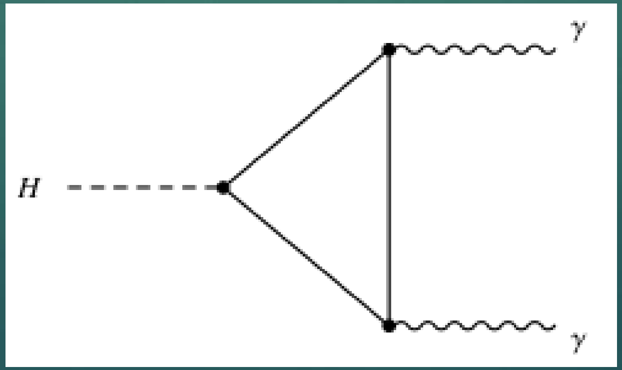
Search for an additional Higgs boson with mass <110GeV and scale energy extraction with the CMS experiment at the LHC

Benjamin Massoteau

Supervisé par Suzanne Gascon

PhD Days 2025 : Groupe CMS



Summary :

anr

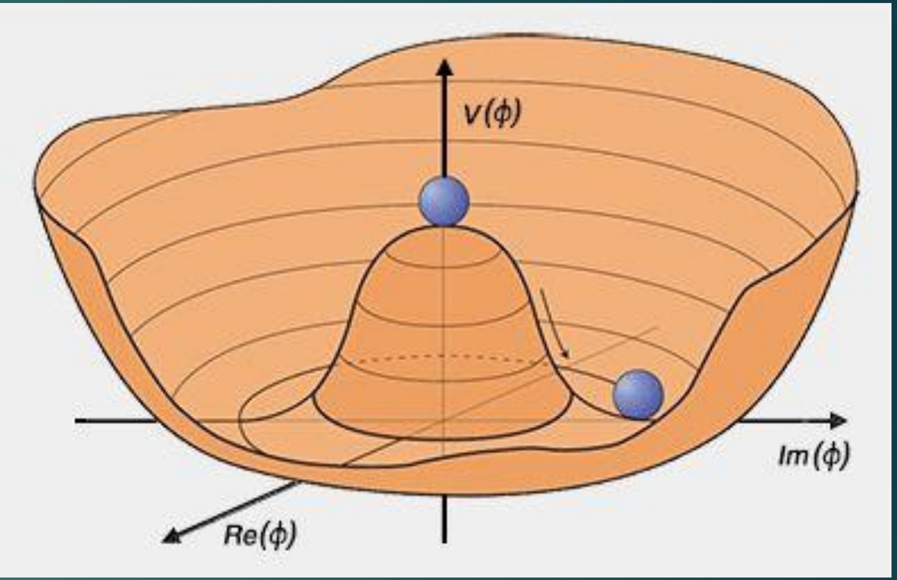
Projet HIGGSENLIGHTENED

The Higgs boson

The Higgs mechanism is responsible for the spontaneous breaking of electroweak symmetry, allowing the W and Z bosons to acquire mass, as well as the fermions, thereby mixing their right-handed and left-handed components.
The Higgs potential is as follows:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$$

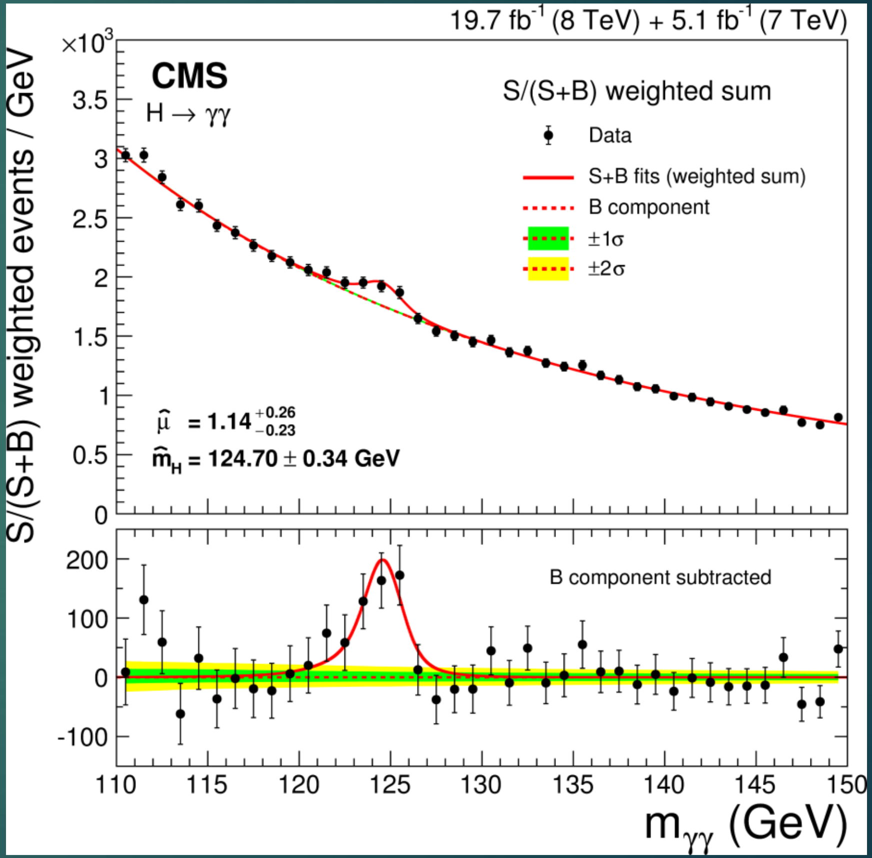
With $\mu^2 < 0$,the minimum of the potential is located at $v = \sqrt{-\frac{\mu^2}{\lambda}}$
This minimum is called the vacuum expectation value (vev). If we expand the Lagrangian of the scalar field around this vev, we notably find a mass term $\frac{m_h^2}{2}h^2$ with $m_h^2 = 2\mu^2$, reflecting the appearance of a scalar boson, a relic of this mechanism.



A Higgs boson discovery at the LHC

Physics Letters B, Vol. 716, Issue 1, 2012, Pages 30-61

In 2012, the discovery of a scalar boson, compatible with the Higgs boson of the standard model, was announced by the ATLAS and CMS experiments at CERN. This discovery was observed in the diphoton decay channel with a significance greater than 4.5σ and greater than 5σ when combined with other channels (ZZ,WW,ττ) providing strong evidence of its existence. This marked a significant milestone in completing the Standard Model of particle physics, confirming the mechanism responsible for particle mass acquisition as predicted by the Higgs mechanism.



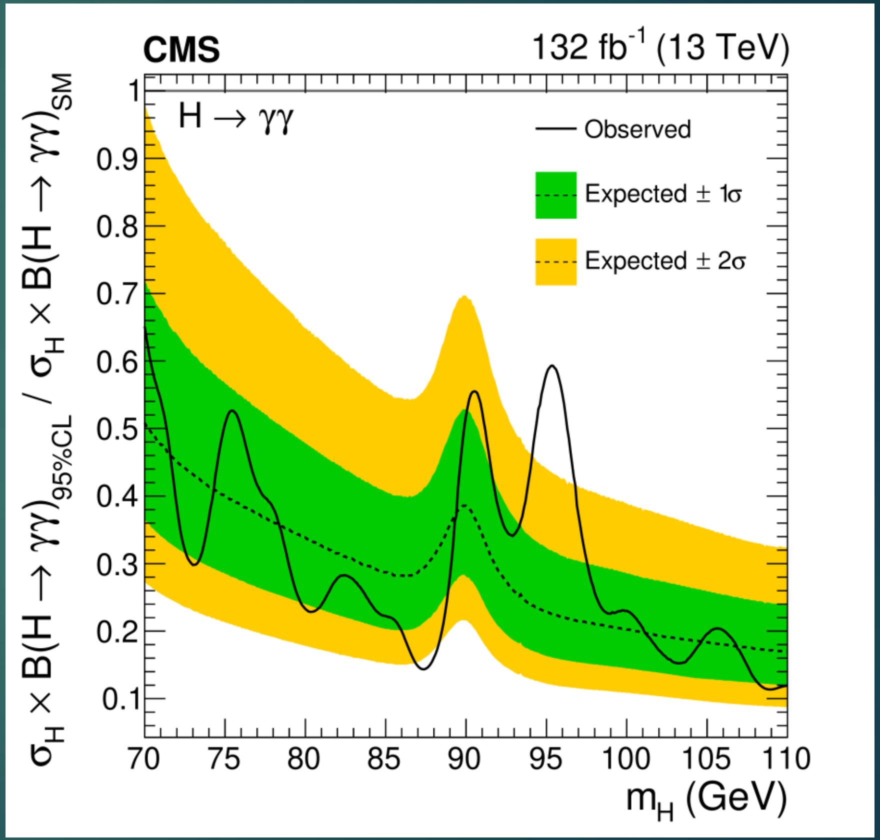
Motivation for the search for an additional Higgs boson

The Standard Model, however, fails to predict and explain certain phenomena:

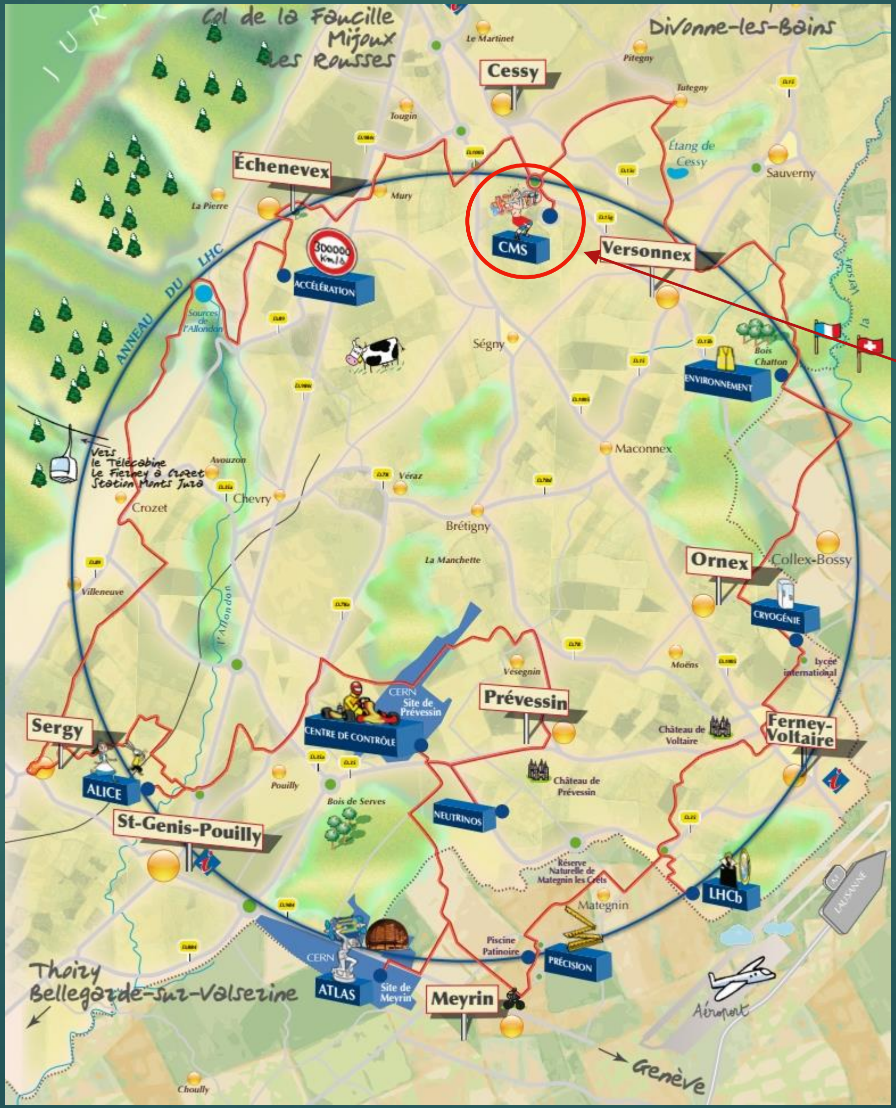
•Dark matter, which is thought to constitute 24% of our universe.
•Gravity at the quantum scale.
•Neutrino oscillations.
•The matter-antimatter asymmetry.
$\Rightarrow$ There must exist a theory beyond the Standard Model that can explain all observed and unexplained phenomena.

