

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 26, 2023

1 Introduction & Motivation

2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

1 Introduction & Motivation

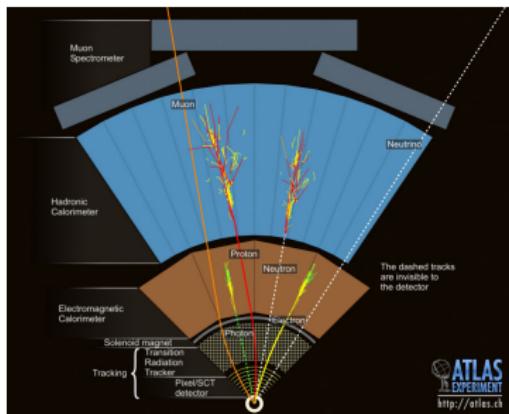
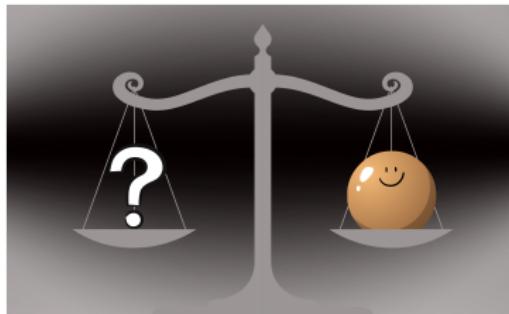
2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

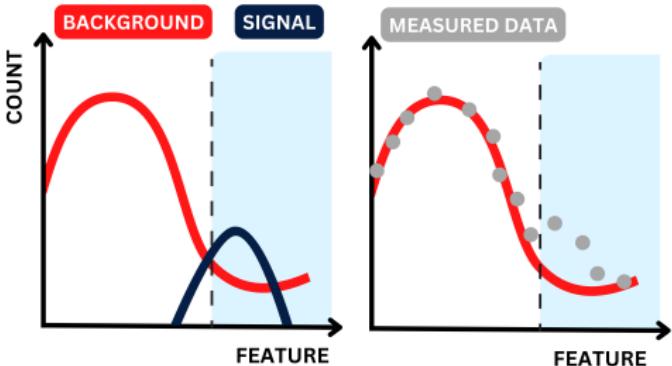
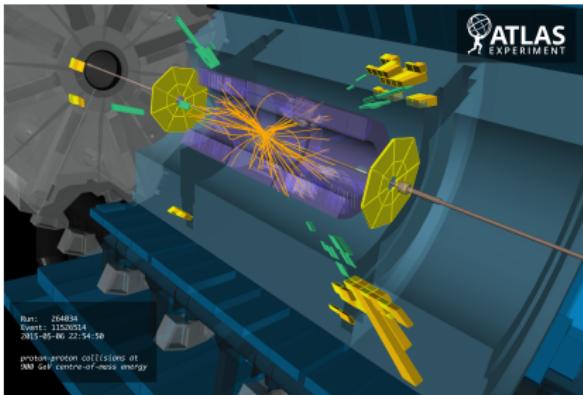
Why apply machine learning to HEP problems?

- 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
 - Data analysis



How do we search for new physics?

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



Aim

- 1 Signal → background
- 2 Use ML to separate signal from background
- 3 Measure performance in Z_{exp}
- 4 Study and compare

