

UiO : **Department of Physics**
University of Oslo

Application of Supervised Machine Learning to the Search for New Physics in ATLAS data

A Study of Ordinary Dense, Parameterized
and Ensemble Networks and their Application
to High Energy Physics

William Hirst

May 25, 2023

1 Introduction & Motivation

2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

1 Introduction & Motivation

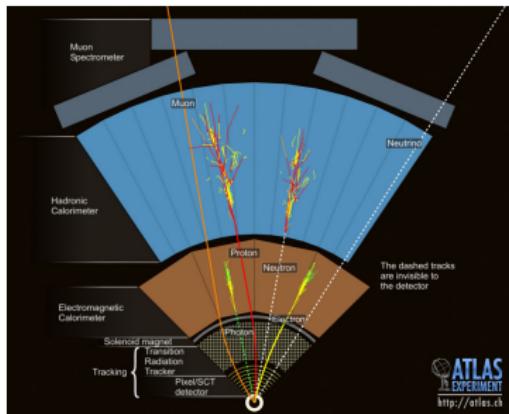
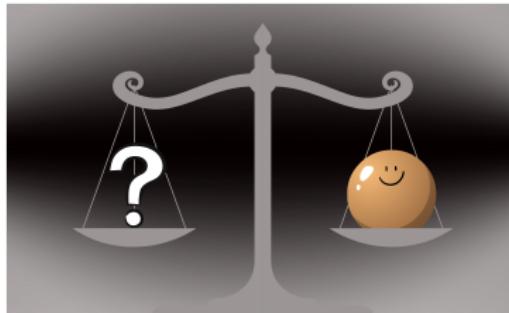
2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

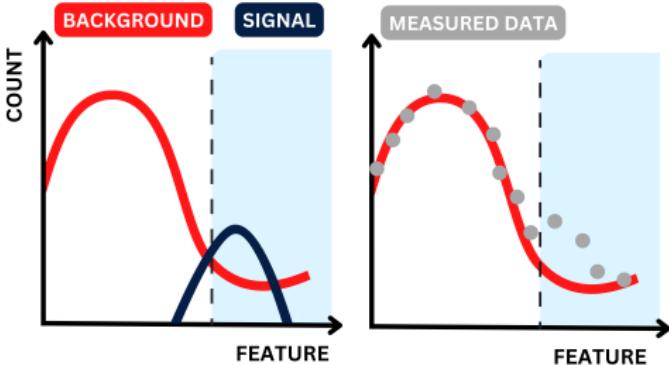
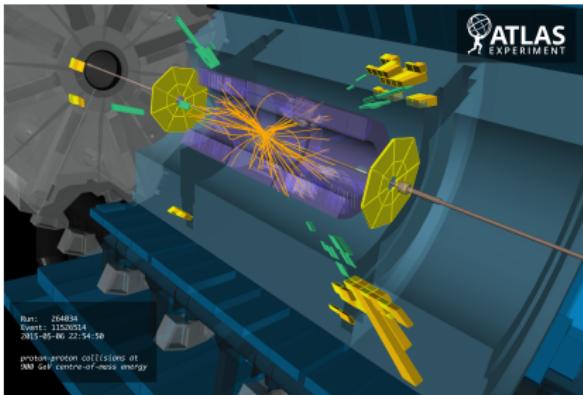
Why apply machine learning to HEP problems?

- The standard model (SM) of particle physics is very successful, but not complete
 - Neutrino masses
 - Hierarchy problem
- Large amount of data
- Machine learning (ML)
 - Event reconstruction
 - Particle classification
 - Creating search regions



How do we search for new physics?

