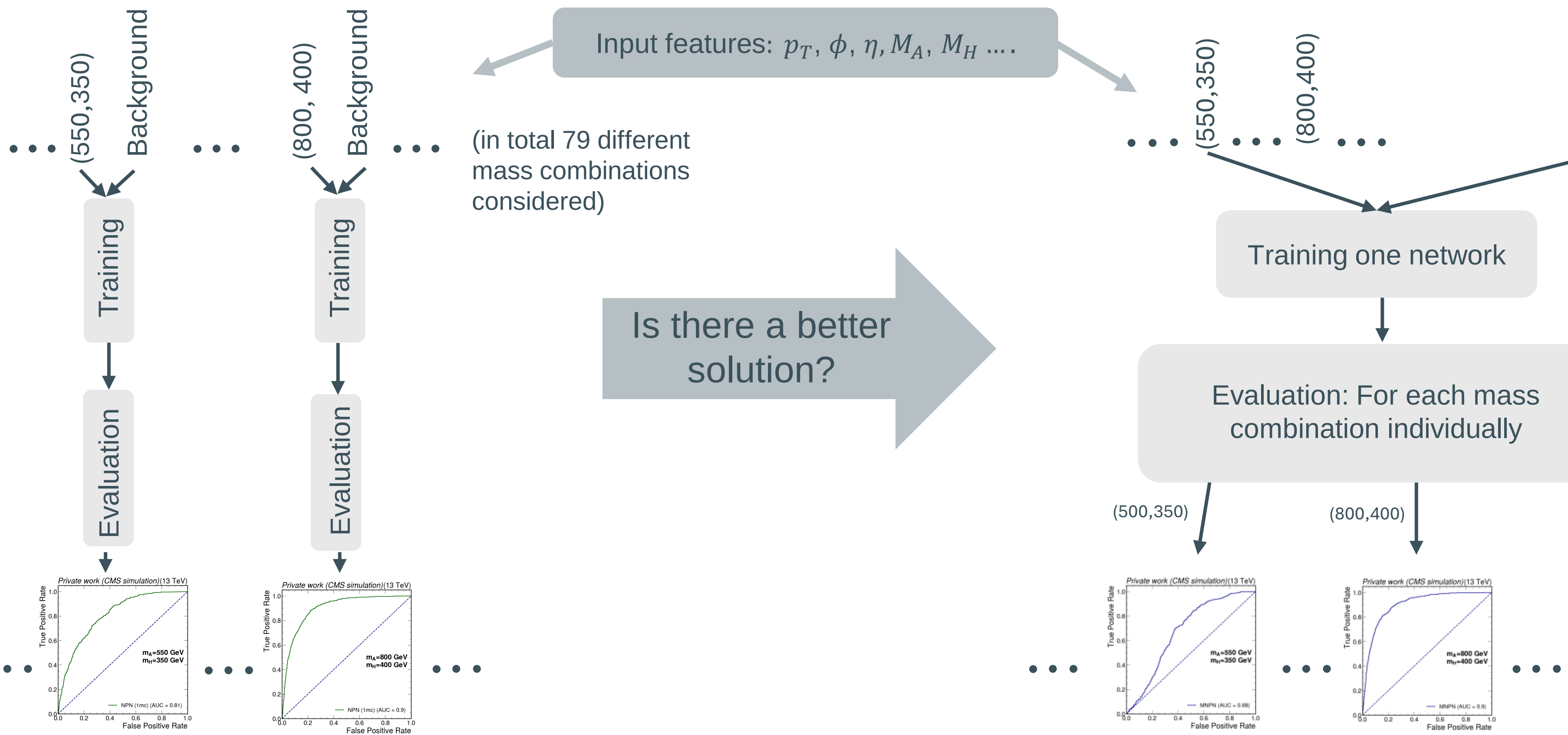
- Performance is good, but it is impractical
- For each mass combination a separate network is necessary

