

DOTTORATO DI RICERCA IN FISICA

CICLO XXXIV

Settore Concorsuale: 02/A1 - FISICA SPERIMENTALE DELLE INTERAZIONI FONDAMENTALI

Settore Scientifico Disciplinare: FIS/01 - FISICA SPERIMENTALE

Application of Deep Learning techniques in the search for BSM Higgs bosons in the $\mu\mu$ final state in CMS

PRESENTATA DA:

Tommaso Diotalevi

COORDINATORE DOTTORATO:

Prof. Michele Cicoli

SUPERVISORE:

Prof. Daniele Bonacorsi

CO-SUPERVISORI:

Dott. Federica Primavera

Dott. Stefano Marcellini

Dott. Gianni Masetti

*Application of Deep Learning techniques in the search for BSM Higgs bosons in the $\mu\mu$ final state
in CMS, May 2022*

Author:

Tommaso DIOTALEVI

University:

Alma Mater Studiorum - Università di Bologna

CONTENTS

Abstract	vii
1 THE HIGGS BOSON IN THE STANDARD MODEL	1
1.1 The fundamental particles	1
1.1.1 Leptons	2
1.1.2 Quarks and Gluons	3
1.2 The structure of the standard model	3
1.2.1 Local gauge symmetry invariance	5
1.3 The Higgs mechanism	9
1.4 The Higgs boson at the LHC	12
1.4.1 Higgs production and decay modes	12
1.4.2 Experimental discovery	16
1.5 Beyond the Standard Model	19
2 SUPERSYMMETRY AND THE MSSM	23
2.1 Phenomenology of the Supersymmetry	23
2.2 The Minimal Supersymmetric Standard Model	25
2.2.1 The Two-Higgs-Doublet Model	26
2.2.2 The Higgs sector in the MSSM	27
2.3 MSSM framework Scenarios	28
2.3.1 The m_h^{max} Scenario	29
2.3.2 The $m_h^{mod\pm}$ Scenario	30
2.3.3 The hMSSM Scenario	30
2.3.4 The Heavy Higgs Scenario	30
2.4 Searches for the MSSM Higgs Bosons	30
3 THE CMS EXPERIMENT AT THE LHC	35
3.1 The Large Hadron Collider	35
3.1.1 The vacuum system	36
3.1.2 Electromagnets	37
3.1.3 Radiofrequency cavities	38
3.2 LHC detectors	39
3.2.1 ALICE	40
3.2.2 ATLAS	40
3.2.3 CMS	41
3.2.4 LHCb	41
3.2.5 Other LHC experiments	41
3.2.6 LHC Operational history	42

3.3	The CMS experiment	43
3.3.1	The CMS detector: concept and structure	45
3.3.2	The tracking system	46
3.3.3	The Electromagnetic Calorimeter	48
3.3.4	The Hadronic Calorimeter	50
3.3.5	The magnet	51
3.4	The Muon System	52
3.4.1	The Drift Tube Chambers	53
3.4.2	The Cathode Strip Chambers	54
3.4.3	The Resistive Plates Chambers	55
3.5	Trigger and Data Acquisition	57
3.5.1	The Level-1 Trigger System	57
3.5.2	The High Level Trigger and DAQ	58
3.6	Global event reconstruction and Particle Flow algorithm	59
4	MACHINE AND DEEP LEARNING	61
4.1	Introduction	61
4.2	Types of Machine Learning	62
4.2.1	Supervised learning	62
4.2.2	Unsupervised learning	62
4.2.3	Reinforcement Learning	63
4.3	Machine Learning formalism	65
4.3.1	Model and training	65
4.3.2	Overfitting	67
4.3.3	Hyperparameters	68
4.3.4	Training and Test set	68
4.3.5	Validate the model	70
4.4	Measuring model quality	71
4.4.1	Accuracy	71
4.4.2	Precision	72
4.4.3	Recall	72
4.4.4	Area under the Receiver Operating Characteristic (ROC AUC)	72
4.5	Some learning algorithms	74
4.5.1	Logistic Regression	75
4.5.2	Gradient Boosting Decision Tree	76
4.6	Neural Networks	79
4.6.1	Optimisation	81
4.6.2	Dropout and regularisation	81
4.7	Implementing a Neural Network	83
4.7.1	TensorFlow	83
4.7.2	Scikit-learn	84
4.7.3	Keras	85
4.8	Machine Learning in High-Energy Physics	87

4.8.1	Separating Signal and Background: the Event Selection	88
4.8.2	Event Reconstruction	89
4.8.3	Fast Simulation	91
4.8.4	Monitoring and Data Quality	92
5	THE MSSM HIGGS BOSONS DECAYING TO $\mu^+ \mu^-$ AND ITS MAIN PHYSICAL OBJECTS	93
5.1	Introduction	93
5.2	Production mechanisms for neutral MSSM Higgs bosons	94
5.3	The dimuon final state	95
5.3.1	Background sources	95
5.4	Data and Monte Carlo samples	95
5.4.1	Run 2 Datasets	96
5.4.2	MC background samples	97
5.4.3	MC signal samples	99
5.5	Object reconstruction	100
5.5.1	Trigger requirements	101
5.5.2	Muons	102
5.5.3	Jets	105
5.5.4	Identification of b jets	105
5.5.5	Missing energy in the transverse plane	109
6	ANALYSIS STRATEGY	111
6.1	Event selection	111
6.2	Event categorisation	113
6.3	Multivariate analysis	113
6.3.1	Control regions	114
6.3.2	Parameterised Neural Network	115
6.3.3	Feature description	118
6.3.4	Feature correlation	127
6.3.5	Data preprocessing	132
6.3.6	Model training	132
6.3.7	Testing the model for the best cut selection	136
6.3.8	Model performances	140
6.3.9	Apply the network discriminator output to the ROOT files	145
7	RESULTS	147
7.1	Signal efficiency	148
7.2	Fitting procedure	154
7.2.1	Signal shape modelling	154
7.2.2	Background shape modelling	154
7.2.3	Signal plus background parameterisation	155
7.3	Systematic uncertainties	155

7.3.1	Efficiency uncertainties	155
7.3.2	Uncertainty on the background parameterisation	156
7.4	Limits calculation	157
7.4.1	Model independent limits	158
8	CONCLUSIONS AND OUTLOOK	161
A	THE HIGH-LUMINOSITY LHC PROJECT	165
B	PARAMETRIC NEURAL NETWORK ADDITIONAL PLOTS	167
B.1	pNN output distributions	167
B.2	pNN ROC curves	176
B.3	pNN PR curves	185
BIBLIOGRAPHY	194	

ABSTRACT

The Standard Model (SM) of particle physics predicts the existence of a Higgs field responsible for the generation of the particles' mass. The excitation of such field, known as the Higgs boson, has been observed for the first time by the ATLAS and CMS Collaborations in 2012. However, some aspects of this theory still remain unsolved, supposing the presence of new physics Beyond the Standard Model (BSM) with the production of new particles at a higher energy scale compared to the current experimental limits. The search for additional Higgs bosons is, in fact, predicted by theoretical extensions of the SM including the Minimal Supersymmetry Standard Model (MSSM). In the MSSM, the Higgs sector consists of two Higgs doublets, one of which couples to up-type fermions and the other to down-type fermions. This results in five physical Higgs particles: two charged bosons H^\pm , two neutral scalars h and H , and one pseudoscalar A .

The work presented in this thesis is dedicated to the search of neutral non Standard Model Higgs bosons decaying to two muons in the context of a model independent MSSM scenario. Proton-proton collision data recorded by the CMS experiment at the CERN LHC at a center-of-mass energy of 13 TeV are used, corresponding to an integrated luminosity of 35.9 fb^{-1} . Such search is sensitive to neutral Higgs bosons produced either via gluon fusion process or in association with a $b\bar{b}$ quark pair.

The extensive usage of Machine and Deep Learning techniques, that I largely developed during my PhD work, is a fundamental element in the discrimination between signal and background simulated events. A new network structure called parameterised Neural Network (pNN) has been implemented, replacing a whole set of single neural networks trained at a specific mass hypothesis value with a single neural network able to generalise well and interpolate in the entire mass range considered.

The results of the pNN signal/background discrimination are used to set a model independent 95% confidence level expected upper limit on the production cross section times branching ratio, for a generic ϕ boson decaying into a muon pair in the range from 130 to 1000 GeV.

The thesis is structured as follows:

Chapter 1 provides a description of the Standard Model framework, highlighting the theoretical structure behind it and focusing on the Higgs mechanism which allows particles to have mass. Also, a summary of the most important results achieved at the LHC on the Higgs boson properties is given.

Chapter 2 provides a general introduction to the supersymmetry theory, in particular with the description of the Minimal Supersymmetric Standard Model extension of the Standard Model. A focus on the Higgs sector is also given, describing the new additional Higgs bosons foreseen by such theory together with the current phenomeno-

logical scenarios, scanning the theory's parameter space to obtain specific limits on their possible values. Finally, a summary on the experimental searches at the CMS experiment is given.

Chapter 3 is dedicated to the experimental apparatus: after a description of the LHC accelerator machine, the important features of the CMS detector are presented. Since the analysis is focused on the decay of the Higgs bosons in two muons, a focus on the muon system is also given, with an accent on the global event reconstruction of a collision event.

Chapter 4 introduces Machine and Deep Learning concepts whose application is also present in the next Chapters. In particular, a focus on the Neural Network architecture is given, with its concepts and terminology, offering also an overview on the implementation of such networks with frameworks deeply used on every scientific field of study. Finally, a brief introduction on the main applications of Machine Learning in High-Energy Physics is also provided.

Chapter 5 introduces the main work of this thesis, showing the MSSM Higgs to $\mu^+\mu^-$ signature and describing the main objects and observables that are employed at the analysis level, focusing also on the Deep Learning applications used for the reconstruction and classification of jet objects. Also, a description of the Monte Carlo simulated samples and the Run 2 dataset used for the analysis is described in detail.

Chapter 6 covers the model independent MSSM $H/A \rightarrow \mu^+\mu^-$ analysis using 2016 data. In particular, the optimisation, training and inference of the parameterised Neural Networks, used for the multivariate analysis strategy, are described in detail. Thus, a detailed description of the network results, with the relative performances, is given.

Chapter 7 provides the description of the obtained results. Starting from the discrimination results of the parameterised Neural Network, the total signal efficiency is computed, and the MSSM model independent 95% CL expected upper limits on the production cross section times branching ratio are obtained, for a generic neutral Higgs decaying to $\mu^+\mu^-$.

Chapter 8 concludes this thesis with an overview of the entire work. Also an highlight on the uncovered aspects is provided, with a discussion of future improvements that will be addressed in future research activities.

THE HIGGS BOSON IN THE STANDARD MODEL

In particle physics, the Standard Model (SM) is a quantum field theoretical model that describes all the fundamental particles, and three of four interactions among them: the electromagnetic, the strong and the weak interactions. Such model is the result of an incredible amount of theoretical achievements acquired during the last sixty years.

According to this model, three main fields exist: matter fields with spin-half, describing matter particles called *fermions*. Spin-1 vector gauge fields, describing the mediators of fundamental interactions, called *vector bosons*. Finally, the Higgs scalar field with spin-0, describing the spontaneous symmetry breaking via the *Brout-Englert-Higgs mechanism*, and allowing the definition of non-zero mass gauge bosons and fermions.

1.1 THE FUNDAMENTAL PARTICLES

Currently, the fundamental particles of the SM are described in Figure 1.1. The fundamental fermions can be divided in two separate groups: the *quarks*, which interact with the strong, electromagnetic and weak interactions, and the *leptons*, which do not interact via the strong force but only with the weak and electromagnetic ones.

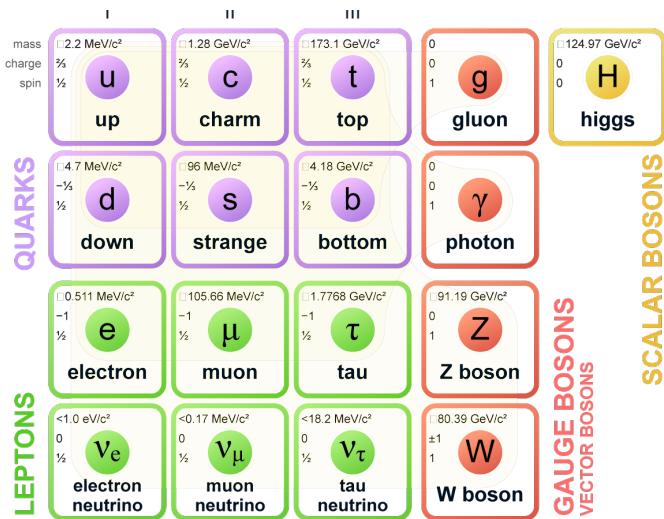


Figure 1.1: Fundamental particles of the standard model [1].

All these particles are described in the SM as point-like and *elementary*, meaning that so far there is no evidence of internal structures. The ordinary matter is composed only by one lepton with negative charge, called *electron* (e) and two types of quarks, the quark *up*

(u) with charge +2/3 and quark *down* (d) with charge -1/3, forming the building blocks of matter: protons (u-u-d) and neutrons (u-d-d). Together with the electronic neutrino (ν_e), they constitute the first generation of fermions (named in roman numbers in the columns of Figure 1.1). The second and third generation have a similar structure and interact in the same way, but have a larger mass. The SM also includes a corresponding antiparticle for each particle with an identical mass, but opposite quantum numbers.

1.1.1 Leptons

In the SM, leptons include the electron (e), the muon (μ), and the tau (τ), with a mass of $0.5 \text{ MeV}/c^2$, $105.7 \text{ MeV}/c^2$ and $1777 \text{ MeV}/c^2$ respectively, each one with its corresponding neutrino (ν_e , ν_μ , ν_τ). They are divided in three weak-isospin doublets, with different weak properties:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}, \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}, \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}.$$

The electron was the first elementary particle to be discovered in the end of the 19th century.

The muon, instead, is an elementary particle with identical properties to those of the electron but with a higher mass. It was discovered by Anderson and Neddermeyer at Caltech in 1936 [2] while studying cosmic radiation, since muons on earth are naturally produced in cosmic rays. Protons arriving from the Sun interact with atomic nuclei in the atmosphere, producing pions (particles composed by a quark/antiquark of the first generation), which subsequently decay into muons. Muons are unstable particles with a mean lifetime of $2.2 \mu\text{s}$ and a single decay channel $\mu \rightarrow \nu_\mu \bar{\nu}_e e$.

The tau lepton was discovered in 1975 at Stanford Linear Accelerator Center (SLAC) [3] by Martin Perl and his team, earning the Nobel Prize in 1995 for its discovery. The τ has the highest mass of all leptons, making it the only lepton with both leptonic and hadronic¹ decay modes. Leptonic decay products comprise either an electron or a muon with accompanying neutrinos, $\tau^- \rightarrow \nu_\tau \bar{\nu}_\tau l^-, l = e, \mu$, with charge conjugate modes implied. Hadronic decay modes, instead, are very rich in variety but most frequent decays comprise final states with one or three charged hadrons and zero or a few neutral hadrons.

Neutrinos, unlike the other charged leptons, are electrically neutral, and are considered to be massless in the SM being their masses experimentally lower than 2 eV [4]. Despite the conservation of lepton flavour is verified in each family for each decay, neutrinos can change their flavour while propagating, with only the total lepton number being conserved². This phenomenon, known as *neutrino oscillation* (Section 1.5), derives from the fact that the flavour and mass eigenstates of the neutrinos do not coincide, causing also the slightly non-zero mass.

¹ An hadron is a subatomic particle made of two or more quarks held together by the strong interaction. It derives from the greek work ἀδρός - "stout, thick".

² A lepton quantum number (L) is associated to leptons, which is additive and assumes values 1 and -1 for particles and antiparticles, respectively.

1.1.2 Quarks and Gluons

Quarks occur in six different flavours: up (u), charm (c), and top (t), that have an electric charge equal to +2/3 compared to the electron charge; down (d), strange (s), and bottom (b), carrying an electric charge equal to -1/3 compared to the electron charge. They are also grouped into three generations, resulting in the following isospin doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}.$$

The masses range from a few MeV of the u and d quarks to the ≈ 173 GeV of the t quark. This large range of masses allow for decays of heavy quarks into light quarks, via the weak interaction which can change the flavour of the quarks³. The top quark, for instance, weakly decays almost exclusively in a bottom quark with a lifetime so short that a bound system of two top quarks cannot be formed.

Similar to the electric charge, quarks have a quantum number called *color*. This charge can assume three different states: *red* (*r*), *blue* (*b*) and *green* (*g*) each one with a positive and negative sign. In fact, anti-quarks have the opposite color charges: the *antired* (\bar{r}), the *antiblue* (\bar{b}) and *antigreen* (\bar{g}). The strong interaction is mediated by 8 gluons with zero-mass, carrying a color and an anticolor charge e.g. $r\bar{b}$, $b\bar{g}$, $g\bar{r}$ and all the other possible combinations. Due to a property of the strong interaction called *color confinement*, free particles must have a color charge of zero: for this reason, gluons and quarks can not exist in isolation but only in a form of composite hadrons and, possibly, in a form of glueballs (gluon self-interactions).

All hadrons are divided into *baryons* or *mesons* based on their baryon number *B*. Quarks have *B*=+1/3 while anti-quarks have *B*=-1/3. A composite quark system with a baryon number of +1 (-1) is called baryon (antibaryon), and a system with a baryon number of 0 is called meson. Therefore, almost all known baryons are made of three quarks while mesons are made of a quark-antiquark couple. However, exotic states are theoretically possible and some of them have been discovered quite recently. For instance, the exotic pentaquark $uuc\bar{d}\bar{c}$ has been discovered by the LHCb collaboration, in a decay of the Λ_b^0 baryon [5]; also, an exotic tetraquark state, composed of $c\bar{c}d\bar{u}$ was discovered by the Belle collaboration [6].

1.2 THE STRUCTURE OF THE STANDARD MODEL

The three interactions described by the SM, the strong, weak and electromagnetic interactions, are introduced into a Quantum Field Theory framework via a local gauge invariance requirement. In theories such the SM, a Lagrangian that describes the field evolution over time must have an internal symmetry under a Lie group and, for each generator of the

³The phenomenon of quark mixing between different generations, with the weak interaction, is directly connected to the non-matching flavour and mass eigenstates.

group, a corresponding field must be introduced (gauge field).

In the SM, the description is based on this gauge symmetry group [7–9]

$$U(1)_Y \otimes SU(2)_L \otimes SU(3)_c. \quad (1.1)$$

$SU(3)_c$ is the strong interaction group, described by the Quantum Chromodynamics (QCD) theory. Quarks are a triplet of this symmetry group and the subscript c stands for color charge. According to the gauge theories, for N possible particles described by a gauge symmetry group there are $N^2 - 1$ corresponding gauge fields, of which $N^2 - N$ charged and $N - 1$ neutral fields. For the $SU(3)_c$ group, therefore, there are 8 possible gauge fields, corresponding to the 8 gluon fields (see Section 1.1.2). As previously mentioned, free quarks cannot be observed in nature and only colorless bound states exist (as baryons or mesons).

The $SU(2)_L$, together with $U(1)_Y$, describe the unification between weak and electromagnetic interactions, the so called *electroweak* theory. Firstly explained in the 60' by S. Glashow, the structure is based on the group symmetry $SU(2)_L \otimes U(1)_Y$. The charge of the $U(1)$ group, in the context of unification, does not coincide anymore with the electric charge but it represents another generator, the weak *hypercharge* (Y_W), linked with the electric charge through the Gell-Mann-Nishijima relation:

$$Q = I_3 + \frac{1}{2} Y_W, \quad (1.2)$$

where Q is the electric charge, and the I_3 generator is the "charge" of the $SU(2)$ weak interaction, the third component of the *weak isospin*. This component may assume three different values: $\pm \frac{1}{2}$ for doublets and zero for singlets. The weak interaction only acts on left-handed particles (hence the subscript L under $SU(2)$), meaning that the spin has to be anti parallel w.r.t momentum i.e. with negative helicity⁴. Right-handed particles are put in isospin singlets. Specifically, the quark sector is described by two singlets (u_R, d_R) and one doublet ($q_L = (u_L, d_L)$) under $SU(2)_L$, while leptons are described by one singlet (e_R) and one doublet ($l_L = (e_L, \nu_L)$). As previously mentioned, this must be copied three times, one per generation. In the mass-less limit, the weak interaction only acts on left fermions: for this reason, right-handed neutrinos (ν_R) are not included in the SM description. Table 1.1 shows the fermionic multiplets in the electroweak interaction.

In terms of gauge fields, to the $SU(2)_L \otimes U(1)_Y$ symmetry group correspond four different vector bosons mediating the electroweak interaction: the massless photon (γ) responsible for the electromagnetic interaction, and the three massive bosons W^+, W^-, Z^0 (two charged and one neutral) mediating the weak interaction⁵.

⁴ The two chiral components, left and right, are expressed by the Dirac gamma matrix $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ with the L operator ($\frac{1-\gamma^5}{2}$) and R operator ($\frac{1+\gamma^5}{2}$). Under the massless limit, the chirality components are equivalent to the helicity spin-momentum projections.

⁵ Hereafter, for the sake of clarity, the weak bosons will be denoted as W and Z bosons.

	Fermionic multiplets			I	I_3	Q	Y_W
Lepton	$(\nu_e)_L$	$(\nu_\mu)_L$	$(\nu_\tau)_L$	1/2	$+\frac{1}{2}$	0	-1
	e_R	μ_R	τ_R	0	0	-1	-2
Quark	$(u')_L$	$(c')_L$	$(t')_L$	1/2	$+\frac{1}{2}$	$\frac{2}{3}$	$\frac{+1}{3}$
	u_R	c_R	t_R	0	0	$\frac{2}{3}$	$\frac{4}{3}$
	d'_R	s'_R	b'_R	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$

Table 1.1: Fermionic multiplets on the electroweak interaction. The left-handed doublets of weak isospin are shown in parenthesis; the right-handed singlets are isolated. For the left-handed quarks, the up-like quarks (u, c, t) correspond to the strong eigenstates, while the down-like (d', s', b') are the weak eigenstates shifted according to the CKM matrix.

1.2.1 Local gauge symmetry invariance

Once the symmetry and the quantum numbers of each field are chosen, the interactions are determined as a local gauge invariance. *Global* gauge invariances are common in particle physics and it is a phase transformation of the matter field ψ :

$$\psi_\mu \rightarrow \psi'_\mu = \psi_\mu e^{i\alpha/\hbar c}, \quad (1.3)$$

where α is a scalar with the same value on each space-time point⁶. The invariance for this transformation implies that the phase of the wave function is arbitrary. The derivative of such function transforms like the function itself

$$\partial\psi'/\partial x^\mu = \partial_\mu\psi' = e^{i\alpha}\partial_\mu\psi. \quad (1.4)$$

The gauge transformation can be generalised considering α as a function of space-time, $\alpha = \alpha(x)$. This is called *local* gauge transformation, since it might differ from point to point. The standard model gauge theories must be invariant also for this transformations. However, writing the local gauge transformation in the form

$$\psi \rightarrow \psi' = e^{i\alpha(x)}\psi, \quad (1.5)$$

the derivative of $\psi(x)$ becomes

$$\begin{aligned} \frac{\partial\psi'(x)}{\partial x^\mu} &= \partial_\mu\psi'(x) = e^{i\alpha(x)}[\partial_\mu\psi(x) + i\psi(x)\partial_\mu\alpha(x)] \\ &\neq e^{i\alpha(x)}\partial_\mu\psi(x), \end{aligned} \quad (1.6)$$

meaning that the derivative does not transform itself like the wave function; therefore, the Lagrangian is not invariant under local gauge transformation. In order to obtain a local gauge invariance, an interaction term must be added defining the structure of the

⁶This transformation is part of the classical electrodynamics (QED) with the $U(1)$ gauge symmetry since it is easier to explain. The same applies to the other groups, with different generators.

interaction.

This is done by introducing the *covariant derivative* D_μ ,

$$D_\mu = \partial_\mu + ieA_\mu, \quad (1.7)$$

and transforming the electrodynamics field A_μ , according to the relation

$$A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) - \frac{1}{e}\partial_\mu\alpha(x). \quad (1.8)$$

At this point, the derivative of the wave function, described in Eq.1.6, becomes

$$\begin{aligned} D_\mu\psi'(x) &= e^{i\alpha(x)}[\partial_\mu\psi(x) + i\psi(x)\partial_\mu\alpha(x) + ieA_\mu(x)\psi(x) - i\psi(x)\partial_\mu\alpha(x)] \\ &= e^{i\alpha(x)}D_\mu\psi(x), \end{aligned} \quad (1.9)$$

with a gauge transformation now equivalent to the function itself. The necessity behind the covariant derivative to obtain such result relies in the interaction of an electric charge with the field A_μ , i.e. eA_μ .

For a free fermion, in fact, the Lagrangian is derived from Dirac's equation

$$\mathcal{L} = \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x) = \mathcal{L}_{free}, \quad (1.10)$$

with γ^μ being the Dirac matrices and $\bar{\psi}(x) = \psi^\dagger(x)\gamma^0$ the Dirac conjugate field. Replacing the standard derivative ∂_μ with the covariant derivative D_μ , Eq.1.10 can be written as

$$\begin{aligned} \mathcal{L} &= \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) \\ &= \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x) - qA_\mu\bar{\psi}\gamma^\mu\psi. \end{aligned} \quad (1.11)$$

With this new formalism, the Lagrangian acquires a new term:

$$\mathcal{L} = \mathcal{L}_{free} - J^\mu A_\mu, \quad (1.12)$$

where $J^\mu = q\bar{\psi}\gamma^\mu\psi$ is a current density of the fermion with charge q , interacting with the electromagnetic field A_μ . This Lagrangian, with some algebraic manipulation, is also invariant for the local gauge transformations of Eqs. 1.5 and 1.8.

To obtain the complete Lagrangian of the QED, another term is missing: the kinetic energy of the gauge field A_μ , that describes the propagation of the particle associated with the field. For the electromagnetic field, it is written in the form:

$$\mathcal{L}_{kinetic} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.13)$$

where $F^{\mu\nu} = \partial^\nu A^\mu - \partial^\mu A^\nu$ is the electromagnetic field strength tensor. Adding every term, the final QED Lagrangian is:

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x). \quad (1.14)$$

The requirement of a local gauge invariance generates the interaction term and defines the structure of the interaction. This requirement also forbids the mass term of the vector field, which has the form $m^2 A_\mu A^\mu$.

Gauge invariance in strong interaction - $SU(3)_c$. The local gauge invariance also fixes the interaction parts in the QCD and electroweak theories (also called *Yang-Mills* theories, from their discovery in 1957). The generators of the $SU(3)$ Lie group, which is a non-abelian⁷ symmetry group of the SM, are connected with the following commutation:

$$[\lambda^a, \lambda^b] = 2if^{abc}\lambda^c, \quad (1.15)$$

where λ^a are the eight Gell-Mann (3x3) matrices and f^{abc} are the structure constants of the group. Similarly to the QED case, the QCD Lagrangian can be written in this form

$$\mathcal{L} = -\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} + i\bar{q}_i(\gamma_\mu D_{ij}^\mu - m\delta_{ij})q_j, \quad (1.16)$$

where $G_{\mu\nu}^a$, compared to $F_{\mu\nu}$ of the electromagnetic field, are eight gluon field strength tensors, q_i are quark fields and D_{ij}^μ is the $SU(3)$ covariant derivative. The a index is referred to the eight different generators, while the (i, j) indexes correspond to the three representation of the $SU(3)$ group. In a local $SU(3)$ gauge transformation, the covariant derivative must be in this form

$$D_{ij}^\mu = \partial^\mu\delta_{ij} + ig_s \frac{\lambda_{ij}^a}{2} A^{a\mu}, \quad (1.17)$$

where the $A^{a\mu}$ corresponds to the gluon fields, with a being the eight kinds of gluon. Moreover, the field strength tensor is given by

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_s f^{abc} A_\mu^b A_\nu^c. \quad (1.18)$$

The non-linear term in Eq.1.18, present in all Yang-Mills theories, allows the gauge fields to be self-interacting and therefore allowing gluons to interact with themselves.

Gauge invariance in electroweak interaction - $SU(2)_L \otimes U(1)_Y$. The structure of $SU(2)_L$ Yang-Mills theory is similar to $SU(3)_c$, but it becomes more complicated due to the *Brout-Englert-Higgs* mechanism, described in Section 1.3. In $SU(2)$ Lie group, the commutation between generator is the following:

$$[\sigma^i, \sigma^j] = 2ie^{ijk}\sigma^k, \quad (1.19)$$

where σ^i are the Pauli matrices and e^{ijk} are the structure constants of the group, the anti-symmetric Levi-Civita symbol in three dimensions.

The $U(1)_Y$ local gauge invariance leads to a single gauge field, denoted B_μ , which can

⁷ A Lie group is called *non-abelian* when the multiplication between generators is not commutative, i.e. $[a, b] = ab - ba \neq 0$

interact on both left-handed and right-handed particles. The weak isospin of $SU(2)_L$ is related to a triplet of gauge bosons (\vec{W} or W^l) as generators. As mentioned in Table 1.1, the fermions are represented by a doublet with $I_3 = \frac{1}{2}$, denoted now as L , indicating a left-handed chirality; and two singlets with $I_3 = 0$, corresponding to right-handed fermions not interacting with the weak \vec{W} bosons. Denoting the up-like fermions and down-like fermions with ψ and ψ' respectively,

$$\begin{aligned} L &\equiv \frac{1 - \gamma^5}{2} \begin{pmatrix} \psi \\ \psi' \end{pmatrix} = \begin{pmatrix} \psi_L \\ \psi'_L \end{pmatrix}, \\ \psi_R &\equiv \frac{1 + \gamma^5}{2} \psi, \\ \psi'_R &\equiv \frac{1 + \gamma^5}{2} \psi'. \end{aligned} \tag{1.20}$$

Changing the chirality of the fermion implies a different field: therefore the fermion mass is not included in the Lagrangian.

The kinetic terms of the gauge fields in the Lagrangian are

$$\mathcal{L}_{kinetic} = -\frac{1}{4} W_{\mu\nu}^l W^{l\ \mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.21}$$

where the field strength tensors are

$$W_{\mu\nu}^l = \partial_\nu W_\mu^l - \partial_\mu W_\nu^l + g \epsilon_{jkl} W_\mu^j W_\mu^k, \tag{1.22}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{1.23}$$

Starting from the free-field Lagrangian described in Eq.1.10, the Lagrangian for the electroweak interaction can be written as

$$\mathcal{L}_{EW} = \mathcal{L}_{kinetic} + \bar{L} i \gamma^\mu \delta_\mu L + \bar{\psi}'_R i \gamma^\mu \delta_\mu \psi'_R + \mathcal{L}_{quarks}. \tag{1.24}$$

The local gauge invariance of $SU(2)_L \otimes U(1)_Y$ requires four gauge fields ($W_\mu^1, W_\mu^2, W_\mu^3, B_\mu$), and the replacement of the classical derivative ∂_μ with the covariant derivative:

$$D_\mu = \partial_\mu + ig' \frac{Y}{2} B_\mu + ig \frac{\sigma_l}{2} W_\mu^l, \tag{1.25}$$

where the coupling constants g and g' arise from the $SU(2)_L$ and $U(1)_Y$ gauge groups, respectively; σ_l are the Pauli matrices, generating $SU(2)_L$, and Y is the hypercharge of $U(1)$. The $SU(2)_L$ part involving the Pauli matrices applies only to left-handed fermions; for the right-handed fermions, the σ_l have to be replaced with a zero. Excluding \mathcal{L}_{quarks} from the computation, for simplicity reasons, the final EW Lagrangian becomes:

$$\begin{aligned} \mathcal{L}_{EW} = & -\frac{1}{4} W_{\mu\nu}^l W^{l\ \mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{\psi}'_R i \gamma^\mu (\partial_\mu + ig' \frac{Y}{2} B_\mu) \psi'_R \\ & + \bar{L} i \gamma^\mu (\partial_\mu + ig' \frac{Y}{2} B_\mu + ig \frac{\vec{\sigma}}{2} \vec{W}_\mu) L + \mathcal{L}_{quarks}. \end{aligned} \tag{1.26}$$

The non-linear term in Eq. 1.22 allows a complex structure of gauge bosons self-interactions, including both cubic ($ZWW, \gamma WW$) and quartic ($ZZWW, WWWW, \gamma\gamma WW$) vertices. The interaction with matter particles is mediated by the exchange of massless gauge bosons, deduced from the absence of a mass term in the \mathcal{L}_{EW} .

The concept of local gauge invariance in the Yang-Mills theories leads to an interaction term mediated by a massless boson. However, experimental observations showed the existence of massive vector bosons, carrying the weak interaction. This fundamental discovery, leading to the Nobel Prize in Physics for the italian Carlo Rubbia and the dutch Simon van der Meer in 1984, also brought to light a theoretical problem behind the concept of symmetry in the SM. In the first part of the 1960s, a few physicists theorised a mechanism that allows gauge fields to have nonnull mass, known as *Brout-Englert-Higgs* mechanism (usually shortened as *Higgs* mechanism). The main concept behind this idea was the introduction of a spontaneous symmetry breaking in the gauge theory, with the formulation of a new scalar field (and relative particle) that allows gauge bosons to acquire mass [10, 11].

1.3 THE HIGGS MECHANISM

Mass terms for the gauge bosons can be introduced by a spontaneous symmetry breaking (SSB), while keeping the theory renormalizable. In a simple case, the spontaneous symmetry breaking requires a scalar field ϕ with the following Lagrangian:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - U(\phi), \quad (1.27)$$

where $U(\phi)$ is a potential term defined as

$$U(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4. \quad (1.28)$$

The energy can be modified only with spatial variation in ϕ , and λ is positive, bringing the Lagrangian to a global minimum. The vacuum value of ϕ is known as the *vacuum expectation value* (VEV) of ϕ (written as $\langle\phi\rangle$). There are two possible outcomes depending on the sign of μ^2 :

- $\mu^2 > 0$: the potential $U(\phi)$ has only one minima at $\phi = 0$, corresponding to the vacuum state, with a clear symmetry;
- $\mu^2 < 0$: there are two degenerate minima at $\phi_0 = \pm v = \pm\sqrt{-\frac{\mu^2}{\lambda}}$, each one possible as vacuum state.