The Implementation

1 Introduction & Motivation

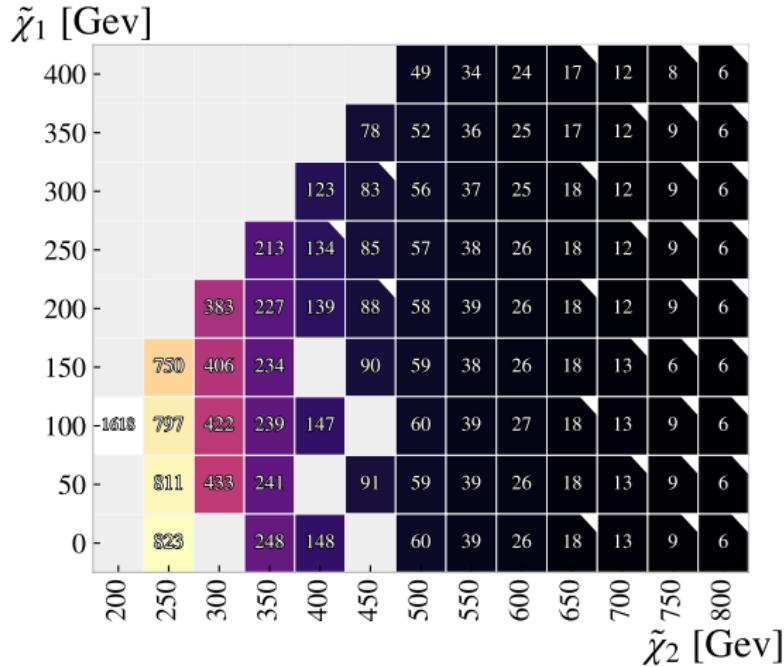
2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

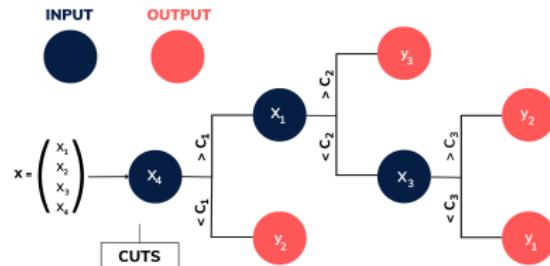
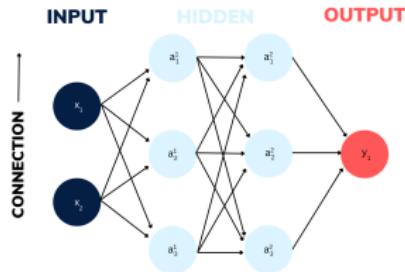
The SUSY signal