Certain models, such as two-Higgs-doublet models or supersymmetry, predict additional Higgs bosons with masses different from the one detected in 2012.
An excess was notably detected around 95.4 GeV, measured with a local significance of 2.9σ by CMS, which could correspond to an additional Higgs boson.
Meanwhile, ATLAS measured a significance of 1.7σ around the same mass https://cds.cern.ch/record/2904053

CMS Collaboration. *Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at √s=13 TeV*. Phys. Lett. B 860 (2025) 139067. https://cds.cern.ch/record/2852907.

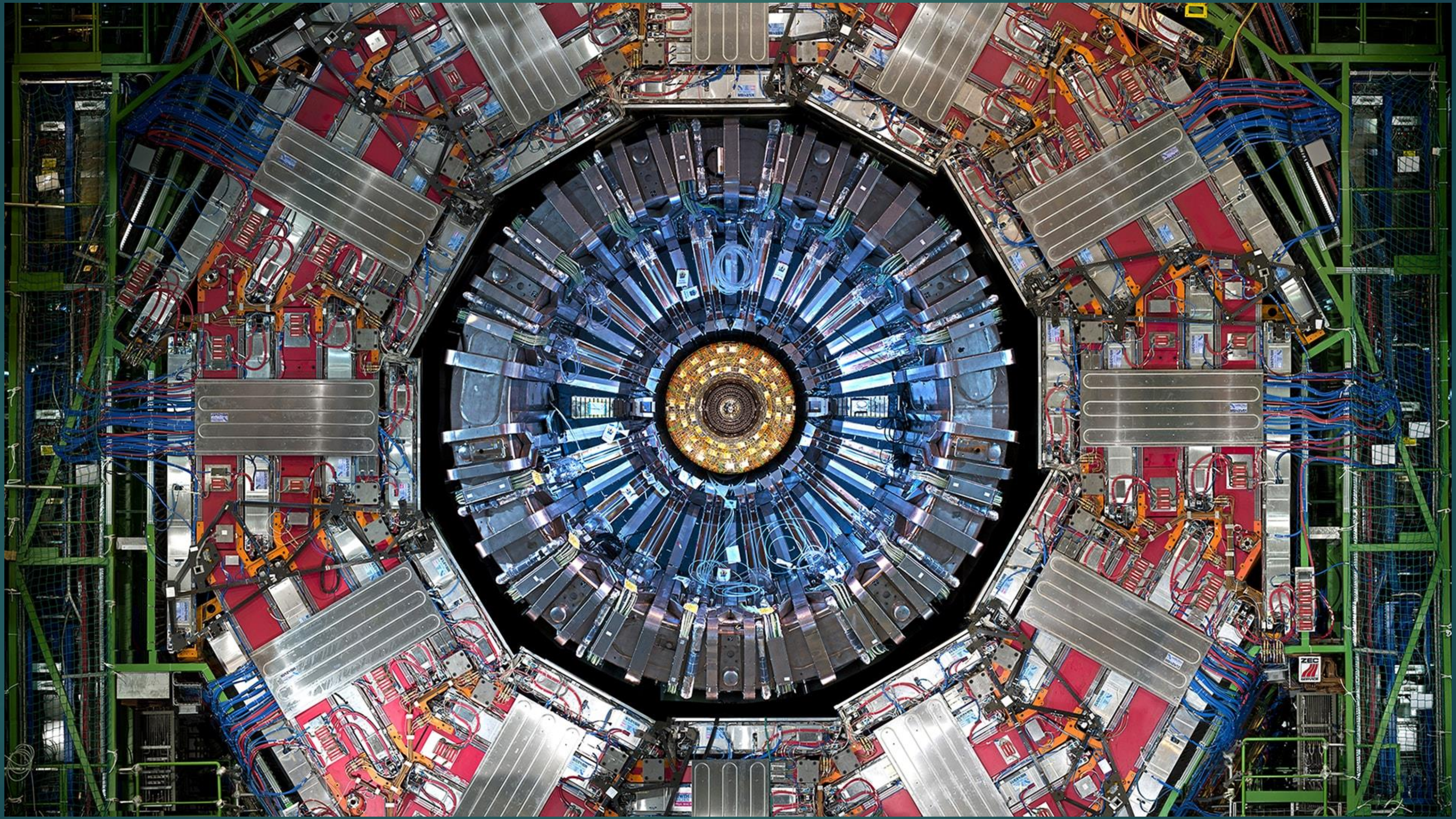


CMS experiment localisation

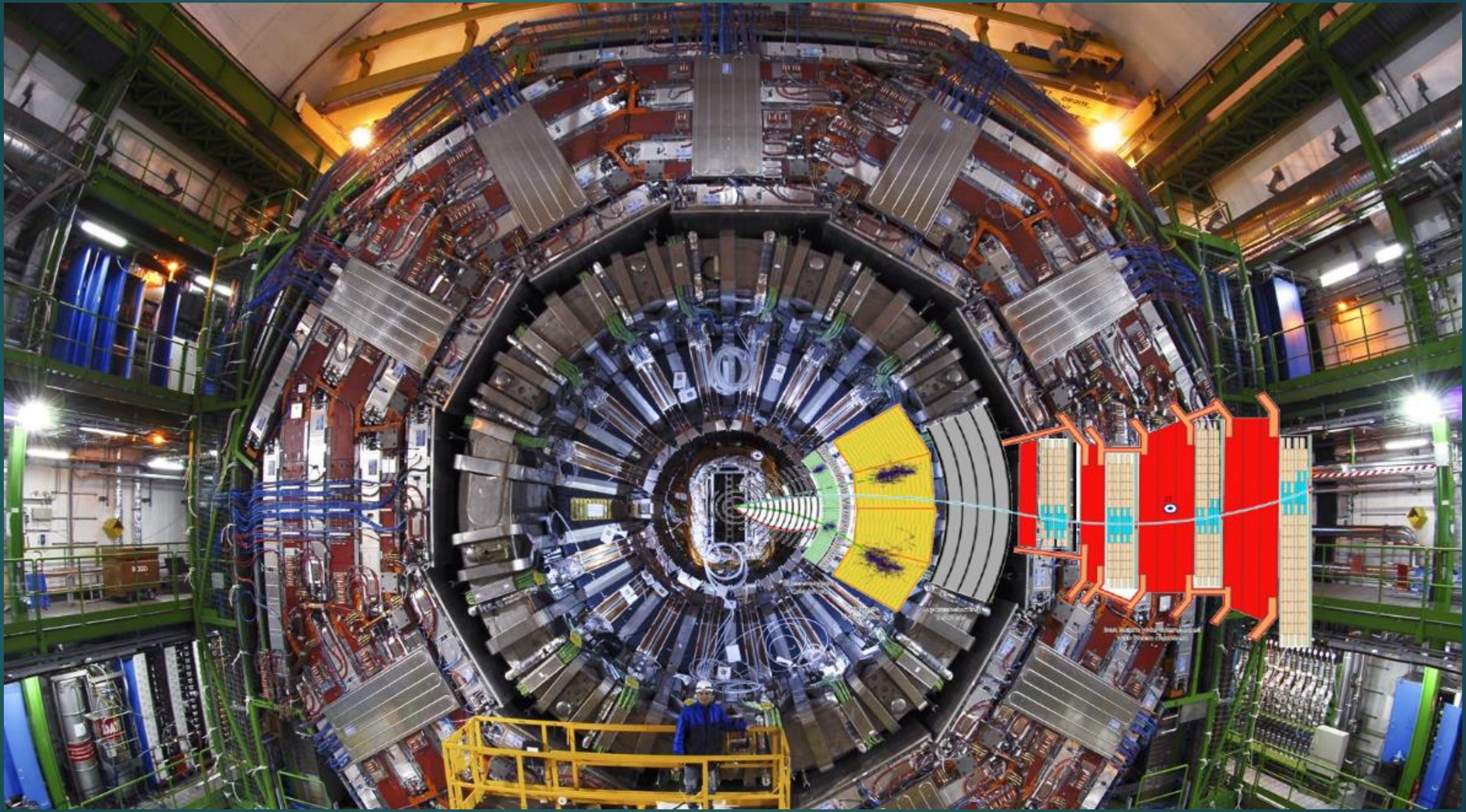


The farthest point from CERN's main site (~20 minutes by car)

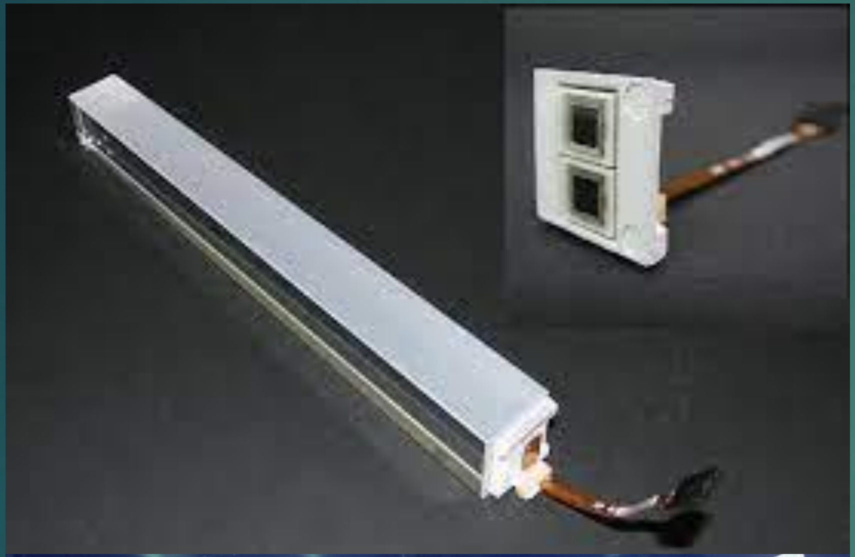
The CMS experiment



The CMS experiment

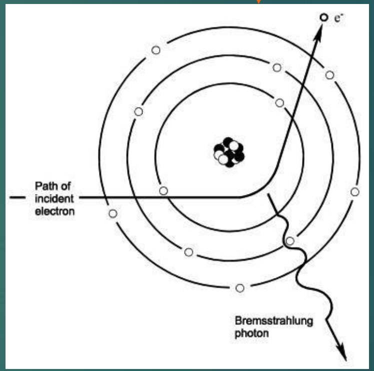


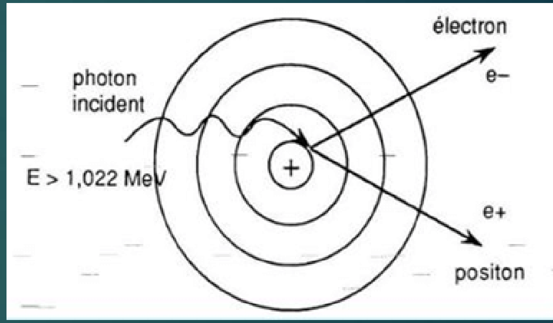
The electromagnetic calorimeter



The high energy resolution electromagnetic calorimeter (ECAL) measures the energy of photons and electrons. It is composed of $PbWO_4$ (lead tungstate) crystals.

