

Reconstruction and identification of close-by e^+e^- in the search for Higgs boson decays into a low-mass dilepton system and a photon final state.

Tesi di Laurea Triennale

A.A. 2025-2026

Giuseppe Morinelli (Matricola 966566)

Relatore: Leonardo Carminati

Correlatori: Laura Nasella, Ruggero Turra

... Physics phenomenon

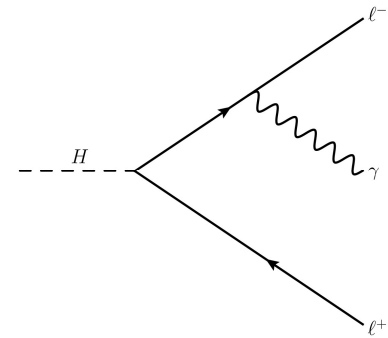
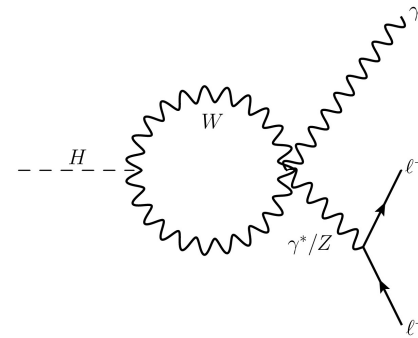
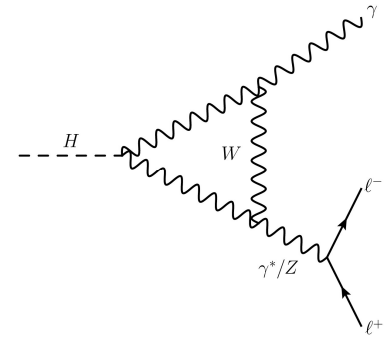
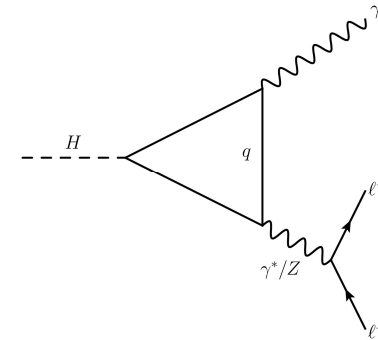
Higgs boson physics

- ATLAS and CMS collaborations proved the existence of the Higgs boson (H) in 2012
- More than ten years later, some aspects of the Higgs sector of the Standard Model need to be cleared out yet
- H bosons decay almost instantaneously ($\sim 10^{-22}$ s) in a large variety channels, most of which have been discovered
- Searching for rare decays of H helps finding coupling constants and possible extensions of the SM

mass →	$\approx 2.3 \text{ MeV}/c^2$	$\approx 1.275 \text{ GeV}/c^2$	$\approx 173.07 \text{ GeV}/c^2$	0	$\approx 126 \text{ GeV}/c^2$
charge →	$2/3$	$2/3$	$2/3$	0	0
spin →	$1/2$	$1/2$	$1/2$	1	0
	u up	c charm	t top	g gluon	H Higgs boson
QUARKS	$\approx 4.8 \text{ MeV}/c^2$	$\approx 95 \text{ MeV}/c^2$	$\approx 4.18 \text{ GeV}/c^2$	0	
	$-1/3$	$-1/3$	$-1/3$	0	
	$1/2$	$1/2$	$1/2$	1	
	d down	s strange	b bottom	γ photon	
LEPTONS	$0.511 \text{ MeV}/c^2$	$105.7 \text{ MeV}/c^2$	$1.777 \text{ GeV}/c^2$	$91.2 \text{ GeV}/c^2$	
	-1	-1	-1	0	
	$1/2$	$1/2$	$1/2$	1	
	e electron	μ muon	τ tau	Z Z boson	
LEPTONS	$< 2.2 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 15.5 \text{ MeV}/c^2$	$80.4 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$1/2$	$1/2$	$1/2$	1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	
					GAUGE BOSONS