- Compare theory with experiment
 - Experiment: Measured
 - Theory: Simulated (background and signal)
- Search regions
- Expected significance
 - $Z_{\text{exp}} \approx \frac{\text{signal}}{\sqrt{\text{background}}}$
- Difficult to separate \rightarrow ML



1 Introduction & Motivation

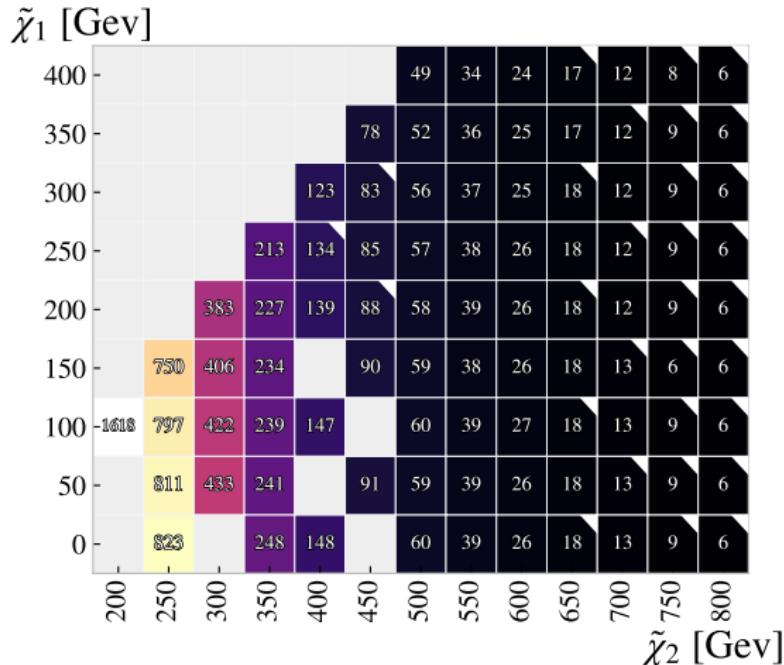
2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

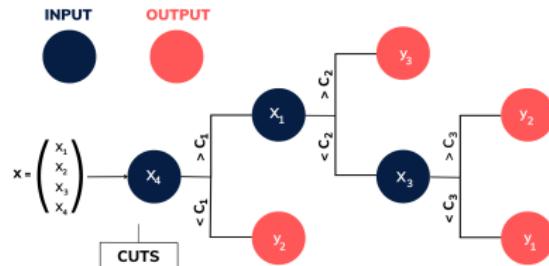
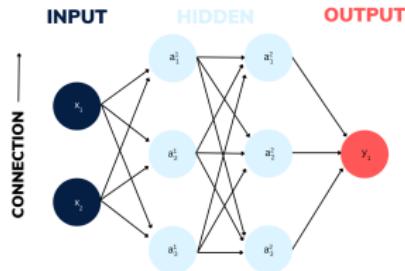
The SUSY signal

- Chargino-neutralino production
- Free parameters → masses
- Nr-of-Events(Mass)



A summary of the applied methods

- Three neural network variants
 - Ordinary dense neural network
 - Ensemble networks utilizing Local-Winner-Takes-All (LWTA) layers
 - Parameterized neural networks (PNN)
- One boosted decision tree method



Training strategy

- Objective
 - Background → 0
 - Signal → 1
- 80% training and 20% validation
- Early stopping criteria
 - Train as long as performance on validation set improves
 - Patience 10 epochs
 - Reset weights to best epoch