mc: number of mass combinations included in the training

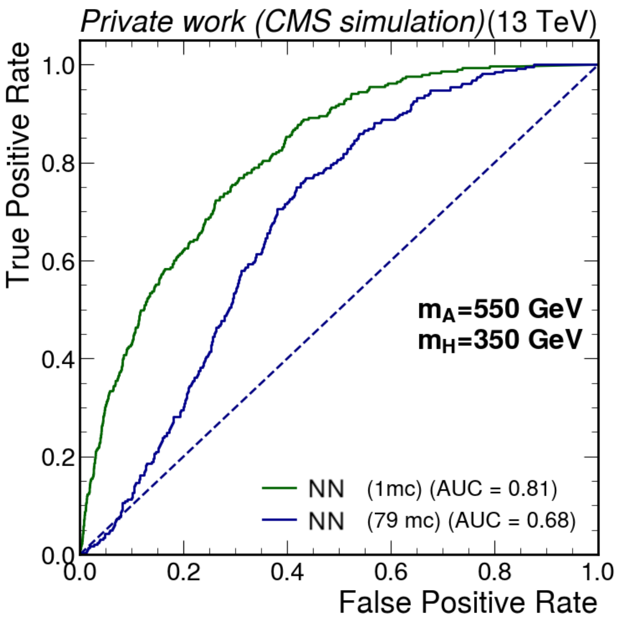
Mass combination used for evaluation of the network

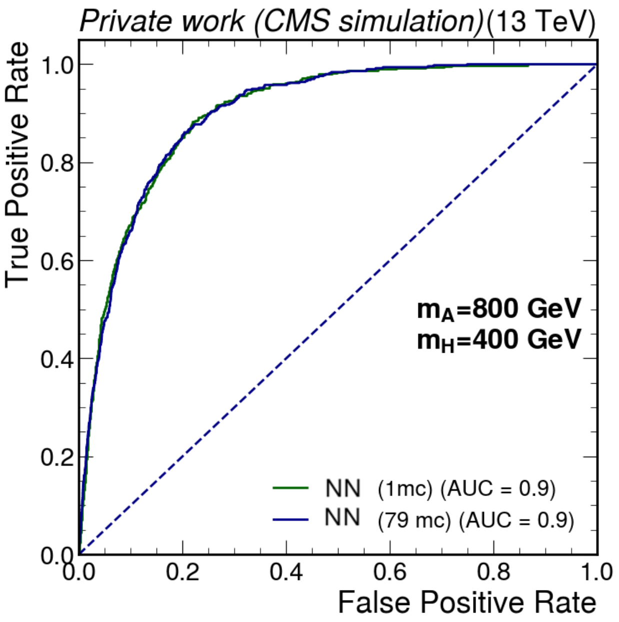
AUC: area under the ROC curve

Training of individual NN



NN trained with all masses





NN (1 mc): trained with 1 mass combination

NN (79 mc): trained with 79 different mass combinations

→ The inclusion of all signal events into a NN leads to a decrease in performance for some mass combinations

Training of individual NN

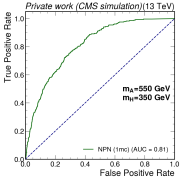
Input features: $p_T, \phi, \eta, M_A, M_H$ ….

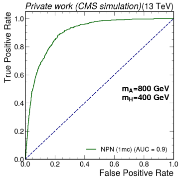
… (550,350) Background … (800, 400) Background …

(in total 79 different mass combinations considered)

Training

Evaluation





Is there a better solution?

… (550,350) … (800,400) … Background

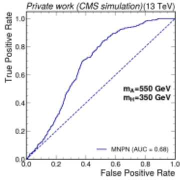
Training one network

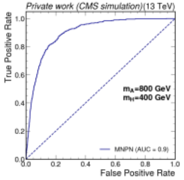
combination individually

Lower performance than before
→ inadequate solution!