The rare decay $H \rightarrow \gamma\gamma^* \rightarrow \ell\ell\gamma$

- The rare channel of interest consists in an H decay in photon pair, where one photon is off-shell and decays in a lepton pair
- The lepton pair could be a muon-antimuon or electron-positron
- It is common for electron pairs to be really close being detected as a single particle **(merged electrons M.E.)**
- This decay has been measured with a significance of 3.2σ in the previous analysis during Run 2 (2015-2018)

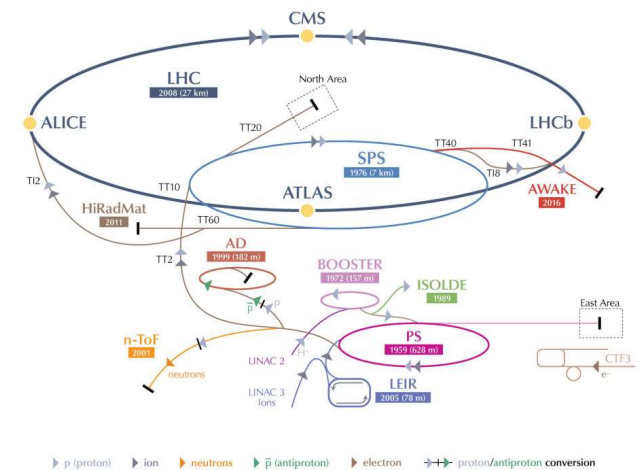


... The ATLAS experiment

The Large Hadron Collider (LHC)

- LHC is the largest particle accelerator in the CERN experimental complex at Geneve, Switzerterland (27 km long ring)
- It is designed to have proton-proton (pp) collisions at a center-of-mass energy of 13.6 TeV at Run 3 (2022-2026)

CERN's Accelerator Complex



LHC Large Hadron Collider SPS Super Proton Synchrotron PS Proton Synchrotron

AD Antiproton Decelerator CTF3 CERN Test Facility AWAKE Advanced WAKEfield Experiment ISOLDE Isotope Separator OnLine Device

