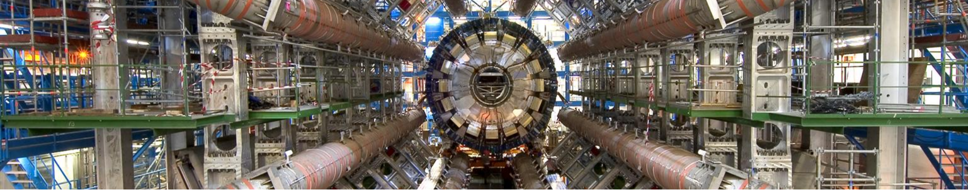


UiO : Department of Physics
University of Oslo

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

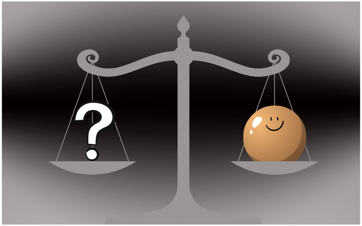
William Hirst

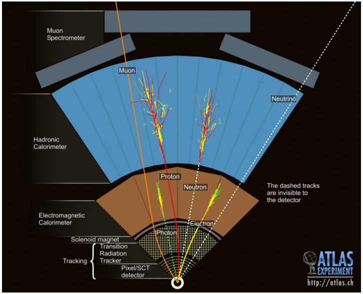
May 24, 2023

Outline

Why apply machine learning to HEP problems?

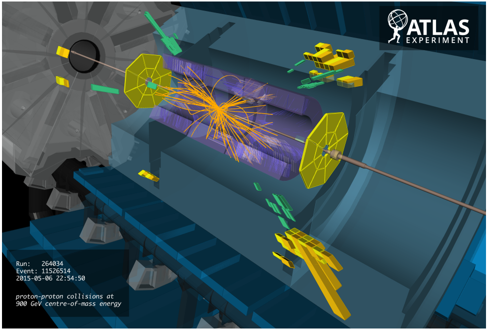
- The standard model (SM) of particle physics is very successful, but not complete
  - Neutrino masses
  - Hierarchy problem
  - Energy-matter density in the universe
- Testing requires progressively larger amounts of data
- Machine learning (ML)
  - Event reconstruction
  - Particle classification
  - Creating search regions

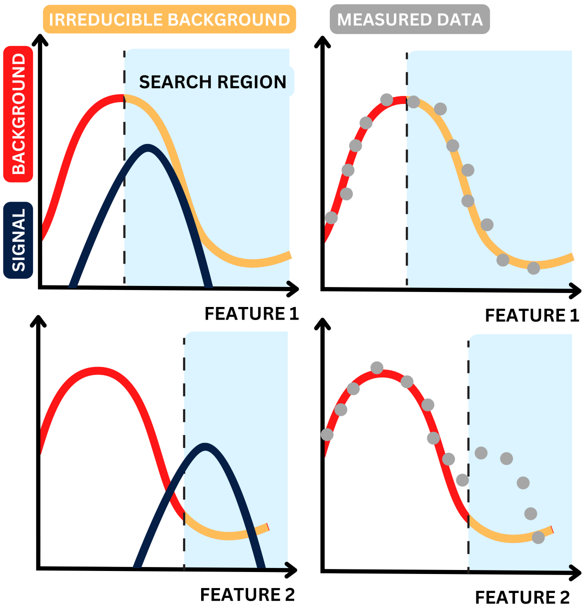




How do we search for new physics?

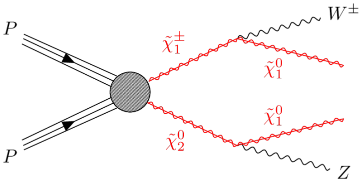
- Compare theory with experiment
  - Experiment: Proton-proton collisions produced at the LHC and measured in the ATLAS detectors
  - Theory: Simulated based on SM physics
- Deviations → New physics (?)
- Measure deviation in significance
  - $Z_{obs} \approx \frac{n_{obs} - bkg_{sim}}{\sqrt{bkg_{sim}}}$
  - $Z_{exp} \approx \frac{sgn_{sim}}{\sqrt{bkg_{sim}}}$




IRREDUCIBLE BACKGROUND
MEASURED DATA
BACKGROUND
SIGNAL
SEARCH REGION
FEATURE 1
FEATURE 1
FEATURE 2
FEATURE 2

This thesis

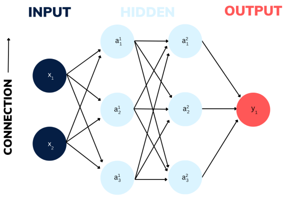
Shed some light on the application of supervised learning in HEP by experimenting and studying a set of ML methods as they search for a set of SUSY signals.