Electron : Bremsstrahlung ; Photon : Pair production





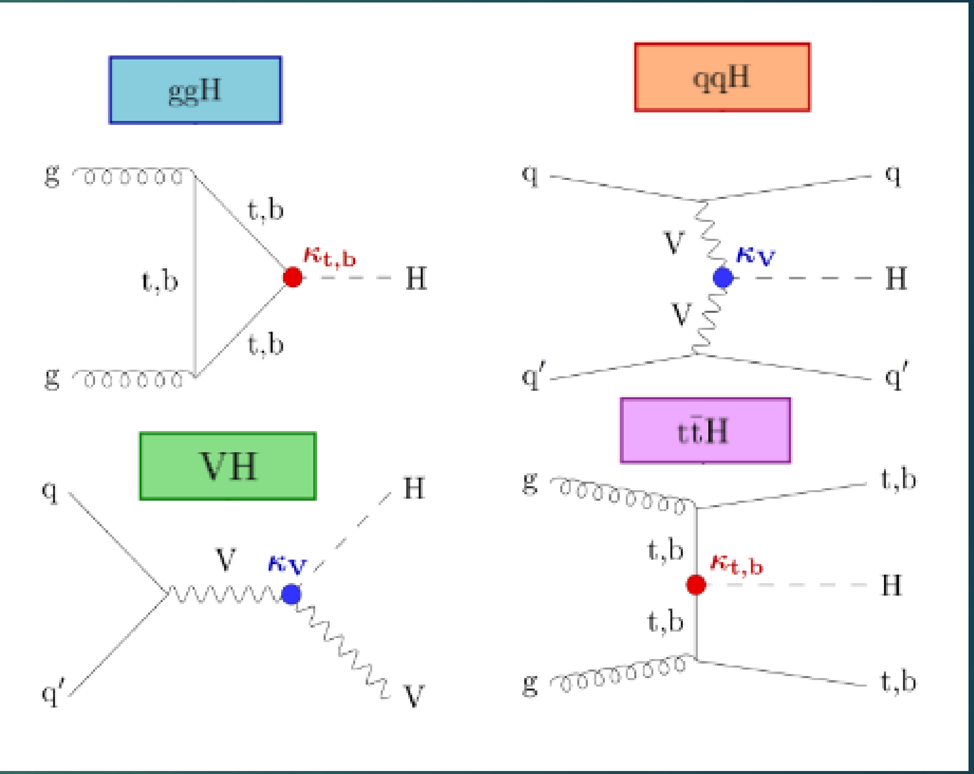
The scintillation light at the end of the particle shower is then detected by photodetectors, allowing the reconstruction of the energy of the incident particle.

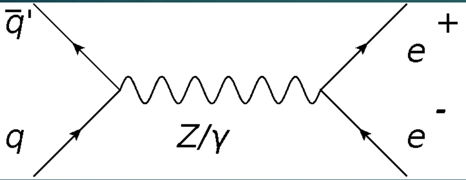
Low mass « standard model-like » h->γγ analysis

- Production mode : ggH, VBF, VH, et ttH

- Diphoton invariant mass :
  - Signal search range: 70-110 GeV
  - Background fiting range: 65-120 GeV
    (Limited by the trigger bandwidth)

- Backgrounds :
  - Irreducible background producing γγ (QCD)
  - γ+jets , jets+jets
  - Drell-Yan : Z/γ*->ee





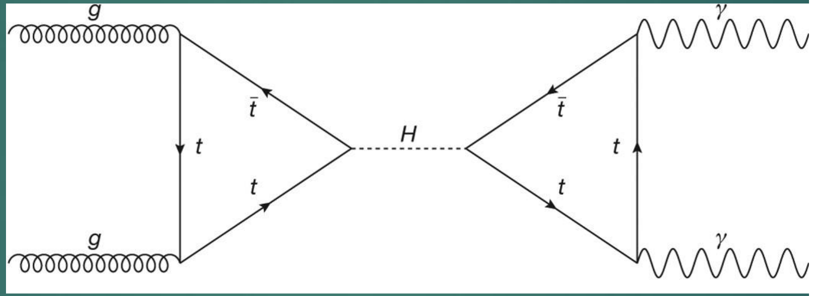
Problem of the Drell-Yan process reduction

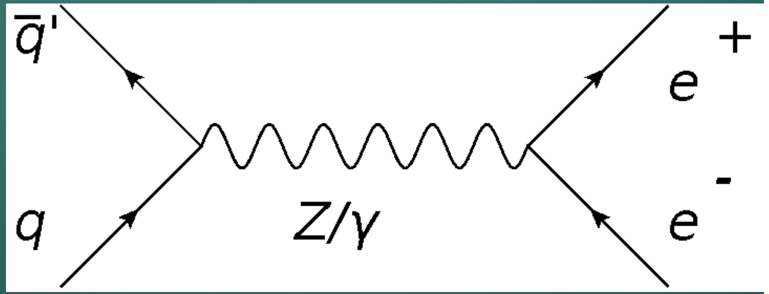
To discriminate photons and electrons in the CMS experiment, both information from tracker and ECAL are used.

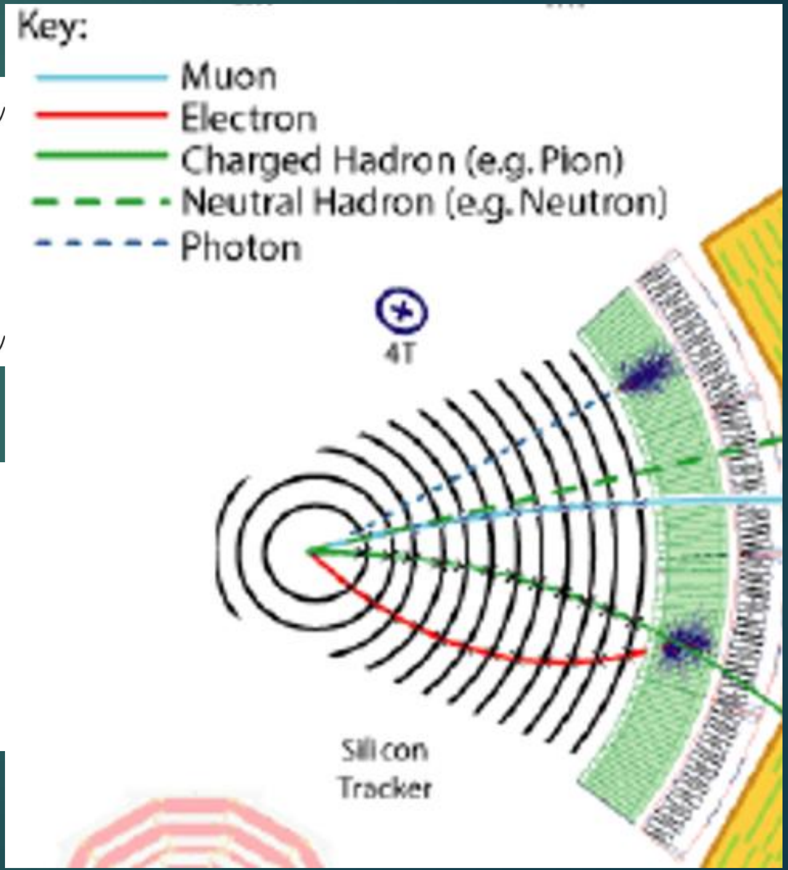
Electron : charged particle that leaves an inner track

Photon : neutral particle that do not have such tracks

Some ineffiencies in the tracker lead to misinterpretation of the electrons as photons, especially coming from the Drell-Yan process (DY).







Problem of the Drell-Yan process reduction

https://cds.cern.ch/record/2852907

(Work started during my M2 internship during the first semester of 2023)

Run 2 (2016-2018) solution : add two selection criteria to preexistent pixel veto

Linear cut in 2D between p_T (transverse momentum) and $\log(\sum p_T^2)$ (Sum of the squares of the transverse momenta of the tracks in the chosen vertex) such that :

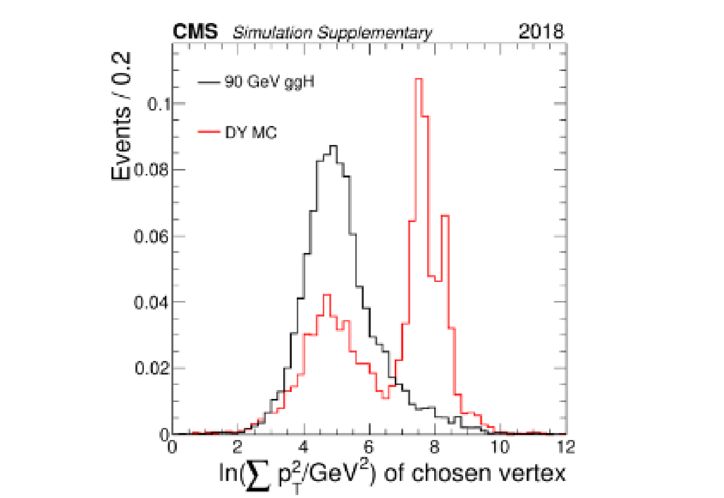
$$\log(\Sigma p_T^2) \leq 0.016 p_T^{\gamma\gamma} + 6.0$$

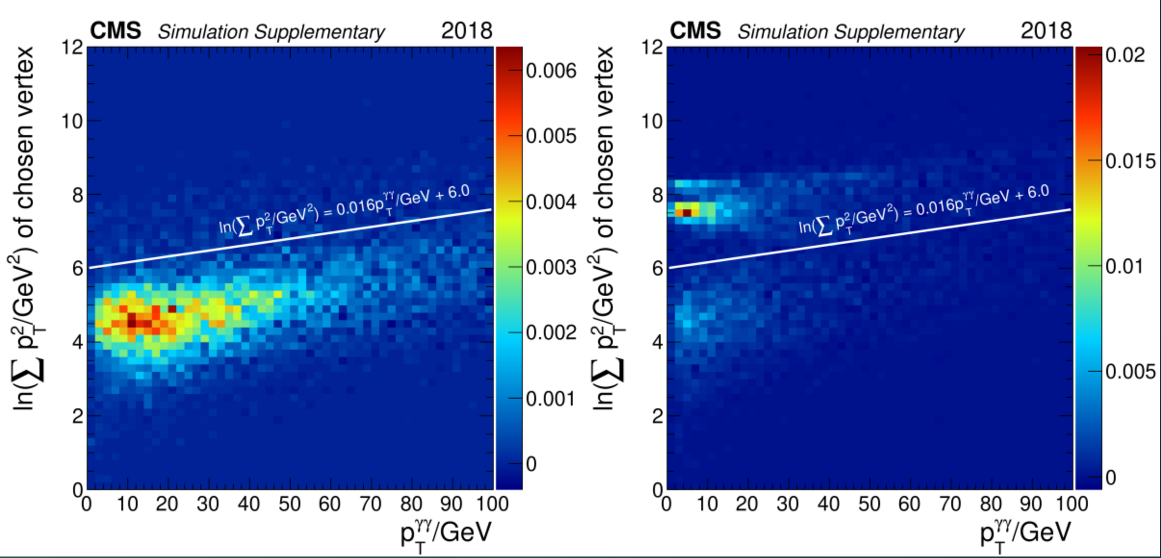
With the additional selection $N_{MatchedEle} = 0$

In this way, we kept ~92% of the signal and ~30% of the DY background.

Challenges for Run 3 :

- Data format has changed from MicroAOD to NanoAOD since Run 3 (2022-2026) and do not have the variables $\log(\sum p_T^2)$ and $N_{MatchedEle}$
- We need the same results as in Run 2, or even better.





Trials with Run 3 variables on Run 2 simulated events

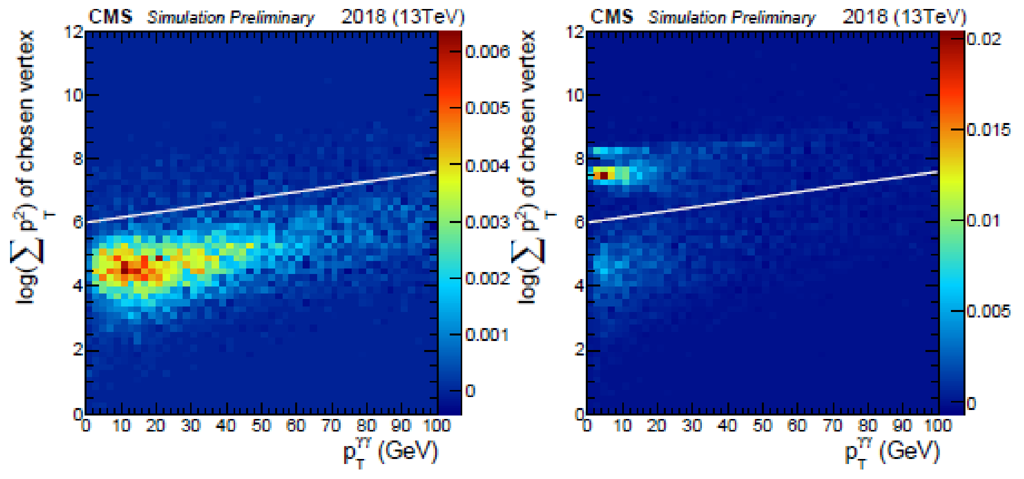
Here, electronIdx indicates whether a photon originates from an electron or not (in the later case, the variable is set to -1).
⇒ Shown to be equivalent to $N_{matched} = 0$

PV_score corresponds to the sum of the transverse momenta of the clustered objects (Not only tracks).