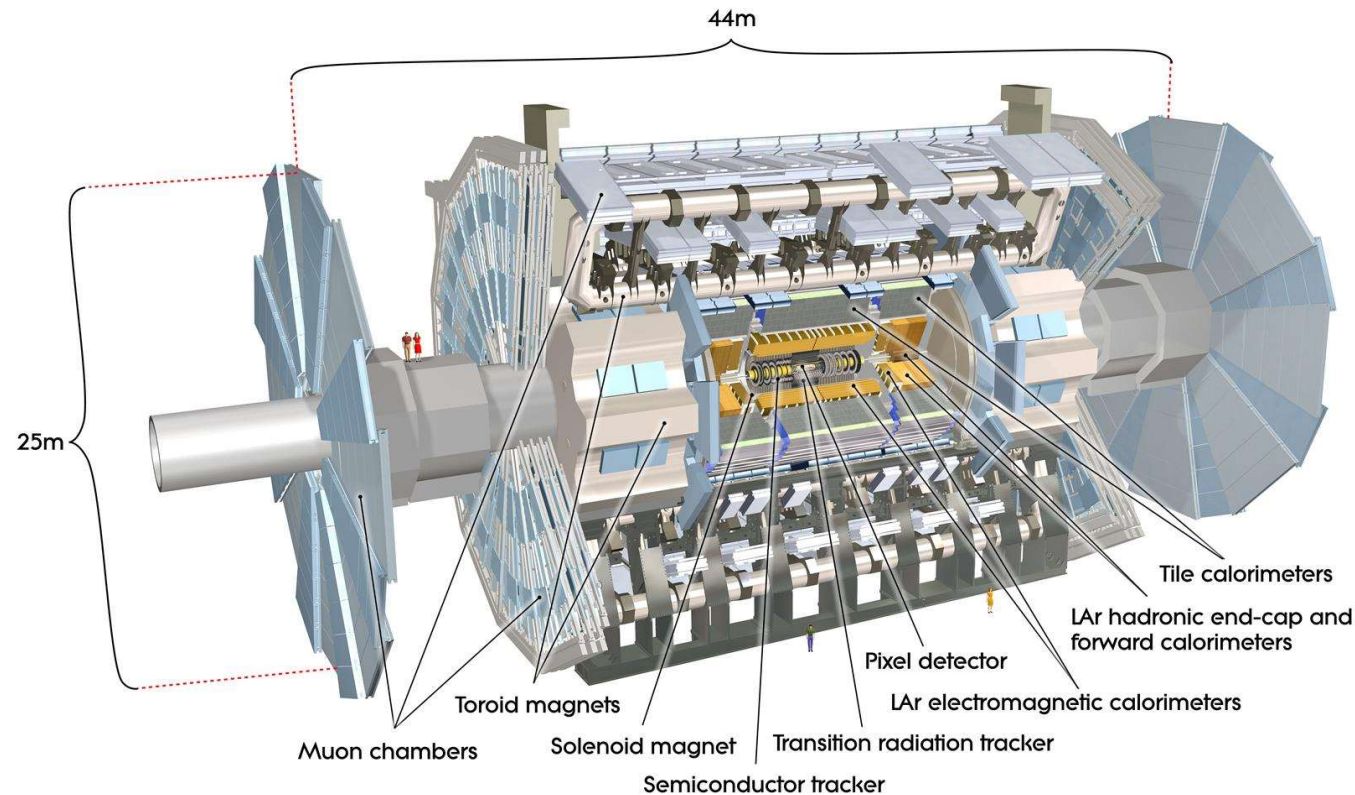
LEIR Low Energy Ion Ring LINAC LINear ACcelerator n-ToF Neutrons Time Of Flight HiRadMat High-Radiation to Materials

© CERN 2013

The ATLAS experiment

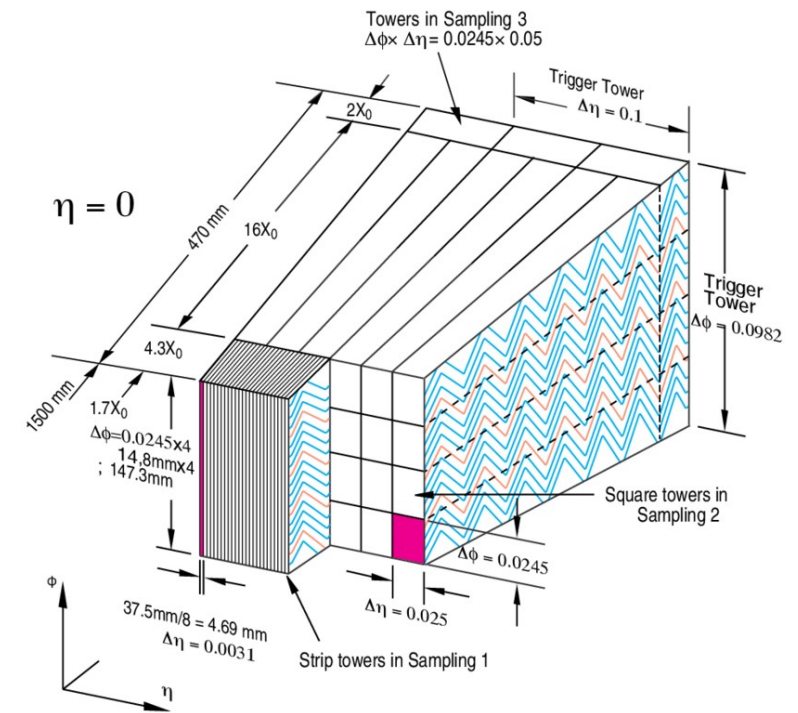
It is composed of many layers of detectors around an interaction point, in cylindrical shape:

1. Inner detector
2. Electromagnetic and hadronic calorimeters (EMC and HC)
3. Magnet system (solenoid and barrel/endcaps toroids)
4. Muon spectrometer



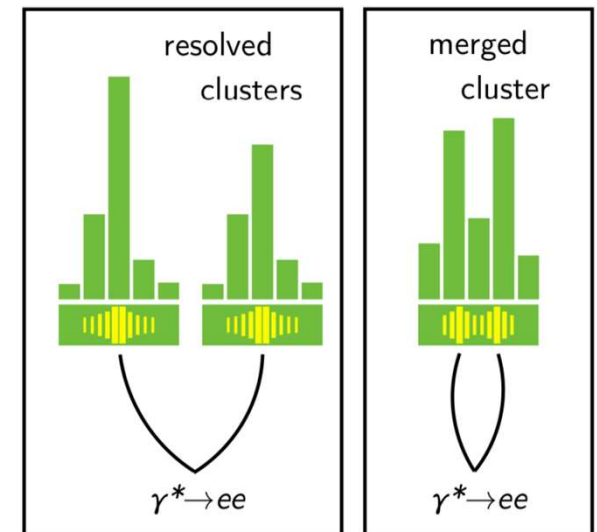
Reconstruction of particles in the EMC

- The EMC is designed to absorb photons and electrons produced by collision events (destructive measure)
- Photons and electrons interact in the EMC generating EM showers of particles, so their energy is deposited in a cluster of spacial cells
- Reconstruction algorithms are used to distinguish between electrons and photons candidates and to measure the energy of the original particles starting from the clusters
- There are ambiguous cases that need to be resolved



Reconstruction of merged electrons

- In case of merged electrons, two clusters of deposited energy overlap
- Merged electrons are reconstructed as particles with a single energy cluster (in the EMC) and two separated tracks (in the inner detector)
- The standard ATLAS electron ID is not optimal for merged electrons
- A new ID could increase performance and help collecting more $H \rightarrow \gamma\gamma^* \rightarrow \ell\ell\gamma$ data



... BDT algorithm

Training data: simulated signal candidates

- Use simulated data because every aspect of generated events is known!
(*truth level* data)

Simulate 13.6 TeV pp
collision



Production of particles
from collision



Select events of
 $H \rightarrow \gamma\gamma^*$ decays



Select events where
Higgs bosons are
produced



Simulate the
interaction with
ATLAS detectors and
reconstruction

Data used to train a BDT model

Same procedure applies for simulated background samples

Simulated background samples

- Fake candidates: objects reconstructed similarly to merged electrons by the detectors
- Jet Filter (JF) sample: it consists of di-jets of hadrons from Higgs decay
- $H \rightarrow \gamma\gamma$ photon decay samples: one sample with Higgs produced via gluon-gluon fusion (ggH), one via vector boson fusion (VBFH)

Selections on training data

Signal sample:

- Kinematic selections, $p_T > 10 \text{ GeV}$ and $|\eta| < 2.37$
- Select candidates matching to γ^* from Higgs decay
- Loose identification

JF bkg sample:

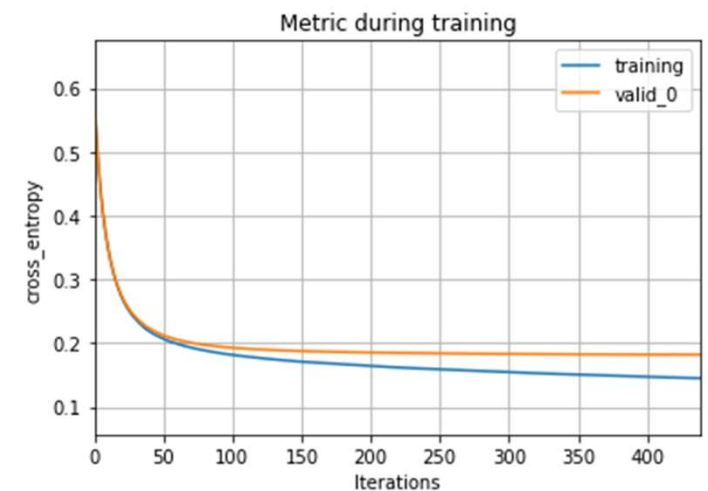
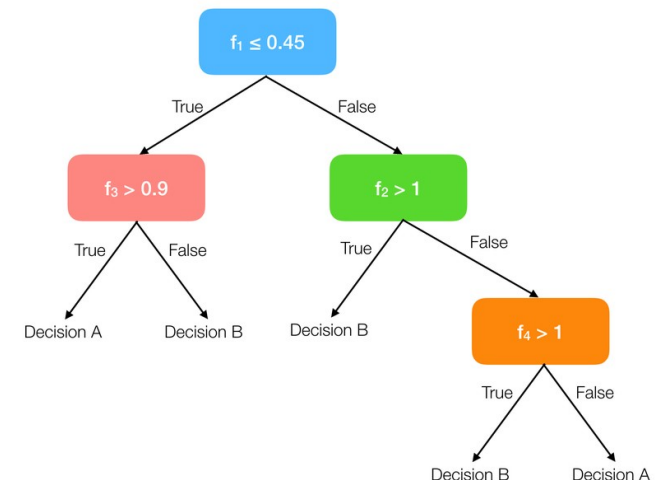
- Kinematic selections, $p_T > 10 \text{ GeV}$ and $|\eta| < 2.37$
- Loose identification
- Select electrons coming from γ : this is done to exclude non-resonant $\gamma\gamma^*$ irreducible bkg

Photon bkg samples:

- Kinematic selections, $p_T > 10 \text{ GeV}$ and $|\eta| < 2.37$
- Loose identification
- Select γ coming from Higgs: this is done to exclude dalitz decays from sample

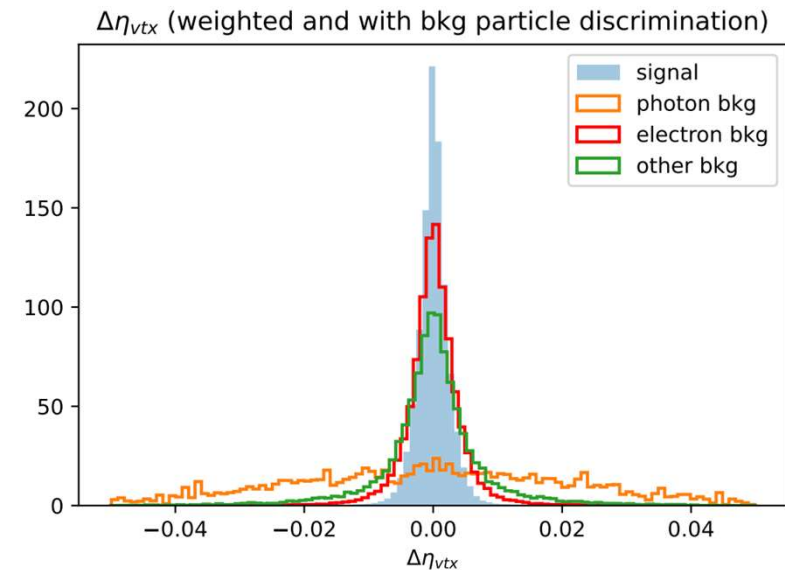
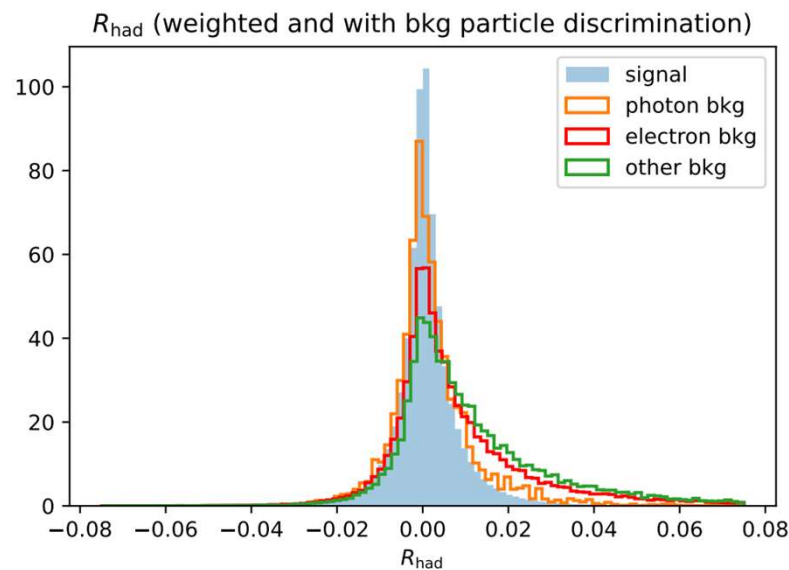
Boosted decision tree (BDT) model