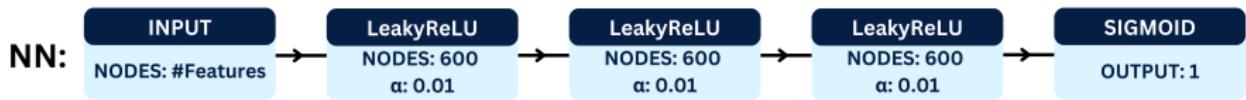
1 Introduction & Motivation

2 The Implementation

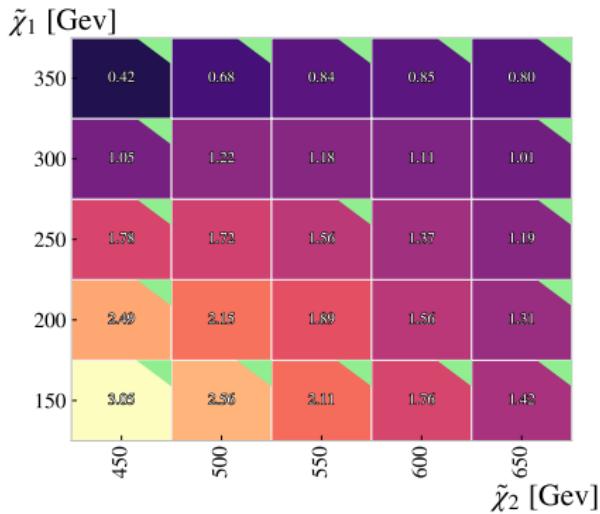
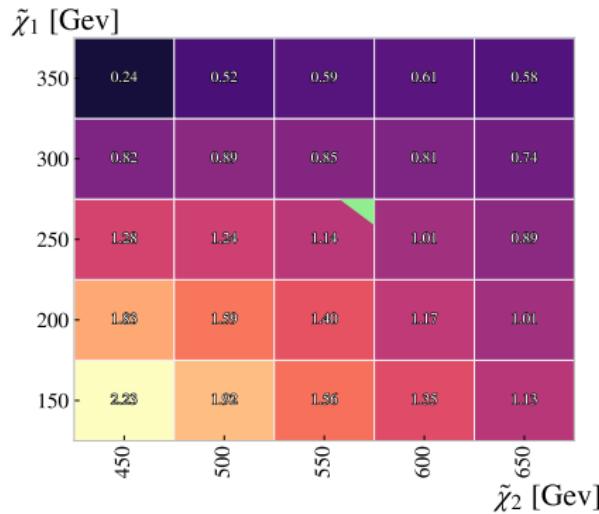
3 Methods & Results

4 Conclusion & Outlook

Ordinary dense neural network



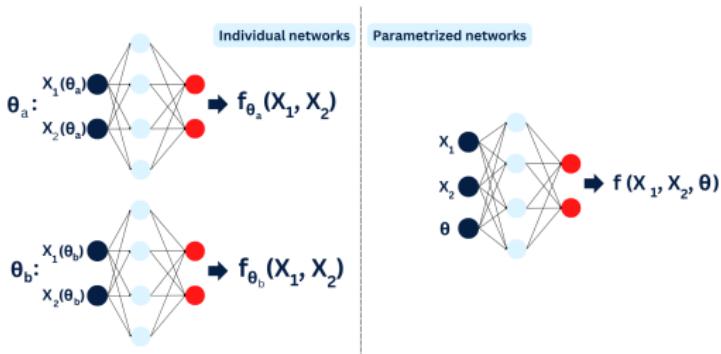
Compare one-mass approach to several-masses approach



Parameterized neural network

- Long-term memory
- PNN → signal includes mass parameter in feature set
- Background assigned parameters randomly using same distribution as signal
- Motivation

- Network will associate parameters with trends in the data



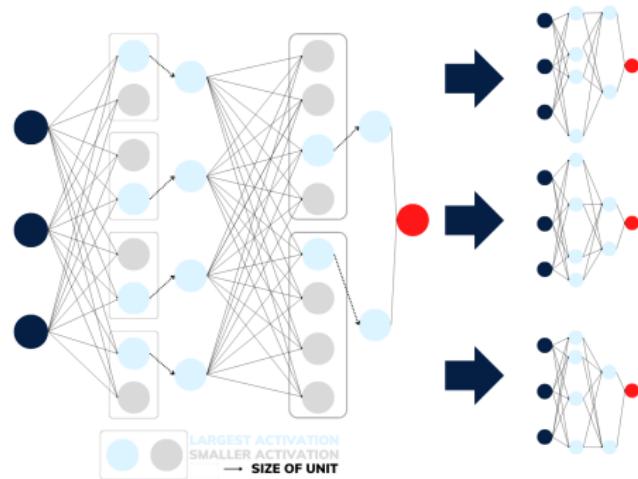
Study the effect of the parameters in the PNN

- Test: Manually assign all the events the same parameters thereby assigning most of the signal the wrong parameters
- Hypothesis: PNN performs better when events are assigned correct parameters
- First test: $\{50, 250\}_{GeV}$
- Second test: $\{200, 300\}_{GeV}$