- 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

- Classification
 - 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

Methods & Results

1 Introduction & Motivation

2 The Implementation

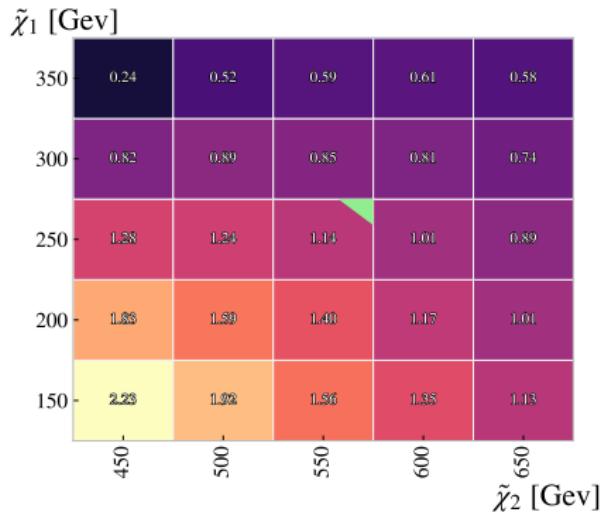
3 Methods & Results

4 Conclusion & Outlook

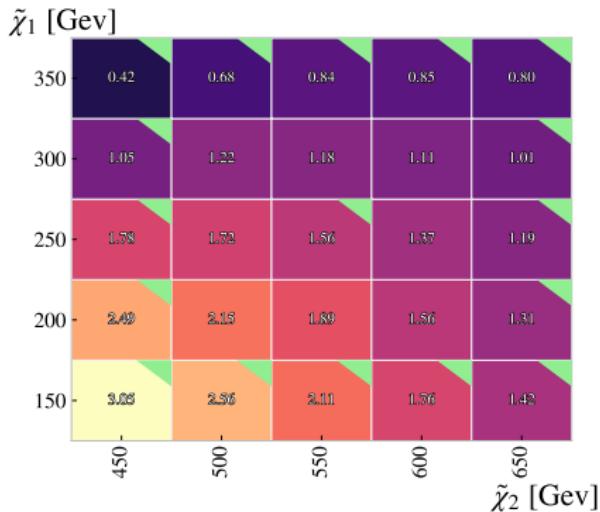
'Traditional' dense neural network



Compare one-mass approach to several-masses approach



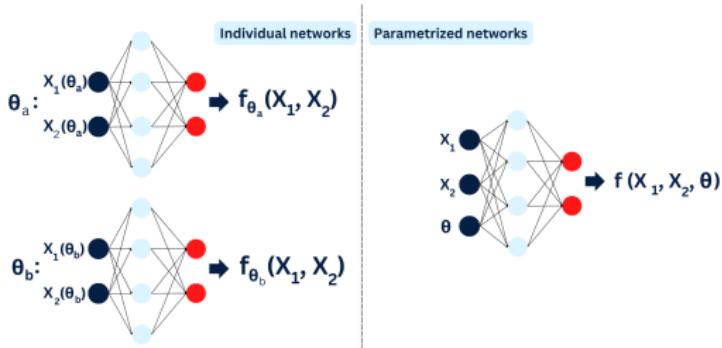
One-mass-model



Several-mass-model

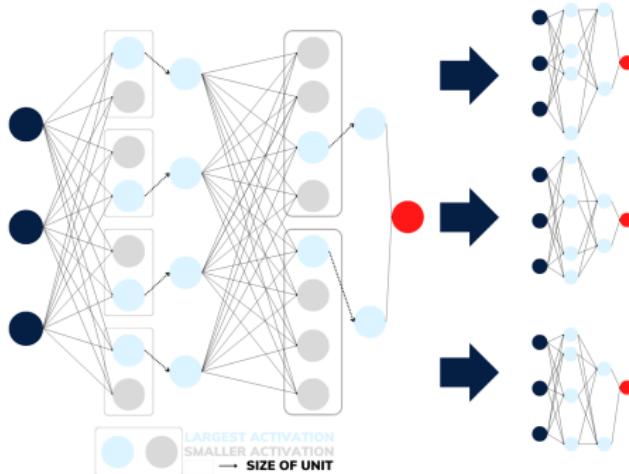
Parameterized neural network

- Long-term memory
- PNN → signal includes mass parameter in feature set
- Motivation
 - Network will associate parameters with trends in the data
- Background assigned parameters randomly using same distribution as signal



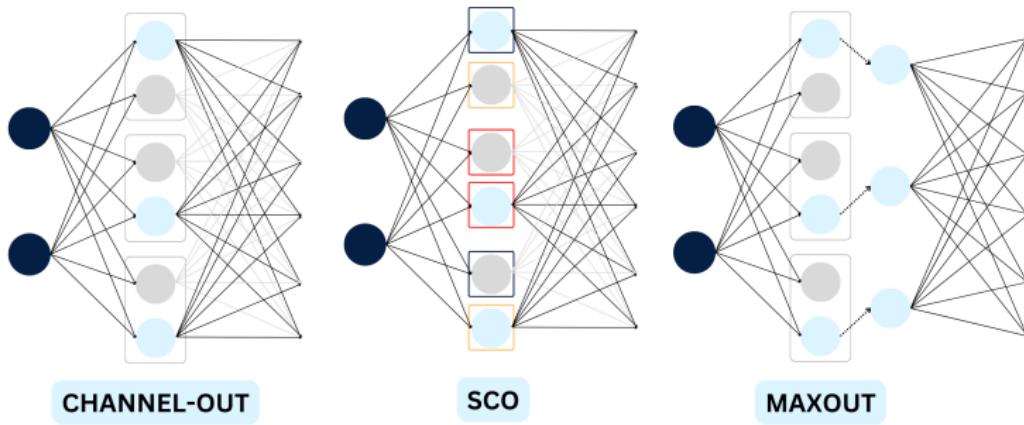
Ensemble methods - LWTA

- Local-Winner-Takes-All
- Competing nodes - Units
- Encode information in pattern specific pathways