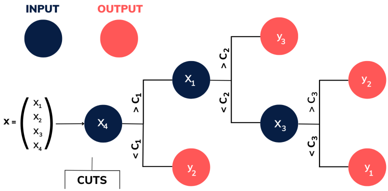


1. Study individual attributes of a set of supervised methods
2. Compare expected sensitivity between methods on a subset of data
3. Attempt to increase sensitivity via feature reduction (PCA)
4. Compare the expected limits achieved by best performing methods to previous ATLAS analysis

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
  - XGBoost



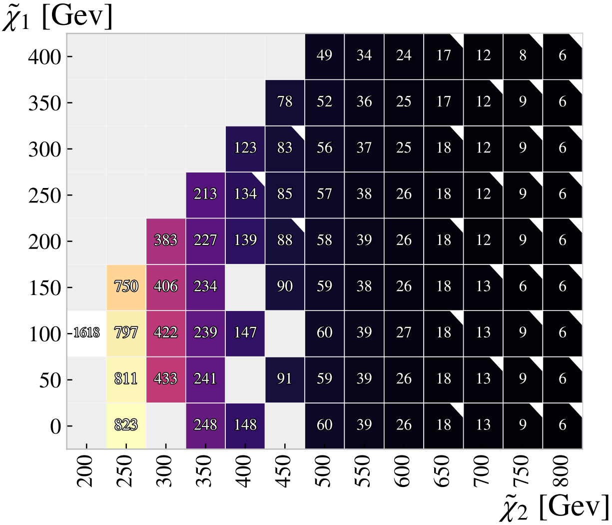


Training strategy

- Train using simulated data
- Objective
  - Background $\rightarrow 0$
  - Signal $\rightarrow 1$
- 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

Mass combinations of the chargino-neutralino pair

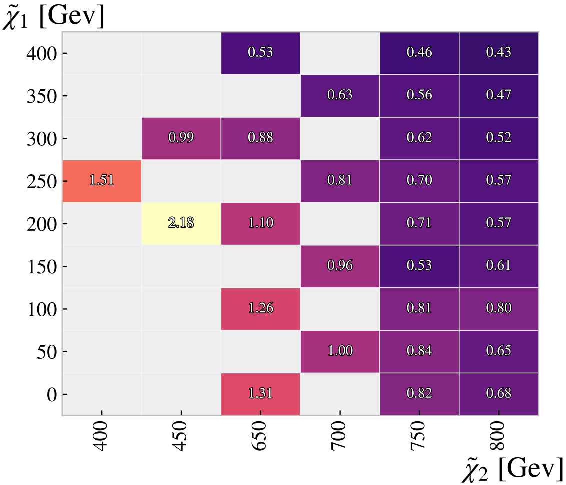
- Full signal grid
  - 89 mass combinations
- Original signal set: white corners
  - 30 mass combinations
- The smaller the masses, the larger the contribution



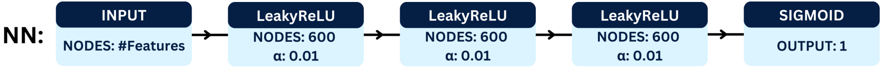
Methods & Results

Boosted decision trees - XGBoost

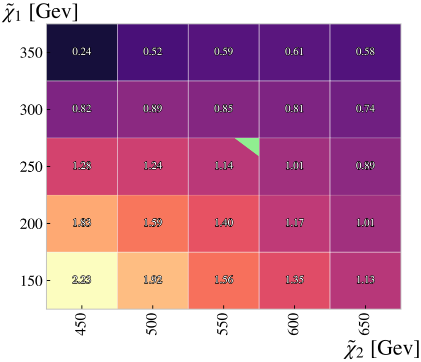
- Used as benchmark
- Trained on original signal set
- Displayed better performance on lower masses