Parameters \ Channel	$\{50, 250\}_{GeV}$	$\{100, 200\}_{GeV}$	$\{150, 300\}_{GeV}$	$\{200, 300\}_{GeV}$	(Background)
(50, 250)	80.8%	45.8%	77.5%	50.1%	2.4%
(200, 300)	77.3%	54.6%	76.3%	59.0%	2.7%

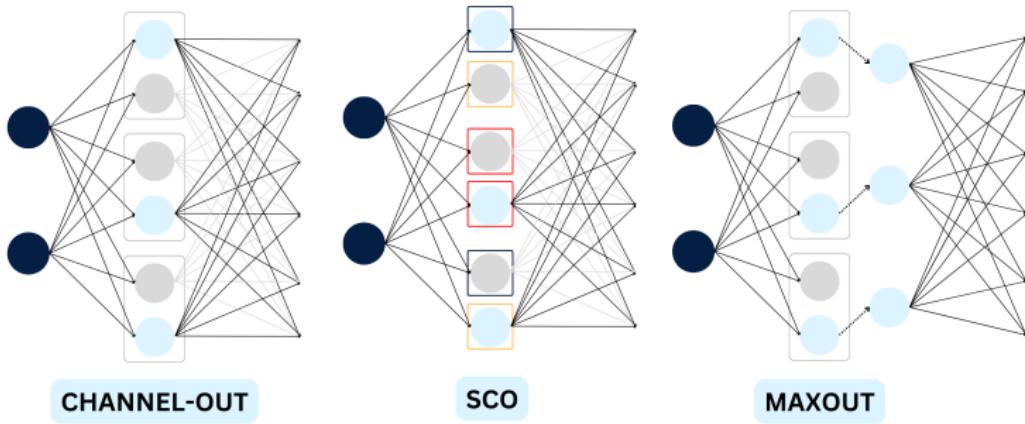
Ensemble methods - LWTA

- Dropout
- What is LWTA?
- Competing nodes - Units
- Encode information in pattern specific pathways



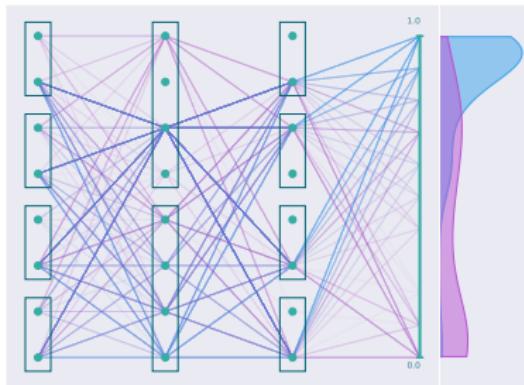
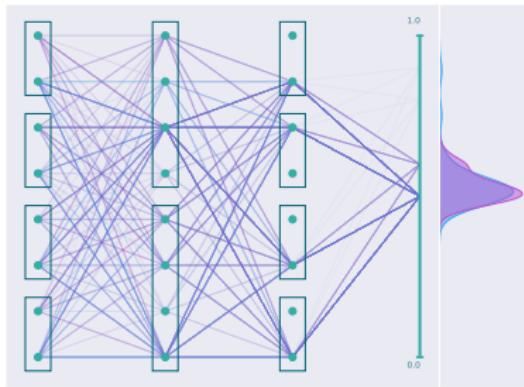
Channel-Out, SCO and Maxout