The specific case with $\mu^2 < 0$ shows a spontaneous symmetry breaking, which is the ground base for the Higgs mechanism. The shape of the potential, in a bi-dimensional case, is shown in Figure 1.2.

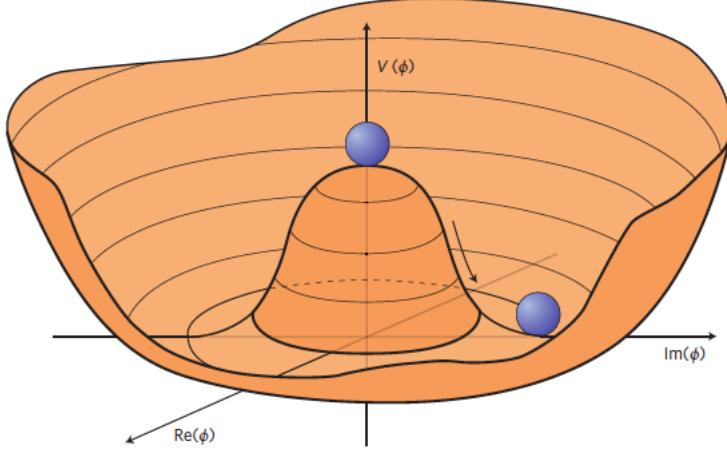


Figure 1.2: Graphical representation of the "Mexican Hat" shaped potential of a 2-scalar field theory. The unstable $\mu^2 > 0$ case is shown on top of the hat, while the degenerate states are shown in the hat's minima.

In the SM case, the spontaneous symmetry breaking is realised via a complex scalar field which breaks the $SU(2)_L \otimes U(1)_Y$ gauge symmetry [12–17]. In particular, a new complex scalar field is introduced:

$$\Phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (1.29)$$

which behaves as a doublet in $SU(2)_L$ and has charge +1/2 under $U(1)_Y$. The additional Lagrangian terms due to the scalar field, therefore will be:

$$\mathcal{L} = (D_\mu \Phi)^\dagger (D^\mu \Phi) + V(\Phi). \quad (1.30)$$

The potential term $V(\Phi) = V(\Phi^\dagger \Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$ will have a minimum VEV at $v = \sqrt{\frac{\mu^2}{2\lambda}}$ and the Φ field, with the appropriate gauge fixing, can be rewritten as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (1.31)$$

where $H(x)$ is the physical Higgs field. The choice of a VEV automatically fixes the masses of three gauge bosons, for now being massless: in fact, the terms in the Lagrangian containing the covariant derivative $(D_\mu \Phi)^\dagger (D^\mu \Phi)$, after the replacement of the derivative and the explicitation of the gauge fields of $SU(2)_L \otimes U(1)_Y$, becomes

$$\begin{aligned} & \frac{1}{2} \partial_\mu H(x) \partial^\mu H(x) + \frac{1}{8} (g W_\mu^3 - g' B_\mu)^2 \cdot (v + H(x))^2 \\ & + \frac{1}{8} g^2 (W_\mu^1 - i W_\mu^2) (W^{1\mu} + i W^{2\mu}) \cdot (v + H(x))^2. \end{aligned} \quad (1.32)$$

From this equation, it is possible to recognise the first term as the kinetic term for the physical Higgs boson field $H(x)$. The second term has a linear combination of the hyper-

charge gauge field B_μ and the third gauge field W_μ^3 , that can be rewritten as the Z_μ field, describing the Z_0 boson:

$$Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}} = W_\mu^3 \cos \theta_W - B_\mu \sin \theta_W, \quad (1.33)$$

where θ_W is the Weinberg angle, representing the weak mixing angle of the two coupling constants g and g' in the electroweak theory: $g' = g \tan \theta_W$. The third term contains the weak gauge fields W_μ^1 and W_μ^2 which are combined to give the W^\pm fields, describing the W^\pm bosons, redefined with the following relation:

$$W^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}. \quad (1.34)$$

The mass of the weak gauge bosons can be expressed as:

$$\begin{aligned} m_Z^2 &= \frac{v^2(g^2 + g'^2)}{4}, \\ m_W^2 &= \frac{g^2 v^2}{4} = m_Z^2 \cos^2 \theta_W. \end{aligned} \quad (1.35)$$

The second and third term also contain the cubic and quartic interaction vertices of Higgs with vector bosons, with a coupling proportional to the square of the vector boson masses.

The Z_μ orthogonal field corresponding to the electromagnetic field A_μ remains massless and represents the photon γ :

$$A_\mu = \frac{gW_\mu^3 + g'B_\mu}{\sqrt{g^2 + g'^2}}. \quad (1.36)$$

The symmetry breaking pattern therefore is $U(1)_Y \otimes SU(2)_L \rightarrow U(1)_{em}$. Three of the four bosons acquire a positive mass, while the photon remains massless, since the $U(1)_{em}$ symmetry remains unbroken. The photon, in fact, does not have an interaction vertex with the Higgs boson, but the $H \rightarrow \gamma\gamma$ decay can happen via loops. The coupling between photons and matter can be derived from the definition of the covariant derivative with a coupling proportional to the common electric charge and with quantum numbers given by the Eq.1.2. Looking at the mass term μ associated with the potential term $V(\Phi)$ ($-\mu^2\Phi^\dagger\Phi$), the Higgs boson mass is given by $m_H = \sqrt{2\lambda v^2}$. The Higgs mass is a free parameter of the model as it depends on the λ introduced by the Higgs potential; on the other hand, the VEV v is known because it can be obtained from the Fermi constant G_F :

$$v = \sqrt{\frac{1}{\sqrt{2}G_F}}, \quad (1.37)$$

obtainable from experimental measures from the muon lifetime.

The presence of the Φ field allows mass terms proportional to the VEV for the fermions, preserving also the local gauge invariance [18, 19]. In the SM Lagrangian, in fact, the fermion mass is not included, as it would take the form $m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$, not invariant

under $SU(2)_L$. The introduction of mass for fermions are introduced via the interaction with the Higgs field, called *Yukawa* interaction. The mass terms becomes:

$$\approx y_e \bar{e}_R \Phi^\dagger e_L + y_e \bar{e}_L \Phi e_R = y_e \frac{v + H(x)}{\sqrt{2}} (\bar{e}_R e_L + \bar{e}_L e_R), \quad (1.38)$$

generating mass terms $y_e v / \sqrt{2}$, and interacting with the $H(x)$ proportional to the lepton mass. A similar term also allows the Higgs boson to give mass both to the down type and up type quarks with Yukawa interactions. The mass terms for the down type quarks are totally analogous to the lepton masses, while in the up type terms the conjugated field $\tilde{\Phi}$ is used.

The incredible success of the spontaneous symmetry breaking and the Higgs mechanism, made the electroweak theory a pillar of particle physics, with a theoretical model later confirmed by experimental results [10, 11].

1.4 THE HIGGS BOSON AT THE LHC

Searching the Higgs boson was one of the main objectives at the Large Electron-Positron Collider (LEP) at CERN [20]. However, only a lower bound on the Higgs mass was determined at 114.4 GeV with a 95% confidence level [21]. After the LEP shutdown, the Higgs search was lead by the CDF and D0 experiments, at the Tevatron collider in the Fermi National Accelerator Facility in the United States, where the dominant production mechanism was the associated production with a W or a Z . The results from these two experiments using data at $\sqrt{s} = 1.96$ TeV were later combined [22], showing an exclusion region in the Higgs mass between 156 and 177 GeV, together with a range of lower masses excluded by LEP. In 2012, at the CERN Large Hadron Collider (LHC), a Higgs-like particle consistent with the SM Higgs at 125 GeV was discovered by the ATLAS [23] and CMS [24, 25] Collaborations.

1.4.1 Higgs production and decay modes

In a proton-proton collider with a center-of-mass energy up to 13 TeV, like the LHC, there are various production mechanisms of the Higgs boson: the most relevant are the gluon-gluon fusion (ggF) production with the highest cross section, the vector boson fusion (VBF), the associated production with a vector boson (WH or ZH), and the associated production with a top quark pair ($t\bar{t}H$). The Feynman diagrams for these production mechanisms are shown in Figure 1.3. The total cross section depends on the center-of-mass energy, as shown in Figure 1.4 considering a Higgs boson from the SM with a mass of 125 GeV.

The **gluon-gluon fusion** (ggF) is the dominant production mode due to the high density of gluons in p-p collisions at the LHC center-of-mass energies ($\approx 85\%$ of the total). It is characterised by an intermediate heavy quark loop, with the largest contribution arising

from the top quark loop (about 90%), which has a mass comparable to the Higgs boson mass. The remaining contribution comes from the bottom quark, around 5-10% of the total cross section.

The **vector boson fusion** (VBF) is roughly ten times rarer than ggF, representing the second dominating contribution to the H boson production at the LHC. The leading diagram consists in the scattering of two quarks in the t or u channel, in a weak process which leads them to exchange a vector boson, W or Z. The fusion of these two vector bosons produces the H boson. Since the exchange of momentum between the quark and the boson is lower than the initial energy of the two quarks, the channel has a clear experimental signature, associated to a pair of high-rapidity quarks in the final state, detectable as forward hadronic jets⁸, at low angles with respect to the beam axis. In addition, the VBF is a pure electroweak process, with low hadronic activity expected in the gap between the two jets, where the Higgs decay modes are typically found.

The **associated production with a vector boson**, also called Higgs-strahlung, is the third dominating mode at the LHC and it is two times rarer than VBF. In this mechanism, the Higgs boson is irradiated from a generic vector boson V leading to a final state with the H boson associated to a W or Z boson. The main contributions to the VH cross section comes from quark initiated processes $q\bar{q} \rightarrow ZH$ and minor contribution to the ZH production comes from gluon processes $gg \rightarrow ZH$. The presence of charged leptons and neutrinos from this specific process is really useful for the $H \rightarrow b\bar{b}$ channel, where is difficult to select events with respect to the fully hadronic final state and the dominant background from QCD.

The **production in association with a top quark pair**, shortened with $t\bar{t}H$, is characterised by a final state including the Higgs and two top quarks, where a t is created by two gluons and the Higgs is produced by the fusion of a $t\bar{t}$ couple. It is the smallest contribution, around 1%, but it is important because it allows a direct measurement of the Yukawa coupling between the Higgs and the top quark (y_t).

It is also worth mentioning another rare production modes, important in the next sections of this thesis, the **b-associated production** or $b\bar{b}H$, where the Higgs is produced together with two bottom quarks. The cross section of the $b\bar{b}H$ process is comparable to the $t\bar{t}H$ process.

The Higgs boson can decay in a large variety of channels, coupling with all SM particles (massive and not) via intermediate loops, allowing the investigation of different couplings to many SM particles. In particular, the Higgs boson can decay into pairs of heavy fer-

⁸ Due to confinement, described in Section 1.1.2, quarks are not detectable as single entities but, soon after their production, they hadronize producing an ensemble of particles, called *jets*.

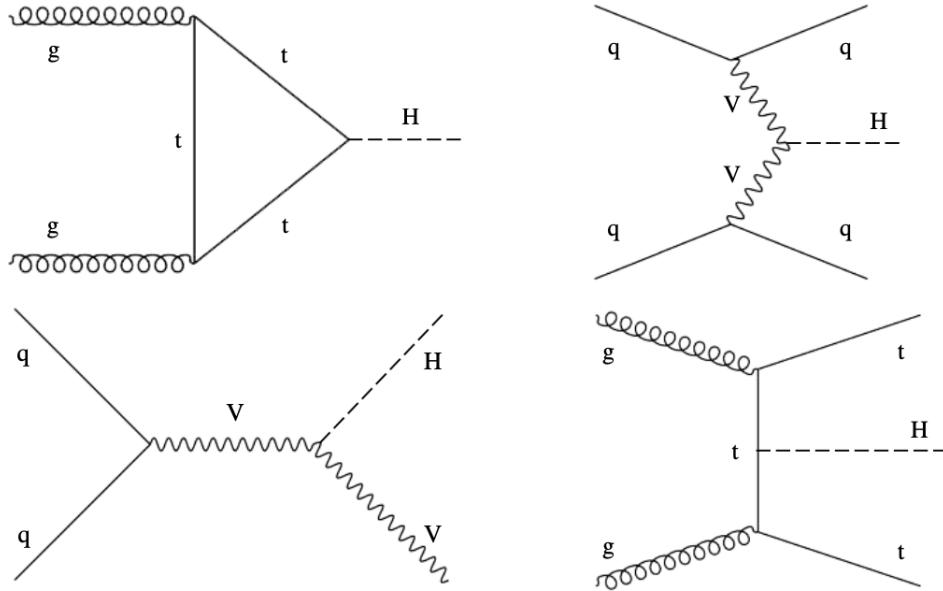


Figure 1.3: The four main mechanism of Higgs production at the LHC energy scale. Gluon fusion (top left), vector boson fusion (top right), associated production with a vector boson (bottom left) and $t\bar{t}H$ production (bottom right).

mions through a Yukawa coupling and pairs of gauge bosons, as shown in Figure 1.5, proportionally to the fermion masses (as shown in the previous section):

$$g_{f\bar{f}H} \propto m_f, \quad (1.39)$$

and, for vector boson couplings, proportionally to the square of the boson mass:

$$g_{VVH} \propto m_V^2 \quad (1.40)$$

The branching ratio, or B.R., defined as the partial width of a decay channel compared to the total decay width:

$$B.R.(H \rightarrow X) = \frac{\Gamma(H \rightarrow X)}{\sum_Y \Gamma(H \rightarrow Y)}, \quad (1.41)$$

are fully determined by the value of its mass (m_H) and are briefly summarised in Figure 1.6. However, the role of each decay mode is not entirely related to the B.R. of the process, but also on an efficient extraction of signal events while rejecting associated backgrounds. For instance, the $H \rightarrow ZZ$ (B.R. 2.6%) and $H \rightarrow \gamma\gamma$ (B.R. 0.2%) decay modes are the rarest channels possible but they played a crucial role in the discovery of the Higgs due to the excellent resolution on the invariant mass as well as the good signal-to-background ratio. Other decay modes, like the $H \rightarrow b\bar{b}$ (B.R. 58%) are dominant but offers a small resolution due to the high amount of background with similar decay products.

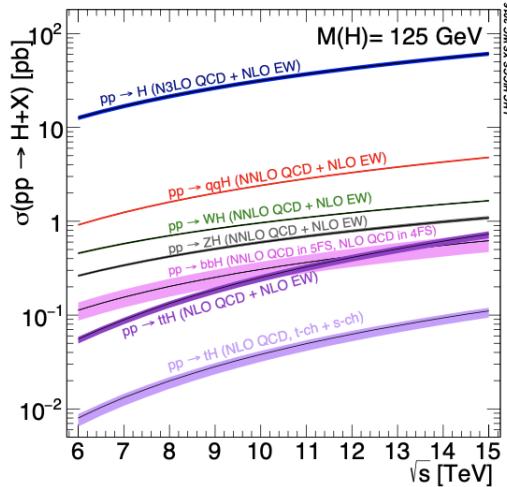


Figure 1.4: In the left plot, the production cross section for \sqrt{s} ranging from 6 to 15 TeV, considering a Higgs boson of 125 GeV. On the right table, the numeric values of cross section, at the LHC centre-of-mass energy $\sqrt{s} = 13$ TeV [26].

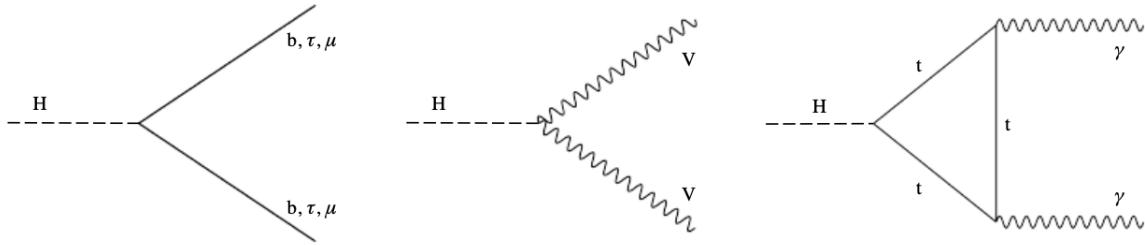


Figure 1.5: Higgs decay vertices with a pair of fermions (left diagram), a pair of vector bosons (central diagram) and a pair of photons through a top loop (right diagram).

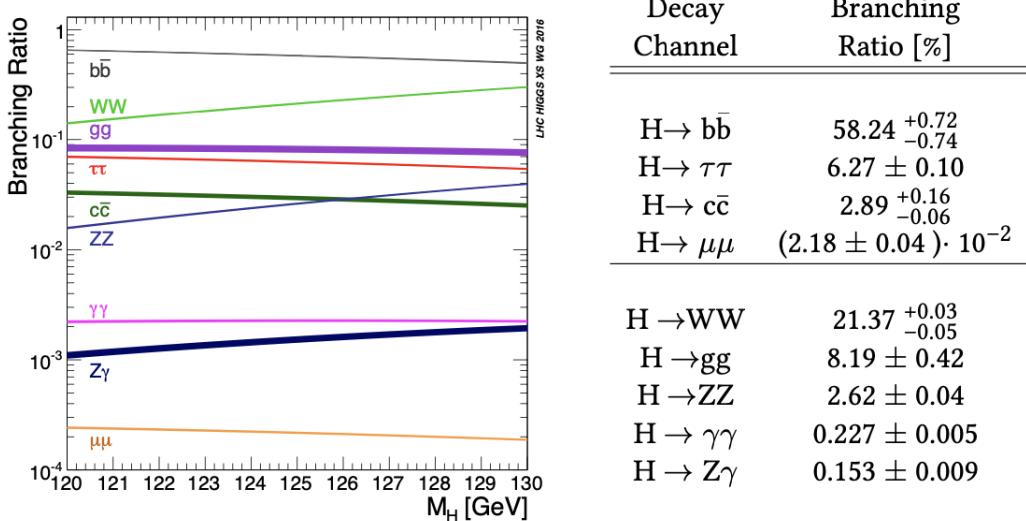


Figure 1.6: Decay branching ratios for the Higgs boson mass in the 120-130 GeV range. In the table, the SM predictions for a Higgs of 125 GeV.

1.4.2 Experimental discovery

In 2012, at the Large Hadron Collider at CERN, with a center-of-mass energy of 7 and 8 TeV, an excess of events peaked around 125 GeV was observed above the expected background reaching a statistical significance of 5σ in both the CMS and ATLAS experiments. The discovery has been obtained from the $H \rightarrow \gamma\gamma$ and the $H \rightarrow ZZ^* \rightarrow 4l$ channels, thanks to the good reconstructed mass resolution in the final states. The invariant mass distribution of the four leptons coming from the $ZZ^* \rightarrow 4l$ decay is shown in Fig. 1.7. The diphoton invariant mass distributions obtained by ATLAS and CMS in the study of the $H \rightarrow \gamma\gamma$ channel are shown in Fig. 1.8.

The LHC Run 1 Higgs results are summarized in two combination papers using data from both experiments [27, 28]. The combined mass measurement performed using the full Run 1 luminosity, 5 fb^{-1} at $\sqrt{s} = 7 \text{ TeV}$ and 20 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$, for the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4l$ is:

$$m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.}) \text{ GeV} \quad (1.42)$$

Several parametrizations were used for the combined fits. The agreement between the SM prediction and the measurement can be described by the signal strength modified μ . For each production mode i and decay channel f , the production and decay signal strengths are defined as:

$$\mu_i = \frac{\sigma_i}{\sigma_{i,SM}} \quad , \quad \mu^f = \frac{B.R.^f}{B.R.^f_{SM}}. \quad (1.43)$$

The global signal strength measurement $\mu = \mu_i \cdot \mu^f$, gives as a result a best-fit value of:

$$\mu = 1.09^{+0.11}_{-0.10} = 1.09^{+0.7}_{-0.7}(\text{stat.})^{+0.4}_{-0.4}(\text{exp.})^{+0.7}_{-0.7}(\text{theory}). \quad (1.44)$$

The fit results are reported in Figure 1.9.

The LHC Run 2 started in 2015 with some initial proton-proton collision data. Later, an incredible amount of data were collected: $\approx 36 \text{ fb}^{-1}$ in 2016, $\approx 41 \text{ fb}^{-1}$ in 2017, and $\approx 60 \text{ fb}^{-1}$ in 2018. This increase in statistics allowed the observation of some of the decays and production modes, not possible during Run 1. The $H \rightarrow \tau\tau$ decay was observed with the 2016 data by ATLAS and CMS [29, 30] independently, after the combined result at 5.5σ with the Run 1 data [28]. Soon after, the $t\bar{t}H$ production mode was also observed by CMS [31], followed by ATLAS later on [32]; the measurement of the Yukawa coupling with the bottom quark [33, 34]; and the recent observation of the decay of the Higgs boson into a pair of muons presented by CMS [35].

Using the data collected at 13 TeV, the study on the signal strength was performed by each experiment, in particular using 35.9 fb^{-1} collected in 2016 by the CMS experiment [36]. The result are compatible with the previous run, improving the previous measurements,

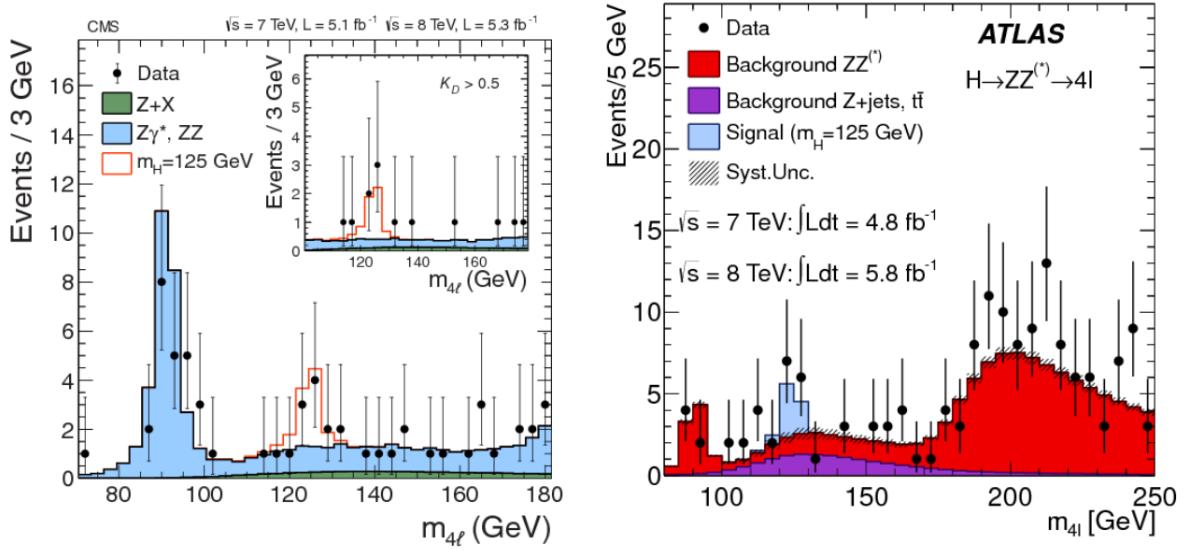


Figure 1.7: Distribution of the four lepton invariant mass for the $ZZ^* \rightarrow 4l$ decay channel. On the left side, the analysis performed by the CMS collaboration [24]: the filled histograms represent the background, the red line histogram shows the signal expectation for a Higgs boson of $m_H = 125$ GeV. On the right side, the analysis performed by the ATLAS collaboration [23]: the histogram shows the background expectation in the 80-250 GeV mass range, with the signal expectation for a SM Higgs with $m_H = 125$ GeV.

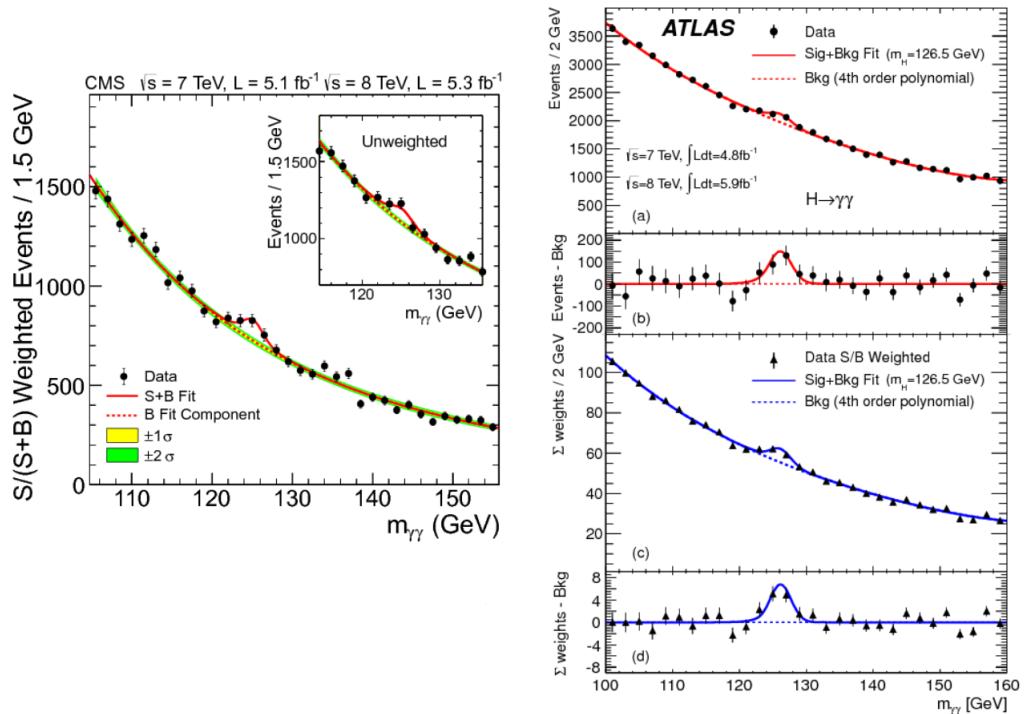


Figure 1.8: The diphoton invariant mass distribution for the $H \rightarrow \gamma\gamma$ decay channel. On the left side, the analysis performed by the CMS collaboration [24]: the lines represent the fitted background and signal, and the colored bands represent the ± 1 and ± 2 standard deviation uncertainties in the background estimate. On the right side, the analysis performed by the ATLAS collaboration [23].

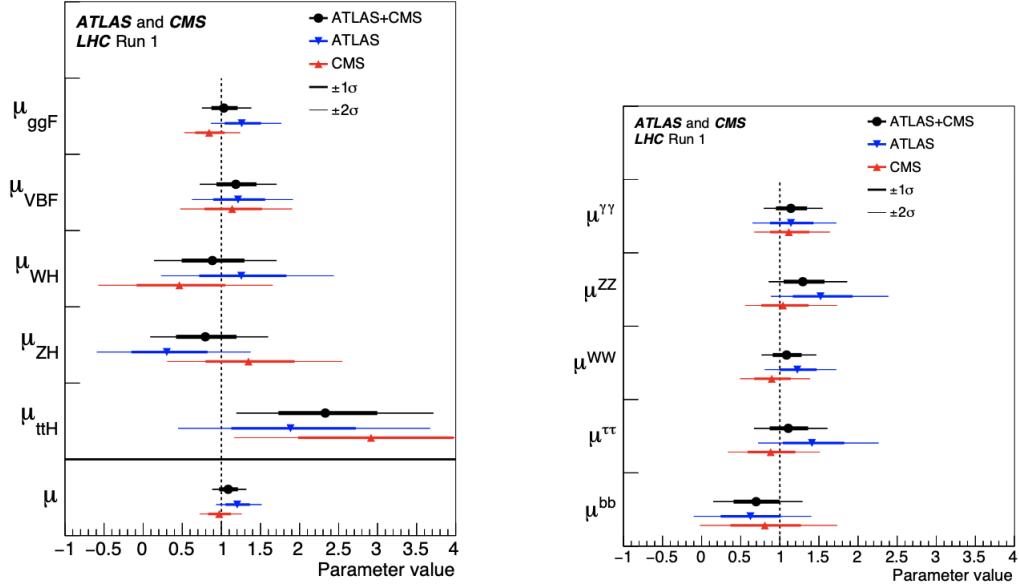


Figure 1.9: Combined CMS and ATLAS measurements on the production signal strength (left side) and branching ratio (right side). Error bars showing the 1σ and 2σ intervals.

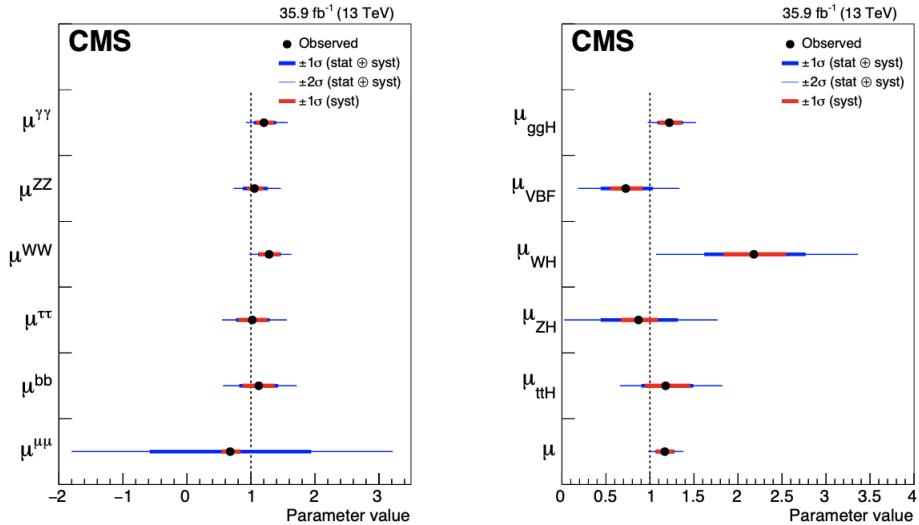


Figure 1.10: On the left plot, the signal strength per decay channel, while on the right plot the signal strength per production mode at $\sqrt{s} = 13$ TeV.

as shown in Figure 1.10 [10, 11].

Right now, the Standard Model describes particles and interactions accurately and with several experimental proofs. However, there is still a lot missing and the idea of something Beyond the Standard Model (BSM) has become an entire field of study. In the next section, an overview of some of the open questions will be addressed.

1.5 BEYOND THE STANDARD MODEL

All the experimental results (except the neutrino mass measurements, coming from the observation of the neutrino oscillations) obtained in High Energy Physics can be explained by the Standard Model. However, there are still many open questions that cannot be explained with the current theory. In this section an highlight on the problems with the SM will be briefly described, although for the moment there is no experimental evidence of physics beyond the Standard Model.

- **Neutrino masses.** Described as massless fermions within the SM, the recent experimental measurements are showing critical differences in their behaviour compared to theory. The observation of the flavour oscillation in solar and atmospheric neutrinos [37, 38], showed that neutrinos have a small mass, which is not generated by the Higgs mechanism. The picture is also further complicated by the nature of neutrinos with non-zero masses: one theory describing them as lepton number violating fermions, also known as *Majorana* neutrinos; the other describing them as lepton number conserving fermions, the *Dirac* neutrinos.
- **Families and masses.** The SM is a theory with about 20 free parameters, too many to consider this theory as fundamental. Currently it is not understood why there are only three families of elementary particles, with only the first family composing the ordinary matter. In addition, the masses of the fermions span over many orders of magnitude, behaviour still not explained by the SM.
- **Dark matter and dark energy.** Many astrophysical observations, such as the cosmic microwave background and the movement of galaxies, are showing that the baryonic matter described by the SM only accounts for roughly 5% the total mass of the universe. The remaining part is expected to be formed for about 25% by a non-relativistic, neutral, non-baryonic matter, the so-called *dark matter*, and a large energy component, called *dark energy*.
- **Gravitational force.** The gravitational interaction is not included in the SM, At the electroweak scale ($m_{\text{weak}} \approx 100$ GeV), gravity is so weak to be negligible, but a complete theory must include it, especially in the so-called Planck scale ($m_P = 10^{19}$ GeV).
- **The hierarchy problem.** There are many orders of magnitude between the electroweak and Planck scales. This problem is usually referred to as the *hierarchy problem*. In the subatomic world, the observable masses of the particles are different from the *bare* mass of their theoretical formulation, since the observable value is affected by radiative corrections. These contributions have to be computed up to the next scale of energy, where their effects become negligible. Regarding the Higgs boson mass, the relation between the observable mass and the value given by the model (*bare*) is given by [39]:

$$m_H^2[\text{obs}] = m_H^2[\text{bare}] + \Delta m_H^2, \quad (1.45)$$

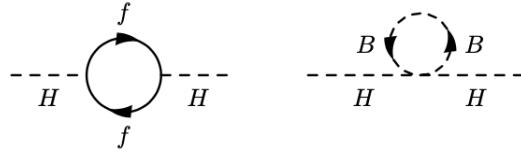


Figure 1.11: First order radiative correction for the Higgs boson corresponding to a fermionic loop (left diagram) and a bosonic loop (right diagram). In the Planck energy scale, the only way in which the computation of radiative corrections up to 10^{19} GeV can lead to such a small value of the observable mass is an extremely precise cancellation between the fermionic terms (negative contributions) and bosonic terms (positive contributions).

where Δm_H^2 represents the radiative corrections, defined as:

$$\Delta m_H^2 = \left(\frac{1}{16\pi^2} g^2 + \frac{1}{16\pi^2} \lambda - \frac{3}{8\pi^2} \lambda_f^2 + \dots \right) \Lambda_{UV}^2 \quad (1.46)$$

where the terms represent the different types of couplings of the Higgs boson. For instance, g is a generic coupling constant for the EW bosons fields, λ is the self-coupling constant of the Higgs boson, λ_f is the Yukawa coupling of the Higgs to fermions, and Λ is the large-scale (UV) cutoff parameter. At incredibly high masses, compared to the LHC energy scale, the radiative corrections to the Higgs mass, shown in Figure 1.11 at first order, tend to push that mass up to the heaviest mass scale of the theory, as visible from Eq. 1.46 when $\Lambda \approx 10^{19}$ GeV. The only way to preserve the hierarchy $m_H^2 \ll M_P^2$ is to fine-tune the bare potential in an unnatural way, in order to cancel the quantum corrections and restore the correct physical mass parameter.