We can try to do a linear cut.

With this method, the efficiency is :
- Signal efficiency= 93%
- DY efficiency= 36,4%

Not quite as good as in Run 2



(h) 2018 90 GeV ggH (i) 2018 DY MC

Neural Network

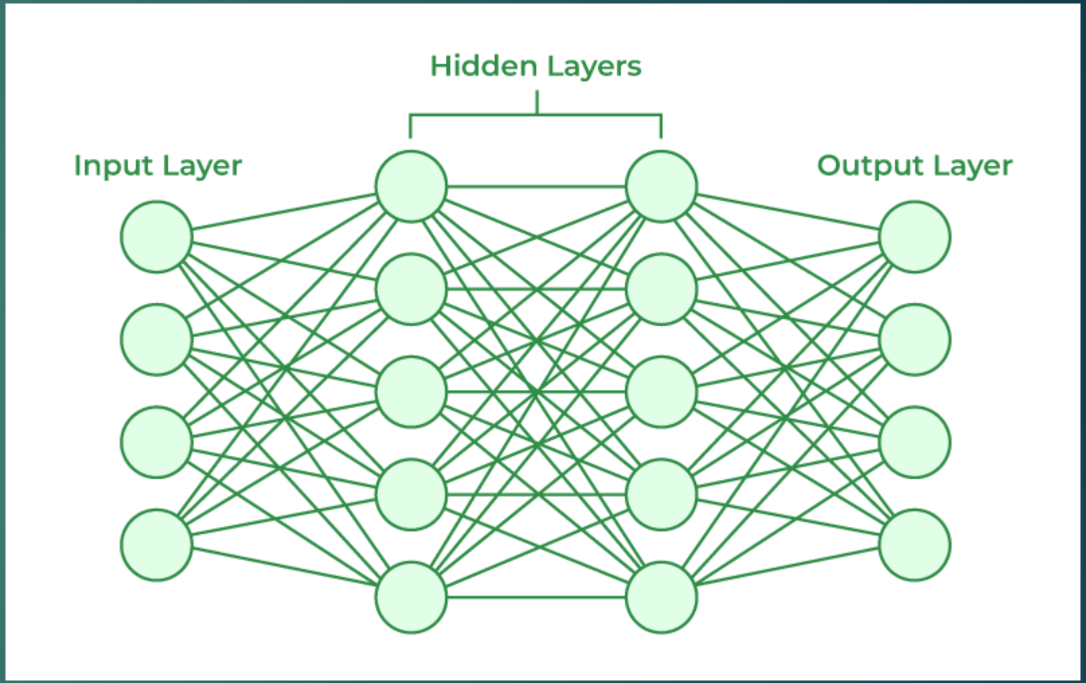
PyTorch

To discriminate the signal from our background, we're investigating using a neural network that can gives us the probability of an event to belong to one or another.

We have a training set and a test set of events that we label (signal = 1 ; background =0).

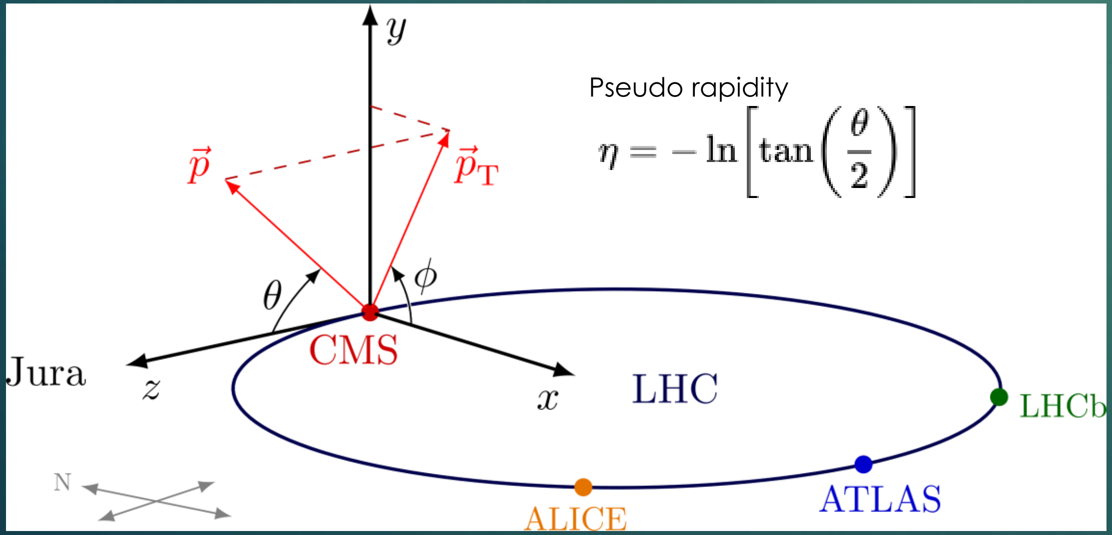
Each event will go into a unit called a neural, which is a matrix filled with scalars called weights.

The output is a probability of an event to be signal or background.



Neural Network

The list of variables used comes from the photon kinematics (eta,phi,p_T) from the most two energetic photons (lead and sublead), information from the primary vertex (PV_*) and from energy deposits in the ECAL and the tracker.



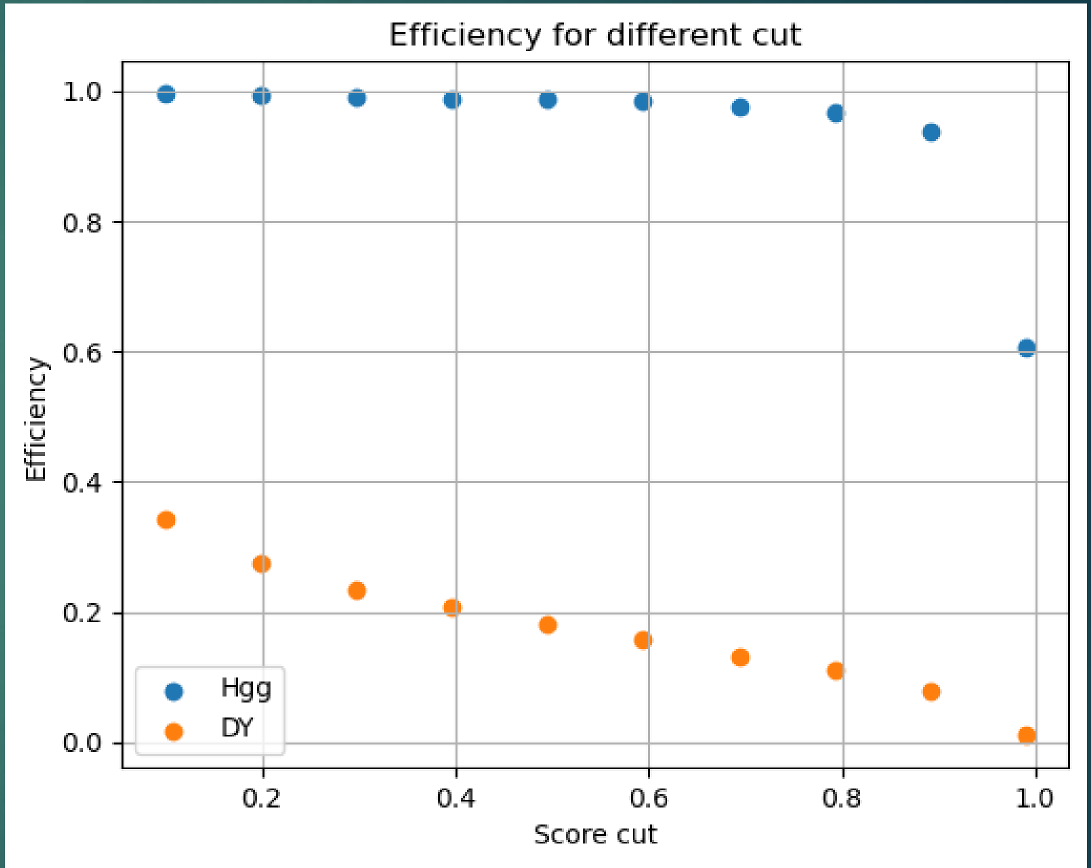
Selection

After training the model, we can calculate the importance of the selected variables by ranking them based on the value of their gradient when a small perturbation is introduced.

We can also calculate the efficiency by performing a selection associated with the score of each event, i.e., the probability that it originates from the signal or background noise.
The best possible cut is 0.90.

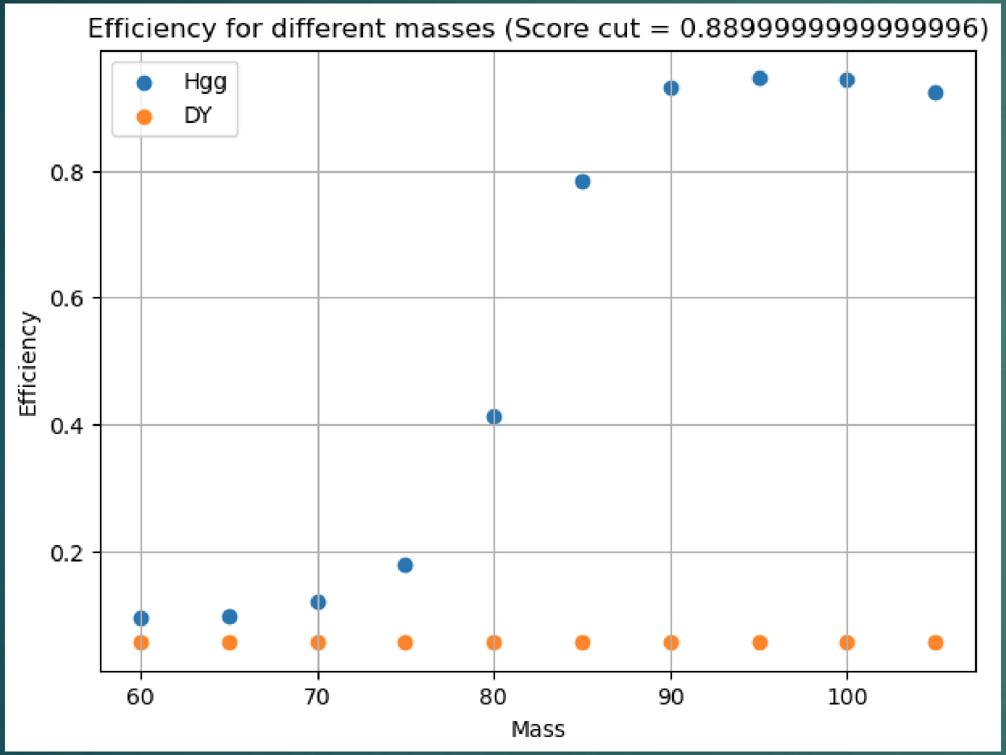
Eff_s = 94%
Eff_bkg = 6%

But this is only true for a signal at 90GeV

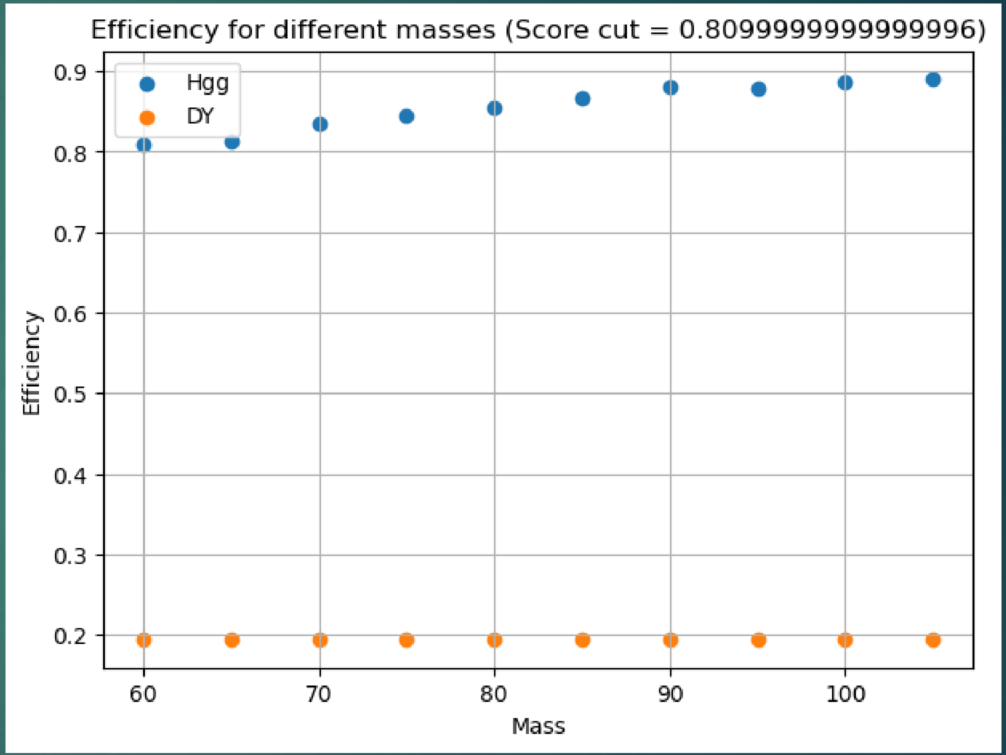


Neural Network

p_T as an input variable



p_T/m as an input variable



We can apply our NN model to every possible mass that we have (from 60GeV to 110GeV) and cut on the distribution output to look at the behavior of the efficiency. We can see that the model is very dependant on the variable p_T and less on p_T/m. Since we don't want any dependance on the mass, using this variable is more likely.