Layer	Separate weights	Static units
Channel-Out	✓	✓
SCO	✓	X
Maxout	X	✓



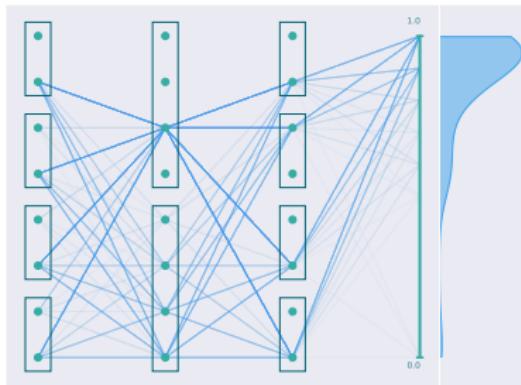
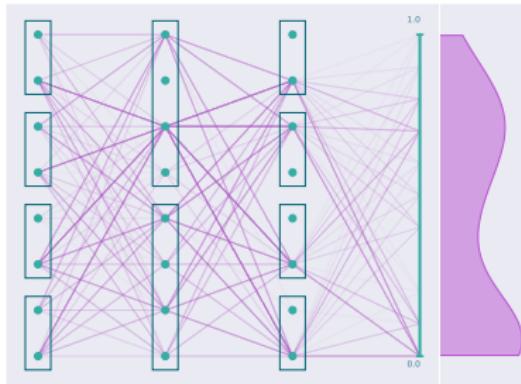
Visualization and study of sparse pathways

- Visualize the activation and paths of randomly sampled events
 - 50 background
 - 50 signal
- The bolder the line the more frequently the path is used.
- Color of lines
 - Pink: SM background
 - Blue: SUSY signal

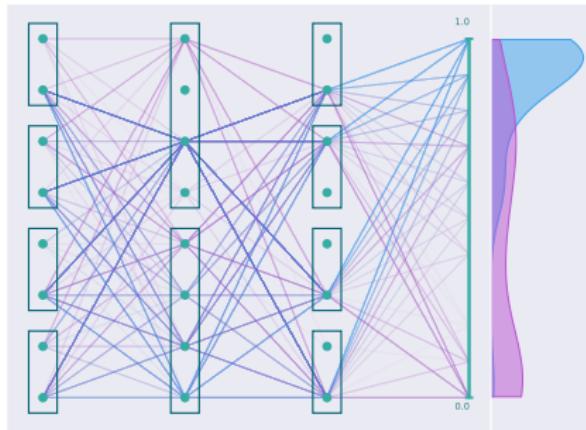


Visualization and study of sparse pathways

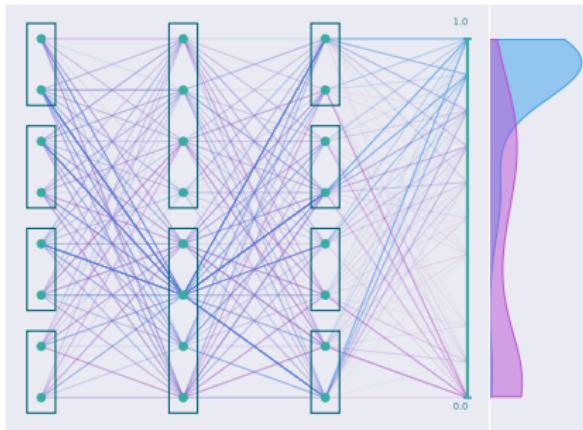
- Visualize the activation and paths of randomly sampled events
 - 50 background
 - 50 signal
- The bolder the line the more frequently the path is used.
- Color of lines
 - Pink: SM background
 - Blue: SUSY signal



Comparing activation of Maxout with SCO



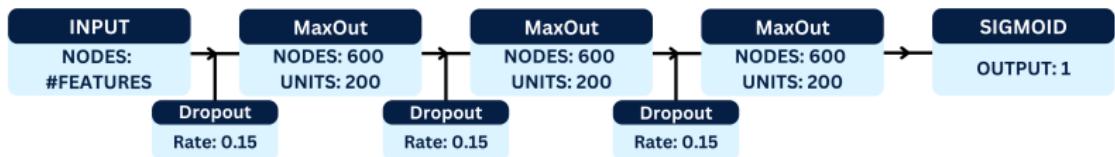
Maxout



SCO

Ensemble network architecture

MaxOut:

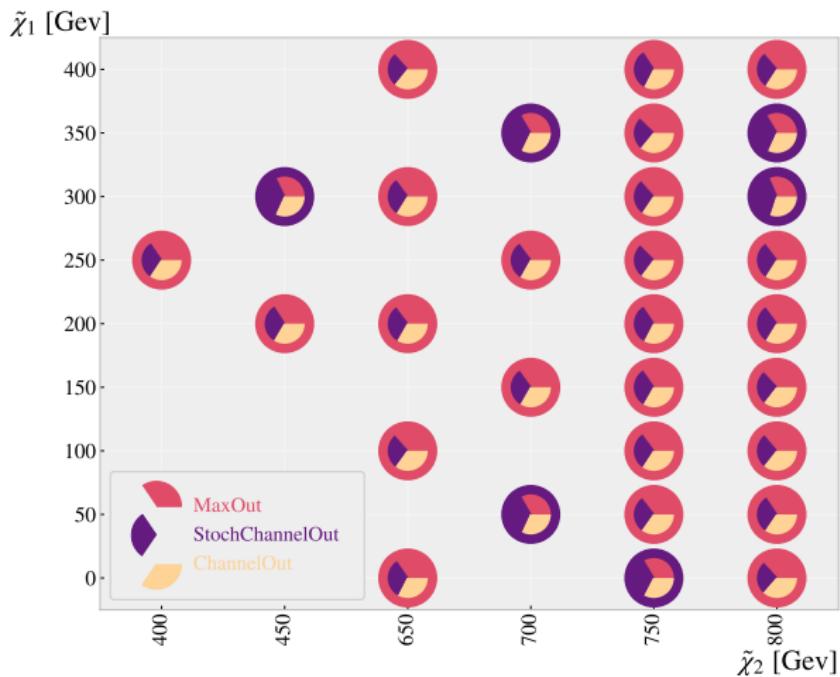


Comparing sensitivity of channel-out, SCO and maxout

■ Maxout: 24/30

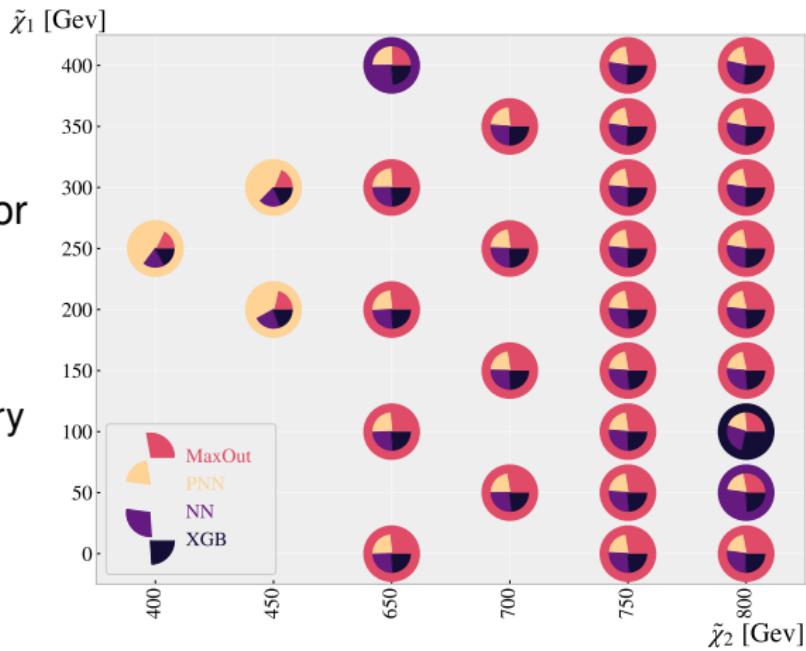
■ SCO: 6/30

- No trend for preferred masses
- Possibly improve without layer on prediction



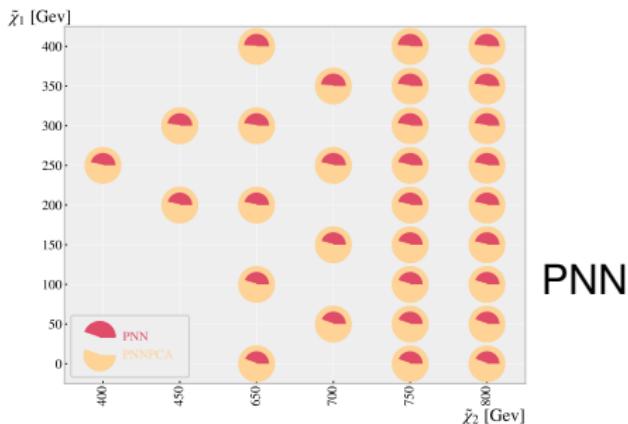
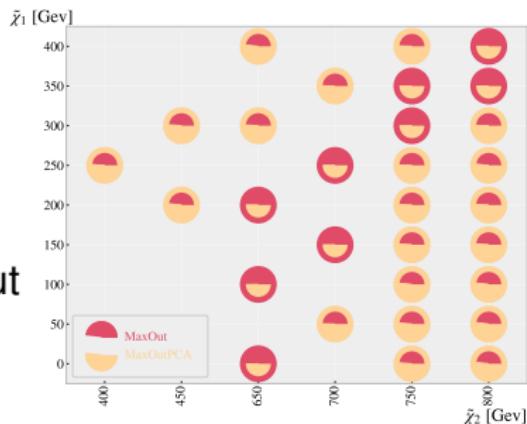
Comparing the sensitivity on a subset of the signal