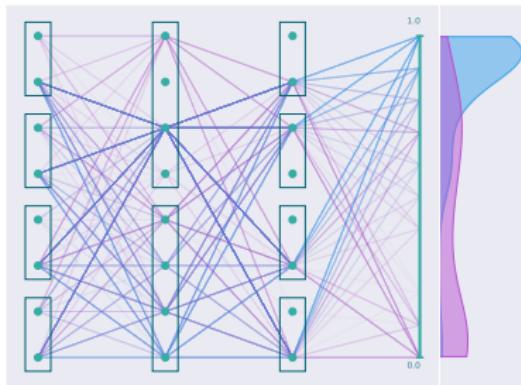
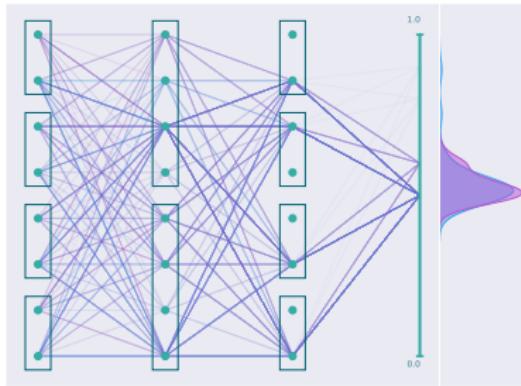
Channel-Out, SCO and Maxout

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



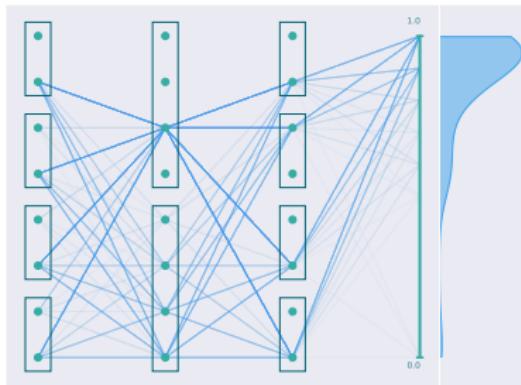
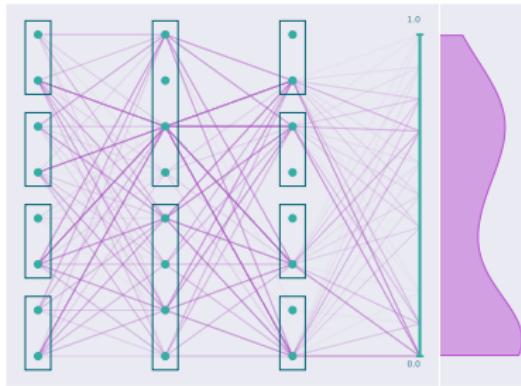
Visualization and study of sparse pathways

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

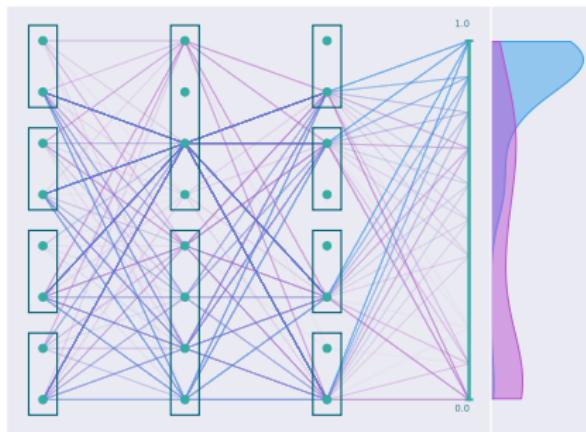


Visualization and study of sparse pathways

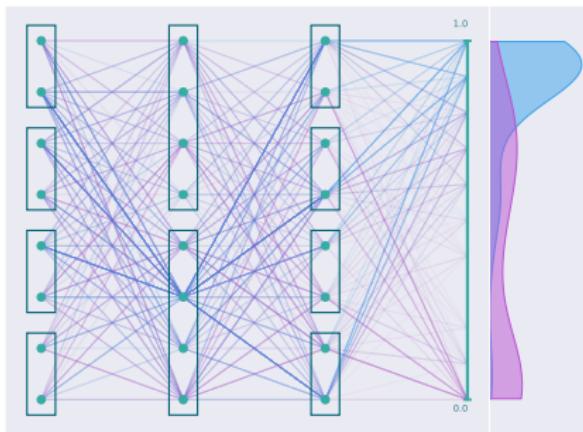
- Visualize the activation and paths of randomly sampled events
- The bolder the line the more frequently the path is used.
- Color of lines
 - Pink: SM background
 - Blue: SUSY signal