Diphoton BDT

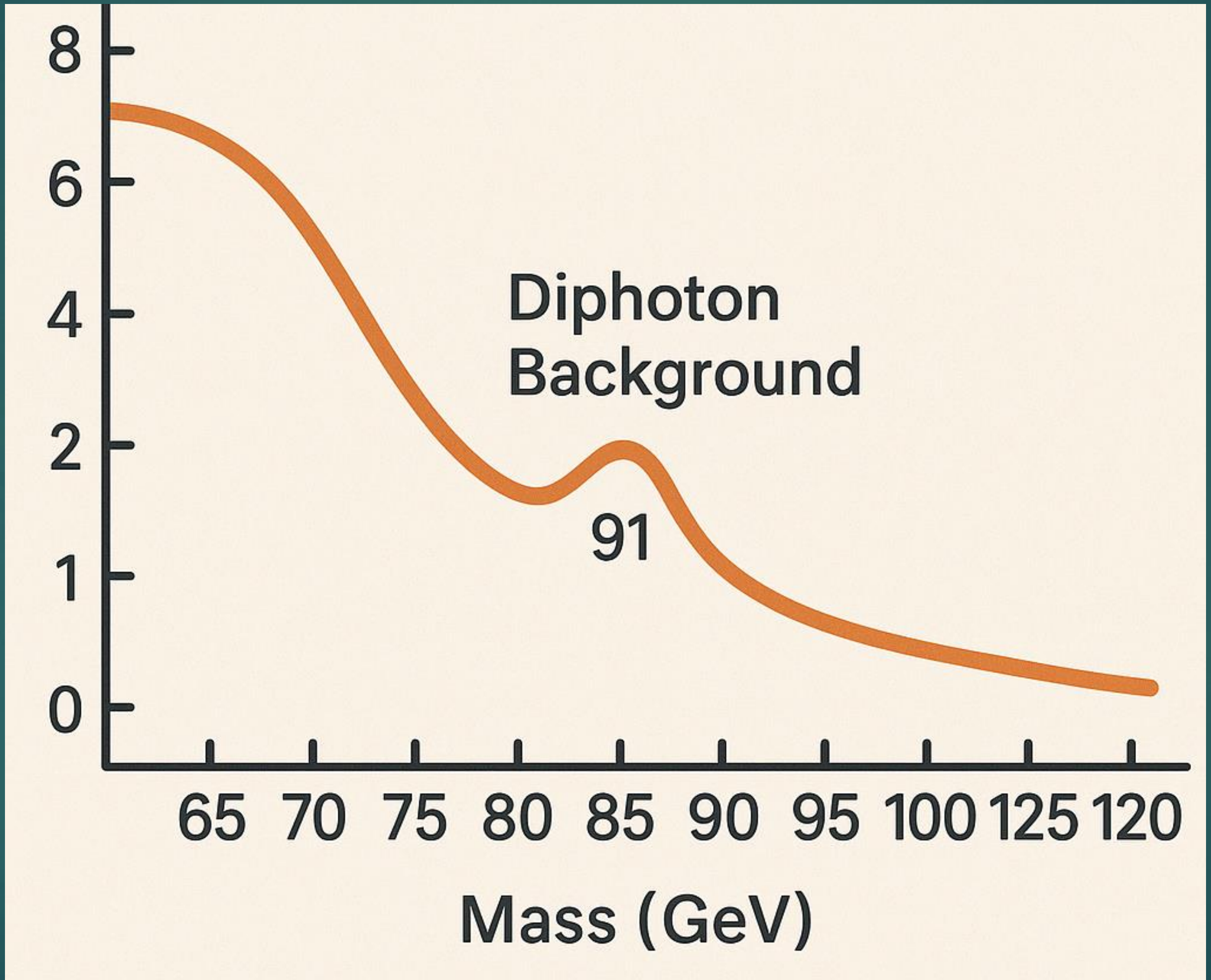
A diphoton BDT (boosted decision tree) have been trained to discriminate prompt diphoton events from the QCD events ($\gamma\gamma$,γ+jet,jet+jet).

The score is then used to divide photon events into 4 classes such that we have the best efficiency for all of them.

Current DY background modeling procedure

78 events	Class 0 Likely photons	Class 1 Possibly photon and electron or jets	418 events
1239 events	Class 2 Possibly photon and electron or jets	Class 3 Likely not photons	1576 events

Total background modeling

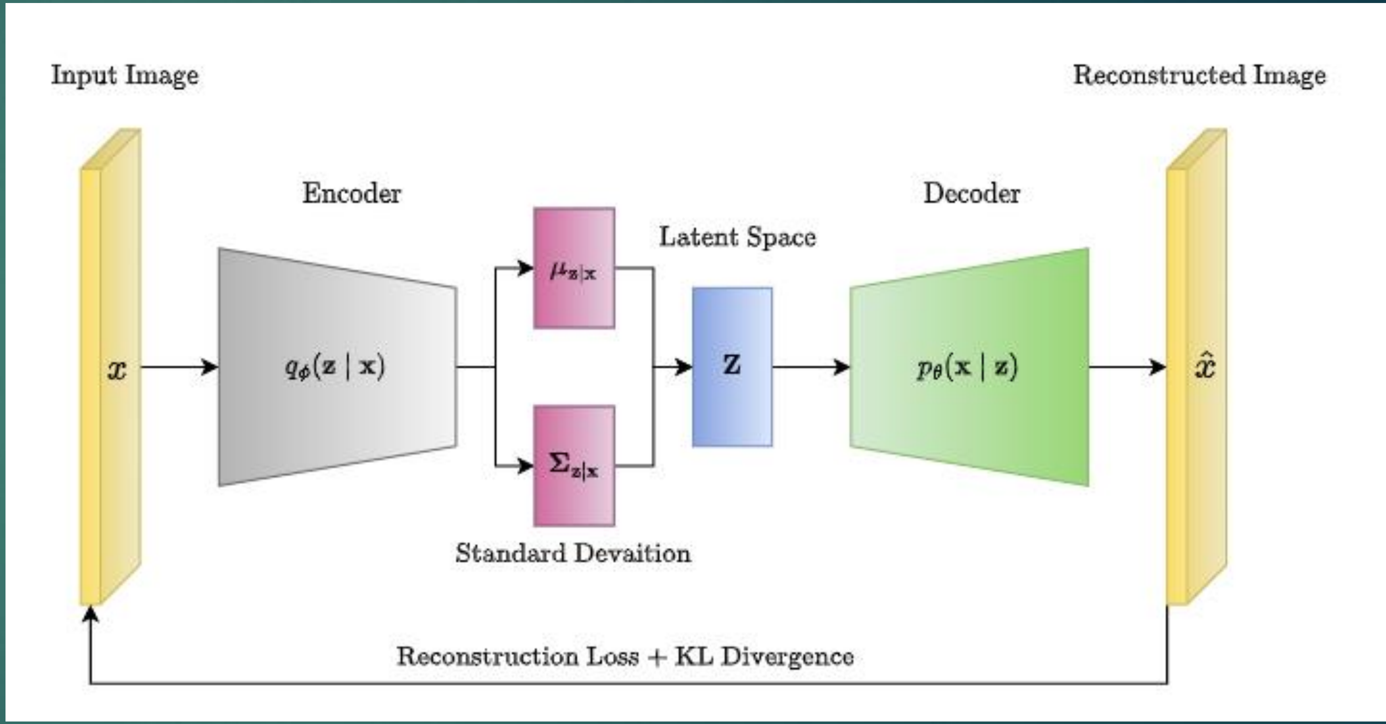


Plot of the total background modeling for each class trying to use different continuum functions

Background modeling challenge

The number of Monte Carlo events is not enough to do a robust modeling so we are investigating using a data augmentation algorithm. The technique investigated so far is called a variational autoencoder.

The idea is to encode a distribution into a latent space and then decode it to generate a similar one with greater statistics.



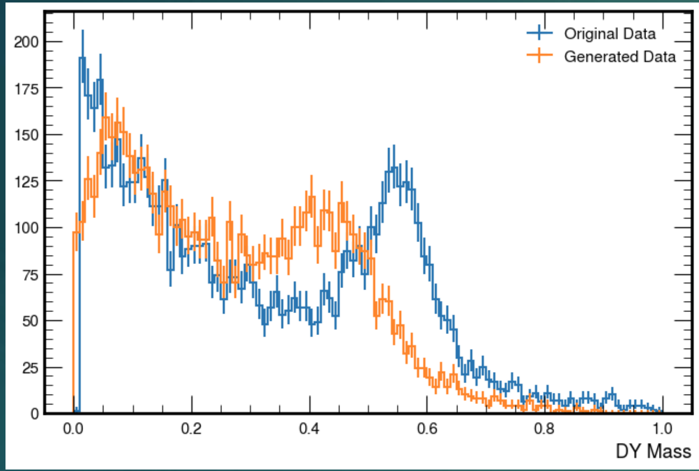
Background modeling challenge

78 events	Class 0 Likely photons	Class 1 Possibly photon and electron or jets	418 events
1239 events	Class 2 Possibly photon and electron or jets	Class 3 Likely not photons	1576 events