- **Matter-antimatter asymmetry.** Currently in the universe, there is a slight asymmetry between matter and antimatter, not explained by the Standard Model, causing the existence of the ordinary matter.
- **The forces unification problem.** The gauge group of the SM - $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ - has three different subgroups, each one with a specific coupling constant. These constants are not fixed, but running with the energy scale and should converge at some scale, referred to as the *unification scale*. However, some experimental studies have shown that such couplings do not merge [40], as shown in Figure 1.12.

The difficulties defined above suggest that the SM might be only a part of a greater fundamental underlying theory, and several extensions have been proposed. One of them, the Supersymmetry, will be addressed in more detail in the next Chapter.

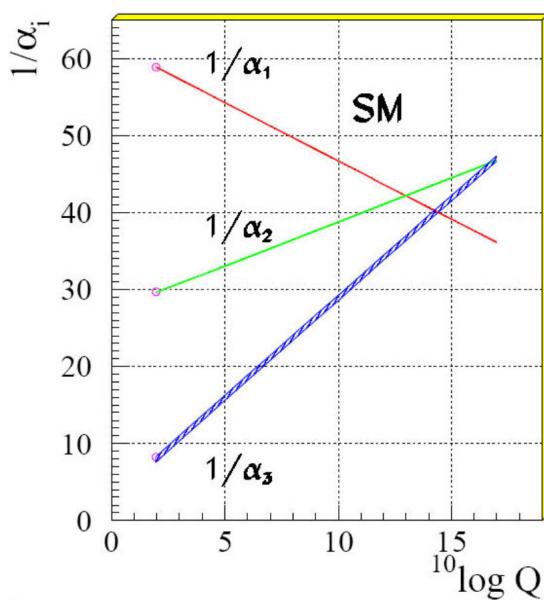


Figure 1.12: The running of the interaction gauge constants in the Standard Model.

2

SUPERSYMMETRY AND THE MSSM

Several theories has been formulated to solve some of the issues of the Standard Model. Currently, none of them is supported by experimental evidence. Among them, the *supersymmetry* (SUSY) despite its current lack of experimental evidence, is interesting, as it introduces new elegant symmetries in the theory, solving a large number of the SM open questions.

2.1 PHENOMENOLOGY OF THE SUPERSYMMETRY

In the SUSY theory [41, 42], fermions and bosons are not independent objects but different flavours of a more fundamental field. According to this assumption, it is possible to have supersymmetric multiplets in which both bosons and fermions coexist. The first consequence is the presence of many new hypothetical particles, not yet experimentally discovered: for every boson in a supersymmetric theory there is a corresponding fermion, and vice versa. The bosons and fermions related to the supersymmetry are also called *superpartners*, belonging to symmetry group representations called *supermultiplets*. In each supermultiplet, the quantum number of their members are all the same, except for their spin, which differs by 1/2. This also means that the masses of superpartners and the couplings with the other particles must be equivalent. For example, this implies phenomenologically the existence of a scalar superpartner of the electron, called *selectron*, which would have the same mass and the same electric charge as the standard electron. In the same way, the SM gauge bosons, that have spin 1 and are massless, would have a supersymmetric partner with spin 1/2 called *gaugini*. In Table 2.1, a summary of the elementary particles theorised by the supersymmetry is shown. The explanation of why there are no such particles can be given in terms of a broken supersymmetry. Two are the possible hypothesis, if SUSY is realised: the superpartners of the known particles are too much heavier for the energy scale accessible by particle accelerators, or their interactions is much weaker than the ones in the SM.

As previously mentioned in Section 1.5, to allow such a small mass for the Higgs boson (compared to the Planck mass) adding all the radiative corrections, a fine-tuned cancellation must happen between the bosonic and fermionic loops. However, if the SUSY theory is correct, the contribution to the radiative corrections of the superpartners, shown in Figure 2.1, would cancel the divergence. This is possible due to the contribution, with similar order of magnitude but opposite sign, from the superpartners, with respect to their SM

Electric charge	SM particles		susy equivalent			
	particle	spin	sparticle	spin	sparticle name	mass eigenstate
-1	e	1/2	\tilde{e}_L, \tilde{e}_R	0	selectron	
-1	μ	1/2	$\tilde{\mu}_L, \tilde{\mu}_R$	0	smuon	
-1	τ	1/2	$\tilde{\tau}_L, \tilde{\tau}_R$	0	stau	
0	ν	1/2	$\tilde{\nu}$	0	sneutrino	
$-\frac{1}{3}, +\frac{2}{3}$	q	1/2	\tilde{q}_L, \tilde{q}_R	0	squark	
0	g	1	\tilde{g}	1/2	gluino	
0	γ	1	$\tilde{\gamma}$	1/2	photino	
0	Z^0	1	\tilde{Z}^0	1/2	zino	
0	$h^0, \mathbf{H}^0, \mathbf{A}^0$	0	$\tilde{H}_1^0, \tilde{H}_2^0$	1/2	neutral higgsino	
± 1	W^\pm	0	\tilde{W}^\pm	1/2	wino	
± 1	\mathbf{H}^\pm	0	\tilde{H}^\pm	1/2	charged higgsino	$\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$

Table 2.1: Elementary particles summary according to Supersymmetry. The ordinary SM sector of the Higgs mechanism with four additional Higgs bosons (H , A and H^\pm , marked in bold), with respect to the SM Higgs h , is a specific requirement of the Minimal Supersymmetric Standard model (MSSM), explained in Section 2.2.

partners. Such contributions are then cancelled by terms of the same order from the SM, thus avoiding a fermion-boson fine-tuning.

To give an insight of the radiative corrections issues, not in the scope of this thesis, the formulation of the SUSY Lagrangian can be approximated at first order, as:

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}, \quad (2.1)$$

where $\mathcal{L}_{\text{SUSY}}$ contains all the SUSY terms and interactions that preserve the symmetry, and $\mathcal{L}_{\text{soft}}$ describes the supersymmetry-broken mass terms, whose scale of energy would be around m_{soft} . The radiative corrections of the Higgs boson, up to the next scale of energy, would then be:

$$\Delta m_H^2 = m_{\text{soft}}^2 \left[\frac{\lambda}{16\pi^2} \ln \left(\frac{\Lambda_{UV}}{m_{\text{soft}}} \right) + \dots \right], \quad (2.2)$$

where the different contributions to the radiative corrections have been merged into a single λ term. It is possible to notice that the divergences have been eliminated, since the corrections are now running on a logarithmic term (to be compared with Eq. 1.46, where the divergence is explicit).

Another fundamental consequence of the SUSY theory is the so called *Grand Unification Theory* (GUT), or the unification of the three fundamental forces. The first unification was achieved with the electroweak theory by Glashow-Weinberg-Salam (described in Section 1.2), but achieving the same with the strong force is problematic, because the QCD and the EW theories are not converging inside the SM theory. If the SUSY interactions are added, the running of the coupling constants could be modified so that they converge at a large energy scale, providing a mechanism for the unification, as visible from Figure 2.2.

In general, SUSY is a theoretical framework that depends on many unknown parameters,

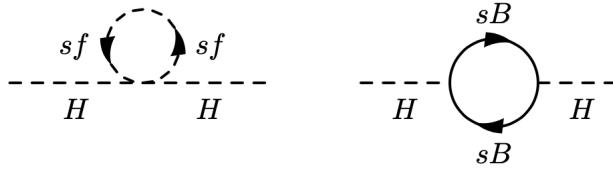


Figure 2.1: Radiative correction at the first order for the Higgs boson corresponding to a sfermion (left diagram) and a bosino (right diagram) loop. These corrections are opposite to those depicted in Figure 1.11.

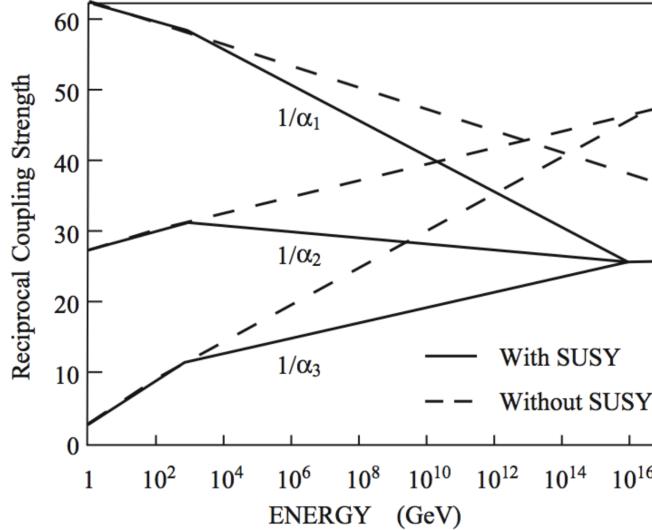


Figure 2.2: Running values with energy of the coupling constants for $U(1)$ α_1 , $SU(2)$ α_2 , $SU(3)$ (α_3) interactions including SUSY (continuous line) and without SUSY (dashed line) particles [43].

therefore can be implemented in several different ways. The simplest SUSY model is called Minimal Supersymmetric Standard Model (MSSM) and will be described with more detail in the next Section.

2.2 THE MINIMAL SUPERSYMMETRIC STANDARD MODEL

The Minimal Supersymmetric Standard Model (MSSM) [42, 44–48] is the simplest supersymmetric extension of the Standard Model. The aim is to add only the minimal amount of new parameters, particles and interactions, while keeping all the current symmetries and observables of the SM. As shown in the previous Section, the symmetry states that each ordinary particle has a supersymmetric partner with the same quantum numbers excluding the spin, shifted by $1/2$. In Table 2.2, on the Higgs boson sector of the SM, there are four additional particles, not present in the current SM. These additional bosons are a specific requirement of the MSSM and the mechanism behind this choice is explained in Section 2.2.1. This approach allows other interactions which have not been observed in the SM yet, in particular processes where the baryon and lepton numbers are not conserved.

So far, every observed process in nature conserve both numbers¹, B and L, and therefore also its difference B-L. In the MSSM, a new symmetry has to be added to suppress such violating processes: the *R-parity*:

$$P_R = (-1)^{-3L+3B-2S} = (-1)^{3(B-L)-2S}, \quad (2.3)$$

where B is the baryonic number, L is the leptonic number and S is the spin. If we assume that these numbers are conserved, then also the R-parity is conserved: it has value +1 for the SM particles and -1 for the SUSY particles. In order to be conserved, the combined P_R has to be positive: the main consequence is that the lightest supersymmetric particle (LSP) is stable, thus ending all the decay chains of heavier SUSY particles. The choice of such LSP is model dependent, but the simplest version of the theory assumes that this particle corresponds to the neutralino $\tilde{\chi}_1^0$. If this assumption is true, a new non-baryonic neutral massive stable particle exists, which could be a good candidate for the composition of the dark matter. Multiple searches looking for a LSP are being performed, currently with no evidence of its existence observed [4].

2.2.1 The Two-Higgs-Doublet Model

The Higgs sector of the MSSM is composed of the SM Higgs doublet and, additionally, another Higgs doublet of the opposite hypercharge, with their respective superpartners. This model is called Two-Higgs Doublet model (2HDM) and it is an extension of the SM that postulate the existence of this second $SU(2)$ Higgs field doublet [49]. This hypothesis is more general than SUSY, since a large amount of theories beyond the SM include the addition of a second doublet: however, for the MSSM, a second doublet is strictly required, being the simplest mechanism compatible with the current observations from the experiments.

The specific case of 2HDM implemented by the MSSM is called Type-II [50], where one of the doublets couples with up-type quarks (H_u) and the other with down-type quarks (H_d), carrying hypercharges $Y = +1$ and $Y = -1$, respectively. Therefore, the Higgs fields of the MSSM has the following $SU(2)_L$ structure:

$$H_u = \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix} = \begin{pmatrix} \phi_u^+ \\ \phi_u^0 \end{pmatrix}, \quad \tilde{H}_u = \begin{pmatrix} \tilde{H}_u^+ \\ \tilde{H}_u^0 \end{pmatrix}, \quad (2.4)$$

$$H_d = \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix} = \begin{pmatrix} \phi_d^{0+} \\ -\phi_d^- \end{pmatrix}, \quad \tilde{H}_d = \begin{pmatrix} \tilde{H}_d^0 \\ \tilde{H}_d^- \end{pmatrix}. \quad (2.5)$$

The total amount of degrees of freedom of the two complex $SU(2)_L$ Higgs doublets is eight. When the electroweak symmetry is spontaneously broken, similarly to the SM, three of them are taken by the vector bosons W^\pm and Z^0 . The remaining five degrees of freedom, therefore, correspond to the physical states of five different Higgs bosons. The linear com-

¹ Processes violating these numbers have not been observed: if allowed, the proton will not be stable and will decay, which is not been observed experimentally. A lower limit on the mean lifetime of the proton has been set in 2016 to $\tau_p > 5.8 \cdot 10^{29} \text{ years}$ [4]

bination of H_d^0 and H_u^0 results in two neutral Higgs bosons: the light \mathcal{CP} -even neutral scalar h , analogous to the SM², and the heavy \mathcal{CP} -even neutral scalar H . In addition, there are three other Higgs bosons: the \mathcal{CP} -odd neutral scalar A and two charged scalars H^\pm . These new particles, despite being postulated by the MSSM framework, are bosons included in the SM, not supersymmetric. Therefore, each of the five bosons would have a supersymmetric fermionic partner, called *higgsinos*, shown in Table 2.1. The masses of all these particles are not known and are considered as free parameters of the theory, lying at any possible value of mass.

2.2.2 The Higgs sector in the MSSM

The Higgs potential in the 2HDM, described by Eq. 2.4 and 2.5 can be described simply as [39]:

$$\begin{aligned} V(\phi_d, \phi_u) = & \mu_d^2(|\phi_d^0|^2 + |\phi_d^-|^2) + \mu_u^2(|\phi_u^0|^2 + |\phi_u^+|^2) - \mu^2(\phi_d^- \phi_u^+ + \phi_d^{0\dagger} \phi_u^0 + h.c.) \\ & + \frac{g_W^2 + g_B^2}{8}(|\phi_d^0|^2 + |\phi_d^-|^2 - |\phi_u^0|^2 - |\phi_u^+|^2)^2 + \frac{g_W^2}{2}|\phi_d^{0\dagger} \phi_u^+ - \phi_d^+ \phi_u^0|^2. \end{aligned} \quad (2.6)$$

Necessary conditions to induce spontaneous symmetry breaking are the following:

1. The determinant of the Higgs mass matrix must become negative,

$$\mu_d^2 \mu_u^2 - \mu^2 < 0; \quad (2.7)$$

2. The potential has a lower limit,

$$\mu_d^2 + \mu_u^2 - 2|\mu^2| > 0. \quad (2.8)$$

The symmetry then is spontaneously broken by the choice of the non-zero vacuum at:

$$\langle \phi_d \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_d \end{pmatrix} \quad \langle \phi_u \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v_u \\ 0 \end{pmatrix}. \quad (2.9)$$

The vev of the SM, v , is recovered by:

$$v^2 \equiv v_d^2 + v_u^2. \quad (2.10)$$

The ratio between the expectation values defines a value that is fundamental in the description of the MSSM, called **tan β** :

$$\tan \beta \equiv \frac{v_u}{v_d}. \quad (2.11)$$

² Currently, the convention in the MSSM framework is to assume that the Higgs discovered at ≈ 125 GeV is the lightest Higgs proposed by the MSSM Higgs mechanism h . There is also the scenario where the discovered particle is the heavier H boson, but it is much less probable although still questionable. Currently the mass limit is $m_H > 92.8$ GeV.

The W^\pm mass relation obtained is the same as the SM, shown in Eq. 1.35, but the Z boson mass will be different, dependent also on the $\tan\beta$ parameter:

$$m_W \equiv \frac{v \cdot g_W}{2}, \quad m_Z \equiv \frac{\mu_d^2 - \mu_u^2 \cdot \tan^2 \beta}{\tan^2 \beta - 1}. \quad (2.12)$$

The formulation of the masses of the five Higgs bosons can be obtained at tree level as:

$$m_A^2 = \frac{2\mu^2}{\sin 2\beta}, \quad (2.13)$$

$$m_{H^\pm}^2 = m_A^2 + m_W^2, \quad (2.14)$$

$$m_{H,h}^2 = \frac{1}{2} \left[m_A^2 + m_Z^2 \pm \sqrt{(m_A^2 + m_Z^2)^2 - 4m_Z^2 m_A^2 \cos^2 2\beta} \right]. \quad (2.15)$$

As visible from the relations above, the properties of the MSSM Higgs sector at tree level depend only on two non-SM parameters, that can be chosen to be the mass of the \mathcal{CP} -odd Higgs boson, m_A , and the ratio $\tan\beta$. At leading order, m_h is bound to be less than the Z boson mass: when higher order corrections are included, the upper bound increases to a maximum of ≈ 135 GeV, in agreement with the measured mass of the observed Higgs boson, if it is assumed to be the h boson.

Beyond the tree level, additional parameters of the MSSM are affecting the phenomenology of the Higgs sector, all unknown and coming from the spontaneous supersymmetry breaking. As this huge amount of free parameters is not practical for experimental studies, the common procedure is to manually set the high order parameters to a particular value, aiming for a specific MSSM phenomenology (called *scenario*), and set experimental limits on the $\tan\beta$ and m_A parameters for these specific scenarios [51–53].

2.3 MSSM FRAMEWORK SCENARIOS

The results of search analysis are usually interpreted in terms of specific benchmark scenarios, scanning the $(\tan\beta, m_A)$ space, to set limits on the allowed values. With every benchmark scenario there is a specific phenomenology of the MSSM Higgs sector: in addition, there are some experimental and theoretical constraints on the SUSY breaking parameters which are relevant to such scenarios. For instance, one of the Higgs bosons described by the MSSM must be the same boson observed at 125 GeV at the LHC. In most of the scenarios, the observed Higgs is interpreted as the light h boson, with $m_h \approx 125$ GeV; also the decays and production rates of the light boson are required to be compatible with the experimental results, indicating that the Higgs particle is exactly like the SM Higgs boson [36].

In these benchmark scenarios, there are some limits on the parameters - from direct searches for SUSY particles - to be taken into account when calculating the Higgs bosons properties, e.g. the limits on the stop and sbottom quarks that are set up to ≈ 1 TeV from recent studies [54, 55]. For this reason the SUSY scale, m_{SUSY} , is usually set to 1 TeV or above.

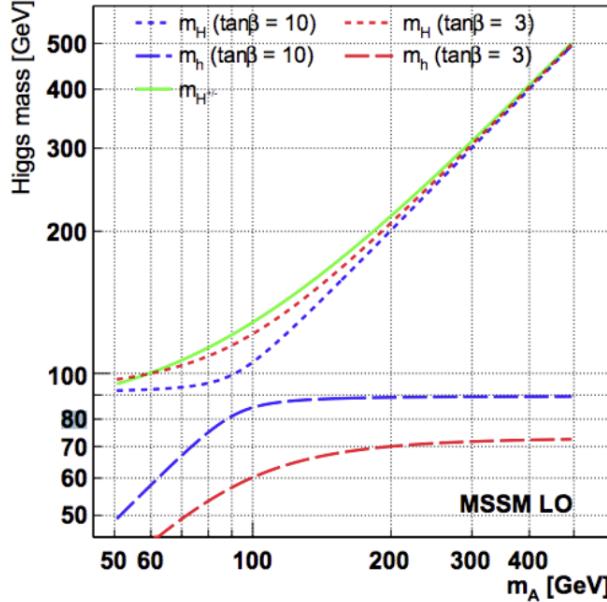


Figure 2.3: Masses of the MSSM Higgs bosons at tree level as a function of m_A for a fixed value of $\tan\beta$ (3 in the red lines and 10 in the blue lines) in the m_h^{max} scenario [59].

As already mentioned, there are many benchmark scenarios, with different phenomenology of the MSSM Higgs sector. Many of those, however, are no longer adopted after the discovery of the SM Higgs boson at the LHC, as they are incompatible with interpreting it as the light h boson from the MSSM. To mention a couple of them as an example, there is the *gluophobic Higgs* scenario, exhibiting large suppression of ggh coupling, or the *no-mixing* scenario, assuming no mixing in the stop sector, which are incompatible with $m_h \approx 125$ GeV. On the other hand, there are a handful of scenarios still used to interpret the SM-like particle and interpret the LHC data [53].

2.3.1 The m_h^{max} Scenario

The m_h^{max} scenario was initially defined to give conservative exclusion bounds on $\tan\beta$ for the LEP searches, by aiming at the heaviest possible light Higgs boson h , at large values of m_A , for a fixed value of $\tan\beta$. The discovery of the Higgs boson at $m_h \approx 125$ GeV constrained significantly the phase parameter of the m_h^{max} scenario, making it possible to be achieved only in a small region of the parameter space, especially for rather low values of $\tan\beta$ ($\tan\beta < 10$). In Figure 2.3, the phase space of the values for the masses, depending on m_A and $\tan\beta$, is computed: in the limit of large m_A , $m_H \approx m_{H^\pm} \approx m_A$, while m_h remains fixed at relatively low values. It is also possible to notice that, differently from the SM Higgs boson mass free parameter, in the MSSM it is constrained from above: at tree level $m_h < m_Z$, refuted by the experimental observations. Only with the radiative corrections it is possible to move the limit up to ≈ 135 GeV, allowing the observed mass of the Higgs boson [56–58].

2.3.2 *The $m_h^{mod\pm}$ Scenario*

With the discovery of the 125 GeV resonance, a couple of modified scenarios, called m_h^{mod+} and m_h^{mod-} , were introduced. Compared to the m_h^{max} , the modified $m_h^{mod\pm}$ have smaller amount of mixing in the stop sector, resulting in a lower mass of the light Higgs boson, compatible with the 125 GeV resonance in a large region of the $(m_A, \tan \beta)$ parameter space. The difference between the two $m_h^{mod\pm}$ scenarios is the relative sign and absolute value of the stop mass matrix elements. In terms of exclusions regions, the difference in the sign does not have a significant effect.

In searches for the heavy Higgs bosons in very sensitive decay modes like $\tau^+\tau^-$, the branching ratio at high values of $\tan \beta$ are highly affected by other decays in the lower regions of $\tan \beta$: decay modes to charginos and neutralinos in this region may reach B.R. of about 70% causing a decrease of the branching fractions of τ leptons. This makes the exclusion of parameters at low and moderate values of $\tan \beta$ quite difficult.

2.3.3 *The hMSSM Scenario*

The hMSSM scenario has a different approach: the description of the MSSM sector is done by assuming the validity of the tree level parameters and setting m_h exactly as the observed value by the ATLAS and CMS experiments, i.e. $m_h \approx 125$ GeV. With this approach, the Higgs sector is again described only by $(m_A, \tan \beta)$. This is true even with the radiative corrections at two-loop order, with a good approximation. The only corrections to the matrix mass are the leading logarithmic terms, arising from top and stop loops. Finally, it assumes that the masses of the other SUSY particles are high enough to escape detection at the LHC, with no effect on the Higgs sector.

Under the hMSSM assumptions, it is possible to determine the production cross sections and decay B.R. of all MSSM Higgs bosons, as a function of $(m_A, \tan \beta)$.

2.3.4 *The Heavy Higgs Scenario*

Other interpretations are also considered as benchmark in the MSSM framework. For instance, the *heavy Higgs* scenario, with considers the 125 GeV Higgs as the heavy \mathcal{CP} -even Higgs boson of the MSSM. In this case, all Higgs bosons would be lighter and the light h would have heavily suppressed couplings with the gauge bosons.

2.4 SEARCHES FOR THE MSSM HIGGS BOSONS

Various searches have been performed by the CMS collaboration in Run 1 and Run 2 of the LHC, covering the Higgs sector of the MSSM [52, 60]. The CMS searches are summarized in the $(m_A, \tan \beta)$ parameter space of the hMSSM scenario, shown in Figure 2.4, and the m_h^{125} scenario, shown in Figure 2.5. Due to the enhanced couplings of the Higgs boson to the down-type fermions at high values of $\tan \beta$, the production and decay modes of the

Higgs bosons depend strongly on the $\tan\beta$ parameter, with consequent different regions in the $(m_A, \tan\beta)$ parameter space for each search channel.

The most important production mode of the H/A neutral Higgs bosons is the gluon-gluon fusion. Additionally, at high values of $\tan\beta$ the b-associated production mode has a significant cross section. The tree level Feynman diagrams for the two production modes are shown in Figure 2.6. Also, searches involving gauge bosons in the final states benefit from VBF production modes.

Charged Higgs bosons, instead, are mostly produced due to interactions with a top quark, as shown in Figure 2.7. Charged Higgs bosons with lower masses are produced in top quark decays, while those with higher masses are produced in association with non-resonant or single-resonant top quarks.

The searches for neutral Higgs bosons H and A generally use decay channels to gauge bosons, third generation fermions and the light Higgs boson: such analyses have been performed both by the ATLAS and CMS collaborations in the $H/A \rightarrow \tau^+\tau^-$ final state [61–67] and the $b\bar{b}$ final state [68–70], both sensitive at high $\tan\beta$ region of the hMSSM parameter space due to enhanced couplings (visible also from Figure 2.4). Other searches, like the Higgs boson decay channel $H \rightarrow WW$ has been performed by CMS [71], which is sensitive at low $\tan\beta$ region where the branching ratios are significantly large. Another one is the decay of the heavy neutral \mathcal{CP} -even Higgs boson H into a pair of light Higgs bosons [72], also sensitive at low $\tan\beta$ and allowed in a kinematical region where the mass of the heavy boson is above the threshold for a light Higgs boson pair production and below the threshold for a top-pair production. The decay channel $A \rightarrow Zh$ [73], is significant only at A mass values lower than the top quark pair threshold, since above that threshold it decays preferably into $t\bar{t}$. Limits on the existence of MSSM Higgs bosons were determined also in pre-LHC experiments, in e^+e^- collisions at $\sqrt{s} = 91 - 209$ GeV at the CERN LEP [74] and in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV at the Fermilab Tevatron [75–78].

The $H/A/h \rightarrow \mu^+\mu^-$ analysis, which includes the work described in this thesis (shown in Chapter 5) in a model independent scenario at $\tan\beta = 5$ where the intrinsic width is smaller compared to the detector resolution, is also sensitive at medium/high values of $\tan\beta$ like the $\tau^+\tau^-$ final states, but has a branching ratio about 300 times smaller. The reason behind the choice of the $\mu^+\mu^-$ final state resides in the full reconstruction of the events, with a dimuon invariant mass measurable with a few percent precision, exploiting the excellent muon momentum resolution of the CMS detector, described in more detail on the next Chapter.

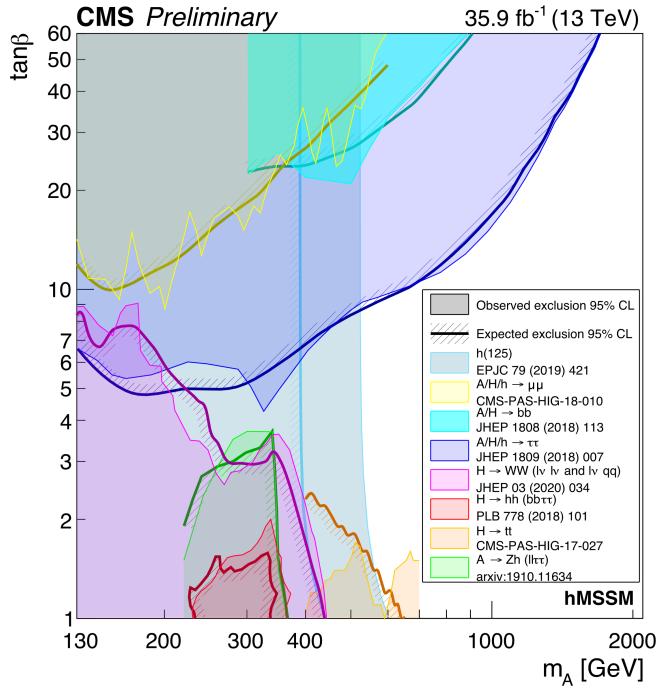


Figure 2.4: Observed and expected 95% CL upper limits for m_A versus the MSSM parameter $\tan \beta$ in the hMSSM benchmark scenario. Updated in March 2020.

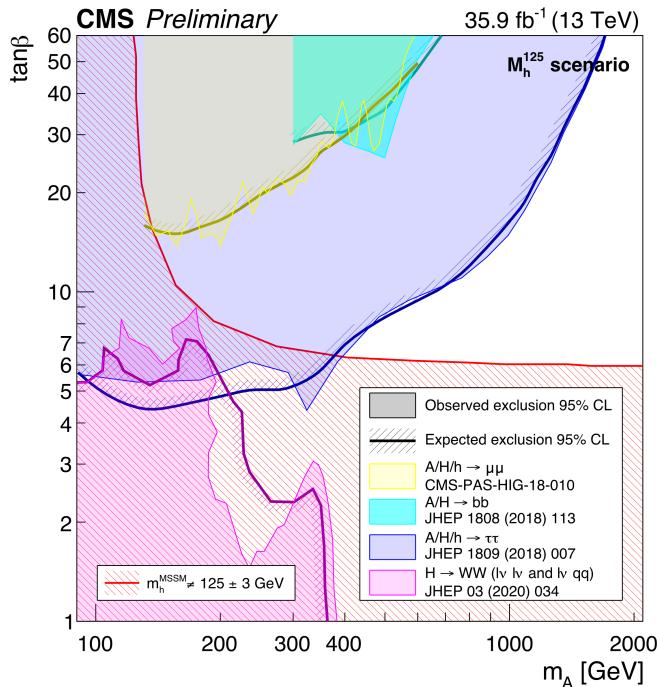


Figure 2.5: Observed and expected 95% CL upper limits for m_A versus the MSSM parameter $\tan \beta$ in the m_h^{125} benchmark scenario. Updated in March 2020.

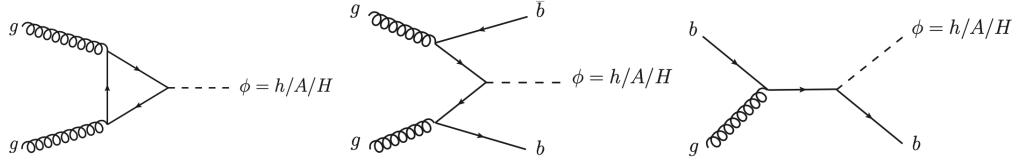


Figure 2.6: Leading order Feynman diagrams for the main production modes of the neutral Higgs bosons in pp collisions. The gluon fusion (left diagram) dominates at low and middle values of $\tan\beta$, while b-associated production (middle and right diagrams) become significant at high values of $\tan\beta$.

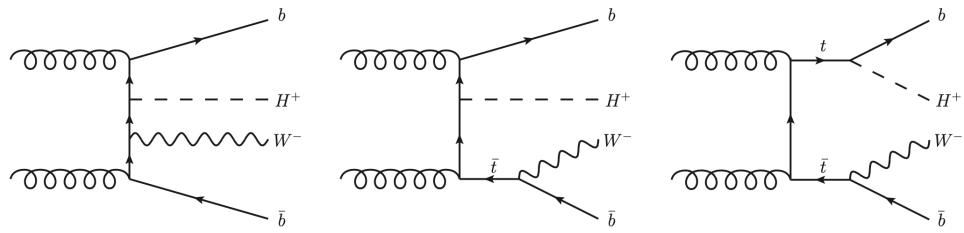


Figure 2.7: Leading order Feynman diagrams for the main production modes of the charged Higgs bosons in pp collisions. The left diagram and middle diagrams are the non-resonant and single-resonant top quark productions which dominate at large H^+ masses. The right diagram is the double-resonant top quark production which dominates at low H^+ masses.

3

THE CMS EXPERIMENT AT THE LHC

Since the analysis described in this thesis exploits data samples collected by the CMS detector during the LHC Run 2, this chapter aims at providing an overview of the main features of the LHC and of the CMS experiment with its subdetectors. A particular focus on the muon system is also covered, due to the specific analysis final state studied.

3.1 THE LARGE HADRON COLLIDER

The Large Hadron Collider (LHC) [79, 80] is part of the CERN accelerator complex (see Figure 3.1) [81], in Geneva, and it is the most powerful particle accelerator ever built. The particle beam is injected and accelerated by each element of a chain of accelerators, with a progressive increase of energy until the beam injection into LHC, where particles are accelerated up to 14 TeV (by LHC design, its nominal center-of-mass energy). It was installed in the existing underground tunnel previously used for the Large Electron Positron collider (LEP), located at the border between Switzerland and France.

The LHC consists of a circular 27 km circumference ring, divided into eight independent sectors, designed to accelerate protons and heavy ions. Particles travel in two separated beams on opposite directions and in extreme vacuum conditions (see Section 3.1.1). Beams are controlled by superconductive electromagnets (see Section 3.1.2), keeping them in their trajectory.