(500,350)

(800,400)





Parameterized neural network (PNN)

Input features: $p_T, \phi, \eta, M_A, M_H$ ….

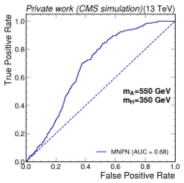
… (550,350) … (800,400) … Background

Training one network

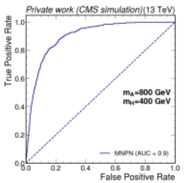
Lower performance than before
→ inadequate solution!

combination individually

(500,350)



(800,400)



Parameterized neural network

Signals: true masses (m_A, m_H)

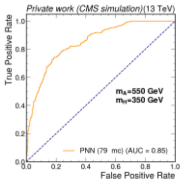
Background: random true mass combinations (m_A, m_H)

… (550,350) … (800,400) … Background

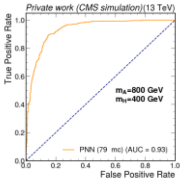
Training one network with parametrization

Evaluation: For each mass combination individually

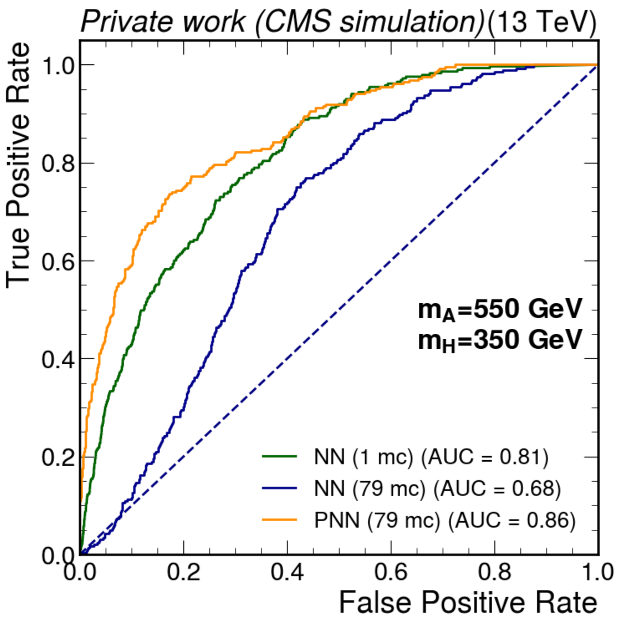
(500,350)

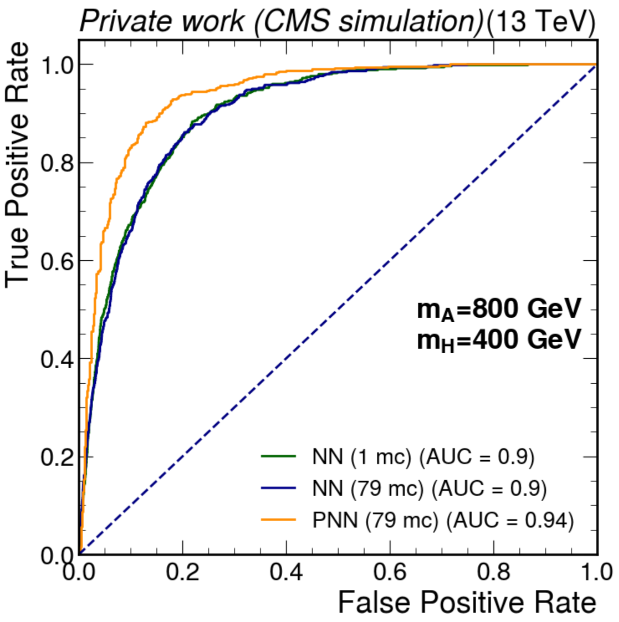


(800,400)