- Split samples in 50% for training and 50% for testing the model
- Set hyperparameters for the model (num_leaves, early_stopping, etc...)
- Train a BDT model using classifier class from LightGBM package, developed by Peking University and Microsoft corporation



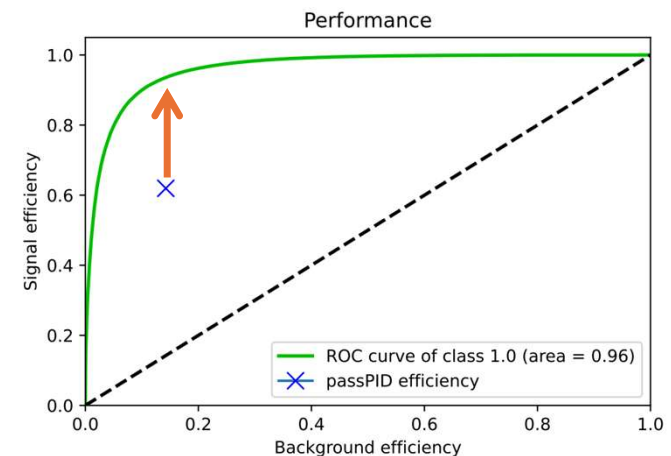
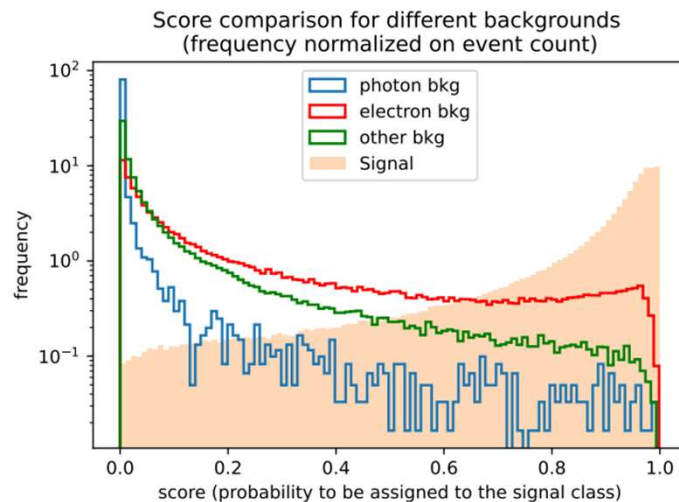
Feature distributions

- 17 discriminating features used: shower shape variables from the EMC, track related variables from the inner detector
- The more the feature distributions are different, the more the BDT model will be capable of discriminating signal and bkg candidates