The progressive acceleration of protons and beam injection in the LHC ring occurs in different steps:

1. Protons are obtained by removing the orbiting electron on hydrogen atoms, with a process called stripping;
2. A linear accelerator, called LINAC2, starts the proton acceleration bringing them to an energy up to 50 MeV;
3. Protons are then injected in the Proton Synchrotron Booster (PSB), where the energy of the beam reaches about 1.4 GeV;
4. Protons are then injected in the Proton Synchrotron (PS), and accelerated up to 25 GeV;
5. The proton beam is sent to the Super Proton Synchrotron (SPS), where they reach an energy of 450 GeV;

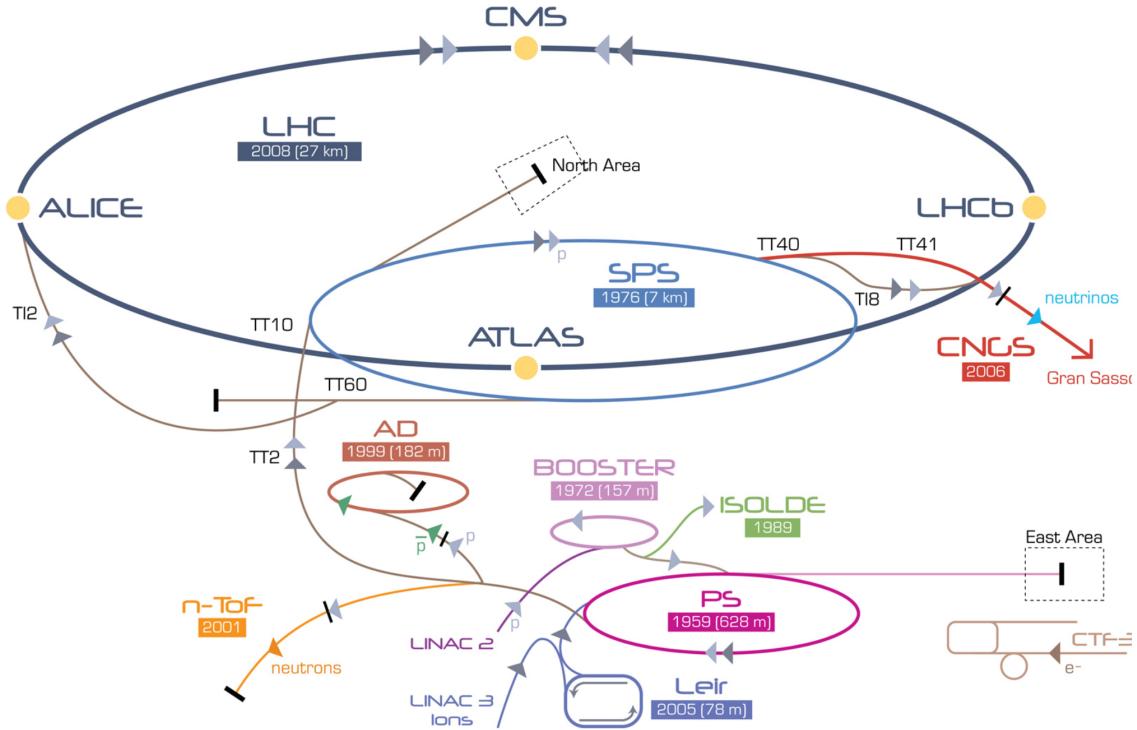


Figure 3.1: Schematic layout of the CERN particle accelerators chain and of the experiments around the circumference of the LHC, which is the last in the accelerator chain (dark blue line).

6. Finally, protons are transferred in a bunch configuration into the two adjacent and parallel beam pipes of the LHC, circulating for several hours around the ring, with one beam in the clockwise direction and the second one in an anticlockwise direction.

Some of the main LHC parameters are shown in Table 3.1.

3.1.1 *The vacuum system*

The LHC vacuum system [82] is, with more than 104 kilometers of vacuum ducts, one of the most advanced in the world. It has two functions: the first is to avoid collisions between beam particles and air molecules inside the ducts, creating an extreme vacuum condition (10^{-13} atm), as empty as interstellar space; the other reason is to cancel heat exchange between components that need low temperatures in order to work properly and maximise the efficiency.

The vacuum system is made of three independent parts:

- an isolated vacuum system for cryomagnets;
- an isolated vacuum system for Helium distribution line;
- a vacuum system for beams.

Table 3.1: Main technical parameters of LHC.

Quantity	value
Circumference (m)	26 659
Magnets working temperature (K)	1.9
Number of magnets	9593
Number of principal dipoles	1232
Number of principal quadrupoles	392
Number of radio-frequency cavities per beam	16
Nominal energy, protons (TeV)	6.5
Nominal energy, ions (TeV/nucleon)	2.76
Magnetic field maximum intensity (T)	8.33
Project luminosity ($\text{cm}^{-2} \text{s}^{-1}$)	2.06×10^{34}
Number of proton packages per beam	2808
Number of proton per package (outgoing)	1.1×10^{11}
Minimum distance between packages (m)	~7
Number of rotations per second	11 245
Number of collisions per crossing (nominal)	~20
Number of collisions per second (millions)	600

3.1.2 Electromagnets

Electromagnets [83] are designed to guide beams along their path, modifying single particles trajectories as well as align them in order to increase collision probability. To bend protons in the LHC accelerator, a magnetic field B is needed with an intensity given by the following equation:

$$p[\text{TeV}] = 0.3 \cdot B[\text{T}] \cdot r[\text{km}], \quad (3.1)$$

where p is the momentum of the beam particle and r is the radius of the LHC ring. The LHC has eight arc sections, where the magnetic fields and vacuum chambers are located, and eight straight sections hosting the collision points with detectors and/or utilities: four collision points, two of which aiming at maximum luminosity; two beam injectors and two beam dump facilities; radiofrequency cavities and the collimation systems. There are more than fifty different kind of magnets in LHC, totalling approximately 9600 magnets.

The most numerous magnets (1232) are dipoles, shown schematically in Figure 3.2, generating a magnetic field with a maximum intensity of 8.3 T. In order to reach such an intense field, a current of 11 850 A is needed. To minimise power dissipation, superconducting magnets are employed, using cables made of niobium-titanium (NbTi). A system of liquid He distribution keeps the magnets at a temperature of about 1.9 K. At this incredibly low temperatures, below that required to operate in conditions of superconductivity, helium becomes also super-fluid: this means an high thermal conductivity, thus an efficient refrigeration system for magnets.

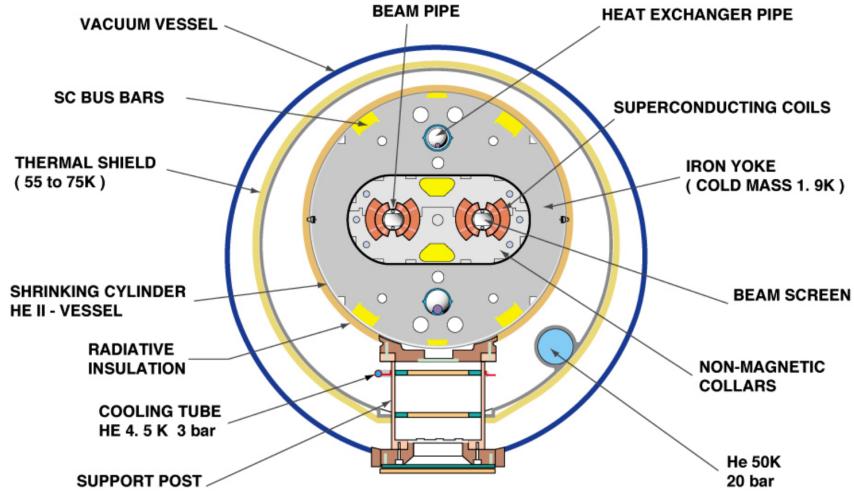


Figure 3.2: Transversal section of a dipole magnet of the LHC [84].

Other important magnets are the quadrupoles (392), which help focusing the beam by squeezing it either vertically or horizontally in order to maximise the chance of two protons colliding head-on. Finally, high order magnets contribute to correct imperfections of the magnetic field in the main ring magnets (dipoles and quadrupoles) and in the interaction region magnets.

3.1.3 Radiofrequency cavities

Radiofrequency cavities [79, 80, 85] are metallic chambers in which an electromagnetic field is applied. Their primary purpose is to separate protons in packages and to focus them at the collision point, in order to guarantee an high luminosity and thus a large number of collisions.

Particles, passing through the cavity, feel the overall force due to electromagnetic fields and are pushed forward along the accelerator. In this scenario, the ideally timed proton, with exactly the right energy, will see zero accelerating voltage when LHC is at nominal energy while protons with slightly different energies will be accelerated or decelerated sorting particle beams into "bunches". LHC has eight cavities per beam: each of which provides 2 MV at 400 MHz. The cavities work at 4.5 K and are grouped into four cryomodules.

At regime conditions, each proton beam is divided into 2808 bunches, each containing about 10^{11} protons. Away from the collision point, the bunches are a few cm long and 1 mm wide, and are compressed down to 16 nm near the latter, increasing the probability of a p-p collision.

The number of bunches affects significantly the *instantaneous luminosity* \mathcal{L} of the machine, defined as follows:

$$\mathcal{L} = f\gamma \frac{n_b N_b^2}{4\pi\epsilon_n\beta^*} F, \quad (3.2)$$

where n_b and N_b are the number of bunches and particles per bunch, respectively, f represents the bunch crossing (BX) frequency, γ the relativistic Lorentz factor of the protons, ϵ_n the transverse emittance describing the shape of the beam and finally β^* the focal length at the collision point. The F factor then takes into account the geometric reduction of the luminosity, depending on the transverse and longitudinal dimensions of the beams σ_{xy} and σ_z at the interaction point, and on the beam crossing angle θ_c :

$$F = \left(1 + \theta_c \frac{\sigma_z}{2\sigma_{xy}} \right)^{-1}. \quad (3.3)$$

At full luminosity, during Run 2, packages are separated in time by 25 ns (half of the bunch crossing interval used during Run 1), corresponding to a frequency of 40 MHz, or 40 million BX per second. The luminosity defined above represents the coefficient of proportionality between the number of events produced per second dN/dt , the *event rate*, and the cross section of the physical process in question σ_p :

$$\frac{dN}{dt} = \mathcal{L}\sigma_p. \quad (3.4)$$

To obtain the total number of collision in a defined time interval, it is often used another value of luminosity called *integrated luminosity*, defined as;

$$L = \int \mathcal{L} dt. \quad (3.5)$$

3.2 LHC DETECTORS

Along the LHC circumference, the particles collide in four intersection points, where the main LHC experiments are located. Each experiment has its own detector, designed and built to gather the fragments of the large number of collisions and to reconstruct all physical processes that generated them.

In particular, the four major experiments installed at LHC are:

- A Large Ion Collider Experiment (ALICE)
- A Toroidal LHC ApparatuS (ATLAS)
- Compact Muon Solenoid (CMS)
- Large Hadron Collider beauty (LHCb)

In addition, there are secondary experiments, among which:

- Large Hadron Collider forward (LHCf)
- TOTal Elastic and diffractive cross section Measurement (TOTEM)

In the following Sections, the LHC detectors are briefly introduced, with a major focus on the CMS experiment.

3.2.1 ALICE

ALICE [86, 87] is a detector specialised in heavy ions collisions. It is designed to study the physics of strongly interacting matter at extreme energy densities, where a phase of matter called "quark-gluon plasma" forms. At these conditions, similar to those just after the Big Bang, quark confinement no longer applies: studying the quark-gluon plasma, as it expands and cools down, allows to gain insight on the origin of the Universe. Some ALICE specifications are illustrated in Table 3.2.

The collaboration counts more than 1000 scientists from over 100 physics institutes in 30 countries (updated to October 2014)

Table 3.2: ALICE detector specifications:

Dimensions	length: 26 m, height: 16 m, width: 16 m
Weight	10 000 tons
Design	central barrel plus single arm forward muon spectrometer
Cost of materials	115 MCHF
Location	St. Genis-Pouilly, France

3.2.2 ATLAS

ATLAS [88, 89] is one of the two general-purpose detectors at LHC. Although its similarities with the CMS experiment regarding scientific goals, they have subdetectors based on different technology choices, and the design of the magnets is also different. Some specifications are illustrated below in Table 3.3.

It is located in a cavern 100m underground near the main CERN site. About 3000 scientists from 182 institutes in 38 countries work on the ATLAS experiment. Around 1200 doctoral students are involved in detector development, data collection and analysis. The collaboration depends on the efforts of countless engineers, technicians and administrative staff.

Table 3.3: ATLAS detector specifications:

Dimensions	length: 46 m, height: 25 m, width: 25 m
Weight	7000 tons
Design	barrel plus endcaps
Cost of materials	540 MCHF
Location	Meyrin, Switzerland

3.2.3 CMS

CMS [90, 91], as well as ATLAS, is a general-purpose detector at LHC. Is built around a huge solenoid magnet with a cylindrical form able to reach a 3.8 T magnetic field. Its main characteristics are illustrated in Table 3.4. In Section 3.3, the CMS experiment will be discussed in much more detail.

Table 3.4: CMS detector specifications:

Dimensions	length: 21 m, height: 15 m, width: 15 m
weight	12 500 tons
Design	barrel plus endcaps
Cost of materials	500 MCHF
Location	Cessy, France

3.2.4 LHCb

The LHCb [92, 93] experiment is specialised in investigating the slight differences between matter and antimatter by studying the quark bottom. Instead of ATLAS or CMS, LHCb uses a series of subdetectors to detect mainly forward particles: the first one is mounted near the collision point while the others are placed serially over a length of 20 meters. Some specifications are illustrated below in Table 3.5. About 700 scientists from 66 different institutes and universities work on LHCb experiment (updated to October 2013).

Table 3.5: LHCb detector specifications:

Dimensions	length: 21 m, height: 10 m, width: 13 m
Height	5600 tons
Design	forward spectrometer with planar detectors
Cost of materials	75 MCHF
Location	Ferney-Voltaire, France

3.2.5 Other LHC experiments

Aside from ALICE, ATLAS, CMS and LHCb, a few details on LHC smaller experiments, LHCf and TOTEM, are given in the following. LHCf [94, 95] is a small experiment which uses particles thrown forward by p-p collisions as a source to simulate high energy cosmic rays. LHCf is made up of two detectors which sit along the LHC beam line, at 140 m either side of ATLAS collision point. They only weights 40 kg and measures (30 x 80 x 10) cm. LHCf experiment involves about 30 scientists from 9 institutes in 5 countries (updated to November 2012).

TOTEM [96, 97] experiment is designed to measure p-p total elastic and diffractive cross section by measuring protons emerging at small angle with respect to the beam lines. Detectors are spread across half a kilometre around the CMS interaction point in special vacuum chambers called "roman pots" connected to beam ducts, in order to reveal particles produced during the collision.

TOTEM has almost 3000 kg of equipment and 26 "roman pot" detectors. It involves about 100 scientists from 16 institutes in 8 countries (updated to August 2014).

3.2.6 *LHC Operational history*

After some major technical problems caused by a magnet quench in one of the sectors of the LHC in 2008, the collider began its research program in the spring of 2010, starting the first phase of operations called in jargon **Run 1**.

Initially, during the first operational run in November 2009, the center-of-mass energy was $\sqrt{s} = 900$ GeV. Then, during the early part of 2010, the energy was increased up to 3.5 TeV per beam. The record of high energy collisions was reached on the end of March 2010 by colliding proton beams at a center-of-mass energy of 7 TeV. By the end of 2011, the CMS experiment had collected a total integrated luminosity of 5.6 fb^{-1} with a record peak instantaneous luminosity of $4 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$. In 2012, the center-of-mass energy was increased to 8 TeV with higher instantaneous luminosities. In total, the luminosity gathered by CMS during this year amounted to 22 fb^{-1} with a record peak luminosity of $7.7 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$. In both years the LHC operated with a bunch spacing of 50 ns corresponding to a collision frequency of 20 MHz. The LHC remained in operation until February 2013, running continuously for three years and delivering a total luminosity of around 30 fb^{-1} . The CMS experiment collected a luminosity of 20 fb^{-1} , achieving the discovery of the Higgs boson together with the ATLAS Collaboration.

At the beginning of 2013, the LHC was shutdown in order to prepare the collider and run at high energy and luminosity. The accelerator was turned on in early 2015, operating at a center-of-mass energy of 13 TeV, and starting the second phase of operations called **Run 2**.

During the years 2016-2018 the majority of the Run 2 data was delivered and collected, with the full 40 MHz collision frequency. The LHC was operating proton-proton collisions from April to November of each year, with increasingly higher instantaneous luminosities. The record luminosity was $1.5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in 2016, and $2.1 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in 2017 and 2018, measured by CMS as shown in Figure 3.3. The total number of collision in 2016 exceeded the number from the whole Run 1 at a higher energy per collision. The integrated luminosities measured by CMS were 41 fb^{-1} in 2016, 49 fb^{-1} in 2017 and 68 fb^{-1} in 2018, as shown in Figure 3.4.

The year 2018 was the end of Run 2. During this data taking period, it was possible to achieve incredible physics results, in particular precision measurements for the constraints of the SM: the masses of the Higgs and the W bosons were computed with greater

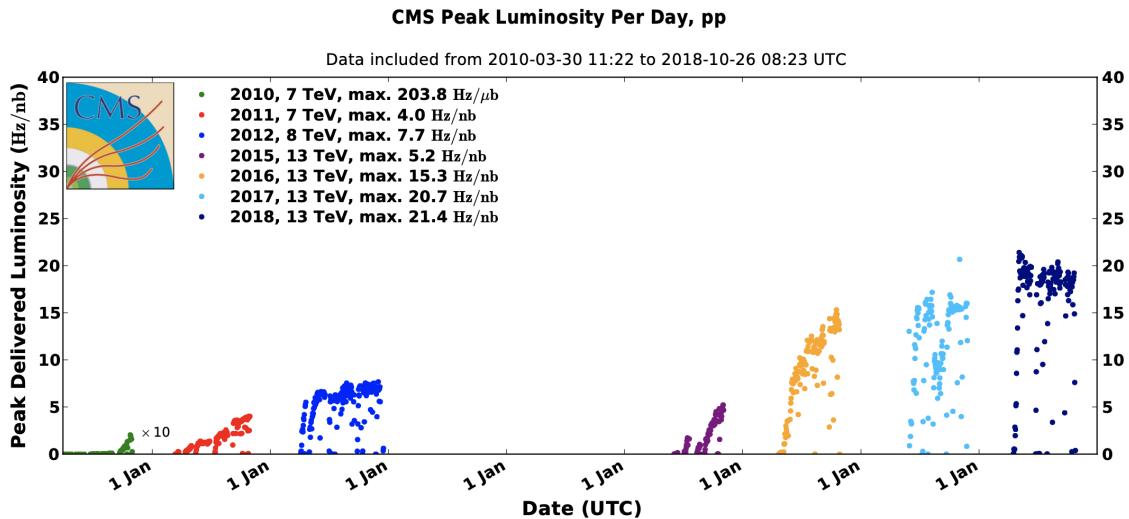


Figure 3.3: Instantaneous peak luminosity recorded by the CMS experiment per year [98]. The luminosity is given in Hz/nb i.e. $10^{33} \text{ cm}^{-2}\text{s}^{-1}$.

precision, new couplings of the Higgs were observed and an improved measurement of the CKM matrix allowed an investigation on the CP violations.

A new shutdown is currently underway. The collider will resume delivering new data from 2022¹ with the new Run 3, aiming at 300 fb^{-1} in the following three years period. After Run 3 and the subsequent long shutdown, around 2026, a new era of the LHC will begin, with a complete redesign of several components of the accelerator and the surrounding experiments, in a phase of operations called *High Luminosity LHC* (HL-LHC). A brief summary of this project is reported in Appendix A.

Such high luminosities have been possible only by squeezing the proton bunches as much as possible at the interaction point. This increases the instantaneous luminosity, but also increases the multiple collisions happening in a single bunch crossing, phenomenon called *pileup*. The distributions of the number of reconstructed interactions, or pileup profiles, are shown per year in Figure 3.5. In general, an high luminosity is advantageous for the physics analysis, thanks to the higher rates of rare interesting processes. However, the unavoidably larger pileup is an obstacle for the data taking and reconstruction.

3.3 THE CMS EXPERIMENT

The CMS main purpose, as stated in the letter of intent presented by the CMS Collaboration in 1992 [99], is to explore the p-p physics at the TeV scale, with a particular focus on the search for the SM Higgs boson. However, thanks to its incredible versatility it allows to cover many other physical processes at the LHC energy scale, with the final goal of probing different unproven models of the elementary structure of matter. It is also well suited for the study of top, beauty, and τ physics at lower luminosity as well as operating on the

¹ The Run 3 initial schedule was middle 2021 but, due to the Covid-19 pandemic, 1 year delay has been added.

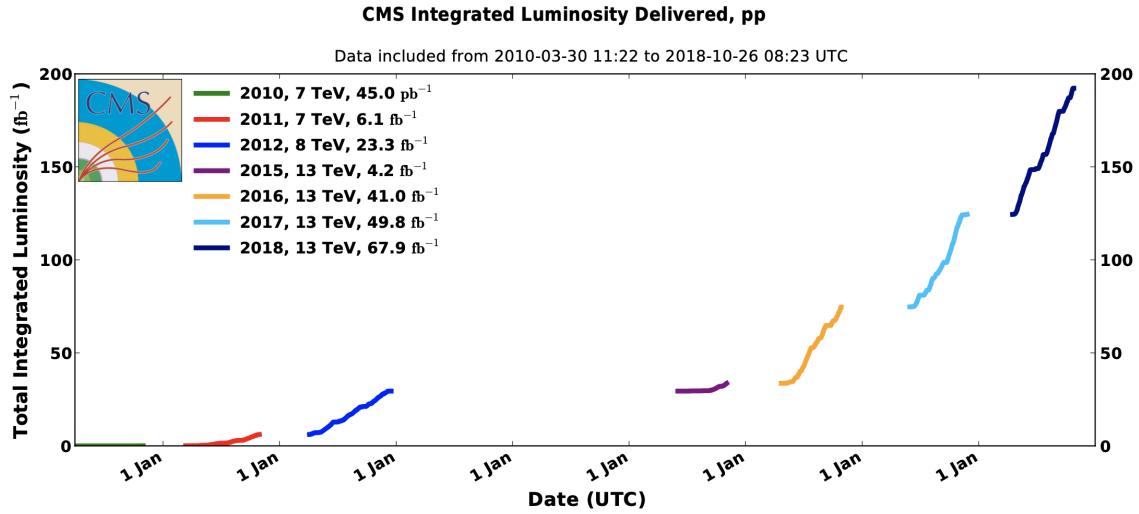


Figure 3.4: Integrated luminosity recorded by the CMS experiment per year [98].

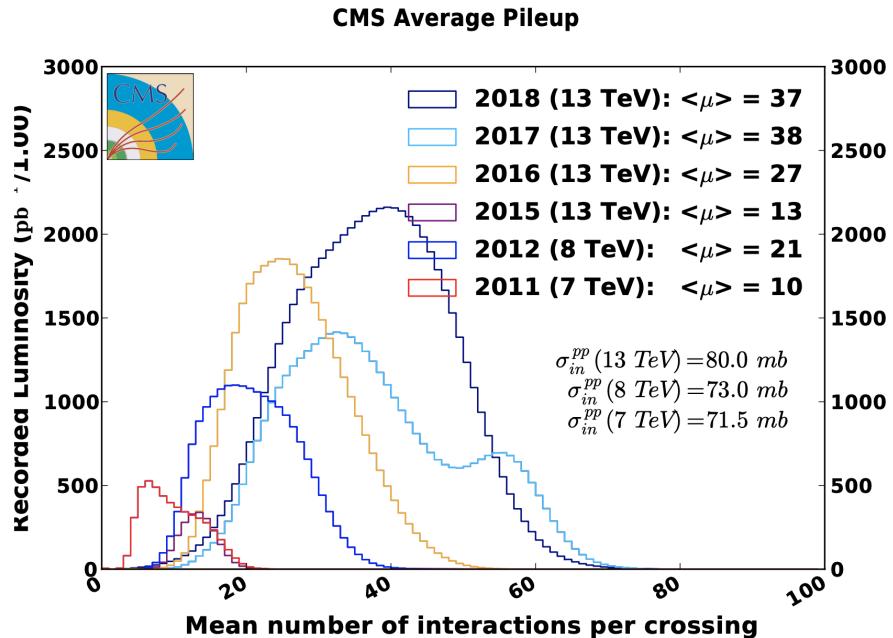


Figure 3.5: Distribution of the number of reconstructed pileup vertices measured by CMS in all the data taking years [98].

heavy ions physics program. It was designed to operate in p-p (Pb-Pb) collisions at a center-of-mass energy of 14 TeV (5.5 TeV), with luminosities up to $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ ($10^{27} \text{ cm}^{-2}\text{s}^{-1}$). Its cylindrical concept is built on several layers and each one of them is dedicated to the detection of a specific type of particle: in particular, it was chosen to identify and measure mainly muons, photons, and electrons with high precision. Therefore, the design is mainly focused on a highly performant muon system, a central tracking system capable of excellent track reconstruction, a high quality electromagnetic calorimeter, and a hadronic calorimeter with a decent energy resolution. The entire detector is then surrounded by an intense magnetic field, produced by a superconductive solenoid, allowing a great muon momentum resolution.

The CMS collaboration (Table 3.4) consists in over 4000 particle physicists, engineers, computer scientists, technicians and students from around 200 institutes and universities from more than 40 countries. Figure 3.6 shows an image of the CMS experiment.

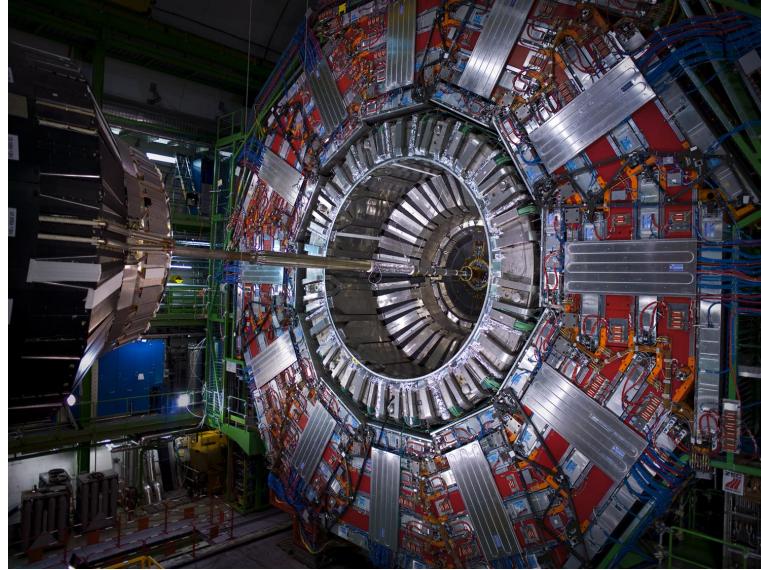


Figure 3.6: A view of the CMS experiment, with one of the endcaps removed.

3.3.1 The CMS detector: concept and structure

The CMS detector [100] consists in a cylindrical barrel, built of five slices, and two disk-like endcaps. The overall detector length is 21.6 m, its diameter is around 15 m and it has a total weight of approximatively 12500 tons. It is made up of different layers, as illustrated in Figure 3.7. Each of them is designed to trace and measure the physical properties and paths of different kinds of subatomic particles. Furthermore, this structure is surrounded by a huge solenoid based on superconductive technologies, operating at 4.4 K and generating a 3.8 T magnetic field.

In CMS, a right handed coordinate system, centred at the nominal collision point, is defined [101]: the *x*-axis pointing radially inward to the center of the accelerator ring, the *y*-axis pointing upward and the *z*-axis parallel to the beam pipe (in direction of the Jura mountains from LHC Point 5 in Cessy, France), as shown in Figure 3.8. Considering the cylindrical structure of the CMS detector, a polar system is often convenient to describe the four momentum of particles: the polar angle θ is measured from the *z*-axis using a $0 \leq \theta \leq \pi$ range, while the azimuthal angle ϕ is measured in the *x*-*y* plane from the *x*-axis in a $0 \leq \phi \leq 2\pi$ range. In a collision, the center-of-mass is boosted along the *z*-axis with respect to the laboratory frame. Therefore, the kinematics are usually described by the coordinates (p_T, y, ϕ, m) , where ϕ is the azimuthal angle, m is the invariant particle mass, p_T the transverse momentum defined as $p_T = p \sin \theta = \sqrt{p_x^2 + p_y^2}$, and y is the rapidity defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_Z}{E - p_Z} \right). \quad (3.6)$$

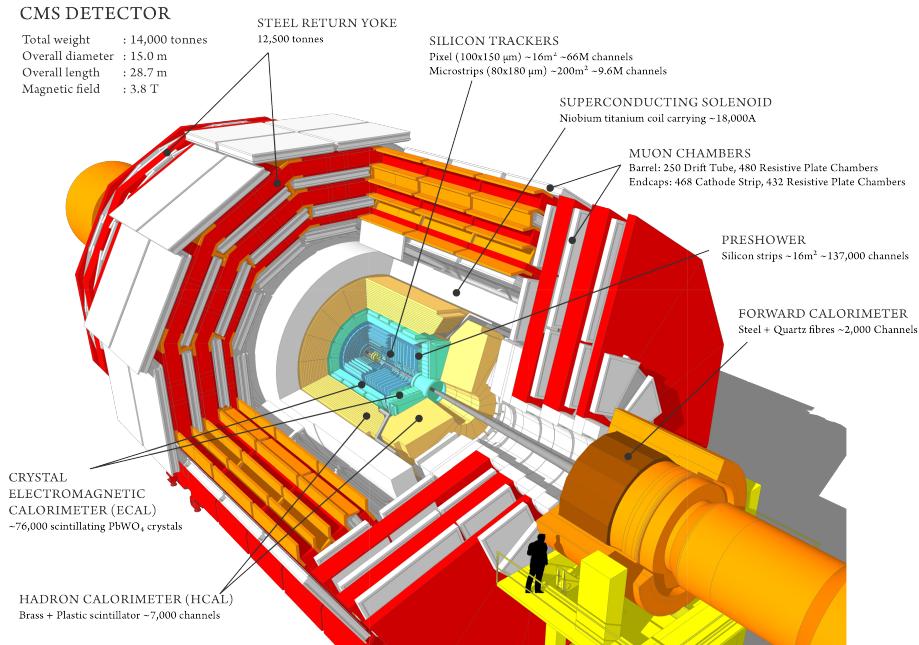


Figure 3.7: A section representation of the Compact Muon Solenoid.

The difference in rapidity between two particles is invariant under boost along the z direction, but since it does not provide an intuitive interpretation it is replaced by another quantity, approximated for ultra-relativistic particles, called *pseudorapidity*:

$$\eta = \frac{1}{2} \ln \left(\frac{|p| + p_z}{|p| - p_z} \right) = -\ln \left(\tan \frac{\theta}{2} \right). \quad (3.7)$$

The first and inner layer of the CMS detector is called Tracker [91, pp. 26–89] (Section 3.3.2): made entirely of silicon, it is able to reconstruct the paths of high-energy muons, electrons and hadrons as well as reconstruct secondary vertices from the decay of very short-lived particles with a resolution of a few nanometers.

The second layer consists of two calorimeters (Section 3.3.3 and 3.3.4): the homogeneous Electromagnetic Calorimeter (ECAL) [91, pp. 90–121], made of PbWO_4 crystals, and a high hermetical brass/scintillator sampling Hadron Calorimeter (HCAL) [91, pp. 122–155], arranged serially. The first measures the energy deposited by photons and electrons, while the second measures the energy deposited by hadrons.

In the end, after the superconducting magnet (Section 3.3.5), a series of muon detectors [91, pp. 162–246] (Section 3.4) are able to track muon particles, escaped from calorimeters.

3.3.2 The tracking system

The Tracker system [102, 103] is a crucial component in the CMS design as it measures particles momentum through their path: the greater is their curvature radius across the magnetic field, the greater is their momentum. As stated above, the Tracker is able to reconstruct muons, electrons and hadrons path as well as tracks produced by short-lived particles decay, such as quark beauty. It has a low degree of interference with particles

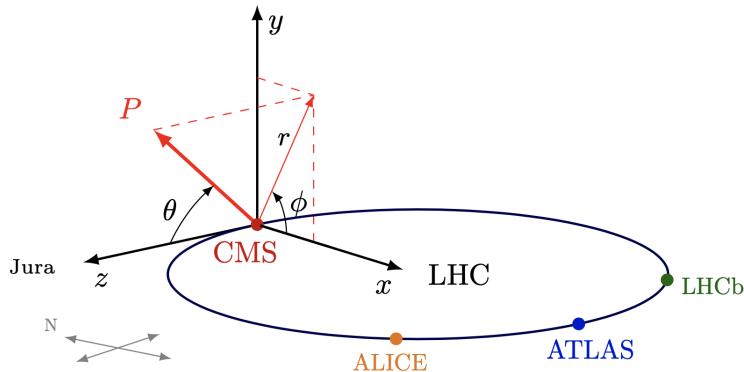


Figure 3.8: The CMS internal coordinate system.

and a high resistance to radiations. It is designed to efficiently detect and measure the trajectory of charged particles whose p_T is above 1 GeV/c, furthermore it has to precisely reconstruct their secondary vertices in order to provide jet-flavour tagging.

At LHC design luminosity, around 1000 charged particles from proton-proton interactions are produced every 25 ns, thus leading to the need of a high granularity and radiation hard system. In order to keep track occupancy low enough to perform efficient and precise measurements, and considering that particle flux quickly decreases with radius, three detection regions can be identified: a fine granularity *pixel detector* system in its innermost parts, and *silicon strips* modules of different pitch in its central and external part. This design allows to have an occupancy of $\approx 1\%$ everywhere during high luminosity p-p collisions, ensuring also a reasonable occupancy level during Pb-Pb ones (1% in the pixels, 20% in the silicon strip detectors).

This high granularity, however, implies an elevated power consumption and, with the low temperatures required to allow a good functioning and to prevent radiation damage (around -10°C), requires an efficient cooling infrastructure. In any case, the total amount of material in the tracker has to be kept as low as possible in order to reduce multiple-scattering and other interactions, therefore a compromise in the tracker design had to be found. This detector (Figure 3.9) is entirely made of silicon: it covers the pseudorapidity region up to $|\eta| < 2.5$ with a radius smaller than 1.2 m and $|z| < 2.7$ m, representing the largest silicon tracker ever built. The total number of sensor elements is 9.3 million strips and 66 millions pixels.

Silicon Pixel Detectors. The pixel detector system consist of finely segmented silicon pixels, whose cell size is of $100 \times 150 \mu\text{m}^2$, placed on a silicon substrate. It is built to ensure a precise 3D vertex reconstruction, for an efficient τ and b jets identification. The system covers a pseudorapidity range up to $|\eta| < 2.5$ and the small pixel size allows to keep single channel occupancy per bunch crossing around 10^{-4} , even in the expected high flux scenario (10^7 particles/s at 10 cm). The final layout consists of three levels of respectively 4.4, 7.3 and 10.2 cm radius, and two disks (radially from 6 to 15 cm) in the endcaps.