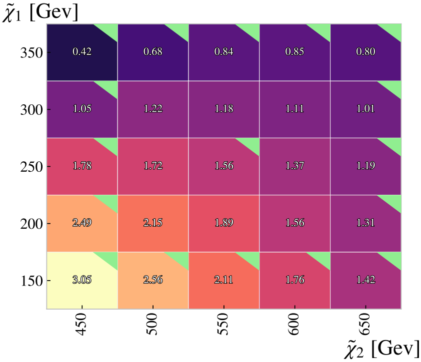


Ordinary dense neural network



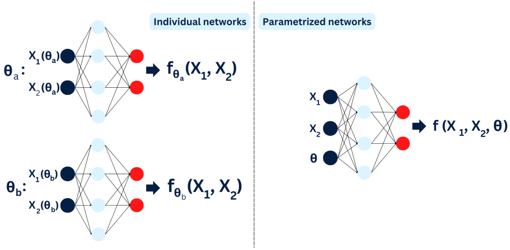
Compare one-mass approach to several-masses approach





Parameterized neural network

- For diverse data set, X, dependent on a parameter, $X(\theta)$
  - Classical approach: One model for each parameter
  - PNN approach: Include θ as feature in feature set
- Signal events using masses $\{A, B\}_{GeV}$ to generate event during simulation will include the parameters A and B in feature set
- Background assigned parameters randomly using same distribution as signal
- Motivation
  - Network will associate parameters with trends in the data



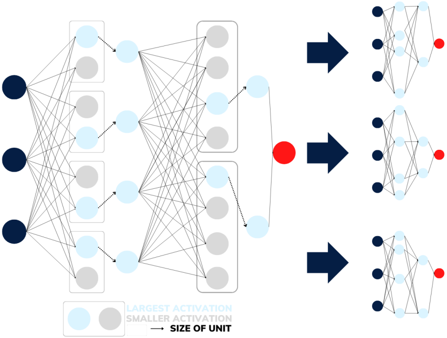
Study the effect of the parameters in the PNN

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

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

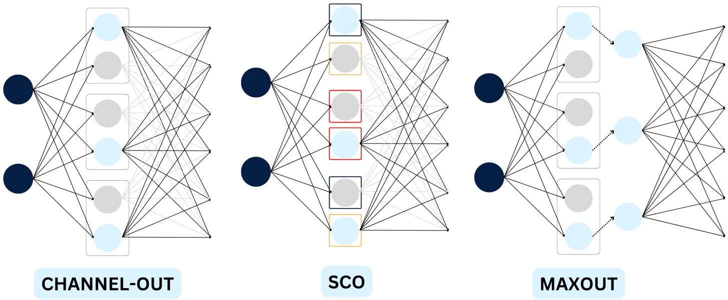
Ensemble methods - LWTA

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



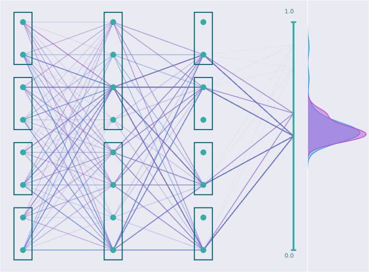
Channel-Out, SCO and Maxout

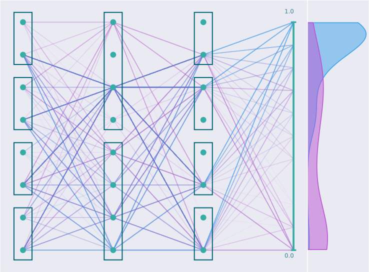
Layer	Separate weights	Static units
Channel-Out	✓	✓
SCO	✓	X
Maxout	X	✓