Comparison of ROC for NN and PNN





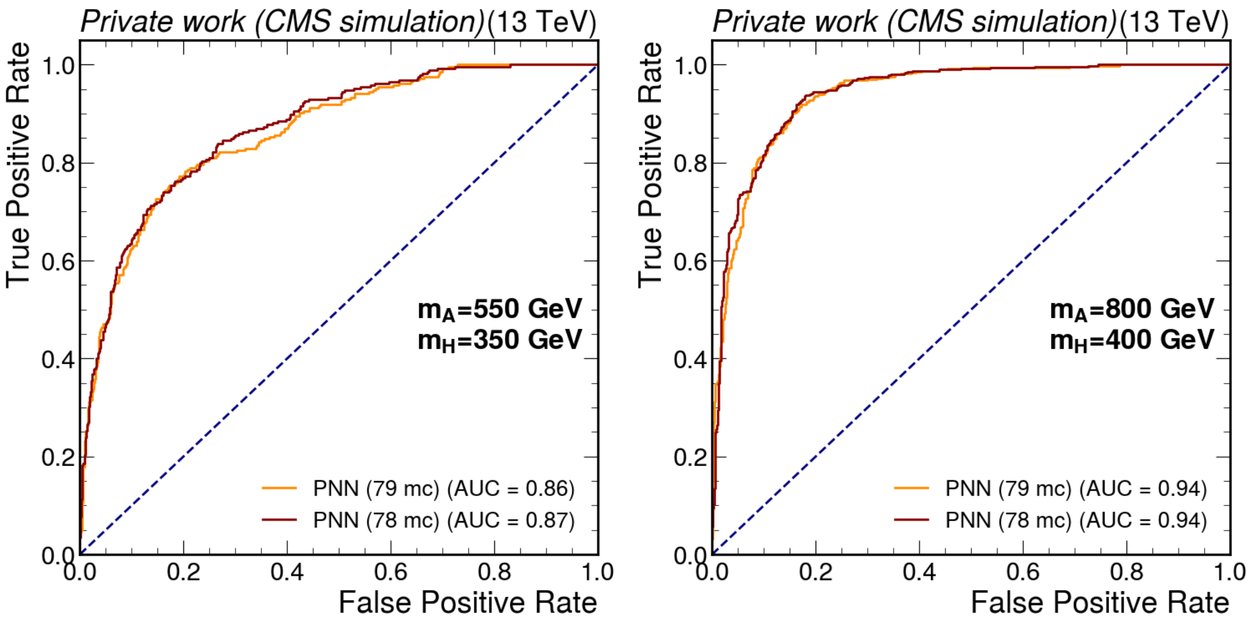
NN (1 mc): trained with 1 mass combination

NN (79 mc): trained with 79 different mass combinations

PNN (79 mc): trained with 79 different mass combinations

→ The inclusion of all signal events with parameterization leads to an increase in performance

Extrapolation for unseen mass combinations



PNN (79 mc): trained with 79 different mass combinations

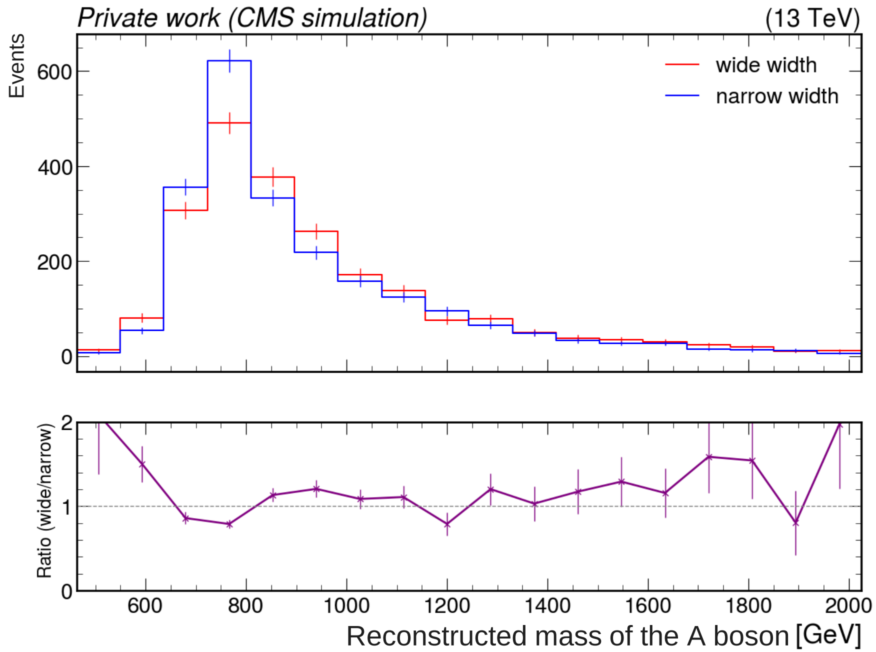
PNN (78 mc): trained with 78 different mass combinations; the mass combination used for evaluation is not included in the training