When a charged particle goes through one of this units, the amount of energy releases

an electron with the consequent creation of an hole. This signal is than received by a chip which amplifies it. It is possible, in the end, to reconstruct a 3D image using bi-dimensional layers for each level.

The power absorption must remain at minimum, because each pixel absorbs about $50 \mu\text{W}$ with an amount of power not irrelevant; for this reason pixels are installed in low temperature pipes.

In the barrel region, a $15 \mu\text{m}$ hit resolution is achieved in the z plane and $10 \mu\text{m}$ in the ϕ plane. In the endcaps, a lower resolution of $15 \mu\text{m}$ and $20 \mu\text{m}$ is expected in the ϕ and z directions, respectively. An upgrade of the silicon pixel detector was performed in 2017 [104].

Silicon Microstrip Detector. The two outermost regions of the tracking system are composed of several layers of silicon microstrip detectors. They consist of ten layers with about 10 million detector strips, divided into 15 200 modules and scanned by 80 000 microelectronic chips in a silicon area of about 200 m^2 , capable of detecting the passage of charged particles from p-p collisions.

The four innermost barrel layers of the system form the Tracker Inner Barrel (TIB) system, while the six outermost ones instead are named Tracker Outer Barrel (TOB). Then the Tracker Inner Disk (TID), made of three layers of disks, is positioned on each side of the TIB. Finally, a group of 9 disks perpendicular to the beam axis, placed after TOB and TID, constitute the Tracker End-Cap (TEC).

Each module is made up of three elements: a set of sensors, a support structure and the electronics necessary to acquire data. Sensors have an high response and a good spatial resolution, allowing to receive many particles in a restricted space; they detect electrical currents generated by interacting particles and send collected data. Also this section of the detector must be maintained at low temperature (-15°C), in order to "freeze" silicon structure damages from radiations and prevent it from perpetuating.

The silicon microstrips have a low occupancy (below 1%) and allow the reduction of the number of channels for readout. The spatial resolution amounts to $40\text{-}60 \mu\text{m}$ in the ϕ plane and $500 \mu\text{m}$ along the z direction.

3.3.3 *The Electromagnetic Calorimeter*

The Electromagnetic Calorimeter (ECAL) [105] is designed to reconstruct electrons and photons position and energy accurately, as well as to perform, in conjunction with the Hadron Calorimeter, precise measurement of hadronic jets. The search for the Higgs boson in the $\gamma\gamma$ channel was possible due to the ECAL design and its resolution of 1% on the diphoton invariant mass. The measurement provided by the tracking system and the calorimetry are often complementary in CMS. However, while the tracker is able to identify only charged particles with a precision inversely proportional to the particle's p_T , the calorimeters can measure both charged and neutral particles with a resolution proportional to the increase of the particle's energy.

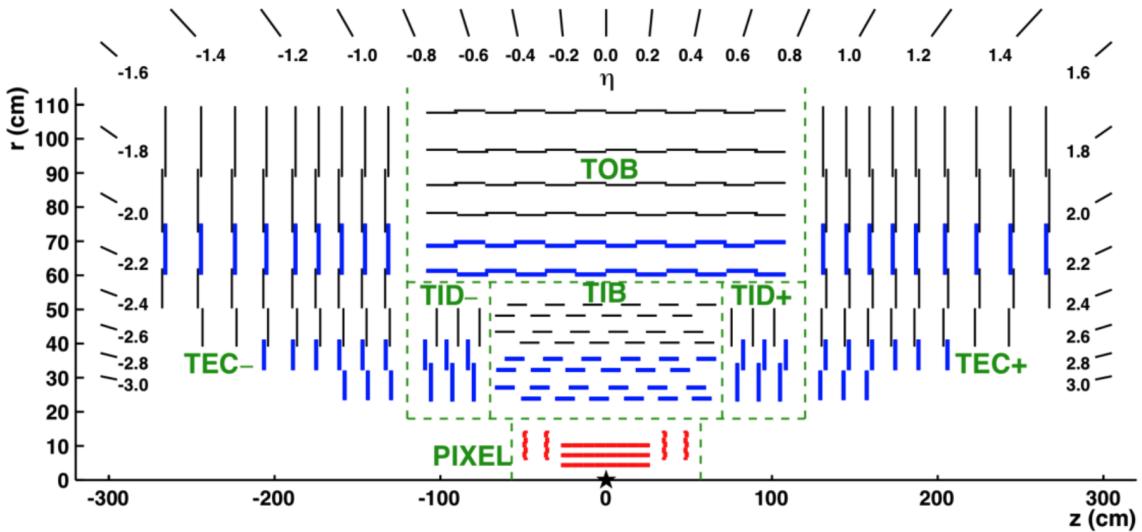


Figure 3.9: Tracker system schematic representation. Each subsystem is divided with the green dashed lines.

In order to achieve such a good resolution, a lead tungstate (PbWO_4) homogeneous, finely segmented, hermetic calorimeter has been developed. Such material is made primarily of metal and it is heavier than stainless steel but, with a touch of oxygen in this crystalline form, it is highly transparent and scintillates when electrons and photons pass through it. PbWO_4 has been chosen because of its radiation-hardness, high density ($\rho = 8.28 \text{ g/cm}^3$) as well as for its small Moliere radius (22 mm) and short radiation length $X_0 = 8.9 \text{ mm}$, resulting in a very compact electromagnetic calorimeter system with a length of approximately $25 X_0$. Moreover these crystals are characterised by a very short scintillation-decay time, that allows to collect about 80% of the light in the read out electronics within a 25 ns time period.

The ECAL can be divided in two regions: the cylindrical barrel (EB) with a pseudorapidity coverage up to $|\eta| < 1.479$, and two endcap disks (EE) providing a pseudorapidity coverage up to $|\eta| = 3.0$. Moreover, a pre-shower system is installed for the discrimination of single photons from $\pi_0 \rightarrow \gamma\gamma$ decays in front of the endcap in the region $1.65 < |\eta| < 2.6$ (ES). It consists of a two-layer sampling calorimeter, where lead radiators are alternated to silicon strip detectors, for a total material thickness of about $3 X_0$.

Figure 3.10 shows the total coverage of the system, with a maximum extension up to $|\eta| < 3.0$. The barrel region is consisting of eighteen supermodules. Every ECAL supermodule is made of 1700 crystals, organised in the $\eta - \phi$ geometry, accounting for 20 crystals along the ϕ direction and 85 crystals in η . Supermodules are further split into four modules, where the first one contains 20 crystals in ϕ and 25 in η , while all the other modules contain 20 crystals along both ϕ and η coordinates. The endcap structure, instead, consists in 3662 crystal complexes in $x - y$ geometry, and are grouped into 5×5 supercrystals. The crystals have a length of 230 mm in the barrel and 220 mm in the endcaps corresponding respectively to 25.8 and $24.7 X_0$. They are trapezoidal in shape with a square front size of $22 \times 22 \text{ mm}^2$ in the barrel and $28.6 \times 28.6 \text{ mm}^2$ in the endcaps.

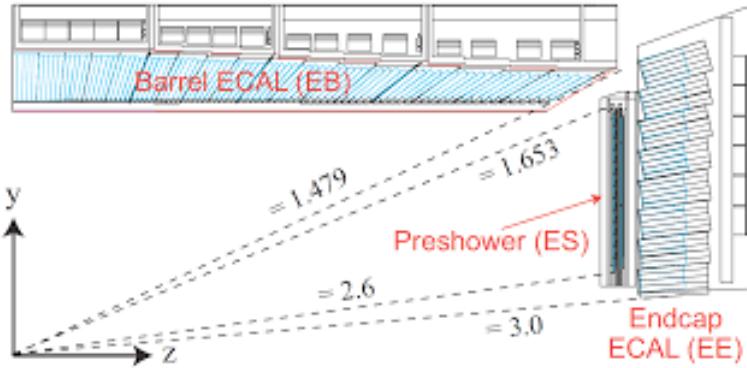


Figure 3.10: The longitudinal view of a fourth of the ECAL subdetector.

The low emitted light output ($4.5 \gamma/\text{MeV}$ at room temperature), requires some photodetectors with an high gain operating in an high magnetic field. Therefore solutions based on Vacuum Photodiodes (VPT) and Avalanche Photodiodes (APD) are thus been adopted in endcaps and barrel respectively. As the latter has a response which is sensitive to temperature, thermal stability up to 0.1°C is required to preserve the energy resolution.

For energy below 500 GeV, the energy resolution for ECAL can be parametrised as follows:

$$\left(\frac{\sigma}{\sqrt{E}}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \quad (3.8)$$

where S is a *stochastic* term due to fluctuations in lateral shower containment, photostatistics and energy deposit in the preshower; N is the *noise* term related to electronics, digitisation and pileup; and C is a *constant* contribution caused by ECAL calibration, non uniformities in the light collection and leakage from the back of the crystals. Studies performed during test beams allowed to estimate these parameters to be: $S = 2.8\%$, $N = 12\%$ and $C = 0.30\%$ [106].

3.3.4 The Hadronic Calorimeter

The Hadron Calorimeter (HCAL) [107] is used, together with the ECAL, to perform direction and energy measurements of hadronic jets and to estimate the amount of missing transverse energy (missing E_T) of each event. The request to perform precise missing E_T measurement implies the development of a very hermetic system, whose design is constrained by compactness requests and by the high magnetic field. In order to achieve such requirements, a sampling calorimeter system based on brass absorber layers alternated to active plastic scintillators has been built. The signal coming from active scintillators is read out with embedded wavelength-shifting fibers (WLS) and transported via clear fiber waveguides to hybrid photodiodes. The choice of brass as absorber material has been driven from its short interaction length (λ_I) and its non-magnetic nature. A longitudinal view of the HCAL layout is shown in Fig. 3.11.

In the barrel region, a barrel calorimeter (HB) covers an η region up to 1.4 and its readout segmentation (of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$) is tight enough to allow proper di-jet separa-

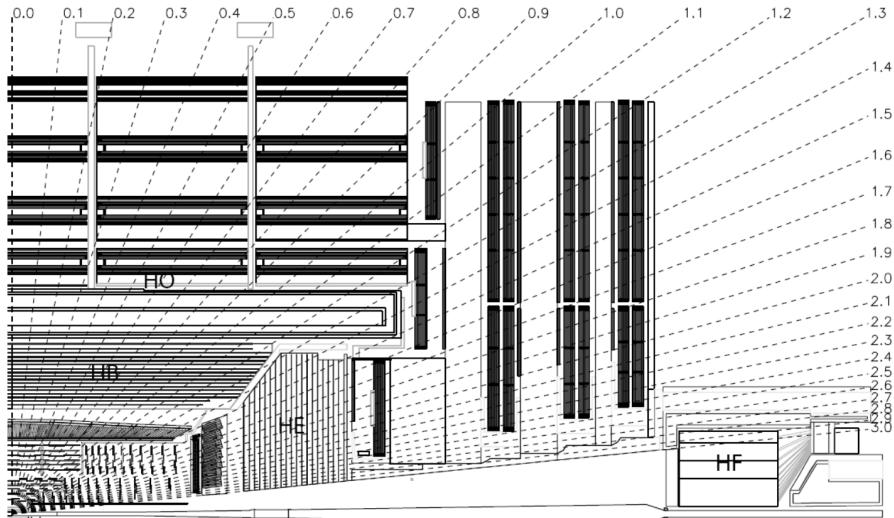


Figure 3.11: The longitudinal view of a fourth of the HCAL subdetector.

tion and mass resolution. The HB total depth increases as a function of η , raising from $5.15 \lambda_I$ at $\eta = 0$ to $10.15 \lambda_I$ at $\eta = 1.3$. Additionally, to obtain a better energy resolution of the barrel calorimeters, an outer calorimeter (HO) is placed outside the magnet coil, extending the total interaction length to about $11\lambda_I$.

In the endcap region, instead, an endcap calorimeter (HE) has been placed inside the magnet bore, covering the $1.4 < \eta < 3.2$ region. Its segmentation overlaps with the HB one and its average depth is about $10.5 \lambda_I$. Outside the magnet a forward calorimeter (HF) covers the η region up to 5.2, guaranteeing the hermeticity of the detector. Due to the harsh radiation environment at high η , hard quartz fibers have been chosen as active medium.

The energy resolution of the system (with E expressed in GeV), is dependent on the imperfect containment of the hadronic shower, resulting in a resolution sampling term up to 100% and a constant term of 5%, according to test beam studies [108]:

$$\frac{\sigma}{E} \approx \left(\frac{65}{\sqrt{E}} \oplus 5 \right) \% \quad \text{for the barrel region,} \quad (3.9)$$

$$\frac{\sigma}{E} \approx \left(\frac{83}{\sqrt{E}} \oplus 5 \right) \% \quad \text{for the HE,} \quad (3.10)$$

$$\frac{\sigma}{E} \approx \left(\frac{100}{\sqrt{E}} \oplus 5 \right) \% \quad \text{for the HF.} \quad (3.11)$$

3.3.5 The magnet

The CMS magnet system [27] is designed to take into account the strict requirements of identification of the charge of high p_T muons and precise measurement of momentum of charged particles requested at the LHC. To fulfil these requirements, with the main objective of keeping the detector as compact as possible, an high field superconductive solenoid has been developed. The CMS magnet is 12.9 m long, with an inner diameter

Parameter	Value
Field	3.8 T
Residual yoke field	2T
Inner bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA

Table 3.6: Parameters of the CMS superconductive magnet.

of 5.9 m, capable of generating a 3.8 T magnetic field. It is made of NbTi cables wrapped with copper, cooled down to 4.5 K, and enclosed in a 12000 tons iron yoke for the magnetic return flux which is also hosting the outer muon detecting system. The 2 T residual field present inside the iron yoke provides enough bending power to perform an efficient p_T based muon trigger selection in a $|\eta| < 2.4$ pseudorapidity region.

A summary of the magnet parameters is given in Table 3.6.

3.4 THE MUON SYSTEM

Many of the interesting physical processes foreseen at the LHC are characterised by final states which will involve the presence of high p_T muons. Hence a robust and redundant muon spectrometer is needed to provide precise muon identification, high resolution p_T measurements and an effective trigger capability. During p-p collisions, high energy muons are produced and, crossing both electromagnetic and hadronic calorimeters, they are specifically detected in the far external group of subdetectors of the CMS experiment, the Muon System (Figure 3.12).

The experimental muon setup [109] consists in three different types of gaseous detectors with a different design, coping with the radiation environment and magnetic field at different values of η :

- 250 Drift Tube Chambers (DT), used in the barrel region (with $|\eta| < 1.2$) where a low residual magnetic field is present and track occupancy is low;
- The endcaps ($0.8 < |\eta| < 2.4$) are equipped with 540 Cathode Strip Chambers (CSC) with a faster and radiation resistant capability in order to cope with a higher particle flux and a non uniform magnetic field;
- To ensure redundancy and improve trigger performances, 610 Resistive Plate Chambers (RPC) complement the DT and CSC in both regions up to $|\eta| < 2.1$, due to their fast response and excellent time resolution but low spatial resolution, improving the precision in the muon trigger on the determination of the bunch crossing (BX) in which the muon has been created.

Figure 3.13 shows the total p_T resolution as a function of η , together with the curve obtained using only information coming from the tracker or the muon system only. For

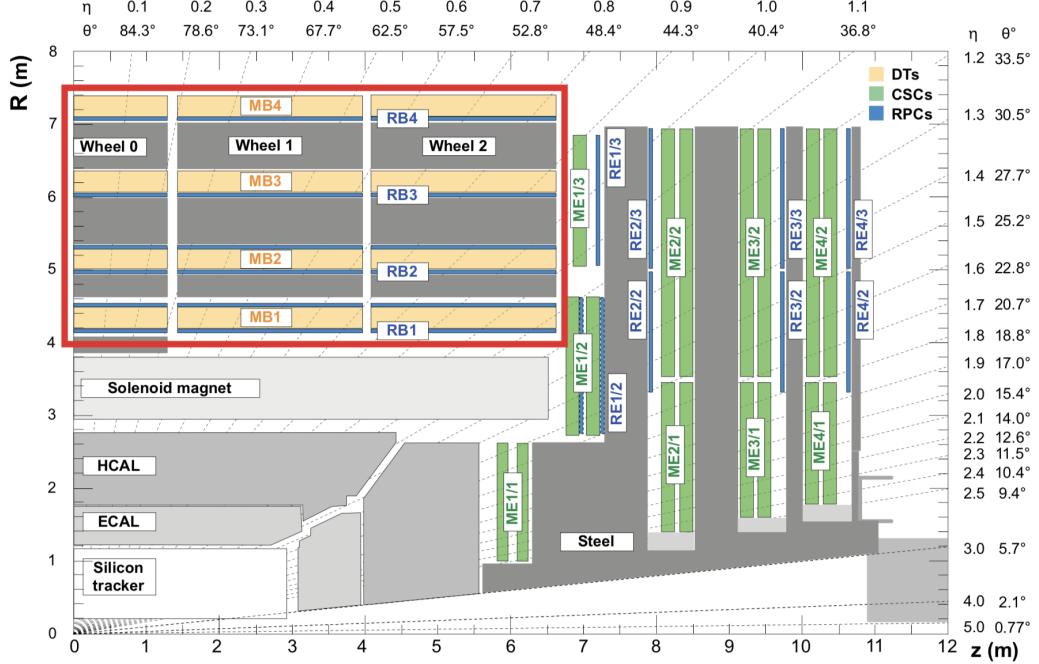


Figure 3.12: A schematic view of the CMS muon system. DTs, CSCs and RPCs subsystems are indicated in orange, green and blue, respectively.

muons with $p_T < 200 \text{ GeV}$, the tracker precision is the most relevant because of the multiple scattering occurring when crossing the calorimeters and the iron yoke of the muon part, while at higher p_T the combination of the two systems improves the overall resolution.

3.4.1 The Drift Tube Chambers

The barrel region of the muon system is characterised by a low residual magnetic field, low occupancy and a large area to be covered. For this reason, Drift Tube chambers have been employed [110].

The detector management follows the yoke segmentation that consists of 5 iron wheels composed of 12 azimuthal sectors, covering an angular region of $\approx 30^\circ$ each and labelled with numbers 1 to 12, starting from the x -axis (Figure 3.14). For each wheel, 4 concentric rings (called *stations*) of DT chambers are installed, named MB1÷MB4 (where MB stands for *Muon Barrel*) respectively from the inner to the outer part. Each station is therefore formed by 12 DT *chambers* with the exception of MB4 which consists of 14 of them.

The basic component of the DT system is the drift cell, shown in Figure 3.15. The cell has a transverse size of $4.2 \text{ cm} \times 1.3 \text{ cm}$ and a length that varies from 2 to 4 m. Each cell is equipped with a $50 \mu\text{m}$ diameter gold-plated stainless steel wire at the center and makes use of 5 electrodes to shape the drift field: the anode wire, 2 cathode strips on the side walls of the tube and 2 strips above and below the wire on the ground planes between the layers, operating at $+3600 \text{ V}$, -1200 V and $+1800 \text{ V}$, respectively, and generating an almost uniform electric field among the cell. The gas used is a mixture of 85%/15% Ar/ CO_2 that allows

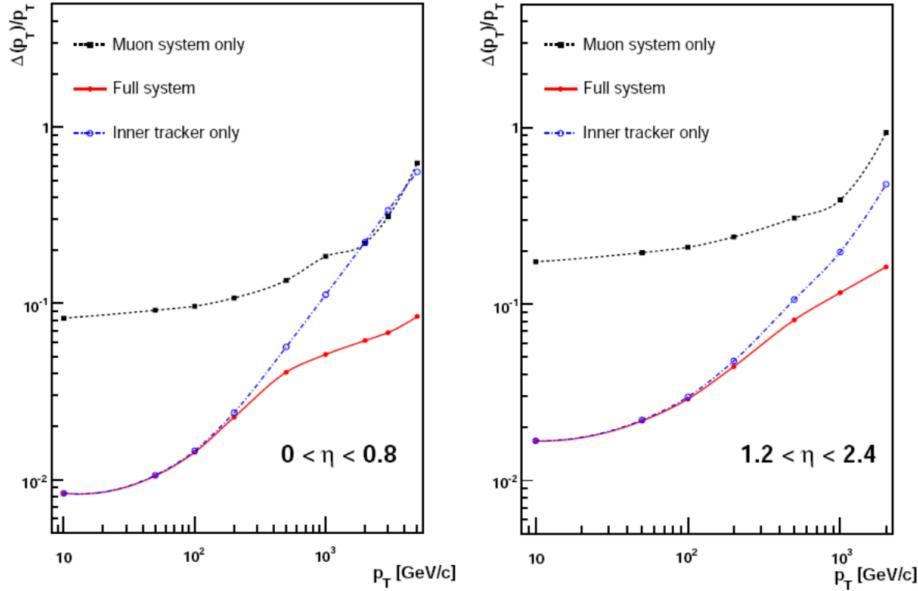


Figure 3.13: Muon transverse momentum resolution as a function of p_T for the inner tracker only (blue dashed line), the muon system only (black dashed line) and the two systems combined (red line). On the left the results for the barrel are shown and on the right for the endcap.

electrons to drift with a saturated velocity of $55\text{ }\mu\text{m/ns}$, corresponding to a maximum drift time around 385 ns.

Four staggered layers of parallel cells form a superlayer (SL). A DT chamber consists of 2 SL that measure the ϕ coordinate plus one measuring the z coordinate. The only exception to this rule applies to the MB4 stations where only the two ϕ superlayer are present.

DT cells have an efficiency of 99.8% and a spatial resolution around $200\text{ }\mu\text{m}$, leading to a resolution of almost $100\mu\text{m}$ for a high quality (i.e. 8 hits) ϕ reconstructed segment.

3.4.2 The Cathode Strip Chambers

The high magnetic field and particle rate expected in the muon system endcaps does not allow to use drift tubes detectors to perform measurements at large η values. Therefore a solution based on Cathode Strip Chambers (CSC) has been adopted [111]. The CSC are gaseous trapezoidal Multiwire Proportional Chambers (MWPC), characterised by a short drift length which leads to fast signal collection. Information about the position of the incoming particle is collected both in the anode wire and on a group of finely segmented cathode strips. The CSC layout is shown in Figure 3.16. Those chambers are arranged to form four *disks* of concentric *rings* placed in between the endcap iron yokes.

Each chamber is composed by 6 layers of 9.5 mm thick arrays of anode wire enclosed between two planes of finely segmented cathode strips for the collection of the ionization signal produced in the 30%/50%/20% Ar/CO₂/CF₄ gas mixture. The wires give information about r coordinate and the strips are used to determine the polar angle. In the first disk, that operates in a region of high magnetic field, the anode wire are tilted by 20° to

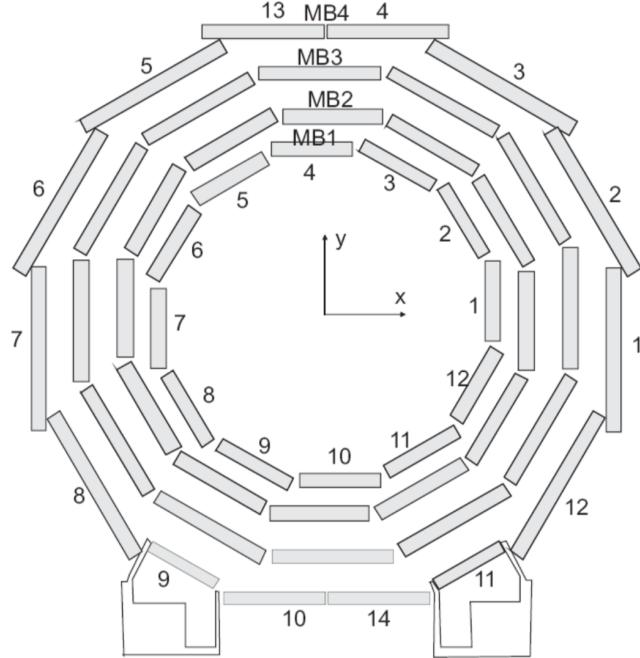


Figure 3.14: Transversal view of the CMS DT system. Station and sector numbers are shown in figure.

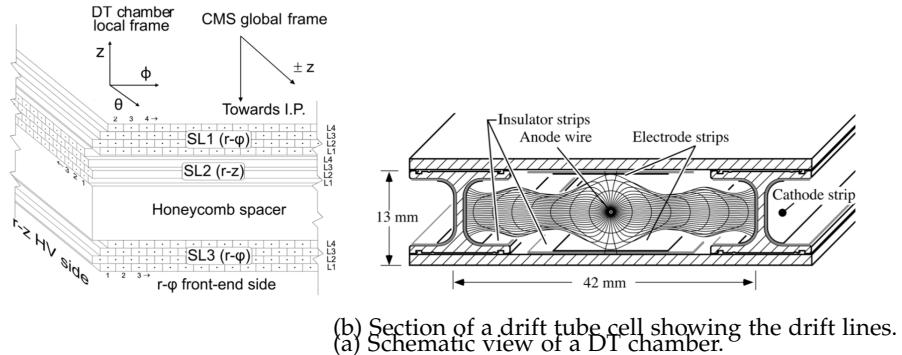


Figure 3.15: DT chamber schematic in the Muon System and a drift tube cell.

compensate for the Lorentz drift effect.

The position resolution measured with the strips varies from $\approx 70 \mu\text{m}$ for the innermost stations to $\approx 150 \mu\text{m}$ for the outermost ones, while r can be determined with a precision of $\approx 0.5 \text{ cm}$.

3.4.3 The Resistive Plates Chambers

Resistive Plates Chambers are used both in barrel and endcaps, complementing DT and CSC systems, in order to ensure robustness and redundancy to the muon spectrometer [112]. RPC are gaseous detectors characterised by a coarse spatial resolution, but may perform precise time measurements, comparable with the ones provided by scintillators. This ensures precise BX identification to the muon trigger system.

CMS uses double-gap RPC chambers composed by 4 bachelite planes forming two 2 mm

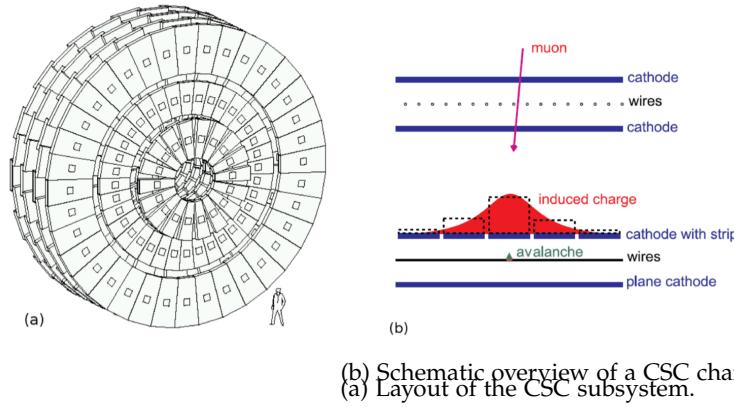


Figure 3.16: The Cathode Strip Chambers of the CMS endcap muon system.

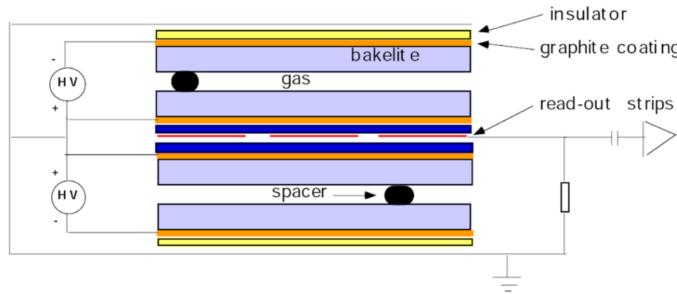


Figure 3.17: Schematic view of a CMS double gap RPC.

gaps, as shown in Figure 3.17, where the crossing particles ionise the 90%/10% mixture of freon ($C_2H_2F_4$) and isobutane (C_4H_{10}), and the electrons are multiplied in avalanche mode. The electrodes, set at a potential of 9.5 kV, are constituted by a graphite coating and the central part of the chamber is equipped with insulated aluminum strips, used to collect the signal generated by the electronic avalanche caused by the crossing particles. The design choice of using double gap chambers is adopted to increase the signal induced on them. In the barrel the strips are rectangularly segmented (12.1 to 41 cm wide and 80 to 120 cm long) and run along the beam axis, whereas the endcaps are equipped with trapezoidal shaped strips covering approximatively the range $\Delta\phi = 5 - 6^\circ$, $\Delta\eta = 0.1$. No measurement is possible in the η coordinate, apart from the constraint imposed by the strip length. In order to sustain higher rates, the detector operates in *avalanche* instead of using the most common *streamer* mode but due to the reduced gas multiplication, improved electronic multiplication is required.

In the barrel region the system layout follows the DT segmentation and two RPC stations are attached to each side of the two innermost DT chambers of a sector, whereas one single RPC is attached to the inner side of the third and fourth DT chambers. This solution ensures to extend the low p_T range of the trigger system in the barrel, detecting even low p_T muons before they stop in the iron yoke.

3.5 TRIGGER AND DATA ACQUISITION

The LHC collisions have a rate of 40 MHz, which means that two proton bunches intersect in the collision points where the experiments are placed every 25 ns = 1 BX. Considering that in the two general purpose experiments ATLAS and CMS each event has a size of about 1 MB, an effective trigger strategy used to select interesting events among all physical interactions, had to be adopted. In order to achieve a rate of events up to ≈ 1 kHz, imposed by the final storage system, a total reduction rate of 10^7 has to be reached. Despite the high rejection factor, trigger algorithms have to be also quite sensitive to different physical processes and their probability: for instance, processes as $Z \rightarrow ll$ and $W \rightarrow l\nu$ have cross sections which are orders of magnitude lower than QCD processes, that would saturate selections based on simple high p_T lepton identification (Figure 3.18).

Furthermore, the BX frequency of 40 MHz requires the development of a system able to take a selection/rejection decision every 25 ns. Since this time is too short to collect all the information coming from all the subdetectors and process it in a single step, a pipelined trigger architecture based on different levels of increasing complexity, has been adopted. In CMS there are two steps, nominally the Level 1 Trigger (L1T) and the High Level Trigger (HLT):

- The L1T system [113], built using dedicated hardware electronics, operates a first selection reducing the rate by a factor of 10^4 , elaborating signals from the muon detectors and the calorimeters;
- The remaining selection is performed by HLT algorithms [114], on the basis of the full information available from all the detectors using a software system running on a filter farm of a thousand commercial processors, reducing by another factor up to 10^3 and satisfying the storage system requirements.

3.5.1 The Level-1 Trigger System

The Level-1 trigger [113] is designed to accept or reject events at a 40 MHz rate (every BX) in a pipeline mode using custom developed programmable hardware. Field Programmable Gate Arrays (FPGA) are used where possible but also Application Specific Integrated Circuits (ASICs) and Programmable Lookup Tables (LUTs) are taken into account to complete each processing step in less than 25 ns. Due to the maximum length for the pipeline buffers fixed at 128 BX ($3.2 \mu s$), taking into account the signal propagation time and the subdetectors latencies, the effective decision operation time for the L1T is less than $2 \mu s$.

The Level-1 Trigger at CMS detector can be further subdivided into three major subsystems: the Muon Trigger, the Calorimeter Trigger and the Global Trigger. The first two systems process informations coming, respectively, from the muon spectrometer and calorimeters and do not have to perform the task of rejecting/accepting events by themselves. On the other hand, they identify and perform sorting on various types of *trigger objects* (electron/photon, jets and muons) and then forward the four best "candidates" of each

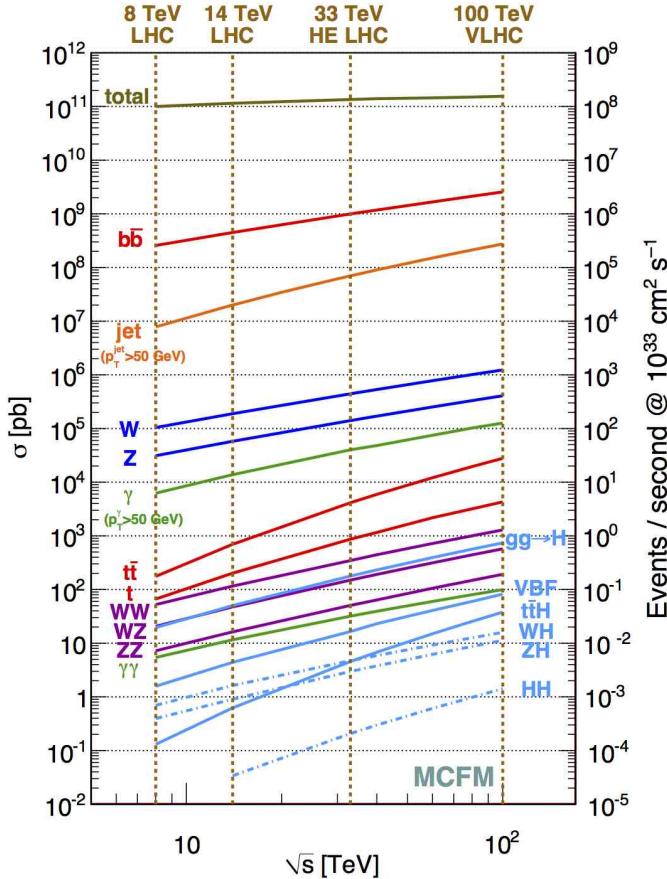


Figure 3.18: Plot showing the different cross sections for the main processes produced at the LHC.

kind of trigger object to the Global Trigger where the final decision is made, as shown in Figure 3.19.

3.5.2 The High Level Trigger and DAQ

The CMS High Level Trigger [114] has the task to further reduce the event rate from the L1 Trigger to ≈ 1 kHz required from the storage system. In order to achieve this requirement, the HLT performs an analysis similar to off-line event reconstruction that relies on a farm of commercial processors.