- NN vs BDT
- Maxout achieves highest sensitivity
- PNN very sensitive for low masses
- Maxout sensitive for high masses
 - Long-term memory

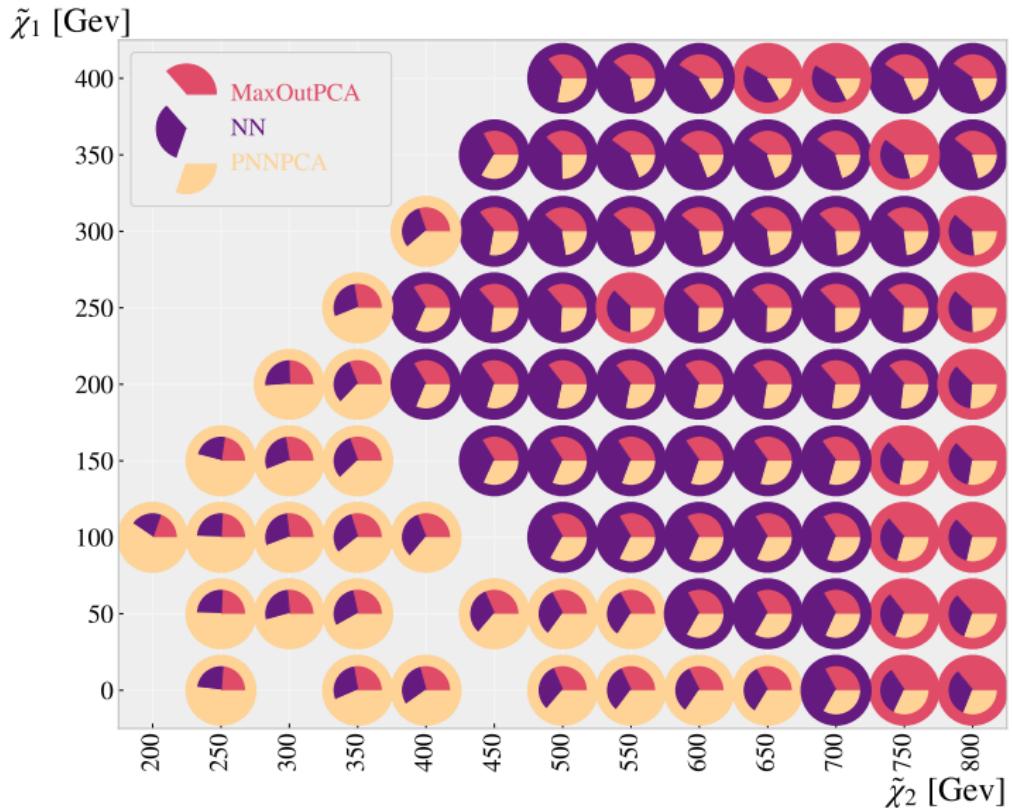


Increasing sensitivity through a PCA

- Dimensionality reduction
- This analysis
 - Demand conservation of 99.9% of variance/spread
 - 5 features removed