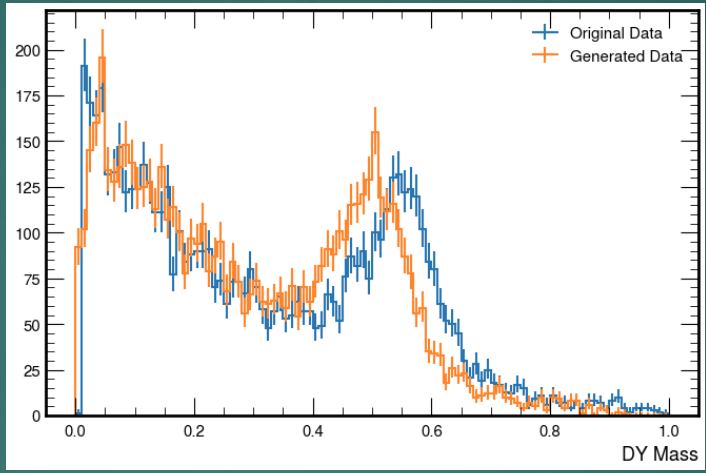
Data Augmentation example

The number of events are the same for both trained and generated events

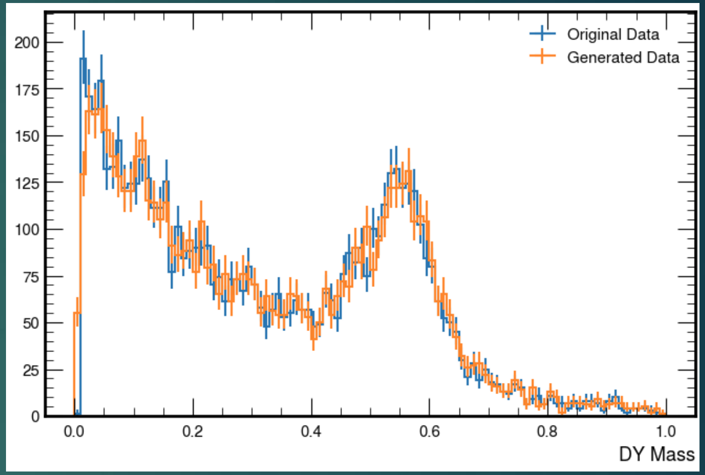
Epoch = 100



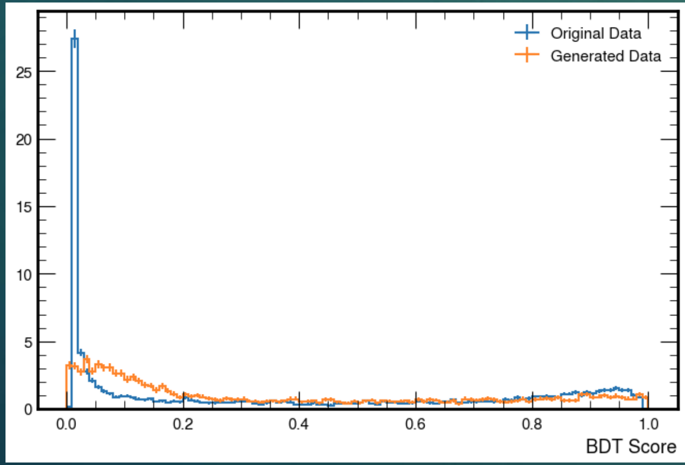
Epoch = 500



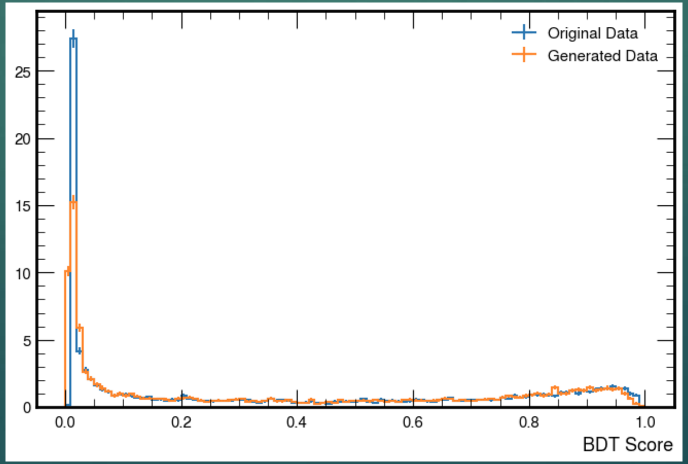
Epoch = 1000



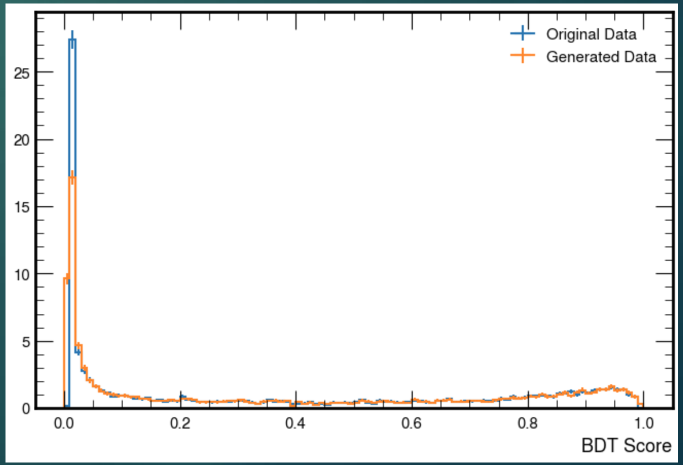
Epoch = 100



Epoch = 1500



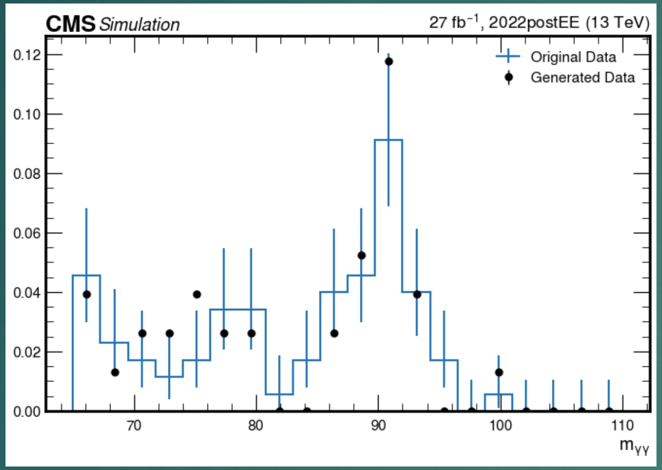
Epoch = 3000

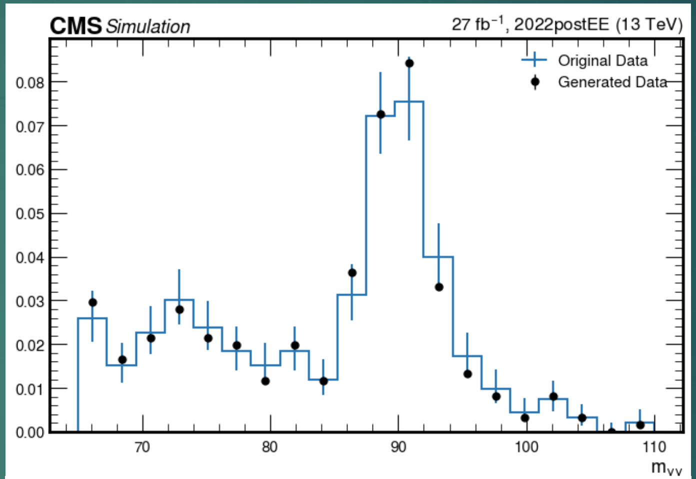


Data Augmentation

Every plot have been normalized to the unity

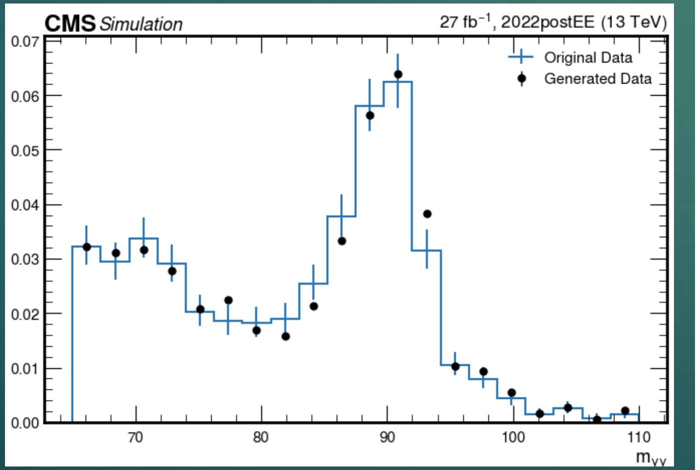
Class 0
Original : 78
Generated : 34818

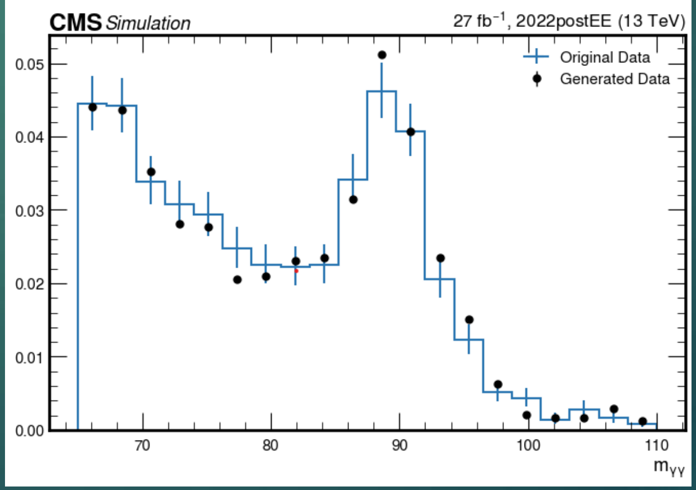




Class 1
Original : 412
Generated : 276 480

Class 2
Original : 1225
Generated : 834 560





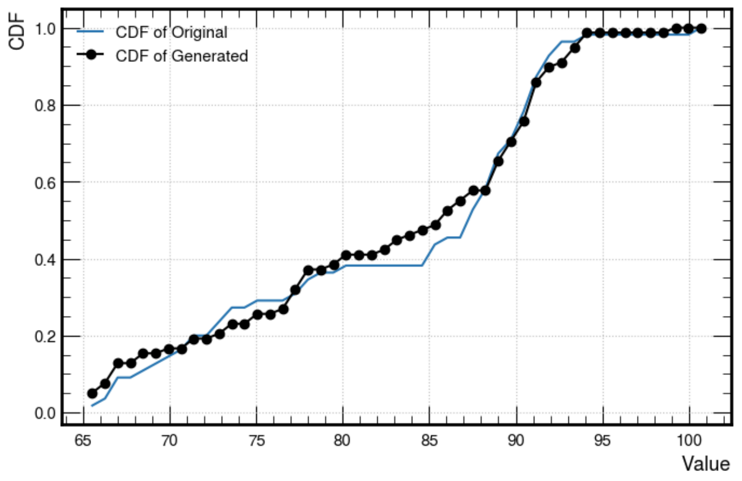
Class 3
Original : 1561
Generated : 1 091 584