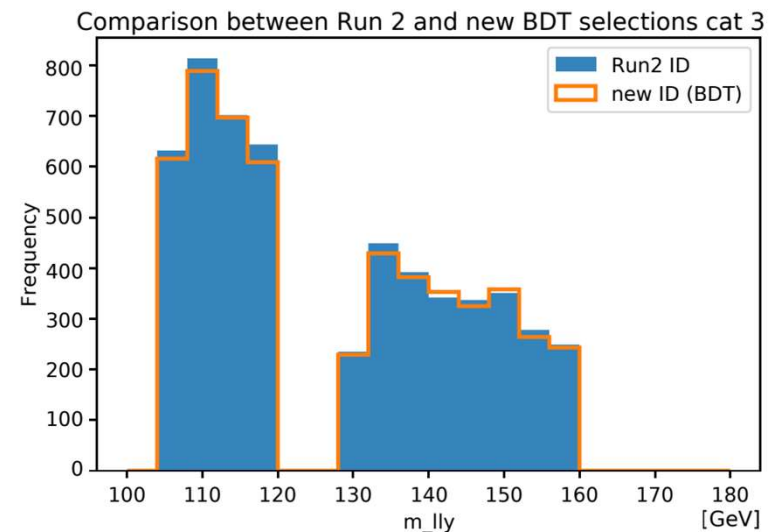
Training results on test sample

- Each candidate has been assigned a prediction score from 0 (bkg) to 1 (signal)
- Set a threshold value for the score to define a BDT identification
- Compare the signal efficiency of the BDT ID with the identification used in the previous Run 2 analysis (passPID)



Significance on sidebands

- Bkg samples used for training not have an arbitrary composition of particle, it does not reflect real data
- New test: BDT ID and passPID are applied to a sidebands bkg sample composed of real data and to a simulated signal sample
- The sample is blinded: events in a range of invariant mass near 125 GeV are removed, in order to exclude possible signal that would induce bias
- Fit are performed to calculate an estimate of the number of events S and B, respectively for signal and bkg, with each ID



$$Z = \frac{S}{\sqrt{B}}$$

Significance Z_{90} results

- The datasets are classified in a total of 9 categories, distinguished by flavor, by kinematics and by the production mode of the Higgs
- Fits and calculation of significance are performed to each category separately
- BDT ID differs from the passPID only on merged electron categories, by construction

Category name	Category code
$\mu\mu$ low- p_{T_t}	1
ee resolved low- p_{T_t}	2
ee merged low- p_{T_t}	3
$\mu\mu$ VBF	4
ee resolved VBF	5
ee merged VBF	6
$\mu\mu$ high- p_{T_t}	7
ee resolved high- p_{T_t}	8
ee merged high- p_{T_t}	9

passPID identification

Cat	B_{90}	S_{90}	Z_{90}
1	2013.42	59.3740	1.323211
2	1061.42	22.5407	0.691847
3	1389.76	33.5220	0.899209
4	6.93411	1.63613	0.621330
5	3.04060	0.55511	0.318346
6	3.60938	1.07378	0.565196
7	45.9867	5.68258	0.837972
8	16.0328	1.69923	0.424373
9	35.4826	4.10044	0.688371

BDT model identification

Cat	B_{90}	S_{90}	Z_{90}
1	2013.42	59.3740	1.323211
2	1061.42	22.5407	0.691847
3	1373.87	43.3654	1.169957
4	6.93411	1.63613	0.621330
5	3.04060	0.55511	0.318346
6	4.85787	1.36507	0.619348
7	45.9867	5.68258	0.837972
9	35.4826	4.10044	0.688371
9	33.2311	5.18831	0.900024

Conclusions

- A working point ensuring the same background efficiency of the one from the passPID merged electron ID has been tuned on the test sample showing an improvement on signal efficiency of 51%
- The performance of the BDT algorithm has been tested on a real data in $m_{\ell\ell}$ sidebands reporting a potential improvement of the 26% in significance (from 1.27σ to 1.60σ).
- When accounting also for muon and resolved electron candidates, that keep their identification method unchanged, the overall increase in expected significance is 8%, 2.28σ to 2.48σ
- The new BDT ID has shown sizeable improvements in identification of merged electrons, and will be used in new analyses