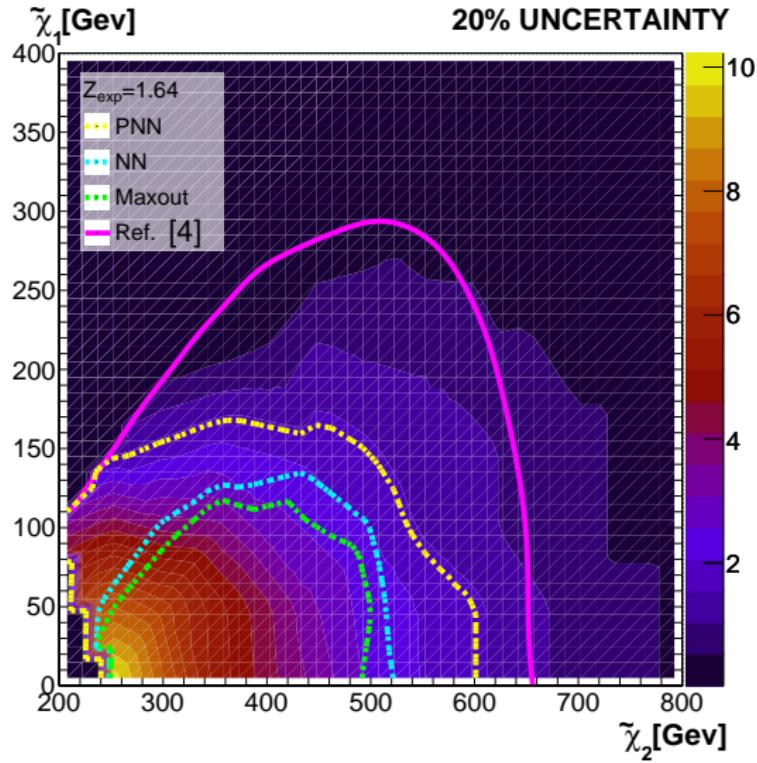


Comparing methods on full signal grid



Comparing the methods to previous analysis

- Compare the expected limits to analysis made by ATLAS in 2021 [4]
- Introduce flat uncertainty (20%, 10%, < 1%)



Conclusion & Outlook

1 Introduction & Motivation

2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

Conclusion & Outlook

- 1 Including a diverse signal set can improve performance
- 2 The LWTA layers improve long-term memory via pattern specific pathways
- 3 All network variants outperformed default settings of XGBoost
- 4 PCA increased sensitivity of PNN and maxout model in original signal set
- 5 None of the networks extended expected limit past previous ATLAS analysis
- 6 PNN exhibited bias towards lower masses, whereas maxout model achieved a more balanced sensitivity
- 7 LWTA layer's increase in long-term memory is promising in future analysis where higher masses are studied

References

-  Maximilien Brice.
‘Installing the ATLAS calorimeter. Vue centrale du détecteur ATLAS avec ses huit toroides entourant le calorimètre avant son déplacement au centre du détecteur’.
<https://cds.cern.ch/record/910381>
Figure on front page
-  Joao Pequenao.
‘Event Cross Section in a computer generated image of the ATLAS detector.’.
<https://cds.cern.ch/record/1096081>
Figure on slide 3
-  ATLAS Collaboration.
‘ATLAS event at 900 GeV - 6 May 2015 - Run 264034 lb 659 event 11526514’.
<https://cds.cern.ch/record/2015238>
Figure on slide 4
-  ATLAS Collaboration [3].
‘Search for chargino–neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector’.
<http://arxiv.org/abs/2106.01676>



William Hirst



**Application of Supervised Machine Learning
to the Search for New Physics in ATLAS data**

A Study of Ordinary Dense, Parameterized and Ensemble Networks and their Application to High Energy Physics