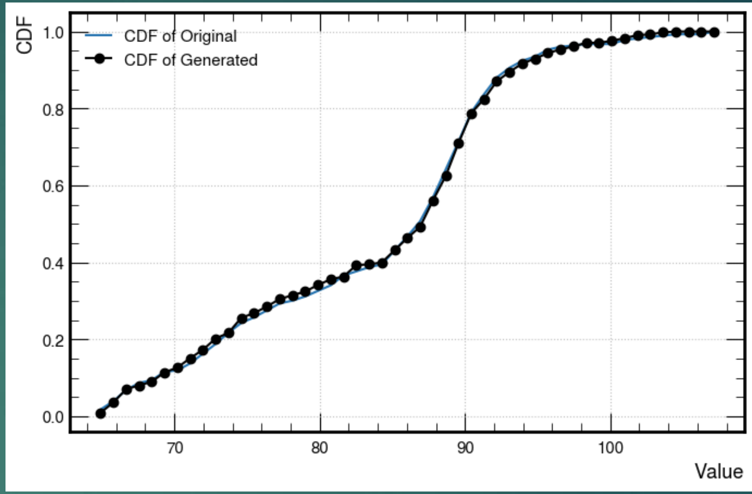
CDF for each class

$F_X(x) = P(X \leq x)$

KS Stat = $\max_{bin}(F_{reel}, F_{VAE})$

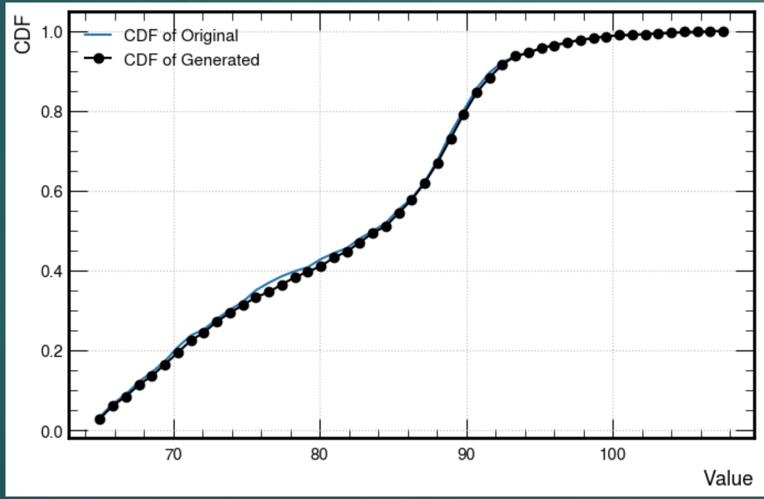
Class 0
KS Stat = 0.096
P-Value = 0.43

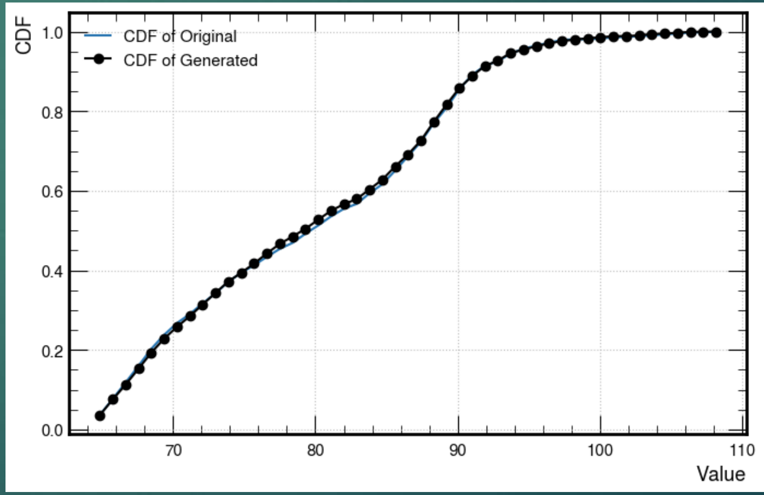




Class 1
KS Stat = 0.050
P-Value = 0.25

Class 2
KS Stat = 0.031
P-Value = 0.17





Class 3
KS Stat = 0.018
P-Value = 0.67

Technique still has to be evaluated and other techniques to investigate (Normalizing flow,Generative Adversarial Network)

Photon energy scale extraction from Z->μμγ Run 3 data as a service task for CMS

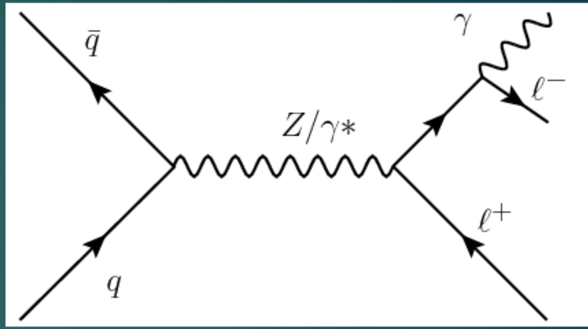
- To correct the energy, we need to calculate :

$$S = \frac{m^2_{\mu\mu\gamma} - m^2_{\mu\mu}}{m^2_Z - m^2_{\mu\mu}} - 1$$

For different region of η (Endcap and Barrel) and for different R9 region (>0,94 or <0,94) and then fit them with a Voigtian function using some percentages from 60 to 100% with a step of 1%. Thus , we keep the fit covering the largest percentage of data that has an acceptable p-value.

Uncertainties :

- Statistical : given by the fit
- Systematics : Quadratic sum of the fit range and fit function uncertainties



The Z boson properties are well known and CMS can measure pricesely muon caracteristcs so the photon energy can be measured and corrected properly thanks to this process

Conclusion

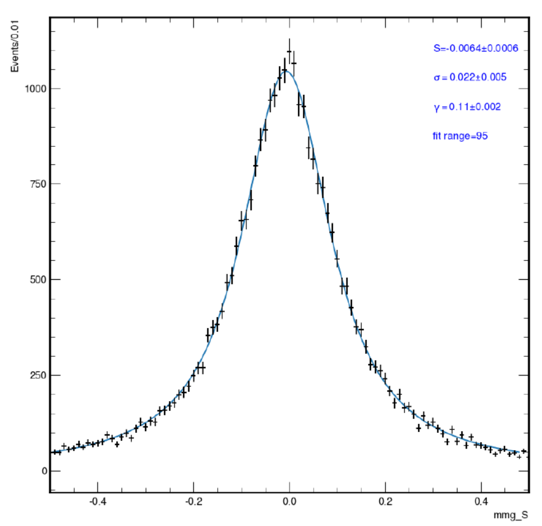
- The strategy of the analysis is still on going :
  - A neural network has been adopted, and a baseline cut defined to discriminate $h \to \gamma\gamma$ events from $Z \to ee$ events
  - The data augmentation technique has to be tested with the input variables of the BDT and the neural network to have more statistics for modeling

THANKS FOR YOU ATTENTION

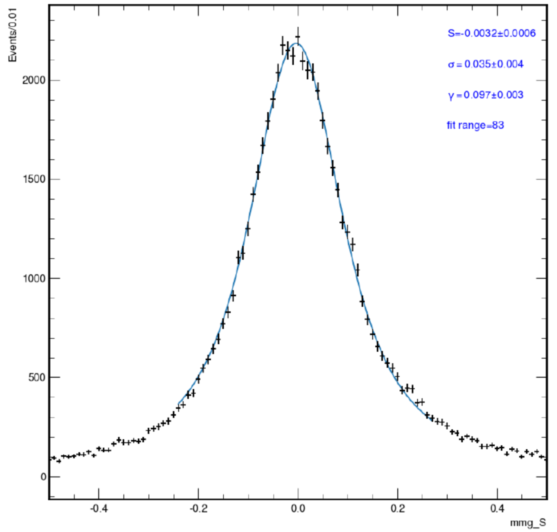
Back ups

Example fits with Voigtian Functions

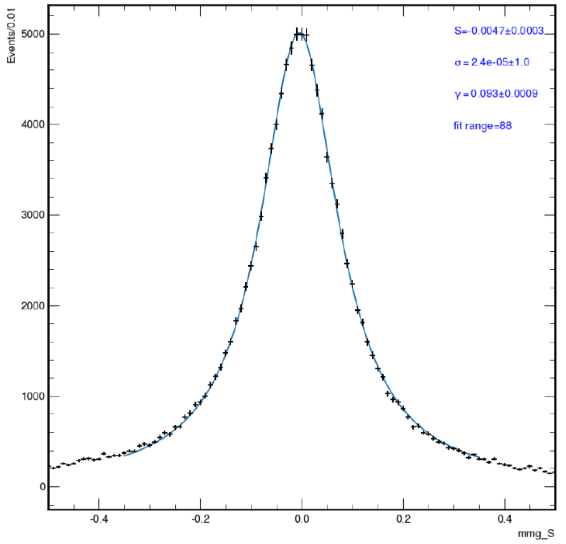
EE High R9



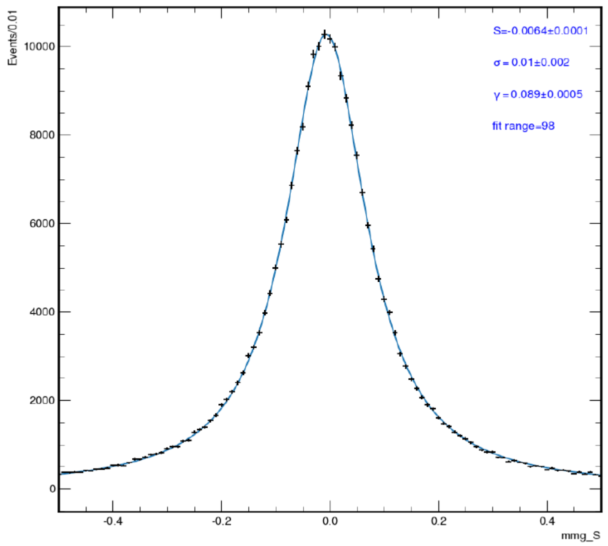
EE Low R9



EB High R9



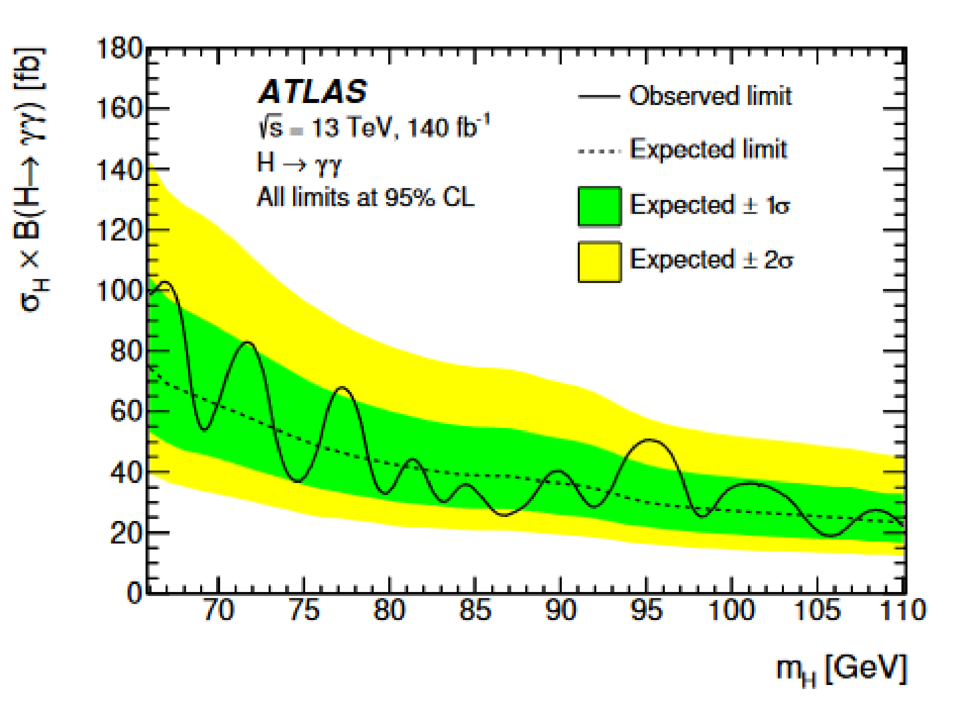
EB Low R9



Plot ATLAS for Run 2 search of low mass Higgs Boson

With the full Run 2 data, ATLAS has found an excess around 95.4GeV with a significance of 1.7σ. Which is less than CMS (2.8σ).

JHEP01 (2025) 053 arXiv:2407.07546



Neural Network

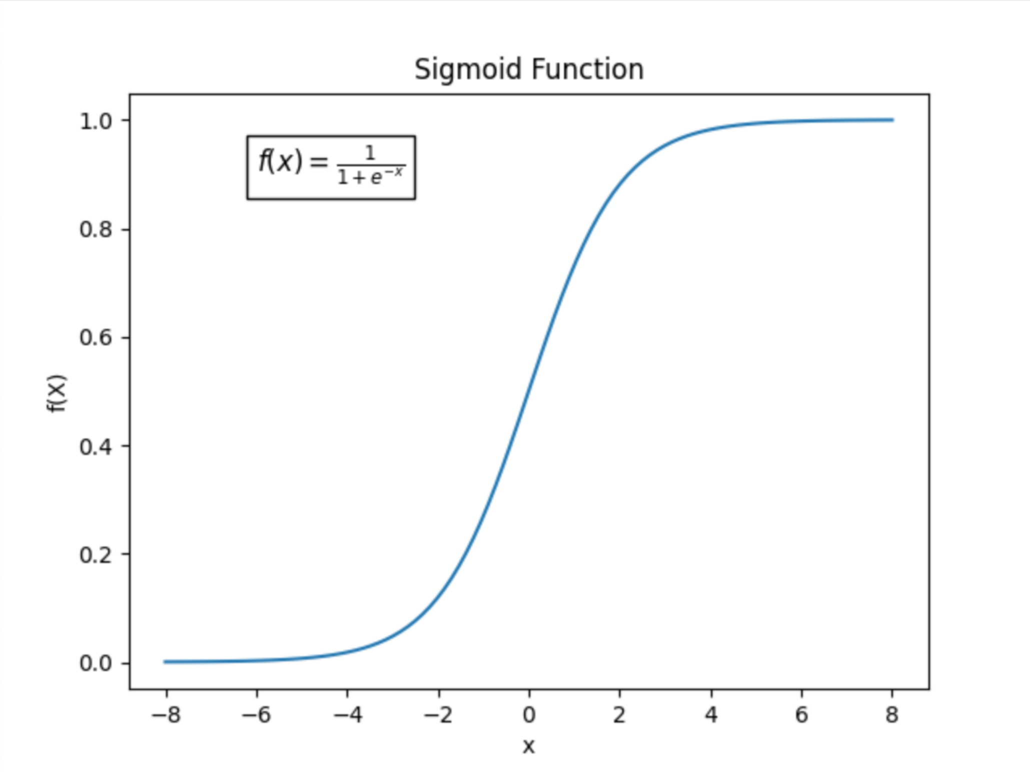
Every output of every neural has an activation function : Identity, ReLU (Rectified Linear Unit), sigmoid, etc…

The ReLU function gives us some non linearity in the model :

$$\mathrm{ReLU}(x)=\max(0,x)$$

The sigmoid is used to have an output between 0 and 1, so our result can be interpreted as a probability.