Visualization and study of sparse pathways

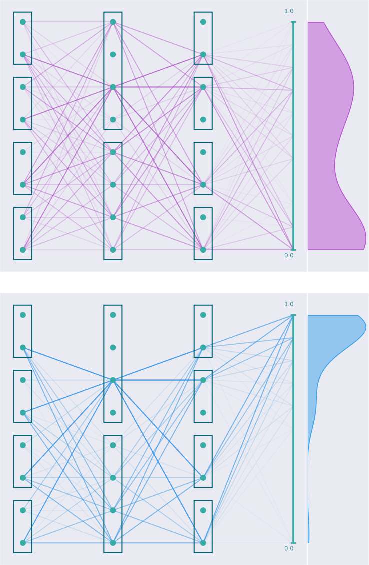
- Visualize the activation and paths of 100 randomly sampled events
  - 50 background
  - 50 signal
- The bolder the line the more frequently the path is used.
- Color of lines
  - Pink: SM background
  - Blue: SUSY signal



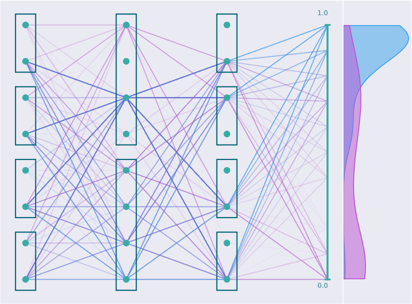


Visualization and study of sparse pathways

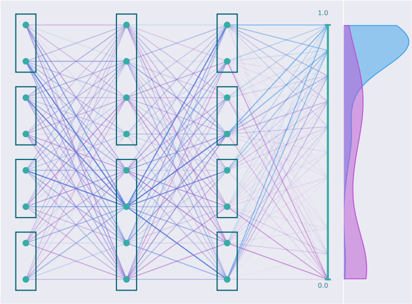
- Visualize the activation and paths of 100 randomly sampled events
  - 50 background
  - 50 signal
- The bolder the line the more frequently the path is used.
- Color of lines
  - Pink: SM background
  - Blue: SUSY signal



Comparing activation of Maxout with SCO



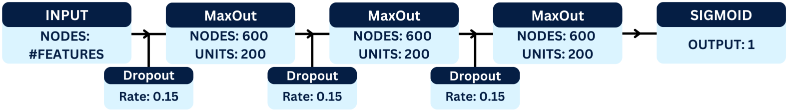
Maxout



SCO

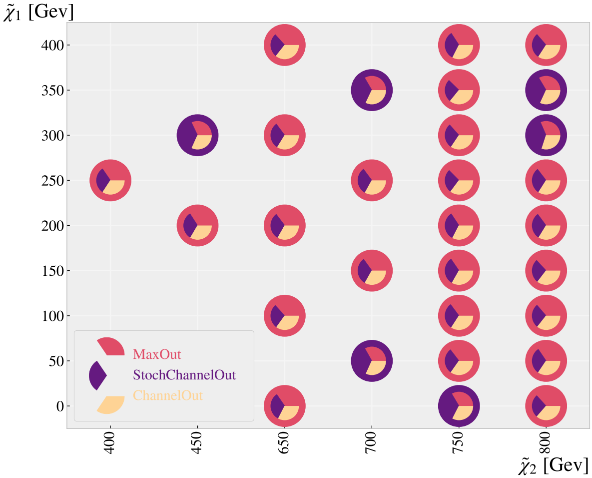
Ensemble network architecture

MaxOut:



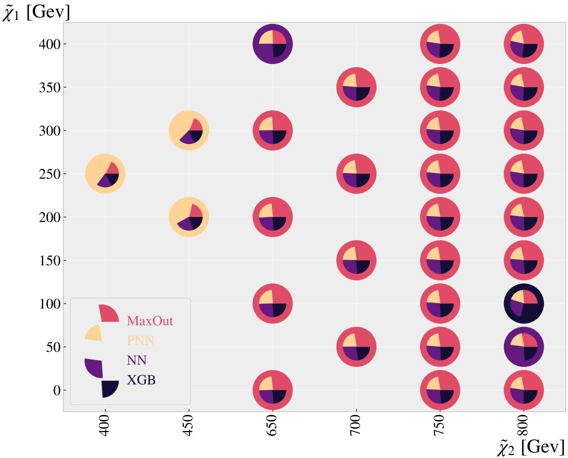
Comparing sensitivity of channel-out, SCO and maxout