Comparing activation of Maxout with SCO



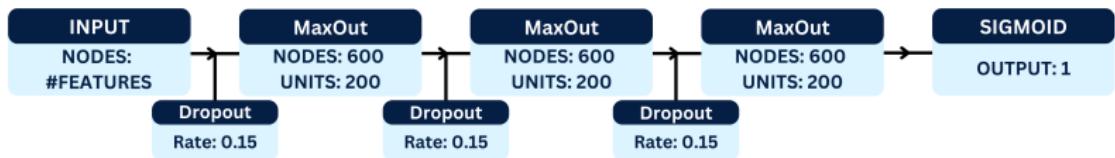
Maxout



SCO

Ensemble network architecture

MaxOut:

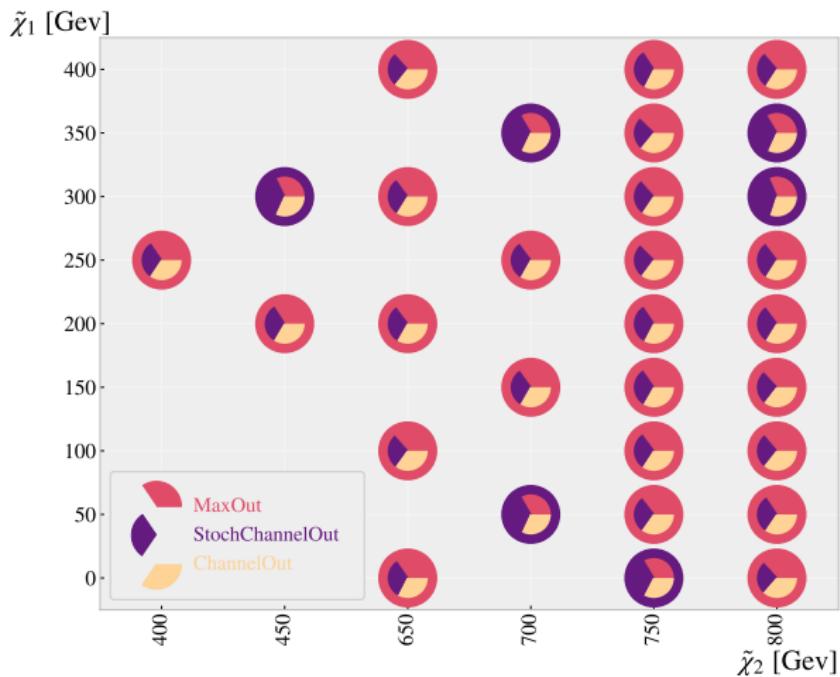


Comparing sensitivity of channel-out, SCO and maxout

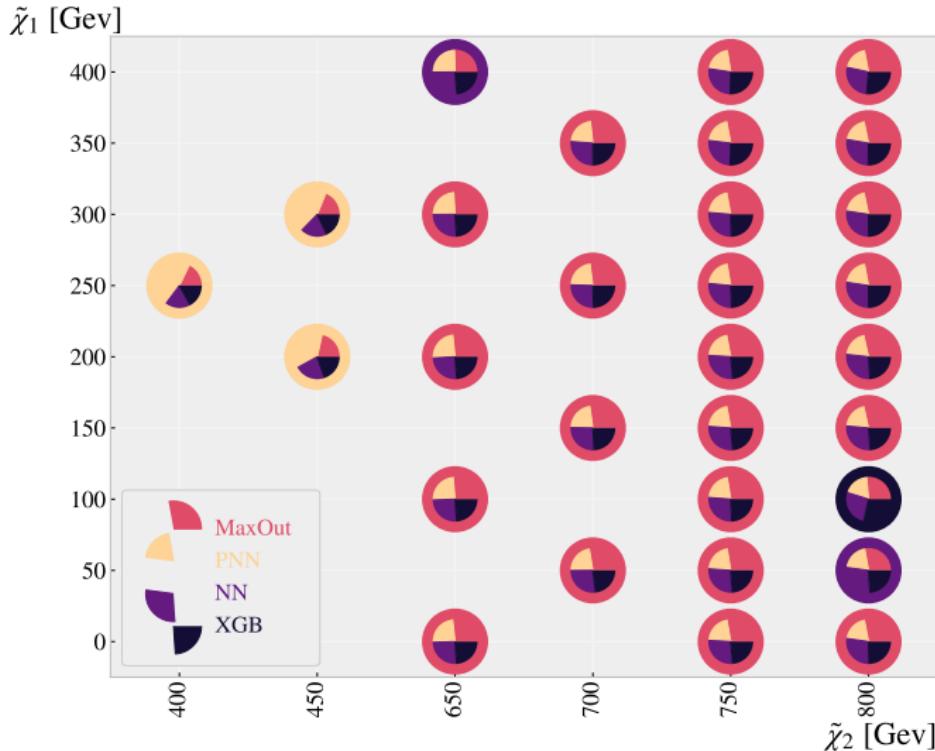
■ Maxout: 24/30

■ SCO: 6/30

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

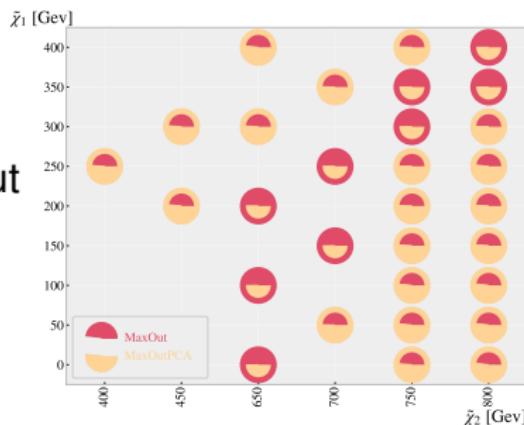


Comparing the sensitivity on a subset of the signal