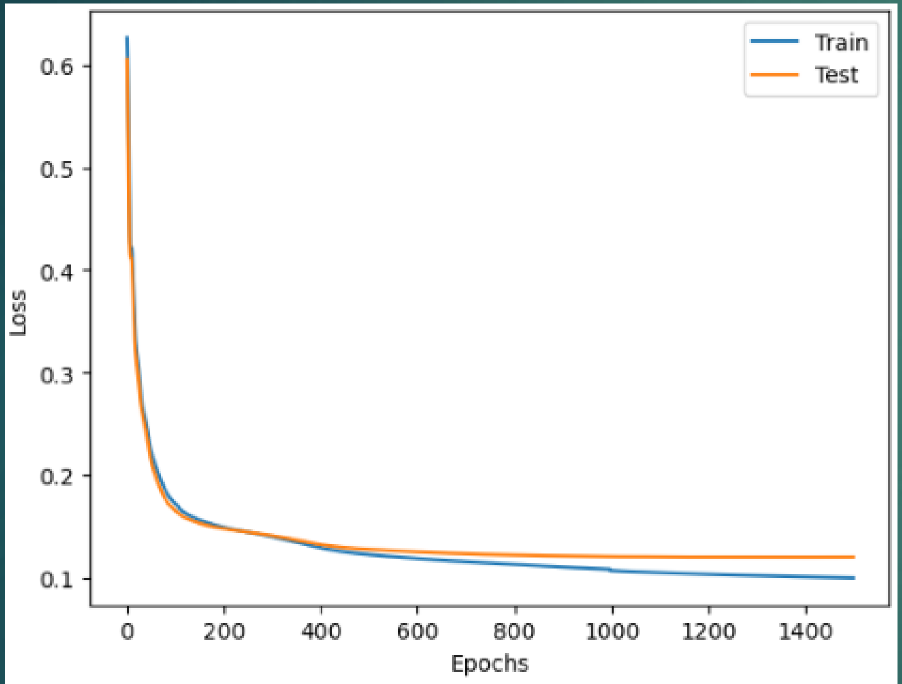
$$\sigma(z) = \frac{1}{1+e^{-z}}$$



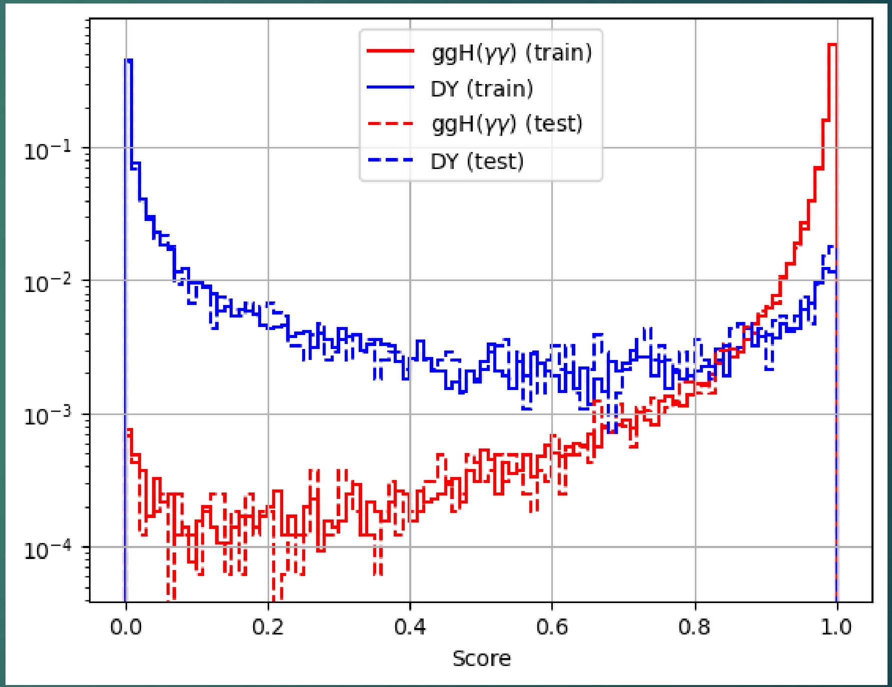
The result is then compared to the training set using a loss function that has to be minimized.

Neural Network

Loss function



Score

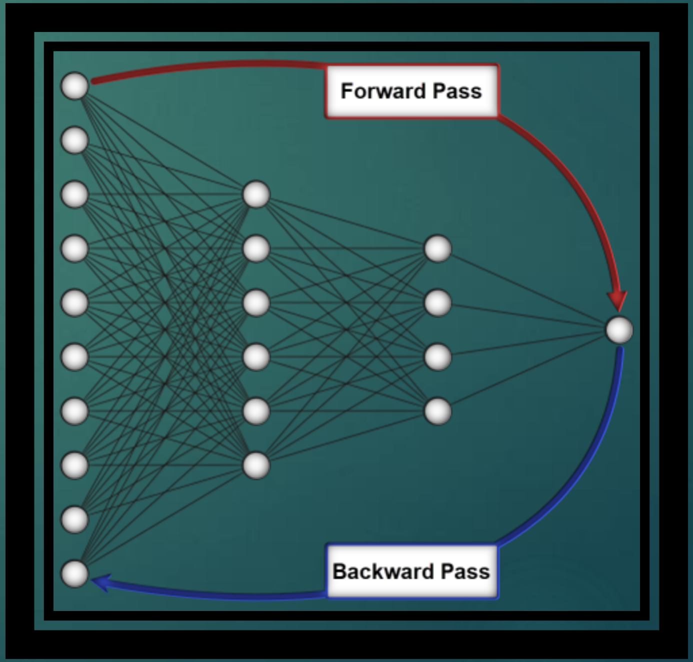


Loss: 0.10602, Acc: 96.63% | Test Loss: 0.09577, Test Acc: 96.95%

Neural Network

For each neural, the corresponding result is $z = \sum_i f(w_i x_i + b)$, we do this for every neural for each layers. This is called the forward propagation.

Then we compute the loss function. For a classifier, it corresponds to the binary cross entropy : $BCE(x) = -\sum \mathrm{y_n} \ln(x_n) + (1 - y_n)\ln(1 - x_n)$. Then, we perform a backpropagation, corresponding to the correction of the weights such that the loss function can be minimized.

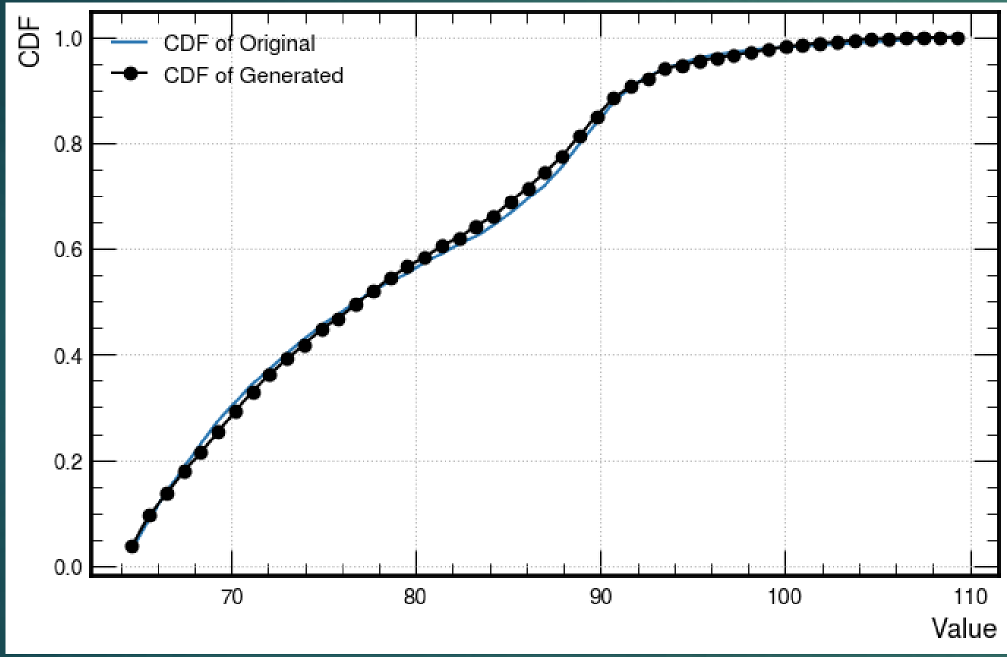


Data Augmentation

$$F_X(x) = P(X \leq x)$$

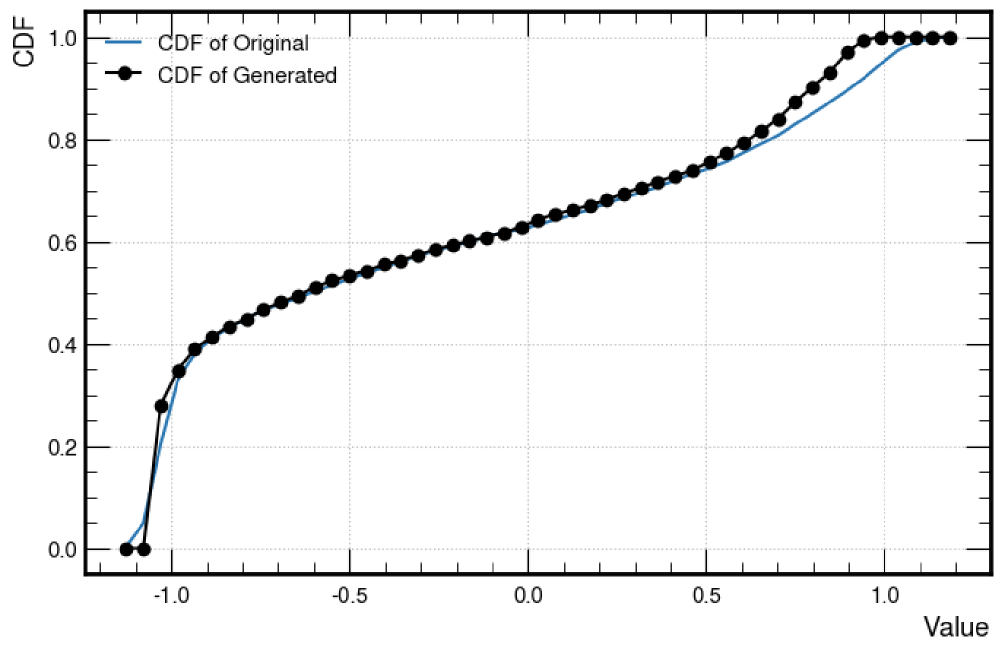
KS Stat = $\max_{bin}(F_{reel}, F_{VAE})$

CDF of the mass distribution



KS Statistics = 0.03
P-value = 0.23

CDF of the BDT distribution



KS Statistics = 0.13
P-value = 10^{-22}

Uncertainties for the scale energy extraction

Statistical Uncertainty: Given by the fit

Range Uncertainty: The maximal difference between the value of S we found and other values of S found within an interval of fit-range 20% wide for which the fits have an acceptable p-value.

Fit Uncertainty: Using the selected fit, 1000 toy models containing as many points as the original sample are generated and fitted with a Cruijff function. The mean of these functions are fitted with a Gaussian and the difference between the mean of the Gaussian and the selected value of S is taken to be the fit uncertainty.

Systematic Uncertainty: the quadratic sum of the range uncertainty and the fit uncertainty