➔ PNN can generalize and extrapolate for mass combinations not included in the training

Advantages of PNNs

- One network can be used to evaluate all mass combinations
- Better performance than NN
- Evaluation of mass combinations not included in the training is possible; performance is better than NN (for considered examples)

Impact of changes of the decay width



Narrow width:

$$\frac{\Gamma_{A/H}}{m_{A/H}} = 0.03$$

Wide width:

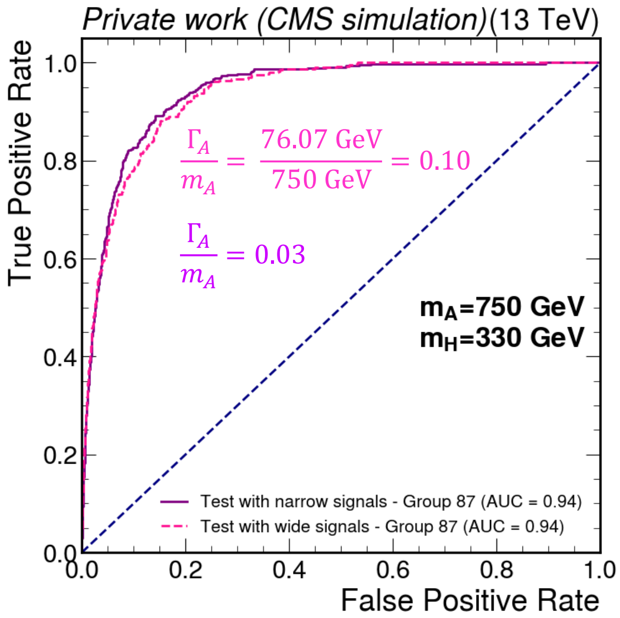
As predicted in 2HDM e.g.

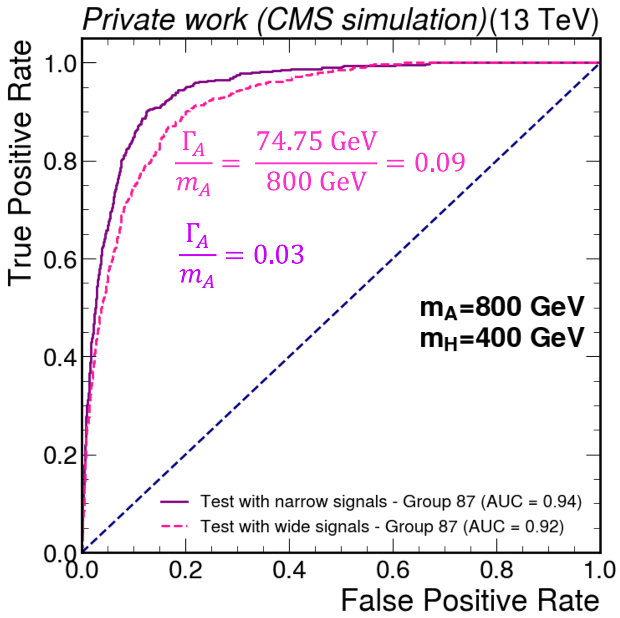
$m_A = 800$ GeV, $\Gamma_A = 74.75$ GeV (for $\tan(\beta) = 2$)

$$\frac{\Gamma_A}{m_A} = \frac{74.75 \text{ GeV}}{800 \text{ GeV}} = 0.09$$

→ Decay width of the A boson is no longer neglectable compared to detector resolution