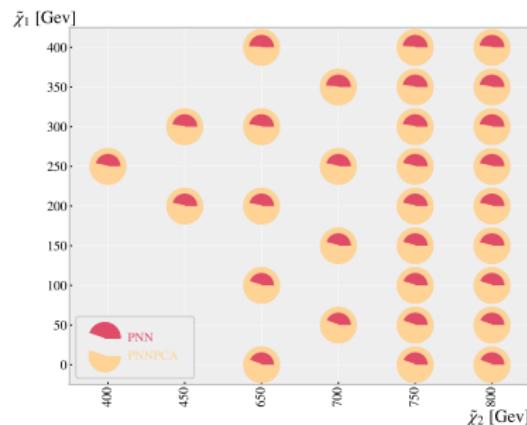


Increasing sensitivity through a PCA

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

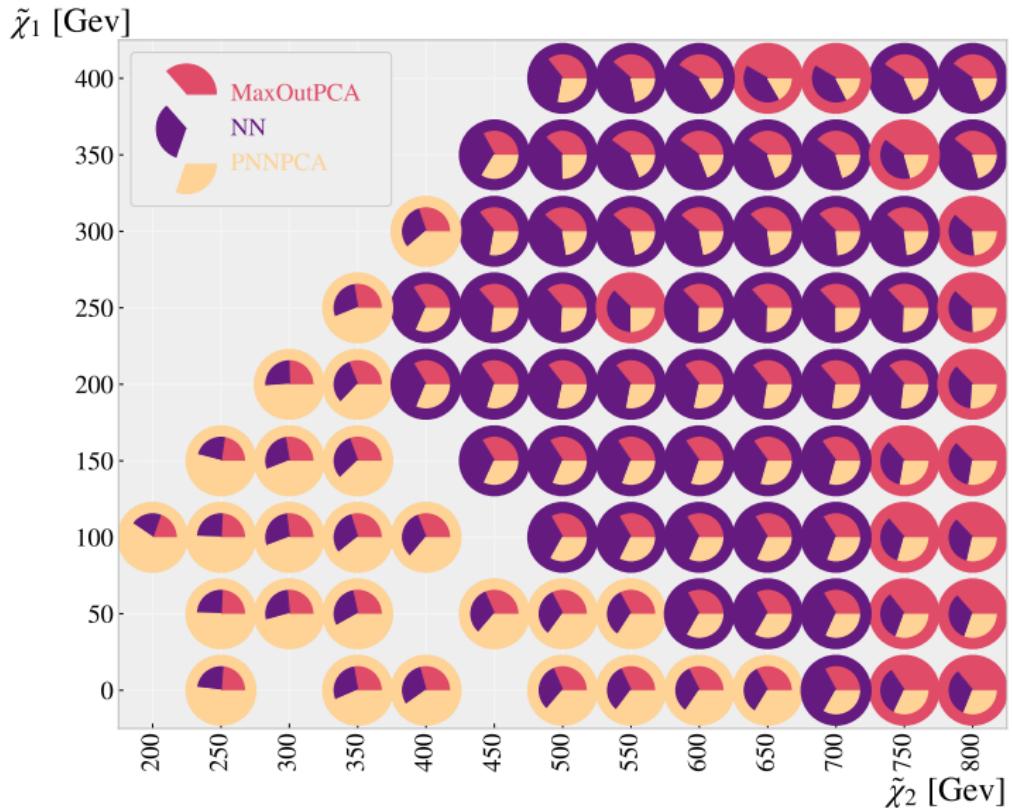


Maxout



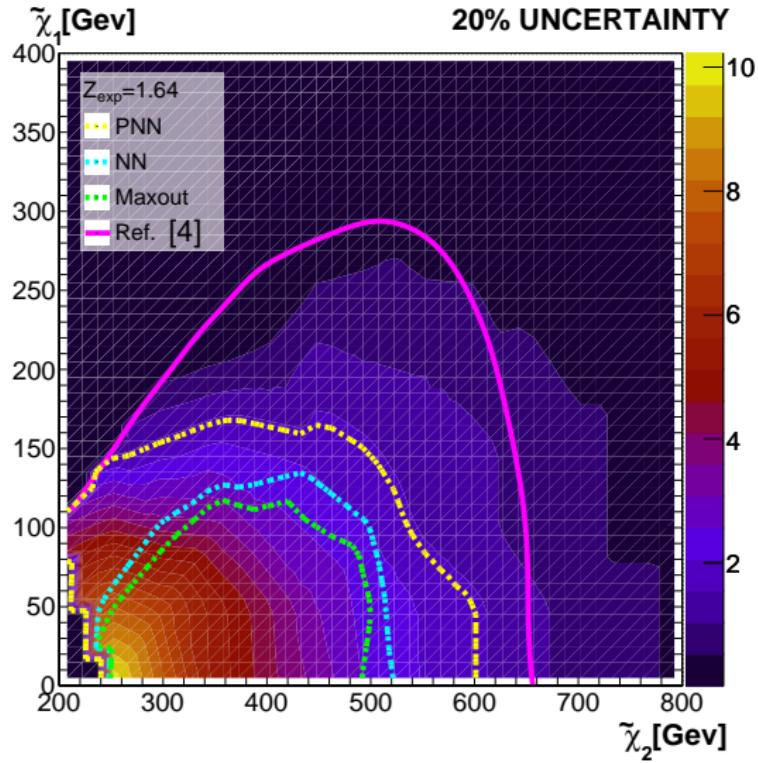
PNN

Comparing methods on full signal grid



Comparing the methods to previous analysis

- Compare the expected limits to analysis made by ATLAS in 2021 [4]



Conclusion & Outlook

1 Introduction & Motivation

2 The Implementation

3 Methods & Results

4 Conclusion & Outlook

Key findings

- 1 Diverse signal → improve performance
- 2 PCA → improve performance of PNN and maxout
- 3 PNN bias towards smaller masses
- 4 Maxout achieved balanced performance
- 5 SCO

The way forwards

- 1 More advanced analysis of ML output
- 2 LWTA promising (SCO)
- 3 Combine PNN and LWTA

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 [4].
‘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