- 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

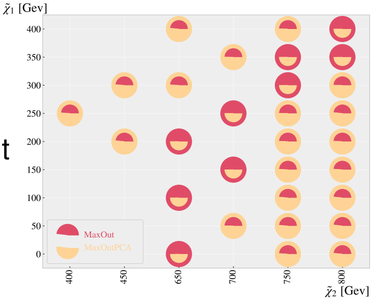
- NN variants outperform BDT
- Maxout model achieves the highest significance on most masses
- PNN very sensitive for low masses
- Maxout (relatively) sensitive for high masses
  - Long-term memory

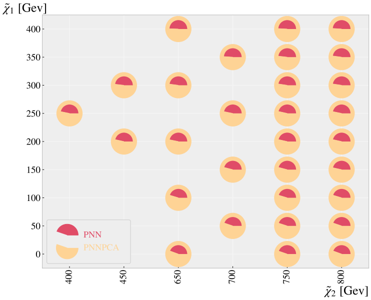


Increasing sensitivity through a PCA

- Dimensionality reduction
- Creates new features using linear combination of original features
- Ranks from most to least variance
- This analysis
  - Demand conservation of 99.9% of variance/spread
  - 5 features removed

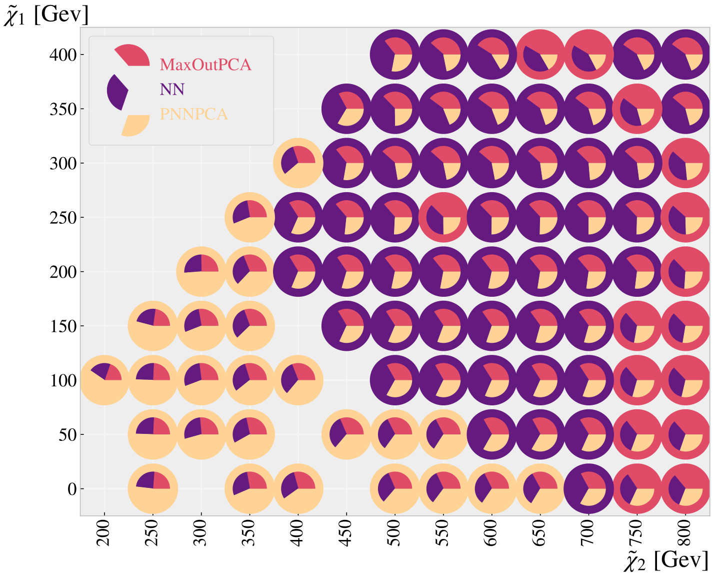
Maxout





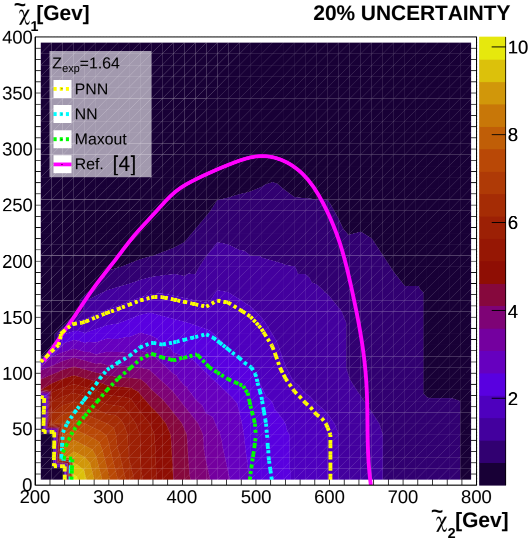
PNN

Comparing methods on full signal grid



Comparing the methods to previous analysis

- Compare the expected limits of three best models to analysis made by ATLAS in 2021 [4]
- Introduce flat uncertainty for realistic comparison (20%, 10%, $< 1\%$)
- Include top performing methods
  - Maxout model with PCA
  - PNN with PCA
  - Ordinary dense neural network without PCA



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

UiO : Department of Physics
University of Oslo

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