As shown in Figure 3.20, data coming from the Level-1 Trigger are initially stored in a pipelined 40 MHz buffer by the Front End System (FES). Afterwards, data are moved by the Front End Drivers (FEDs) to the Front End Readout Links (FRLs) which are able to get information from two different FEDs. Event fragments coming from different FRLs are then sent to the *Event Builder* system in charge to build up the full event which has two different purposes: transport data from the underground to the CMS surface buildings and than begin the reconstruction phase. After the assembly phase is completed, the event is sent to the *Event Filter* where HLT algorithms, together with some Data Quality Monitoring operations, are performed. Filtered data are then separated into several on-line streams,

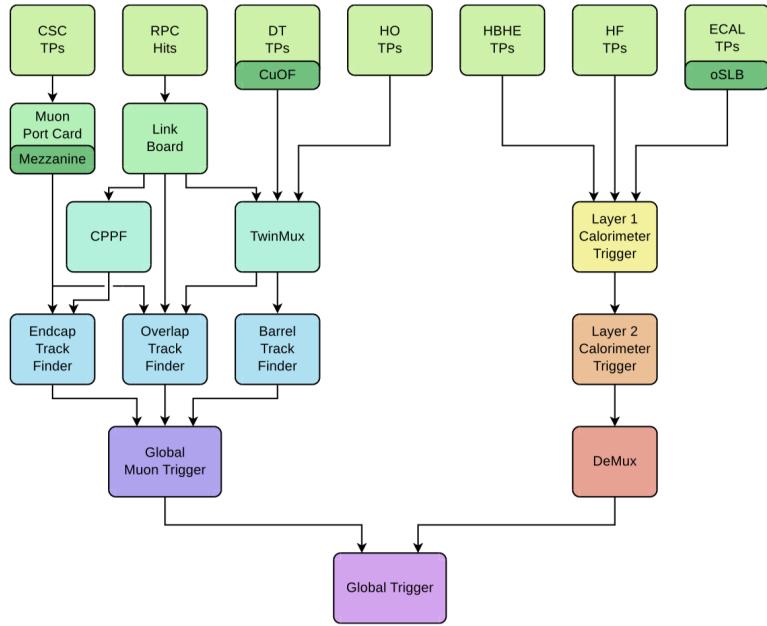


Figure 3.19: Schematic diagram of the CMS Level-1 trigger system.

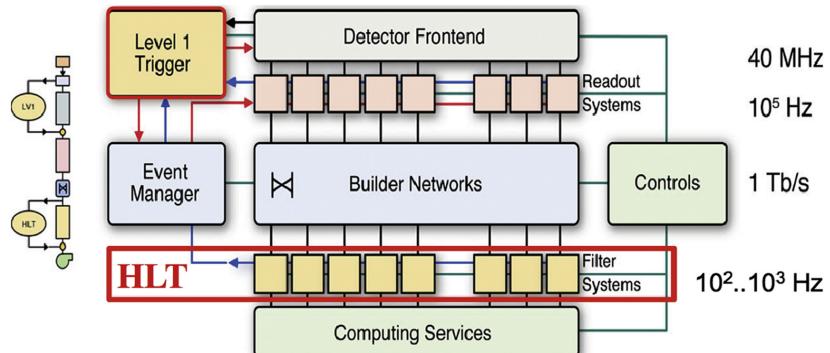


Figure 3.20: Schematic diagram of the CMS High Level Trigger System.

whose content depend on trigger configuration, and are sent to a local storage system before being migrated to the CERN mass storage infrastructure.

3.6 GLOBAL EVENT RECONSTRUCTION AND PARTICLE FLOW ALGORITHM

This section aims at concluding the description of the CMS experiment with an overview on the identification and reconstruction of physics objects candidates coming from each collisions event, using a *particle flow* (PF) technique [115].

The PF approach relies on the combination of information coming from the CMS subdetectors, in order to give a global and coherent description of the events, under the form of a reconstructed particle candidate. Events collected by CMS are centrally processed with reconstruction algorithms referred to as "event reconstruction", starting from raw data and giving as output a collection of detected particles with properties like momentum or angle. Firstly, the individual particles are labelled into mutually exclusive types: muons, electrons,

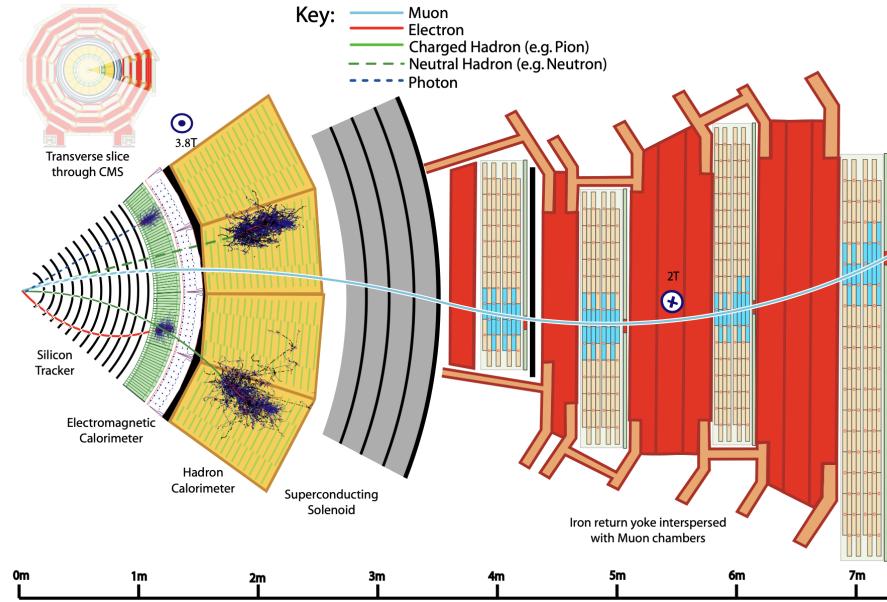


Figure 3.21: A transversal representation of the CMS subdetectors, with a simplified path of each type of particle candidate.

photons, charged hadrons, and neutral hadrons. The signatures of the main categories of particles, passing through the CMS detector, are sketched in Fig 3.21.

The track trajectory is reconstructed starting from the hits in the tracker, and then linked to deposits in the ECAL only (for electrons) or in the HCAL as well (for charged hadrons), while the passage of a photon is identified as ECAL energy clusters not matched to the extrapolation of any charged particle trajectory from the tracker. Muons, instead, are identified firstly as tracks from the central tracker and then matched with either tracks or several hits in the muon system, potentially associated with calorimeter deposits. Charged and neutral hadrons create hadronic showers in the ECAL with a subsequent absorption in the HCAL, whose clusters are then used to estimate the energy and direction. An indirect measurement of non-interacting, uncharged particles provided by the calorimeters is also crucial to compute the missing energy (MET) which could be a signature of new particles and phenomena.

Finally, higher level physics object e.g. jets, MET, τ leptons and lepton isolation can be built starting from the PF candidates:

- All particles are clustered into jets, with the $\text{anti-}k_T$ algorithm. The jet momentum and spatial resolution, are quite good thanks to the excellent ECAL granularity and the high quality of tracking detectors;
- The τ lepton, characterised by a very short lifetime, can be identified thanks to its hadronic decays through the reconstruction of the intermediate resonances;
- The MET vector is reconstructed as the opposite of the transverse momentum sum of all final state particles reconstructed in the detector.

4

MACHINE AND DEEP LEARNING

This chapter provides an introduction on the Machine and Deep learning concepts with their application in High-Energy Physics. In particular, a focus on *neural networks* will be provided, intensively applied in the analysis described in the next Chapters.

4.1 INTRODUCTION

The name *machine learning* was coined in 1959 by Arthur Samuel [116] and, over the past two decades, Machine Learning (ML) has become one of the pillars of information technology and quite adopted into daily technologies: users all around the globe, each day, use a learning algorithm without even knowing. Every time a user performs a Google search or watches a video on YouTube or Netflix, a complex structure of learning algorithms records all our choices and preferences in order to build a customised environment. Email providers also adopt spam filters that are the result of hidden ML algorithms and an increasing improvement of *pattern recognition* techniques allows social networks (Facebook, Instagram etc...) to identify different people in photos but also to create 3D images of human tissues and organs for medical diagnosis.

Currently, the rise in learning algorithms in many sectors is mainly related to an increasing quantity of data available combined with a technological progress in storage and computational power, together with lower maintenance and material costs. From Section 4.8, a more detail look into the applications of Machine and Deep Learning for High Energy Physics will be provided.

The most common tasks where ML is employed are *classification* and *regression*:

- In classification tasks, the computer algorithm has to separate data in different categories; it can be a binary classification, e.g. an email spam filter which has to decide if a message is or is not fraudulent or, in a physical use case, the signal/background discrimination of a decay process, or it can be multiclass if the categories are more than two. To solve this task, the learning algorithm adopts a function of the form $f : \mathbb{R}^n \rightarrow [0, 1]$.
- In regression tasks the learning algorithm is asked to predict a continuous numerical value given some inputs e.g. an house prizing algorithm or the prediction of the muon p_T given the track information from the muon chambers [117]. To solve this task, the learning algorithm is asked to output a function of the form $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

4.2 TYPES OF MACHINE LEARNING

There are several types of learning algorithms. The main two types are called *supervised* learning, *unsupervised* learning and *reinforcement* learning.

4.2.1 Supervised learning

Supervised learning algorithms learn to produce an output that is known a priori, called in jargon *target* or *label*, based on an input distribution of variables, usually referred to as *features*. Supervised learning algorithms are therefore built to estimate a probability $p(y|x)$. The algorithm can often be seen also as a function mapping the inputs into the outputs. The algorithm consists in optimizing the function based on a criterion and find the best parameters.

In order to understand this concept, a brief example is shown [118]. Let's consider a dataset containing the US housing prices with a summary plot shown in Figure 4.1. In this example we have the price of the house on the vertical axis and the size of the house in the horizontal axis. In order to make a prediction for a 750 ft² house, a plausible learning algorithm might put a straight line through the data (pink line in Figure 4.1). In this case, the corresponding price would be 150 k\$. However, this is not the only algorithm we could use: instead of a straight line (linear regression), a n-order polynomial curve may be used (blue curve in Figure 4.1). Each possible curve with a different prediction on the housing price.

Visual recognition is an application domain that highly relies on supervised ML algorithms. For instance, a system might need to learn to identify pedestrians on a street in a automotive application for self-driving cars: to do so, it is trained with millions of short videos about street scenes, with some of the videos containing no pedestrians at all while others having up to dozens. The presence or not of pedestrian is known a-priori, hence the learning is supervised: a variety of learning algorithms are trained on such data, with each having access to the correct answer.

Many other decision support mechanisms that have at disposal large quantity of data on which to train a ML system, could be the base of a supervised ML approach: e.g. historical data on medical exams' output may drive a supervised ML system to learn and prompt the probability of suffering from a disease, etc.

4.2.2 Unsupervised learning

Unsupervised algorithms are those that experience only "features" but not a supervision signal, such as the target or label, with the goal of underline the possible structure of data and therefore learn the possible outcomes. Given this dataset, an unsupervised learning algorithm may decide to perform a clustering selection, as shown in Figure 4.2.

Clustering algorithms like this one are used in many web-based application such as newspaper websites. Using thousands of stories as input data, the learning algorithm auto-

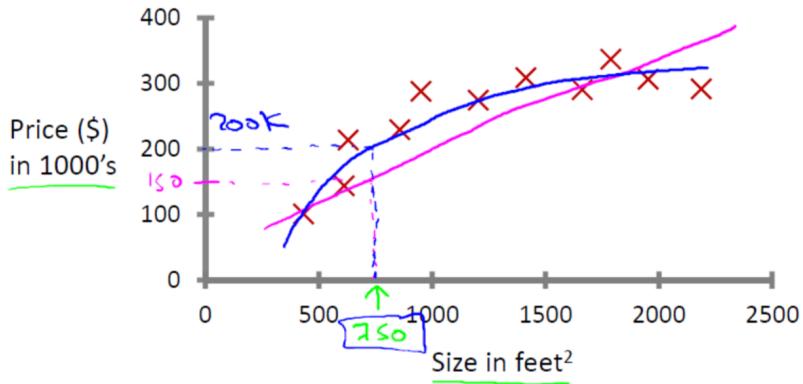


Figure 4.1: US housing prices example. The information is represented by the red crosses while the prediction curves are plotted with coloured lines.

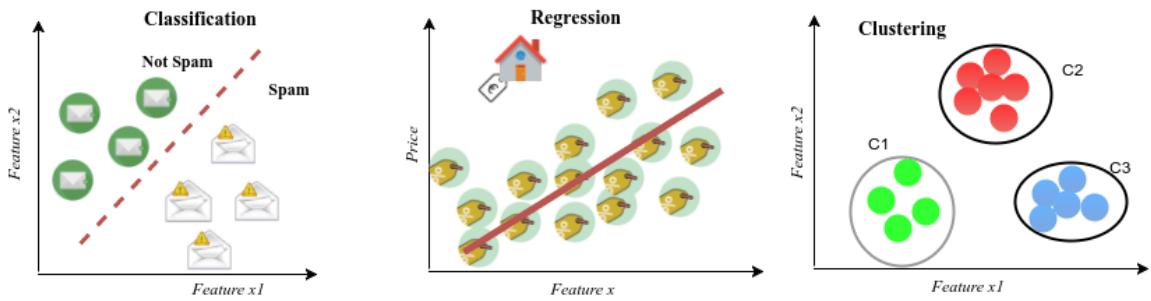


Figure 4.2: Examples of real-life problems in the context of supervised and unsupervised learning tasks: Spam filtering as a classification task and House price estimation as a regression task are part of supervised learning; Clustering is part of unsupervised learning in which customers are grouped into three different categories based on their purchasing behaviour (image from [119]).

matically cluster the information together based on topic or user preferences, organizing information in a coherent and customizable way.

Unsupervised learning is used also for a bunch of other applications. It is used to organize large computer cluster, deciding which machines tend to work together in order to put those machines together and increase the overall efficiency. Another application is on social network analysis identifying between hundreds of friends the closest one, based on private communications or recent social activities.

4.2.3 Reinforcement Learning

Reinforcement learning [120] is an important type of ML, in which an agent learns how to behave in a environment by performing actions and deciding about the next actions based on the outcome of the previous ones. In recent years, a lot of improvements were observed in this area of research. Most popular examples include DeepMind and the Deep Q learning architecture (dated 2014), which culminated in the media when the champion of the game of Go was beaten by AlphaGo in 2016.

Reinforcement learning can be understood using the concepts of Agents, Actions, Environments, States, and Rewards (note that this description is intended to be short and provided

just for completeness in describing possible ML types, and no deep discussion of its characteristics and implementation will be given, as it would go beyond the research scope of this thesis).

- An "Agent" is the component that takes actions, e.g. a video game character navigating in its virtual environment, as well as a drone making a delivery.
- An "Action" is one amongst the set of all possible moves/choices the agent can make. In a reinforcement learning environment, agents can choose only among a predefined list of possible actions. E.g. in video games, the list might include moving right or left, jumping or not, jumping high or low, or crouching, or standing still; in the stock markets, the list might include buying, selling or holding any title among a list of financial product.
- The "Environment" is the world through which the Agent moves. The environment takes as input the Agent's current State and its selected Action as input, and returns as output the Agent's Reward and next State. E.g. in a model in which the Agent would be myself, the Environment could well be the entirety of the physics laws, plus the rules of the society I live in and within which filter all my Actions would be processes and the consequences of them determined.
- The "State" is a concrete situation in which the Agent happens to put itself. E.g. it could be a specific moment in time and/or place in space, a local and instantaneous configuration that puts the Agent in contact and relation with its Environment (e.g. tools, obstacles, prices).
- The "Reward" is the feedback based on which you measure the success or failure of the Agent's choices. E.g. in a video game, a reward could indeed be the gain of a prize when a special object is captured, and similar. Every time an Agent does something in the Environment that foresees a possible Reward, the Agents sends output in form of Actions to the Environment and the Environment returns the Agent's new state as well as the obtained Rewards (or lack of them).

In a nutshell, as shown in Figure 4.3, reinforcement learning judges actions by the results they produce. It is fully goal oriented, as its aim is just to learn sequences of actions that will eventually lead an Agent to achieve a predefined goal, in terms of maximizing its objective function. In the video game example, the final goal is to finish the game with the maximum score, so each additional point obtained throughout the game will affect the Agent's subsequent behaviour (i.e. an Agent might learn that making some combination of Actions will lead to a point to it will try to redo the same actions effectively in case a similar situation will be offered next in the Environment). In a robotics example, a robot might have as a goal to travel from A to B: every millimetre the robot gains that makes it closer to the spatial objective B is counted as additional Reward, so the robot will learn the direction to go and eventually reach the final destination.

The implementation of reinforcement learning models requires a lot of data. For this reason, it has historically been associated with domains in which plenty of simulated data

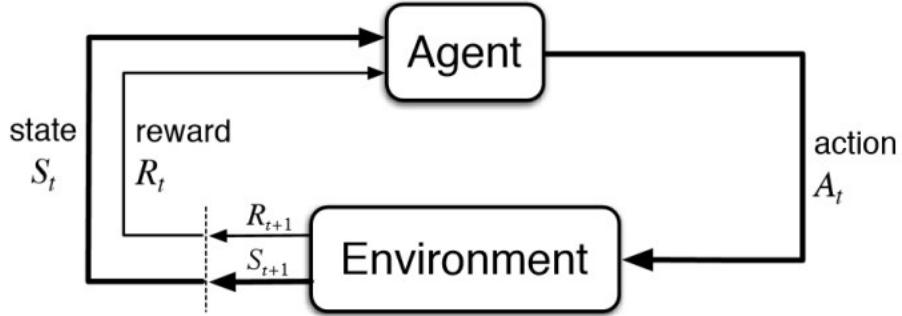


Figure 4.3: Schematic workflow of a reinforcement learning algorithm.

is available (e.g. video games and robotics, as in the examples above). One other aspect to mention is that - with respect to other ML types - it is far from easy to take results from academic research papers and implement them in real-world applications. Also, the results reproducibility is challenging, and this is a key issue: as ML gets deployed in mission-critical situations, the reproducibility and the ability to estimate errors become essential, and this aspect is particularly important for reinforcement learning.

Nowadays, applications of RL can be seen e.g. in high-dimensional control problems, like robotics: they have been the subject of research (in academia and industry) for many years, and now start-ups are beginning to use this ML type to build products for industrial robotics applications.

4.3 MACHINE LEARNING FORMALISM

Let us define two domains, the features domain \mathbb{X} and the target \mathbb{Y} (or *label*). Let the features domain be a real vector space, so that we may call an object belonging to that domain a vector of features. Target domain can instead be either a real vector space or a finite set. Call a 2-tuple (x, y) with the first element from the features domain and the second element from the target domain an *example*. Let us also have a probability distribution over examples with the probability density $p(x, y)$, which is unknown to us. A sample from this unknown distribution i.e. a series of points in (x, y) (example) space, is called *training dataset*¹.

4.3.1 Model and training

The problem searching for the best model in the abstract function space, is not computationally feasible for a computer. Therefore, a ML model is commonly defined as a parametric function $f(\mathbf{x}, \boldsymbol{\theta})$, where \mathbf{x} is an element of the features domain and $\boldsymbol{\theta}$ are the parameters of the model, and f is a fixed function. In order to infer a quantity \mathbf{y}_{true} from the true label, such parametric function f is used to define $\mathbf{y}_{pred} = f(\mathbf{x}, \boldsymbol{\theta})$, where \mathbf{y}_{pred} is

¹ With this notation, unsupervised and reinforcement learning are omitted. In this thesis, only supervised learning techniques have been used.

the value of \mathbf{y} predicted by the model.

The degree of correspondence between a model and data is defined in terms of some error (or *loss*) metric. This metric is usually written as a function of the two arguments from the target domain, $L(\mathbf{y}_{pred}, \mathbf{y}_{true})$. It is desirable to match the loss function and the utility of the model for the purpose it is being created, e.g. the uncertainty of a branching fraction measurement. Since it is quite difficult to derive such relationship, if not impossible, a computational-friendly loss function is often used.

An example is the *mean square error* (MSE): the averaged squared loss per example over the whole dataset. To calculate the MSE, sum up all the squared losses for individual examples and then divide by the number of examples:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{(x,y) \in D} (y_{true} - y_{pred}(x))^2, \quad (4.1)$$

where (x, y) is an example and D is a dataset containing many labelled 2-tuple examples. The objective of training is to find the parameters' values that minimise the expected loss over all possible examples:

$$\boldsymbol{\theta} = \operatorname{argmin}_{\boldsymbol{\theta}} [E_{(x,y)} L(f(\mathbf{x}, \boldsymbol{\theta}), \mathbf{y})]. \quad (4.2)$$

The expectation is estimated using the training dataset:

$$\boldsymbol{\theta} = \operatorname{argmin}_{\boldsymbol{\theta}} \left[\frac{1}{N} \sum_{i=1}^n L(f(x_i, \boldsymbol{\theta}), y_i) \right], \quad (4.3)$$

where N is the number of examples in the training dataset and x_i, y_i are examples from the training dataset features and labels, respectively. Initially, all the parameters are randomly set and, at each iteration, the value of the loss function generates new values for $\boldsymbol{\theta}$. This learning procedure continues until the algorithm discovers the model parameters with the lower loss possible. When the loss stops changing (or changes really slow), the model has *converged*. Once the model has converged, the loss function reaches a *minimum* for a given value θ_j . Therefore, this kind of regression problems will have a convex plot of loss vs parameters. However, calculating the loss function for every value of θ_j is an inefficient way of finding the convergence point. Another technique - quite common in ML - is called *gradient descent*. After choosing a starting value for the parameters, typically random, the gradient descent algorithm calculates the gradient of the loss curve at the starting point. Mathematically speaking, a vector of partial derivatives with respect to the parameters, is computed:

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} L(\boldsymbol{\theta}), \quad (4.4)$$

where α is a scalar value known as *learning rate* (or *step size*), one of the model's *hyperparameters* (see Section 4.3.3) to determine the next point. If the gradient is negative, the algorithm performs the next step (shown in Figure 4.4), edging closer and closer to the

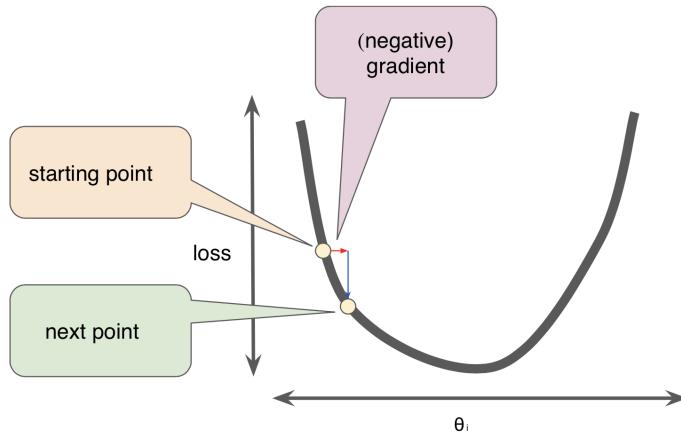


Figure 4.4: Schematic representation of the gradient descent mechanism (image from [121]).

minimum.

Tuning the learning rate is a tricky problem: if a learning rate too small is selected, the learning procedure will take too long; conversely, if it is too large, the next point will exceed the minimum θ value and the correct value will never be reached.

In case of loss function minimisation with gradient descent, when large volumes of data come into play, other parameters in the gradient descent strategy become very relevant for a fast convergence of the minimisation algorithm towards finding the global minimum, e.g. the gradient descent *batch size*: it corresponds to the total number of examples you use to calculate the gradient in a single iteration. By default, the batch matches the entire dataset. However, datasets may often contain millions and millions of examples as well as a huge number of features. In these cases, a batch can be enormous and also the corresponding computing time. Smaller batch sizes are therefore suggested: in the *stochastic gradient descent* (SGD), explored in more detail in Section 4.6.1, for example, the batch size is brought to one single example per iteration. Given enough iterations, SGD works but results quite noisy.

An intermediate approach among batch GD and SGD is the so called mini-batch GD: the original training set is subdivided into smaller batches (usually of the order of 10, but can be much more depending on the size of the dataset) - these are chosen at random and allow parallelisation of the entire process, which might be extremely useful in case of heterogeneity of hardware devices available for training - and will ultimately provide GD algorithm convergence in shorter time.

4.3.2 Overfitting

A common problem with ML is *overfitting*. This particular issue occurs when a trained model - e.g. based on high-order polynomial hypotheses - matches the training dataset in a very precise way, thus achieving a high accuracy on that dataset but a poor generalisation capability over new, previously unseen data. In Figure 4.5, an example is shown. A model, that is too complex for the problem and whose performance on the training dataset it

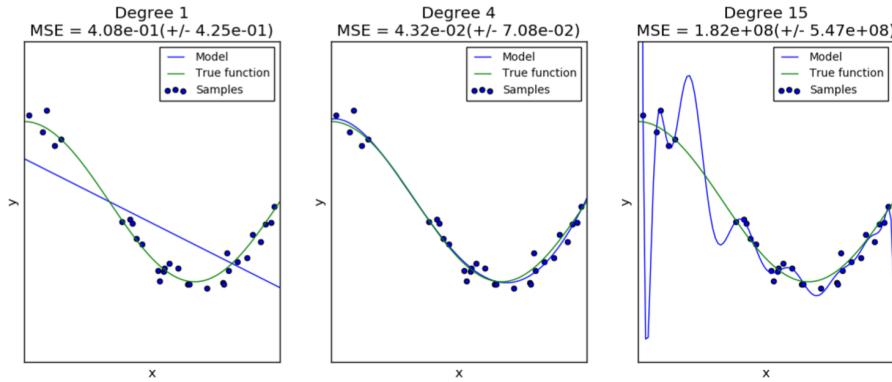


Figure 4.5: An example of underfitting and overfitting. Both features x and target y are real scalars. The polynomial of degree 1 is a case of underfitting, of degree 15 overfitting. MSE in the title refers to the mean squared error, $MSE = \sum_{i=1}^N (y_{predicted} - y_{true})^2 / N$, where N is the number of examples. Reproduced from [122].

much higher than on examples it has never seen, is commonly called *overfitted*. A model that is not flexible enough for the problem performs similarly poorly on training, and new data is known as *underfitted*. Therefore, the less complex a ML algorithm is, the more likely an empirical result is not caused by peculiarities of the sample.

4.3.3 Hyperparameters

Usually, a ML model is a member of a specific parametric function family. The parameters that define this particular model from its family, e.g. the degree of a polynomial for polynomial regression, along with the parameters of the optimisation procedure are commonly called *hyperparameters*. Hyperparameters are not changed during model training. If there is an automated procedure for finding the optimal value of hyperparameters, than from the point of view of that procedure, they become just model parameters.

4.3.4 Training and Test set

A ML model aims to make good predictions on new, previously unseen data. However, when the model is built starting from a specific dataset, there are no remaining examples with new unseen data. There are 4 different techniques in order to fix this common problem for a ML study:

1. Split into a Training and Test Set;
2. *k-fold* Cross-Validation;
3. Leave One Out Cross-Validation;
4. Repeated Random Test-Train Splits.

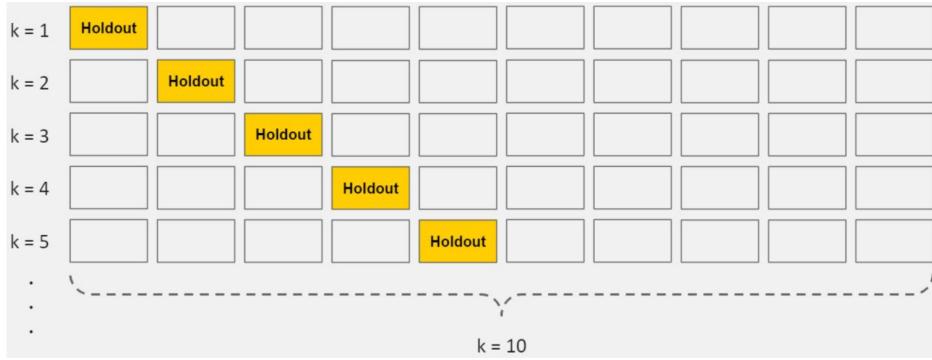


Figure 4.6: Pictorial representation of 10-Fold Cross Validation [123].

Training and Test split. The fastest technique is splitting the dataset into a train and test subsets. The model is trained on the training dataset, and the performance is evaluated on the test dataset. It has the following pros and cons:

- Pros: It is ideal for large datasets (hundreds of thousands of records). Splitting a large dataset into large sub-datasets allows that, first, each split of the data is not too tiny, and second, both are representative of the underlying problem. Because of the high speed, it is useful to use this approach when the algorithm you are investigating is slow to train.
- Cons: A downside of this technique is that it can have a high variance. This means that differences in the training and test dataset can result in meaningful differences in the estimate of accuracy. Furthermore, in order for this split to work, the test set has to be representative of the dataset as a whole, i.e. the test set should not have different characteristics than the training test.

***k*-fold Cross-Validation.** *K*-fold cross-validation is another approach that can be used to estimate the performance of a ML algorithm with less variance than a single train-test set split. The data set here is divided into k subsets (called "folds"): for each fold, one of the k subsets is used as the test set and the other $k - 1$ subsets are put together to form the training set (see Figure 4.6). Then, the average error across all k trials is computed, together with the standard deviation.

This has the following pros and cons:

- Pros: This method is less sensitive to how the data division happens to be made: every data point gets to be in a test set exactly once, and gets to be in a training set $k - 1$ times. Therefore, the variance of the resulting estimate is reduced as k is increased.
- Cons: The disadvantage of this method is that the training algorithm has to be rerun from scratch k times, which means it takes k times as much computation to make an evaluation, thus resulting evidently slower.

Leave One Out Cross-Validation. The *Leave One Out Cross-Validation* is a simplified version of the k -fold where the size of the fold is 1 (i.e. $k = n$). The result is a large number

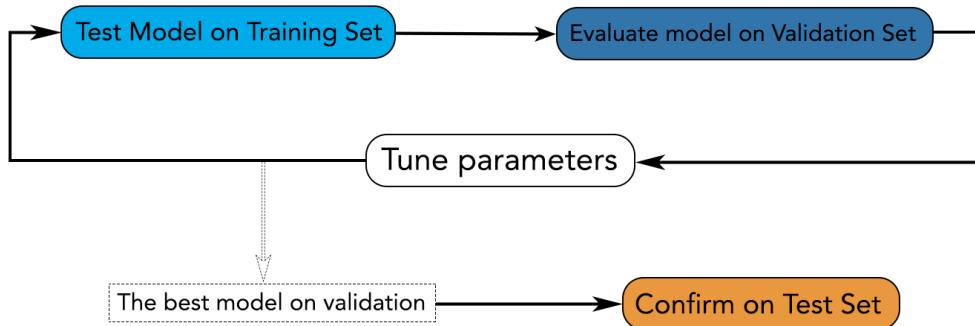


Figure 4.7: Workflow using a validation and test subset.

of performance measures that can be summarized in an effort to give a more reasonable estimate of the accuracy of your model on unseen data. An evident downside of this procedure is the computational power which is much more expensive than k -fold CV.

Repeated Random Test-Train Splits. Another variation on k -fold cross-validation is to create a random split of the data like the train/test split described above, but repeating multiple times the process of splitting and evaluation of the algorithm, like a cross-validation. This has pros and cons:

- Pros: it has the speed of using a train/test split and the reduction in variance in the estimated performance of k -fold cross-validation. It is also possible to repeat the process many more times as needed to improve the accuracy.
- Cons: repetitions may include much of the same data in the train or the test split from run to run, introducing redundancy into the evaluation.

4.3.5 Validate the model

In the previous topic, the concepts of training and validation datasets are introduced in order to perform a check on the trained model. This workflow is repeated many times: each time a parameter is changed i.e. the learning rate or the addition of an extra feature. At the end, the best model is used for the test set.

Another way to do this, reducing the chances of overfitting, is the partition of the dataset into three subsets: the *training set*, the *validation set* and the *test set*. With this data splitting, the validation set is used to evaluate results coming from the training set. Then, the test set performs a double-check on the evaluation after the model has passed the validation set. In Figure 4.7, an example of workflow is shown: the training set is firstly checked by the validation set, which provides the model's optimal parameters for another cycle of training; then, the best model is tested again using the testing subset to obtain a reliable model.

4.4 MEASURING MODEL QUALITY

Choosing the quality metric is quite important and non-trivial. Say, for example, that the true target values are $[1, 0, 0]$. Which model is better, the one that predicts $[0, 0, 0]$, or $[0.6, 0.4, 0.4]$? Ideally, the quality metric must be defined to measure the performance of the whole data analysis system. For instance, in the Higgs Machine Learning Challenge [124], the approximate median significance was used, i.e. an estimate of the significance of the result of an analysis that uses the model. Yet, for a sufficiently complex data analysis system, establishing a direct relationship between the model output and the overall system performance, given in terms of a single number, becomes quite difficult. There are several general-purpose metrics that are commonly used by ML users, described in more detail in the next Sections [125].

4.4.1 Accuracy

For a classification problem, the model output is transformed into a defined class label. The prediction label is defined as:

$$y_i^{pred} = \operatorname{argmax}_k f_i^k, \quad (4.5)$$

where f_i^k is the model prediction of the k class for the example i . The accuracy, therefore is defined as:

$$\text{Accuracy} = \frac{\sum_i^N [y_i^{pred} = y_i]}{N}, \quad (4.6)$$

where y_i is the true class label for the example i , N is the total number of examples and $[y_i^{pred} = y_i]$ is equal to 1 when $y_i^{pred} = y_i$ and 0 otherwise. This definition is usually reported in another form, based on the definition of *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) and *False Negative* (FN), shown in Figure 4.8. With this new definition, the accuracy becomes:

$$\text{Accuracy} = \frac{TP + TN}{N}, \quad (4.7)$$

where N is the total number of examples or the sum of every outcome: $TP + FP + TN + FN$. While natural from a common-sense point of view, as a practical measure, accuracy has some major drawbacks. First, it discards information: the accuracy scores of a model that predicts 0.7 and 0.999 in a binary classification problem are the same. Second, it is misleading in the case of one class being much more common than the others. The accuracy score of a trivial model that always predicts the most frequent class will be the fraction of said class in dataset.