Impact of changes of the decay width





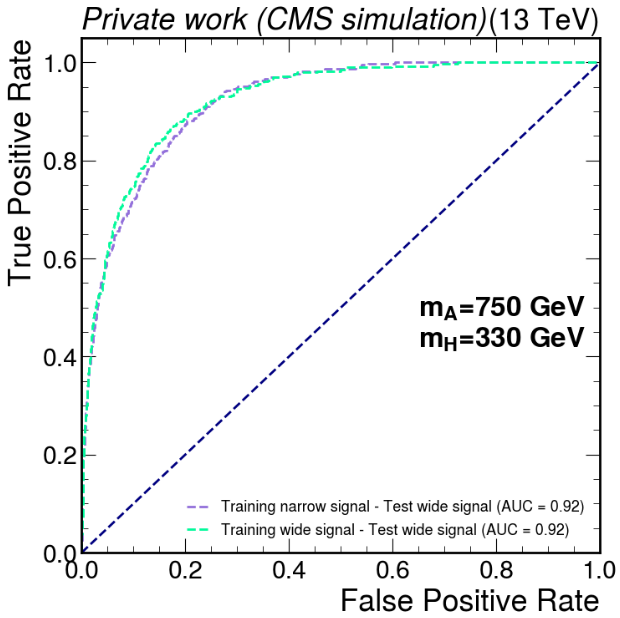
Training: narrow signal events

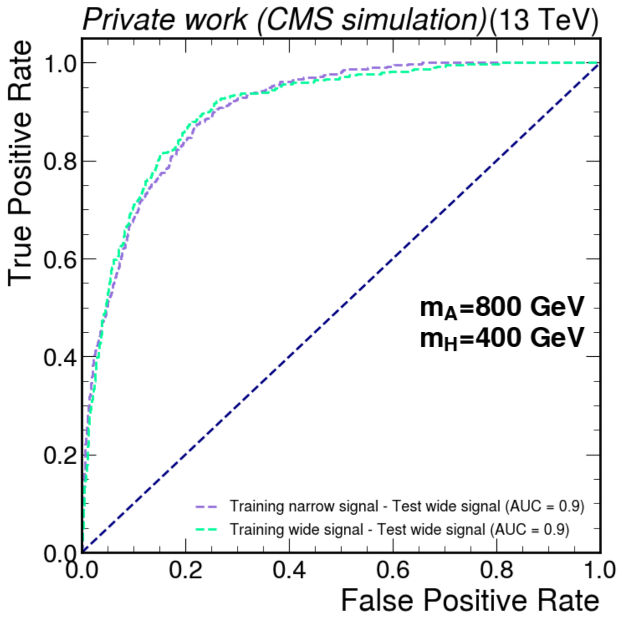
Testing: narrow or wide signal events in comparison

→ Even if we train the network with narrow signal events, the network performs nearly similar for testing with wide signal events

→ Neural network performs robust when the signal width is changed

Impact of changes of the decay width





Training: narrow or wide signal events in comparison

Testing: wide signal events

→ Using wide signal events for training does not lead to an increase in performance compared to training with narrow signals when testing with wide signal events

Summary

- Search for heavy Higgs bosons ($A \rightarrow ZH \rightarrow l\bar{l}\,t\bar{t}$)
- Studied neutral networks for signal/background classification
- Network performance measured using ROC and AUC score
- Parameterized network trained on several mass points performs better than several individual networks
- Parameterized neural network is robust to the effects of a change in the width of the A boson