Accuracy is sometimes used [126] for abstract comparison of ML algorithms, due to its interpretability. This is especially relevant in the case of more than two classes.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figure 4.8: Definition of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). This matrix is often called *confusion matrix*

4.4.2 Precision

Precision is another frequently used metric in ML, and is defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} , \quad (4.8)$$

which can be defined as the fraction of positive predictions actually being correct. This particular metric, in a binary classification for signal/background discrimination of an high-energy physics analysis, is usually called *purity* since it measures the number of correct signal predictions divided by the total amount of signal predictions (both correct and wrong).

4.4.3 Recall

The Recall, also known as *sensitivity* or True Positive Rate (TPR), is defined as:

$$\text{Recall} = \text{TPR} = \frac{TP}{TP + FN} , \quad (4.9)$$

which can be defined as the fraction of actual positive correctly identified. In high-energy physics terms it is usually called *signal efficiency* since it measures the correct signal predictions divided by the total amount of signal events.

4.4.4 Area under the Receiver Operating Characteristic (ROC AUC)

In a binary classification problem, the feature space $\mathbb{Y} = [0, 1]$. Let $f_i \in \mathbb{R}$ be the model output for i -th example: the greater it is, the more likely the example is of class 1. If we need to make a binary decision using this prediction, e.g. keep an event for further analysis, a threshold t is defined, and the continuous values f_i are transformed into the

definite class labels $y_i^{pred} = [f_i > t]$ (1 if the condition is satisfied, 0 otherwise). From a statistics point of view, when classifying, a decision is made under the null hypothesis (background only/SM) that the example is class 1 (or signal hypothesis). Such decisions will suffer from two kinds of errors: false positives (or type I) errors, where a background example is wrongly selected as signal, and false negatives (or type II) errors, where a signal example is rejected as background. By varying this threshold, the trade-off between the error types can be adjusted. The choice of the threshold (also called *working point*) depends on the specific classification problem and the costs of making the errors of each type. In high-energy physics, for example, the working point is usually selected with the threshold value that maximises the so called *significance*, defined as:

$$S = \frac{N_s}{\sqrt{N_b}}, \quad (4.10)$$

where N_s is the number of selected signal examples and N_b the number of selected background examples.

For each threshold value, the true positive rate (TPR) or *signal efficiency* is computed (see Section 4.4.3), together with another quantity called False Positive Rate (FPR), defined as:

$$FPR = \frac{FP}{FP + TN}, \quad (4.11)$$

in high-energy physics terms, also called *background efficiency* (or *1-background rejection*). Thus, the TPR is the fraction of the signal examples that pass the selection threshold, while the FPR is the fraction of background examples that erroneously pass the selection threshold.

If there is no priori information to decide the decision threshold, a common measure is to plot the TPR as a function of FPR. This curve is also called *Receiver Operating Characteristic*² or ROC. While allowing for maximum flexibility of evaluation, such curve is not a convenient scalar performance score. To solve this, a commonly used summary statistic is the Area Under the ROC Curve (AUC). The AUC therefore takes the role of the average value of TPR for all possible values of FPR. If some information about the desired threshold region is available, e.g. the significance value described above, it can be incorporated in the form of taking the area under a part of the ROC curve, or just a point on the curve.

The AUC has mathematical properties that makes it attractive for the comparison of different classifiers. First of all, the AUC is a finite quantity, lying in a well-defined interval. For a perfect classifier, all predicted values for signal examples are greater than for all background examples, therefore the AUC = 1. For a totally random classifier, the predictions are distributed equally for signal and background examples, and the ROC curve is a straight line from (0,0) to (1,1) with an AUC=0.5. An illustration between TPR, FPR and AUC is presented in Figure 4.9.

The AUC is defined for a binary classifier. If there are several classes to distinguish, the

² This name derives from the military radar receivers operators in 1941 during the Second World War, which adopted this metric for the first time.

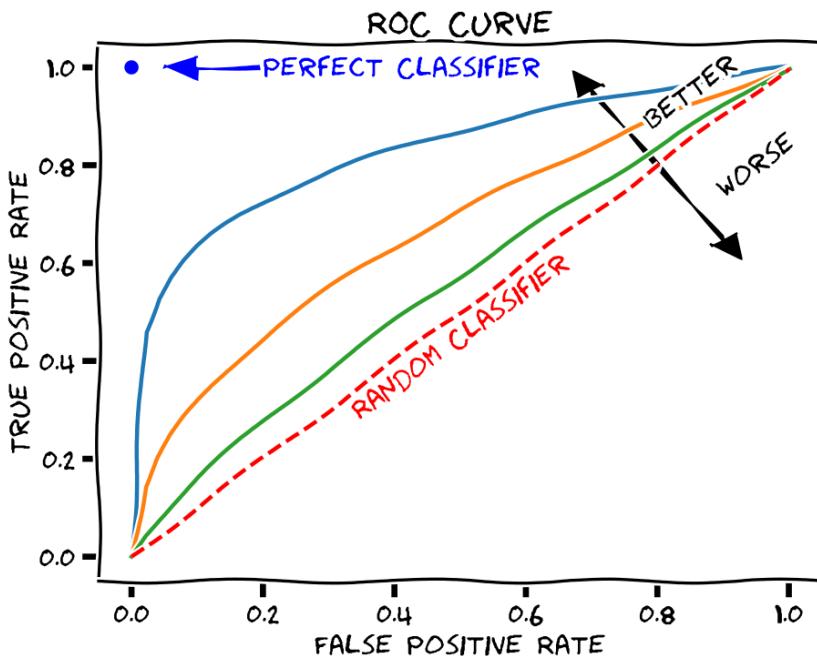


Figure 4.9: An illustration of a Receiver Operating Characteristic curve (ROC), and the area under the curve (AUC). In red, the ROC curve of the worst possible model giving random predictions ($AUC=0.5$). An higher AUC corresponds to a better classifier (coloured curves). The blue dot is the perfect classifier, with 100% correct predictions.

single ROC AUC metric can be replaced with a set of numbers. There are two main ways to transform a n-class classification into sets of binary classification problems:

- one-vs-rest: the classification is reduced to n binary classification problems. For each class, it considers the value of the metric, computed with the other classes collapsed into a single virtual class;
- one-vs-one: the classification is reduced into $n(n - 1)/2$ binary classification problems. Each class is evaluated against every other class.

4.5 SOME LEARNING ALGORITHMS

In the last 15 years, the combination of powerful ideas and huge amounts of data and computing resources caused a significant boost in a plethora of ML algorithms. In the section below, an introduction on the most famous ones is provided, based on the classical textbooks [127, 128], with a particular focus on Neural Networks, deeply used in this thesis [125].

4.5.1 Logistic Regression

Consider the problem of binary classification. Let $x_i \in \mathbb{R}^M$ be feature values for the i -th training example, M the dimensionality of the feature space, and $y_i \in [0, 1]$ the class label. The simplest model possible, is made out of a M dimension hyperplane:

$$y(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + w_0, \quad (4.12)$$

where $y(\mathbf{x})$ is the model decision function ($y > 0$ means the prediction is class 1, $y < 0$ class 0), $\mathbf{w} \in \mathbb{R}^M$ the slope (or *weight* vector) and $w_0 \in \mathbb{R}$ the intercept (or *bias* term). The question is to find the optimal values for \mathbf{w} and w_0 . Theoretically, it is possible to use the formulation used so far, e.g. maximise the number of correctly classified examples:

$$\mathbf{w}, w_0 = \underset{\mathbf{w}, w_0}{\operatorname{argmax}} \left[\sum_i H(y(\mathbf{x}_i)) + \sum_i H(-y(\mathbf{x}_i)) \right], \quad (4.13)$$

where H is the Heaviside function:

$$H(n) = \begin{cases} 0 & n < 0 \\ \frac{1}{2} & n = 0 \\ 1 & n > 0. \end{cases} \quad (4.14)$$

The Eq.4.13 is not convenient for optimisation, since no closed-form solution is available and the gradient of the argmax argument is zero almost everywhere. However, if a sigmoid function σ is applied to the model output, the output can be interpreted as the posterior probability of the example belonging to the class 1:

$$f_i = p(\text{i-th example is class 1} | \mathbf{x}_i, \mathbf{w}, w_0) = \sigma(y(\mathbf{x}_i)) = \left(1 + e^{-\mathbf{w}^T \cdot \mathbf{x}_i - w_0}\right)^{-1} \in (0, 1). \quad (4.15)$$

This model is called *logistic regression*, despite it is used as a classification model. To find the parameters of the model, the maximum likelihood method is used. The posterior probability that i-th example has been correctly classified can be therefore formulated as:

$$p(\text{i-th example is correctly labelled} | y_i, f_i) = f_i^{y_i} (1 - f_i)^{1-y_i}. \quad (4.16)$$

Since examples are sampled independently, the posterior probability is a product of per-example probabilities:

$$\mathbb{L}(\mathbf{w}, w_0) = p(\text{all train examples are correctly labelled} | \mathbf{w}, w_0) = \prod_i f_i^{y_i} (1 - f_i)^{1-y_i}. \quad (4.17)$$

To simply the optimisation, the loss is defined as the negative logarithm of the likelihood, leading to the so called *cross-entropy* error function:

$$L(\mathbf{w}, w_0) = -\ln \mathbb{L}(\mathbf{w}, w_0) = -\sum_i [y_i \ln f_i + (1 - y_i) \ln(1 - f_i)]. \quad (4.18)$$

This loss function has continuous gradients and thus can be optimised with gradient descent algorithms described in Section 4.3.1.

4.5.2 Gradient Boosting Decision Tree

This particular learning algorithm did not experience - during the last years - a particular bursts in terms of advancement but, for many problems with tabular features, they achieve the same or better quality as the most advanced learning models.

Decision Tree. Tree-based methods partition the feature space into a set of rectangles, and then fit a simple model (usually a constant) in each one [129]. In terms of high-energy physics scenarios, they can be viewed as automated multilevel cut-based analysis. Decision trees have been used in the same untouched form from at least 1950s [130, 131]. An illustration on how tree-based algorithms work is shown in Figure 4.10.

In this section, only binary trees are considered, with two alternatives on each point. Having splits with more alternatives is, of course, possible, but would complicate training and is not used in practice. A tree V is a tuple consisting of internal nodes u , leafs v , predicates β_u and leaf values $c_u \in Y$. Y depends on the problem: for a regression problem, it is equal to the target domain \mathbb{Y} ; for a classification problem, instead, Y can be equal to \mathbb{Y} . In this case, the model will only output the predicted class label, without any score indicating the uncertainty of the prediction. An internal node has two children, left and right, which can be either leaves or other internal nodes. The first node is u_0 and $\beta_u : \mathbb{X} \rightarrow [0, 1]$ is a function from the features domain into $[0, 1]$. For an example with features x , if its value is 0, the example is assumed to fall into the left subtree, otherwise into the right. Finally, a leaf is an end node where the prediction is stored.

β_u can be any function: in practice, in ML algorithms, the function compare the value of some feature to some threshold:

$$\beta_u(\mathbf{x}) = [x_{j_u} < t_u], \quad (4.19)$$

where the bracket $[x_{j_u} < t_u]$ is 1 if the condition is satisfied, and 0 otherwise. j_u is the index of the feature that is used for the split, and t_u is the threshold. When training a decision tree, an algorithm selects splits step-by-step, on each iteration choosing the split that offers the most obvious immediate performance improvement. It does not think ahead about possible future splits. This approach requires three things: a loss function quantifying the performance of a split, a stopping criterion and the leaf value assignment procedure. The loss function $L(X, j, t) \rightarrow \mathbb{R}$ estimates how good is a split of the dataset X by feature number j with threshold t . Leaf assignment usually minimises some loss. Finally, a stopping

criterion looks the state of the algorithm and decides whether the current node should be a leaf, or additional splits must be made.

There are multiple viable choices for loss function and the stopping criterion. The idea of the loss function is to promote splits that result in leaves with best purity - least variance of target value in the set of examples that fall into the leaf after the split. The impurity can be defined as:

$$H(R) = \min_{\mathbf{c} \in \mathbb{Y}} \frac{1}{|R|} \sum_{(\mathbf{x}_i, \mathbf{y}_i) \in R} L(\mathbf{c}, \mathbf{y}_i), \quad (4.20)$$

taking the loss function $L(y_{pred}, y_{true})$ and finds the leaf value \mathbf{c} that minimise the average loss for this set. The value of c in a leaf is selected to minimise the impurity. If the problem is a regression, the usual loss to use is the mean squared error and $c = E(\mathbf{y}_i | \mathbf{y}_i \in R)$. For classification, the logical choice for c is to be a vector of class probabilities. There are two primary choices for the loss function: Gini index and cross-entropy. The cross-entropy loss here is derived from Eq.4.18:

$$L_{\text{cross-entropy}} = - \sum_k c_k \ln c_k, \quad (4.21)$$

where c_k is both the fraction of examples belonging to the k -th class, and the predicted probability of an example belonging to the k -th class. Gini impurity, instead, is defined as follows:

$$L_{\text{Gini}} = \sum_k c_k(1 - c_k). \quad (4.22)$$

Consider a leaf that has a set of examples. Randomly choose an example and randomly assign it a label according to the distribution of labels in the leaf. Gini impurity is the probability that the example is mislabeled. With this choice, if all the labels in the leaf are the same, the resulting impurity is equal to zero. On the other hand, if each label is unique, the impurity would equal to $1 - 1/n$, where n is the number of examples in the leaf.

The stopping criterion has a large impact on model quality. If it stops too early, the model is underfitted, otherwise the model is overfitted. Practically, there are a few choices possible for such criterion:

- Maximum tree depth. Results in a computationally convenient uniform tree, but disregards the variation among the leaves;
- Minimum number of objects in leaf. Much like for bin of a histogram, the smaller is the leaf, the noisier will be the estimate constructed from it;
- Maximum number of leaves in the tree;
- Stop when improvement gains drop below a certain threshold.

Trees algorithms usually have, in addition to the basic algorithm outlined in this section, various additions for preventing overfitting. Decision trees are sometimes used in the simplest implementation, mostly for the reason that they are human-interpretable. However, their performance is almost always inferior to the ensemble methods, e.g. *boosting*,

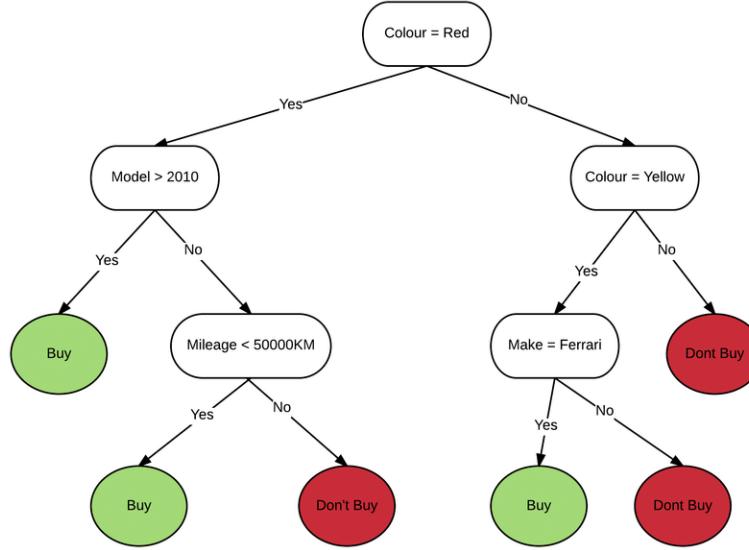


Figure 4.10: An example of decision tree for deciding whether to buy a car or not. Reproduced from [132]

explained in the next paragraph.

Boosting. The main idea behind boosting is the training of many weak models sequentially, so that each next model corrects the error of the ones previously constructed. In line of principle, any model can be boosted, but decision trees are usually the ones implemented with this technique.

As already mentioned, boosting relies on building a sequence of weak learners. Those are formally denoted as a set of predictors $H = \{h_j : \mathbb{X} \rightarrow \mathbb{Y}\}$. Consider, for instance, a regression problem with a single-dimensional target domain, a MSE loss, and a training dataset (\mathbf{x}_i, y_i) consisting of N examples. On each iteration, the differences between the current model predictions and the true label values are computed, a learner is fitted to correct these differences, and the learner is finally added to the model.

Let a_t be the state of the model on iteration t . Let $h \in H$ be the weak learner of the model on such iteration. The objective of the t -th boosting iteration is to minimise the loss:

$$\underset{\gamma \in \mathbb{R}, h \in H}{\operatorname{argmin}} \sum_{i=1}^t L(a_{t-1}(\mathbf{x}_i) + \gamma h(\mathbf{x}_i), y_i). \quad (4.23)$$

To do so, the gradient descent technique introduced in Section 4.3.1 can be exploited. The quantity s_i is computed:

$$s_i = -\frac{\delta L(y_{pred}, y_i)}{\delta y_{pred}} \Big|_{y_{pred}=a_{t-1}(\mathbf{x}_i)} \quad (4.24)$$

and train a weak learner to estimate s_i . This procedure is often called *gradient boosting* [133].

There are several open-source packages that implements GBDT: XGBoost [134], LightGBM [135], CatBoost [136]. All these libraries have an integrated python API for implementation

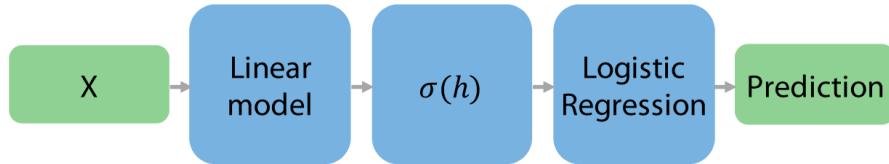


Figure 4.11: A simple working principle for a Neural Network.

as well as supporting multithreading and GPU training.

In conclusion, Gradient Boosting Decision Tree (GBDT) are a powerful learning method, used deeply in data science applications as also in high energy physics (mainly for classification in physics analysis). It has quality competitive with other learning algorithms on tabular data, where there is no known structure in the features, especially with smaller datasets.

4.6 NEURAL NETWORKS

A neural network (NN) is a computing system vaguely inspired by the biological neural connections that constitute a human brain, specifically designed to non-linear learning problems. Currently, NNs are one of the main tools used in ML [137]. The schematic working principle is described in Figure 4.11. The simplest NN has prediction described by:

$$p(\text{example is class } 1|x) = \sigma \left[\mathbf{w}_0^T \cdot \sigma(\mathbf{W}_h \cdot \mathbf{x} + \mathbf{b}_h) + b_0 \right], \quad (4.25)$$

where \mathbf{W}_h and \mathbf{b}_h are the parameters of the first linear model, \mathbf{w}_0 and b_0 are the parameters of the second model (*weights* and *bias* of a logistic regression). \mathbf{W} is a $N \times M$ matrix, where M is the dimensionality of \mathbf{x} and N the dimensionality of the linear transformation output. More complex models can be created, by stacking more intermediate transformations.

NNs have their own terminology:

- A model consisting of a sequence of linear transformations and elementwise non-linear transformation (see Figure 4.11) is called a multilayer perceptron (MLP) or a feedforward fully-connected NN with a single hidden layer;
- The transformations inside a NN are called *layers*. A linear transformation is a dense layer, as each output element depends on each input element. The dimensionality of the linear transformation output is also defined as the number of *neurons*;
- Non-linear transformations (e.g. σ in Figure 4.11) without learnable parameters, are called *activation functions*;
- The first layer is usually named *input layer*; the last layer, instead, is named *output layer*. All layers in between are named *hidden layers*;

In Figure 4.12, a graphical representation of a NN is displayed, in order to understand better all these principles described above.

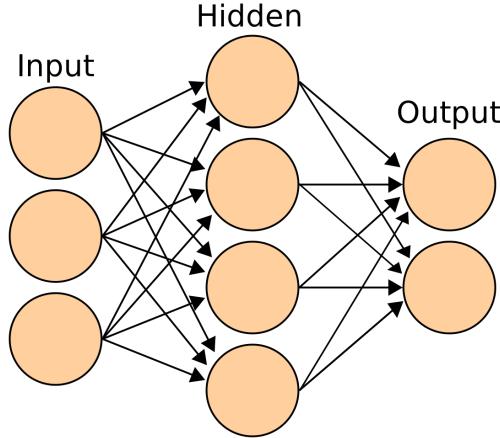


Figure 4.12: Neural Network basic working principle.

Activation function. To achieve a non-linearity to the model, a σ activation function is required. The most common are:

- $\text{ReLU}(x) = \max(0, x)$. It is zero for negative values of the argument and increases linearly for the positive values [128]. This activation function is the default one and it is recommended for use with most feed-forward NN. Applying this function to the output of a linear transformation, yields a non-linear transformation. However, the function remains close to linear, in the sense that it is mainly a piecewise linear function made with two linear pieces. Since ReLU units are nearly linear, they are able to preserve many of the properties that make linear models easy to optimise with gradient-based methods. They also preserve a good generalisation, possible with many of the linear models. One drawback to ReLU units is the impossibility to learn via gradient-based methods on examples where the respective activation is zero.
- $\text{LeakyReLU}(x) = \max(0, x) + \alpha \min(0, x)$ [138]. The α value is usually around 0.01. It is a piecewise linear function with a small slope for the negative argument values and slope equal to 1 for positive. It alleviates the problems caused by the zero activation of ReLU. However, due to some inconsistent gains [139], such activation did not manage to replace the standard ReLU.
- $\text{sigmoid}(x) = (1 + e^x)^{-1}$. Such non-linearity is usually used in the last layer of a network for binary classification, since its value lies in interval $(0, 1)$. The value tends to 0 as the argument approaches negative infinity and to 1 as the argument approaches positive infinity. Its logarithm is always defined and finite, which helps to alleviate the computational instability of the cross-entropy loss. Sigmoid is, however, a poor choice for the activations of the hidden layers, as it suffers from a dying gradient issue.

4.6.1 Optimisation

Optimising a NN is done via gradient techniques, as described in Section 4.3.1. Given a function $L(\theta)$, its minimum will be at a point where its gradient is zero. To arrive at such a point, iteratively follow the inverse gradient:

$$\theta^{\tau+1} = \theta^\tau - \eta \nabla L(\theta^{(\tau)}), \quad (4.26)$$

where τ is the iteration number and $\eta > 0$ is the learning rate. In a ML setting, to compute the loss $L(\theta)$ and its gradient as a function of the model parameters, model predictions must be computed for every example in the dataset. That can be prohibitively expensive in terms of computing power. To address the computational complexity, stochastic gradient descent (SGD) [140] is commonly used. Usually, it is organised as follows. The dataset is shuffled and arranged into a queue. For each iteration, only a subset (or *batch*) of examples is extracted from the queue and used to compute the gradients. A pass over the whole dataset is called an *epoch*. After an epoch, the dataset is reshuffled and processed again. With this technique, the gradient is computed only with the examples from the batch, for the estimation of the model parameters:

$$\theta^{\tau+1} = \theta^\tau - \eta \frac{1}{m} \sum_{i=1}^m \nabla_{\theta} L(f(x_i, \theta^{(\tau)}), y_i), \quad (4.27)$$

where $f(x_i, \theta)$ is the model prediction for the i -th example. Stochastic gradient descent follows noisy estimates of the true gradient. This slows down convergence, as shown in Figure 4.13, but eventually reaches a similar minimum with less computational power.

4.6.2 Dropout and regularisation

Neural Networks, during their learning process, exploit the loss function to establish how to set their parameters, i.e. *weights* and *bias*. To handle problems like overfitting, described in Section 4.3.2, a few techniques have been developed that, acting on the loss function, helps reduce the effects of overtraining a NN. These techniques are called *regularisation* techniques.

Generally, such techniques add a term, dependent on the weights, after the loss function:

$$\text{loss function} = \text{original loss function} + \lambda \cdot \text{regularisation function}. \quad (4.28)$$

Regularisation can be seen as a compromise between finding small weights and minimise the loss function. The λ parameter is often called *regularisation rate* and it is required to find the optimal "balance" of such compromise: when λ is small, then the minimisation of the loss function is dominant, otherwise the necessity of finding small weights is dominant. The most used regularisation is called *L2 regularisation* or *weight decay*: the squared sum of the weights is used as the regularisation function. Another technique, called *L1 regularisation*, is similar to the L2 but adopts the linear sum of weights, instead of the squared one.

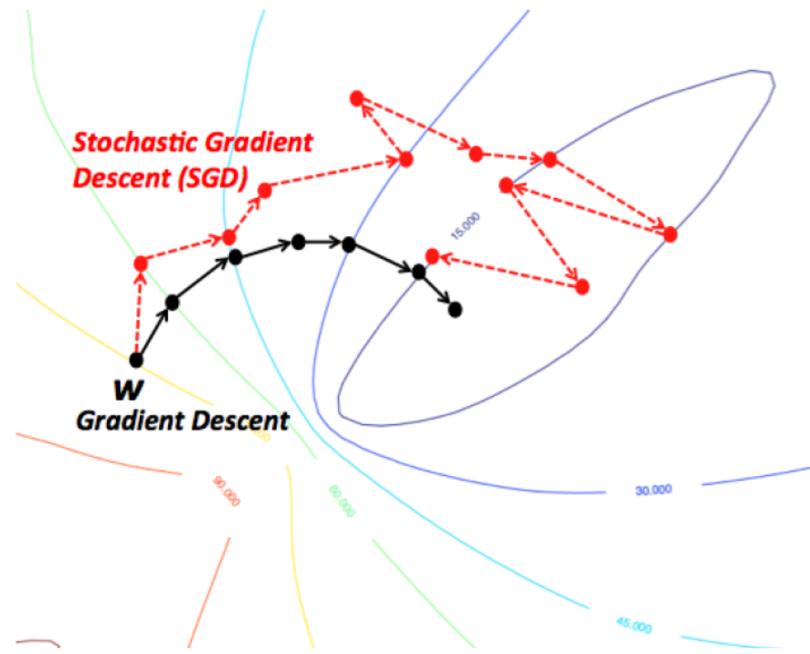


Figure 4.13: Optimising a learning algorithm with gradient descent (black line) and stochastic gradient descent (red dashed line).

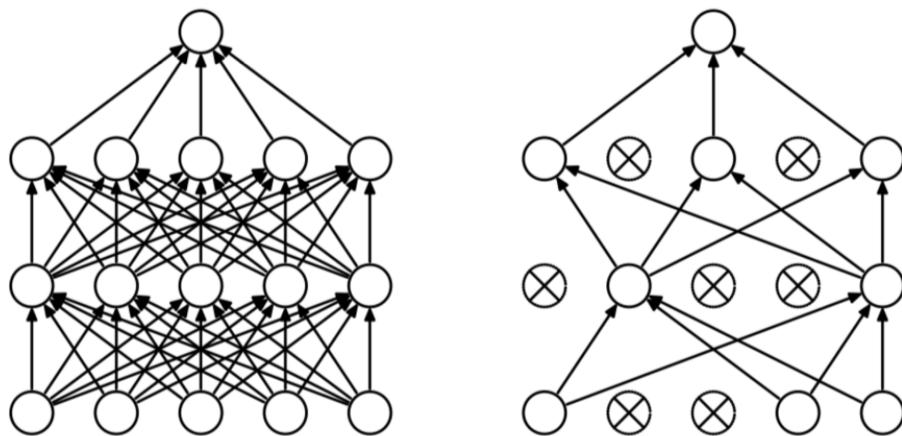


Figure 4.14: Graphical representation of the dropout technique. For each epoch, the network is different, considering different neurons.

A *dropout* technique, instead, is a different technique since it does not change the loss function of the network, but the network itself. Basically, the dropout removes some neurons in each hidden layer: for each epoch, some neurons are randomly discarded before starting the network training, as shown in Figure 4.14. The procedure is repeated on each epoch with different neurons and, in the end, an averaged result is used to tune the weight matrix. This procedure is performed only during training: during the real usage, the network is considered in its entirety.

4.7 IMPLEMENTING A NEURAL NETWORK

In these recent years, Machine and Deep Learning have become an important part of users' web experience but also a difficult task for developers creating new models. In order to simplify this process, several ML Frameworks have been created allowing developers to build ML models without getting into the complex underlying algorithms.

In this section, a few of these frameworks will be listed with an increasing attention to those involved in this thesis.

4.7.1 *TensorFlow*

TensorFlow is an open-source software library for dataflow programming across a range of tasks. It is a symbolic math library for numerical computation that uses data flow graphs and is also used for ML applications, such as NN. It was developed by the Google Brain team [141] for internal use and released under the Apache 2.0 open-source licence on November 2015 [142]. Currently it is one of the most utilised frameworks, due to its completeness and reliable libraries.

The architecture is flexible so to allow users to deploy computation to one or more CPUs (or GPUs) in a desktop, server, or even mobile device with a single API. In a not exhaustive list of features, one could mention:

- TF runs on Windows, Linux, and macOS, and also on mobile devices, including both iOS and Android;
- TF provides a very simple Python API, which is called TF.Learn (*tensorflow.contrib.learn*), compatible with Scikit-Learn;
- TF provides another simple API called TF-slim (*tensorflow.contrib.slim*) to simplify building, training, and evaluating NNs;
- TF's main Python API offers much more flexibility (at the cost of higher complexity) to create all sorts of computations, including any NN architecture you can think of;
- TF includes highly efficient C++ implementations of many ML operations, particularly those needed to build NNs. There is also a C++ API to define your own high-performance operations;
- TF provides several advanced optimization nodes to search for the parameters that minimize a cost function: TF automatically takes care of computing the gradients of the functions one defines, i.e. implements automatic differentiating (or autodiff);
- TF also comes with a great visualization tool called TensorBoard that allows you to browse through the computation graph, view learning curves, and more;
- Once a model is done with TF, computations can be deployed to one or more CPUs or GPUs, as needed.

An example code usage for the creation of a TensorFlow Gradient Descent model is listed above:

Listing 1: TensorFlow example code.

```
import tensorflow as tf
import numpy as np

n_epochs = 10000
learning_rate = 0.01
X = tf.constant(<MY_FEATURES>, dtype=tf.float32, name="X")
y = tf.constant(<MY_TARGET>, dtype=tf.float32, name="y")

theta = tf.Variable(tf.random_uniform([n+1,1],
-1.0,1.0, seed=42), name="theta")

y_pred = tf.matmul(X, theta, name="predictions")
error = y_pred - y
mse = tf.reduce_mean(tf.square(error), name="mse")
gradients = 2/m * tf.matmul(tf.transpose(X), error)

training_op = tf.assign(
theta, theta - learning_rate * gradients)

init = tf.global_variables_initializer()

with tf.Session() as sess:
    sess.run(init)

    for epoch in range(n_epochs):
        if epoch % 100 == 0:
            print("Epoch", epoch, "MSE =", mse.eval())
        sess.run(training_op)

    best_theta = theta.eval()
```

4.7.2 Scikit-learn

Scikit-learn was initially developed by David Cournapeau as a Google summer of code project in 2007. It provides a range of supervised and unsupervised algorithms via a consistent interface in Python and is distributed under many Linux distributions, for academic and commercial use [143].

With few lines of code, scikit-learn provides lots of algorithms for classification and regression problem e.g. SVM, Linear Regressor, BDT, NN, etc...

The following example shows how to load a dummy dataset containing Iris flowers characteristics:

Listing 2: Scikit-learn example code.

```
# Sample Decision Tree Classifier
from sklearn import datasets
from sklearn import metrics
from sklearn.tree import DecisionTreeClassifier
# load the iris datasets
dataset = datasets.load_iris()
# fit a CART model to the data
model = DecisionTreeClassifier()
model.fit(dataset.data, dataset.target)
print(model)
# make predictions
expected = dataset.target
predicted = model.predict(dataset.data)
# summarize the fit of the model
print(metrics.classification_report(expected, predicted))
print(metrics.confusion_matrix(expected, predicted))
```

In the example shown above, the model is created in Line 8 and trained in the next line. At last, predictions are made in Line 13 and the scorers of such predictions in the last two lines.

4.7.3 Keras

Keras is a high-level NN API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano [144]. It was developed with a focus on enabling fast experimentation. It is a framework with a simple interface, written in python, minimizing the number of user actions required for common use cases and providing clear and actionable feedback upon user error.

Keras is one of the most favourite frameworks among developers, as shown in the number of mentions in scientific papers uploaded to arXiv.org (Figure 4.15). The core data structure of Keras is a model, a way to organize layers. The simplest type of model is the *Sequential* model, a linear stack of layers.

Here is an example of sequential model:

Listing 3: Create a sequential model.

```
from keras.models import Sequential
model = Sequential()
```

Stacking NN layers is easily achieved by:

Listing 4: Stack layers in a Neural Network.

```
from keras.layers import Dense
```

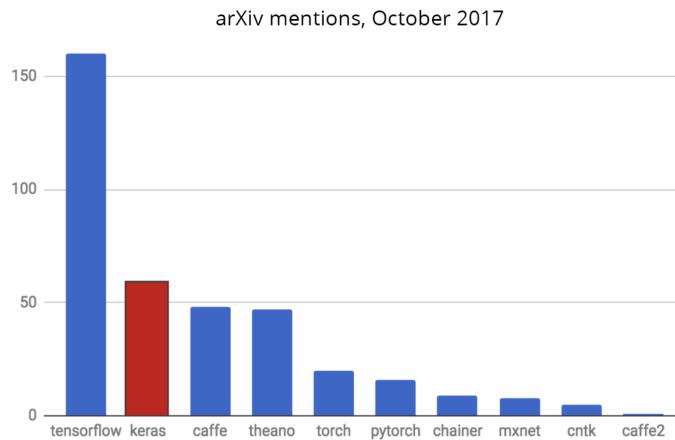


Figure 4.15: Keras popularity in arXiv mentions.

```
model.add(Dense(units=64, activation="relu", input_dim=100))
model.add(Dense(units=10, activation="softmax"))
```

In Line 3 a hidden layer is created: *units* corresponds to the number of nodes (or neuron) for that specific layers, *activation* is the activation function (explained in Section 3.3) required. The first hidden layer requires also the dimension of input layer which is given in the *input_dim* parameter.

Once the model looks good, the configuration of its learning process is done by:

Listing 5: Configure learning.

```
model.compile(loss='mean_squared_error', optimizer='sgd')
```

This method takes as input the *loss* function (mean squared for this example but many more can be used) and the configuration of the *optimizer* (in this case the Stochastic Gradient Descent, introduced in Section 3.2).

Training the model in batches is than performed:

Listing 6: Training iteration in batches.

```
model.fit(x_train, y_train, epochs=5, batch_size=32)
```

At last, the predictions on new data with a simple line:

Listing 7: caption=Predictions on new data.

```
classes = model.predict(x_test, batch_size=128)
```

This basic example shows the simplicity in the construction of a primitive NN. Using a more complex structure, the analysis presented in this thesis is mainly done in Keras (using TensorFlow engine as backend).

4.8 MACHINE LEARNING IN HIGH-ENERGY PHYSICS

Standard Model has had a resounding success, including almost all subnuclear particle physics phenomena in nature. However, each problem of interest must be always accompanied by some assumptions and simplifications, including the results of proton-proton collisions at the LHC. In particular, at the LHC experiments such complexity must be multiplied by the scale of a building-sized particle detector. It is also important to mention that physical considerations and studies only allow to solve a direct problem: how does a process X look in a Y detector. However, in order to draw conclusions from any experiment, the inverse problem must be solved: given some readout from a detector Y, what is occurring in terms of a physical process X? Most of the utility of ML for particle physics comes from being able to solve this inverse problem [145, 146].

Compared to traditional, expert-designed algorithms, ML has two advantages: first of all, it is able to deliver a better quality; secondly, a learning algorithm saves effort, by replacing an HEP-specific manual algorithm solution with an application of a general method, generally "stolen" from an AI research and adopted also in different fields of study. Such quality and convenience, unfortunately, come at a cost. Most of the ML algorithms are in fact considered "black boxes". The main problem is not simply the lack of human comprehension, but rather the specific mathematical setting: the training data must be distributed the same as the data to which the algorithm is to be applied. This is almost never the case: ML methods, in fact, provide little formal guarantees of behaviour in a case of such data shift. In many non-scientific applications, this might not matter that much. But scientific applications require extreme rigour. There are of course methods that are used to verify the data analysis techniques, but an expert judgement on a case-by-case basis is pretty much always required.

In general, HEP community approaches validation of ML methods in the same way and case-by-case physical considerations are used at every stage of the problem solving to arrive at an acceptable result:

1. Training data selection. The data must cover the entire desired phase space and all the physical processes that we want to learn;
2. Features selection. Only the features that correspond to the physics involved are used;
3. Validating on different data samples. For an algorithm trained on simulated data, its performance must be performed on a different sample for calibration;
4. Manual inspection of physically-meaningful distributions. Training features must be selected according to their physical distribution: in case of correlations, for instance, a possible bias may be introduced and therefore affect negatively the training procedure.

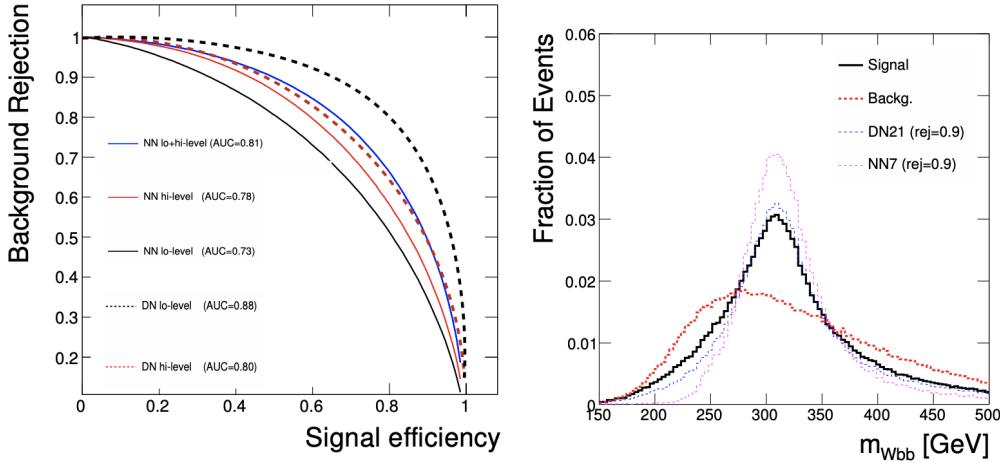


Figure 4.16: On the left, deep networks (DN) performance in signal-background classification compared to shallow networks (NN) with a variety of low- and high-level features [152]. On the right, a comparison of the distributions of invariant mass of events selected by a deep network (DN21) using only object momentum to a shallow network (NN7) that has been trained using this feature, at equivalent background rejection.

4.8.1 Separating Signal and Background: the Event Selection

Selecting events that contain interesting processes is a fundamental requirement of high-energy physics experiments and probably is the most established area of application of ML in HEP. Most analyses consist of measuring the fraction of events that contain a specific decay channel. The usual way of doing that relies on building an event selection algorithm, estimating its efficiency selecting signal and rejecting background and measuring the count of events passing it. The traditional way of selection is the so-called *cut-based* selection, building a decision tree manually and using either physical considerations, Monte-Carlo simulations, or both. This procedure can, of course, be fully automatised via ML algorithms, both in the final statistical analysis [147] or at the initial trigger decision [148]. These ML tools have found high-profile application in single t quark searches [149], early Higgs boson searches [150], and the Higgs boson discovery itself [151].

In the past few years, several studies have also demonstrated that the traditional shallow networks (with a few hidden layers) based on physics-inspired engineered ("high-level") features are outperformed by deep networks (with multiple hidden layers) based on the higher-dimensional features which receive less pre-processing ("lower-level") features. This confirms also the idea that feature engineering, applying physics knowledge to construct high-level features, is often overestimated. An early study [152], in particular, compared the performance of shallow and deep networks in distinguishing a cascading decay of new exotic Higgs bosons from the dominant background. This study, using a structured data set in which a large set of basic low-level features (object four-momenta) were reduced to a smaller set of physics-inspired high-level engineered features, showed that the performance of the deeper network with lower-level data are greater than the shallow network with higher-level physical features, as reported in Figure 4.16.

4.8.2 Event Reconstruction

Reconstruction is a process by which the raw detector readout is transformed into physically-meaningful objects, namely particle tracks, particle types and vertices. For ATLAS, CMS and LHCb experiments, this is separated into three distinct operations. First, the charged tracks are reconstructed using the input from the tracker. Second, the information from the particle identification subsystems is used to assign the particle types to the charged tracks. Third, the calorimeters are used to find some of the neutral particles that escaped the magnetic field. This operation can be considered as the inverse of simulation: for the latter, the response is computed given the particles, while in the former, the particles that have caused a given detector response are found.

ML is a natural way to approach the problem, especially considering its purely algorithmic nature. The most straightforward way is to simulate something, use the detector response as features for a ML algorithm, and the Monte-Carlo truth as the labels. Of course there are a lot of problems arising, mainly of algorithmic nature: an high-energy physics event is an highly complex and structured object, and standard ML algorithms are not well suited for this kind of operations. Each reconstruction therefore must be tackled by system experts, depending on the object type.

Tracking. Track-reconstruction algorithms are among the most central processing unit (CPU) and data intensive of all low-level reconstruction tasks. The initial stage of track reconstruction involves finding the hits, or the points where some charge is deposited on a sensing element in the detector. In the case of the pixel sensors that form the innermost layer of the detector, neighbouring hits are clustered into pixel clusters, which then form track seeds. These seeds are used to form a starting point for a Kalman filter, which extends the seeds into full tracks. ML has proven useful in several aspects of track reconstruction. In cases where multiple tracks pass through the same pixel cluster, ATLAS relies on NNs to return a measurement for each track rather than assigning each to the cluster center [153, 154].

Thanks to these algorithms and careful tuning, track reconstruction is now nearly 100% efficient and spuriously reconstructed tracks are rare, meaning that the clustering aspect of tracking is largely solved. However, reducing the CPU overhead remains a significant problem, especially within high-level trigger farms. In ATLAS and CMS, these cluster are made of $\approx 10^4$ processors that reconstruct $\approx 10^5$ events per second [155]. To keep tracking CPU costs manageable, the experiments reconstruct tracks only in limited regions of the detector. These regions are selected on the basis of their proximity to muons or to calorimeter energy deposits that are consistent with relatively rare physical signatures like leptons or high- p_T jets.

Jet tagging. ML has been applied to a wide range of jet classification problems, to identify jets from heavy (c , b , t) or light (u , d , s) quarks, gluons, and W , Z , and H bosons. Traditionally these classification problems have been grouped into flavour

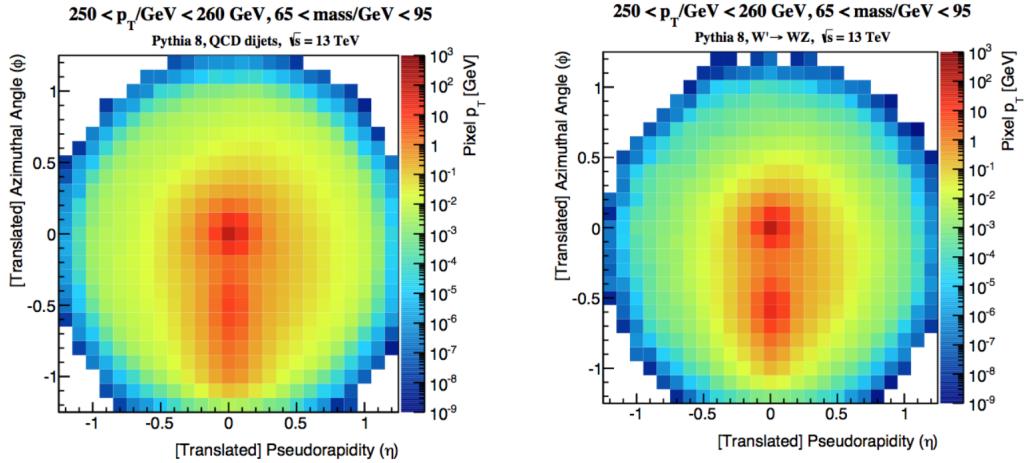


Figure 4.17: Example jet image inputs from the jet substructure classification problem. The background jets (left image) are characterized by a large central core of deposited energy caused by a single hard hadronic parton, while the signal jets (right) tend to have a subtle secondary deposition due to the two-prong hadronic decay of a high- p_T vector boson. Use of image-analysis techniques such as convolutional neural networks allow for powerful analysis of this high-dimensional input data.

tagging, which discriminates between b, c, and light quarks, jet substructure tagging, which discriminates between jets from W, Z, t and H, and quark-gluon tagging. In 2014, a study [156] recognized that the projective tower structure of calorimeters present in nearly all modern HEP detectors was similar to the pixels of an image (Figure 4.17). This representation of the data allowed physicists to leverage the advances in image classification such as Convolutional Neural Networks (CNN). While the image-based approach has been successful, the actual detector geometry is not perfectly regular; thus, some preprocessing is required to represent the jet as an image. Both ATLAS and CMS have since commissioned flavour-tagging NNs that rely on individual tracks or, in the CMS case, particle-flow candidates. CMS's DEEPJET [157], for instance, adopts NNs embedding first each flow candidate (see Section 3.6) with a shared transformation, then it combines the high-level variable candidates in a single dense network. A more detailed description of this algorithm will also be provided in Section 5.5.4.

Fast inference on designed hardware. With the advent of the High-Luminosity phase of the LHC (HL-LHC), explained in more detail in Appendix A, the instantaneous luminosity of the Large Hadron Collider at CERN is expected to increase up to $\approx 7.5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Therefore, new strategies for data acquisition and processing will be necessary, in preparation for the higher number of signals produced inside the detectors. Following the rapid rise of ML through deep learning algorithms, the investigation of processing technologies and strategies to accelerate deep learning and inference is well underway. In particular, a lot of effort has been made to convert such learning models into specific firmware code [158], able to run on fast hardware like Field Programmable Gate Arrays (FPGA). In CMS, new reconstruction algorithms, aiming for an improved performance, are being developed: concerning the tracking of muons in the muon trigger, for instance,

one of the figures that is being improved is the accuracy of the transverse momentum (p_T) measurement [159]. The implementation of ML models on FPGAs is beneficial for two main reasons: first, such models are able to predict with an improved precision the p_T , since they exploit much more information collected by the detector; second, FPGA hardware promises a smaller latency with respect to traditional inference algorithms running on CPU, which is an important aspect for a trigger system. Another study, from the ATLAS collaboration, concerns instead the implementation of a CNN in a FPGA for the identification of significant energy deposits in the Liquid-Argon (LAr) calorimeters [160]: in this way, the fast inference enabled by the hardware solution might be capable of handling the enormous amount of signals, coming from the HL-LHC increased pileup.

High- p_T muons. Another application of ML techniques in physical object reconstruction, regards the high- p_T muon reconstruction. In the current phase of operations, such measurement is performed by the *TuneP* algorithm which, based on a small set of track-quality parameters, chooses the best reconstruction among different refit techniques: *inner-track* with tracker-only information (used mainly at low momentum), *tracker-plus-first-muon-station* (TPFMS) using the inner tracker and the innermost muon station containing hits, *picky* designed to handle cases where electromagnetic showers generate a high multiplicity of hits in the muon chambers, and *dynamic-truncation* (DYT) for cases when radiative energy losses cause significant bending of the muon trajectory. The adoption of learning techniques tree-based and NNs will be able to identify the optimal refit, choosing among the existing ones, based on a larger set of input variables than the ones used during the *TuneP choice*. Training will be achieved by comparing the resolution in terms of q/p_T relative residual with respect to the generated information coming from the MC simulations. Preliminary results [161] are showing an important reduction of the q/p_T relative residual distribution tails (outliers) of $\sim 60\%$ in the barrel region and $\sim 27\%$ in the endcap region, with the adoption of BDTs implemented using the XGBoost library. Results are also indicating that further improvements are achievable if a larger statistics could be available for the training process.

4.8.3 Fast Simulation

Simulation is the most intensive CPU operation in High-Energy Physics. A fast simulation, therefore, is really valuable since the full simulators, which faithfully describe the low-level interactions of particles with matter, are very computationally intensive and consume a significant fraction of the computing budgets of current experimental collaborations. A promising approach is based on Generative Adversarial Networks (GANs). The training of such generative model G is accomplished through the competition with an adversary network A. While G generates simulated samples, the adversarial network A has the task to determine whether a given sample is from the generative model or from the full simulator. The two networks are put against each other: A attempts to identify differences between the traditional samples and those generated by G, while G attempts to fool A into accepting

its events, and in doing so learns to mimic the original sample generation. The stability of such training arrangement, however, can be difficult to achieve and expert knowledge is usually required to construct an effective network. Currently, some GAN approaches are used in the simulation of the electromagnetic showers in a calorimeter [162], reporting computational speed-ups while achieving a reasonable energy deposition model. Similar results are also applied in simulating jet images [163].

4.8.4 Monitoring and Data Quality

The LHC systems and detectors are quite complex machines and, therefore, are equipped with monitoring systems. They are constantly checking that every parameter on all level is in an acceptable range: from voltages to reconstructed masses from known decays. A big challenge for the monitoring system is that it must be able to recognise legitimate changes of data taking conditions from equipment malfunctions. If an observed variable distribution are different from the corresponding reference, the operator investigates the discrepancy; if it has not been flagged as addressed before, the incident is recorded and the relevant expert is contacted. The system is also equipped with automatic alarms, if some large discrepancy is detected. False alarms occur when the reference is not updated in time and the discrepancy is due to a legitimate change in data taking conditions. In general, ML algorithms can be exploited to monitor the detector's conditions and also predict eventual future anomalies. This is also called *anomaly detection* and it is used widely in data science applications [164, 165]. Some efforts has been already made for the CMS Data Quality Monitoring system, using unsupervised learning with dimensionality reduction and clustering techniques [166].

5

THE MSSM HIGGS BOSONS DECAYING TO $\mu^+ \mu^-$ AND ITS MAIN PHYSICAL OBJECTS

Stable particles from proton-proton collisions are identified combining the detector information with the Particle Flow (PF) algorithm. The combination is used to build physics objects and observables to be employed at the analysis level. Such objects are then used to constitute the inputs of the physical analyses: in particular, the work presented in this thesis is focused on the study of MSSM neutral Higgs bosons decaying to $\mu^+ \mu^-$.

5.1 INTRODUCTION

In Chapter 1, the Standard Model of Particle Physics has been described (Section 1.1) with a particular focus on the Higgs mechanism and relative field/particle H (Section 1.3). The discovery of the scalar Higgs boson at the LHC in 2012 [23, 25] (Section 1.4.2), with a mass of $125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV consistent with the SM predictions [13–15], finally completed the description of the fundamental interactions (excluding gravity) and provided a working mechanism for the creation of the masses of the fundamental particles. However, several arguments are in favour of theories that extend the SM (Section 1.5), and which predict extended Higgs sectors, described in more detail in Chapter 2.

One extension of the SM is the Minimal Supersymmetric Standard Model (MSSM), explored in Section 2.2. In the MSSM the Higgs sector consists of two Higgs doublets, in the so called 2HDM described in Section 2.2.1, one of which couples to up-type fermions and the other to down-type fermions. This results in five physical Higgs particles: two charged bosons H^\pm , two neutral scalars h and H , and one pseudoscalar A ¹.

At tree level, the Higgs sector described by the MSSM can be described by only two parameters, chosen to be m_A , the mass of the neutral pseudoscalar A , and $\tan \beta$, defined as the ratio of the vacuum expectation values of the neural components of the Higgs doublets (both are described in more detail in Section 2.2.2). The other four bosons can be then expressed as a function of these parameters. Beyond the tree level, the MSSM Higgs sector depends on other additional parameters, which can enter only via higher order corrections in the perturbation theory. These parameters are usually fixed to values motivated by experimental constraints and theoretical assumptions. For a fixed combination of such parameters, the definition of a benchmark scenario is made (described in Section 2.3). The

¹ From now on, the notation ϕ will be used to represent any of the three neutral particles, unless explicitly mentioned.

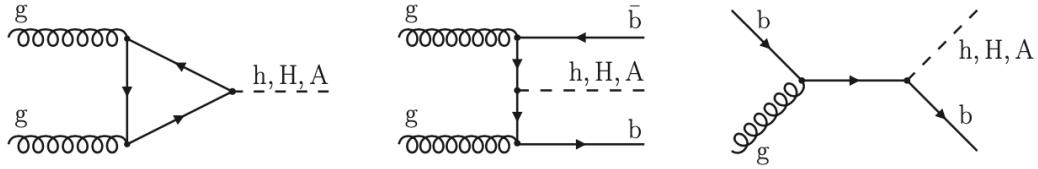


Figure 5.1: Leading order Feynman diagrams for the MSSM production of the neutral Higgs bosons: gluon fusion, on the left, and b-associated production, on the center and right.

relevant scenarios are those consistent with a mass of the neutral boson of 125 GeV for the majority of the probed $(m_A, \tan \beta)$ space [167], and not ruled out by other existing measurements. In particular, the m_h^{mod+} scenario (Section 2.3.2) [168] constrains the mass of the h boson in the interval 122 to 128 GeV for a wide h range of $(m_A, \tan \beta)$ values. The phenomenological hMSSM model (Section 2.3.3) [57, 169, 170], instead, also incorporates the observed neutral boson at 125 GeV, interpreting it as the h boson. When the intrinsic width of the signal is smaller than the experimental detector resolution, typically at low values of $\tan \beta$, a *model independent* scenario can be defined: in this scenario, no assumption is made on the cross section, mass and width of any MSSM ϕ neutral boson, which is sought as a single resonance with mass m_ϕ .

The work described in this thesis, following the analysis already published by the CMS collaboration [171], is inserted in the model independent search for the MSSM neutral Higgs bosons in the dimuon final state, in pp collisions at center-of-mass energy of 13 TeV in the mass range between 130 and 1500 GeV, with the 2016 Run 2 data. Similar searches have been already performed by the CMS collaboration using data collected at 7 and 8 TeV [172]. Searches of the neutral Higgs boson in the framework of the MSSM different scenarios have also been performed by the CMS experiment in the τ -pair [67, 173] and $b\bar{b}$ [69] final states. Searches performed by the ATLAS experiment can be found in [64].

5.2 PRODUCTION MECHANISMS FOR NEUTRAL MSSM HIGGS BOSONS

At LHC centre-of-mass energies, the dominant production mechanisms for neutral MSSM Higgs bosons is the gluon fusion (or gg, shown in Section 1.4.1), where the Higgs is produced by a virtual loop of b or t quarks, or in association with real b-quarks (b-associated production or bb). The Feynman diagrams for these production mechanism are shown in Figure 5.1. The gluon fusion mechanism is more relevant for small and medium values of $\tan \beta$, whereas at large $\tan \beta$ (> 30) an enhanced coupling of the Higgs with down-type fermions is predicted, resulting in b-associated production becoming increasingly dominant.

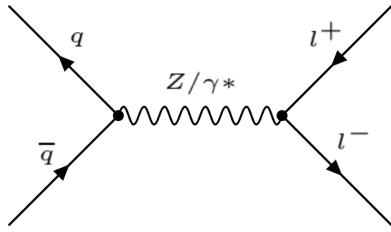


Figure 5.2: Leading order Feynman diagram for the Drell-Yan process, where $l^\pm = \mu^\pm$.

5.3 THE DIMUON FINAL STATE

Since the coupling of the neutral Higgs to muons is enhanced in this region of the parameter space, the dimuon final state is considered as an interesting probe of the MSSM. Although the cross section times the branching ratio for the dimuon final state is a factor 10^3 smaller than the corresponding $\tau^+\tau^-$ final state, the muon final state can be fully reconstructed, and the invariant mass can be measured by exploiting the excellent muon momentum resolution of the CMS detector (explored in detail in Chapter 3).

The common experimental signature of the two processes is a pair of oppositely charged muons with high transverse momentum (p_T) and a small imbalance of the vectorial sum of the total p_T in the event. The b-associated process is characterised by the presence of additional jets originating from b quarks, whereas the events with only jets from light quarks or gluons are sensitive to the gluon fusion production mechanism. The presence of a signal would be characterised by an excess of events over the background in the dimuon invariant mass, which would then correspond to the mass of the ϕ boson.

5.3.1 Background sources

The main source of background for ϕ bosons is the Drell-Yan process $Z/\gamma^* \rightarrow \mu^+\mu^-$, shown in Figure 5.2. When the Z/γ^* is produced in association with a quark anti-quark pair, the process is an irreducible background for b-associated production. Another relevant background source comes from opposite-sign dimuon pairs produced in semileptonic decay of the top quark in $t\bar{t}$ and single top events, shown respectively in Figure 5.3 and 5.4. Other less relevant contribution comes from the diboson production processes $W^\pm W^\mp$, $W^\pm Z$ and ZZ , shown in Figure 5.5. These events contribute very little to the dimuon invariant mass for masses larger than 130 GeV, where the Higgs signal is searched.

In the following sections, the main physical objects needed to classify the events and the raw detector information used to reconstruct them will be described.

5.4 DATA AND MONTE CARLO SAMPLES

The analysis adopts a data driven approach for the estimation of the background and relies on the simulation only for the signal modelling. The simulated Monte Carlo (MC) samples

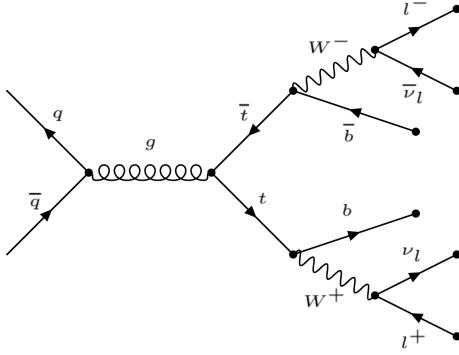


Figure 5.3: Dominant Feynman diagram for the $t\bar{t}$ process decaying semileptonically into the dimuon final state, where $l^\pm = \mu^\pm$.

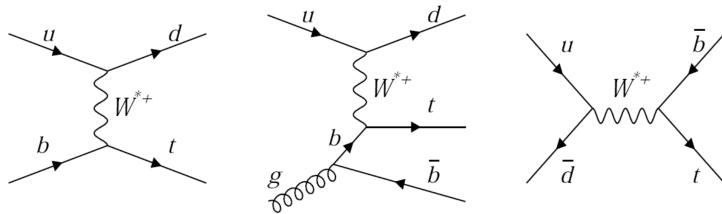


Figure 5.4: Feynman diagrams for electroweak single top quark production: on the left, the leading-order t-channel process; on the middle, the next-to-leading order t-channel process; on the right, the leading-order s-channel process. Reproduced from [174]. The top then decays semileptonically to leptons, as shown in Figure 5.3.

are used to optimise the event selection in order to maximise the significance of the results. The MC samples are generated for collisions at $\sqrt{s} = 13$ TeV, with the pileup conditions expected for the Run 2 data taking of ≈ 30 collisions per bunch crossing, with a spacing between bunches of 25 ns. For each sample, both data and background/signal MC, the latest offline reprocessing has been used.

5.4.1 Run 2 Datasets

The data sets used for this analysis are listed in Table 5.1. For the purpose of this thesis, however, only the 2016 year has been considered. For each set, the corresponding run range and its integrated luminosity is also reported. For 2016, data are split in two different reconstruction versions: the letters from B to F are reconstructed including the HIP mitigation, in short called "HIPM" or "APV", while the letters from G to H, also known as "no-HIPM" or "no-APV", are reconstructed with the default track reconstruction [175]. The data format used is the NanoAOD format [176], containing the essential per-event information required by the analysis.

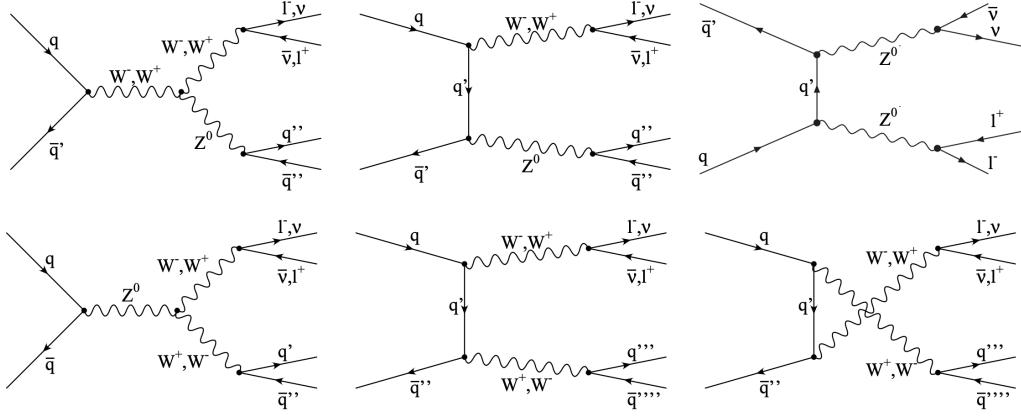


Figure 5.5: Leading order Feynman diagrams for the $W^\pm W^\mp$, $W^\pm Z$ and ZZ diboson processes, where $l^\pm = \mu^\pm$.

Dataset	Integrated lumi (fb^{-1})
2016	
/SingleMuon/Run2016B-HIPM	5.751
/SingleMuon/Run2016B-HIPM	
/SingleMuon/Run2016C-HIPM	2.576
/SingleMuon/Run2016D-HIPM	4.242
/SingleMuon/Run2016E-HIPM	4.025
/SingleMuon/Run2016F-HIPM	3.106
/SingleMuon/Run2016F	
/SingleMuon/Run2016G	7.576
/SingleMuon/Run2016H	8.651
Total	35.927
Luminosity mask: Cert_13TeV_Legacy2016_Collisions16_JSON.txt	

Table 5.1: Single muon datasets collected during proton-proton collisions at $\sqrt{s} = 13$ TeV by the CMS experiment.

5.4.2 MC background samples

As previously mentioned in Section 5.3.1, the main sources of background are the Drell-Yan process, opposite-sign dimuon pairs produced in the decay of $t\bar{t}$ and single top events, and diboson production processes.

Single top processes, excluding the s-channel, top pair processes and the diboson processes are generated with **POWHEG** box [177], while the Drell-Yan and the single top s-channel processes are generated with **MADGRAPH_v5** [178] at next-to leading order (NLO) with **aMC@NLO** [179]. Spin effects in some processes are also simulated using **MADSPIN**. The parton-showering (PS) and hadronisation are finally modelled by the **PYTHIA8** generator [180], using the Tune CP5 [181].

The event samples used for this analysis are listed in Table 5.2, with their corresponding

Background MC	Generator + PS	Number of Events	Cross section (pb)
Drell-Yan			
DY($\ell\ell$)+Jets	MADGRAPH+PYTHIA	202178343 (no-APV)	6080
		187634918 (APV)	
Single Top			
single top, tW antitop	POWHEG+PYTHIA	3654510 (no-APV)	35.85
		3176485 (APV)	
single top, tW top	POWHEG+PYTHIA	3361253 (no-APV)	35.85
		3294673 (APV)	
single top, t-channel top	POWHEG+PYTHIA	46422500 (no-APV)	136.02
		46969400 (APV)	
single top, t-channel antitop	POWHEG+PYTHIA	22799100 (no-APV)	80.95
		23036400 (APV)	
single top, s-channel, leptonic decays	MADGRAPH+PYTHIA	6275000 (no-APV)	3.36
		5518000 (APV)	
DiBoson			
WW($\ell\nu\ell\nu$)	POWHEG+PYTHIA	2900000 (no-APV)	12.2
		3018000 (APV)	
WZ($\ell\nu\ell\ell$)	POWHEG+PYTHIA	904000 (no-APV)	4.43
		1080000 (APV)	
ZZ($\ell\ell\nu\nu$)	POWHEG+PYTHIA	15928000 (no-APV)	0.564
		16862000 (APV)	
ZZ($\ell\ell\ell\ell$)	POWHEG+PYTHIA	18155696 (no-APV)	13.74
		19622315 (APV)	
TopPair			
$t\bar{t}(\ell\nu\ell\nu)$	POWHEG+PYTHIA	43546000 (no-APV)	88.34
		37505000 (APV)	

Table 5.2: Summary of the background MC samples by process. The generator and PS simulator used are listed in the second column of the table. The number of events and cross sections used in the analysis are specified in the third and fourth column, respectively.

cross sections and number of events. In the same way as the data, only 2016 MC simulated background has been considered for this thesis, divided between APV and no-APV. The simulated background events are used only to optimise the selection criteria, and not to compute the exclusion limits. The background properties, their amount and shape, are entirely measured from the data.

To increase the statistics in the high invariant mass region, to allow a better neural network training, an additional set of MC background samples has been considered, in a binned fashion. The binned event samples are listed in Table 5.3, also with their corresponding cross sections and number of events.

Background MC	Generator + PS	Number of Events	Cross section (pb)
Drell-Yan mass binned			
$Z(\mu\mu)$, $50 < m_{\mu\mu} < 120$ GeV	POWHEG+PYTHIA	2955000 (no-APV)	2116
		2996000 (APV)	
$Z(\mu\mu)$, $120 < m_{\mu\mu} < 200$ GeV	POWHEG+PYTHIA	100000 (no-APV)	20.58
		100000 (APV)	
$Z(\mu\mu)$, $200 < m_{\mu\mu} < 400$ GeV	POWHEG+PYTHIA	100000 (no-APV)	2.89
		100000 (APV)	
$Z(\mu\mu)$, $400 < m_{\mu\mu} < 800$ GeV	POWHEG+PYTHIA	98000 (no-APV)	0.2515
		98000 (APV)	
$Z(\mu\mu)$, $800 < m_{\mu\mu} < 1400$ GeV	POWHEG+PYTHIA	100000 (no-APV)	0.01709
		100000 (APV)	
$Z(\mu\mu)$, $1400 < m_{\mu\mu} < 2300$ GeV	POWHEG+PYTHIA	100000 (no-APV)	0.00137
		100000 (APV)	
TopPair mass binned			
$t\bar{t}$, $700 < m_{t\bar{t}} < 1000$ GeV	POWHEG+PYTHIA	33502717 (no-APV)	66.85
		23727154 (APV)	
$t\bar{t}$, $m_{t\bar{t}} > 1000$ GeV	POWHEG+PYTHIA	23673116 (no-APV)	16.42
		23208356 (APV)	

Table 5.3: Summary of the additional background MC samples, binned in mass, by process. The generator and PS simulator used are listed in the second column of the table. The number of events and cross sections used in the analysis are specified in the third and fourth column, respectively

5.4.3 MC signal samples

The signal samples have been produced according to the values of m_A and $\tan \beta$ reported in Table 5.4. Only 2016 MC simulated signal has been considered for this thesis. The A and H bosons produced via gluon fusion and in association to b-quarks have been generated for each $(m_A, \tan \beta)$ point². The h boson samples have been instead generated only for one arbitrary $(m_A, \tan \beta)$ value, since its physical properties (mass, width and kinematic) are predicted to remain constant. At generator level, a tail of events in a off-mass shell is generated: this fraction of events, in the low-mass part of the distribution, increases with m_A and $\tan \beta$. This feature is unphysical, and it is related to the presence of an almost flat Breit-Wigner description far from the mass peak. When such Breit-Wigner distribution is convoluted with the large gluon luminosity at low x values, it produces a large excess of off mass-shell events. This feature is enhanced at large m_A and $\tan \beta$ because the intrinsic width of the Higgs boson increases as a function of these two quantities.

The events at very low mass values would affect the determination of the signal efficiency (described in Section 7.1), since they fall very far from the mass region where the presence

² At the time of writing this thesis, the gluon fusion samples for the A boson are still not available for the analysis due to an intense workload on the central CMS MC production system. Although they will be ready in time for the final analysis publication, in the following work such samples will be omitted.

of the signal is searched. Moreover, the signal cross section refers to the on-mass-shell Higgs bosons. In order to overcome the problem, the Higgs boson events are generated with a mass restricted to be within $\pm 3\Gamma$ of the nominal Higgs boson mass, where Γ is the intrinsic width. The values of Γ strongly depend on m_A and $\tan \beta$, being, for example, $\Gamma = 0.2(2.7)\%$ of the nominal Higgs boson mass at $m_A = 150(550)$ GeV and $\tan \beta = 10(40)$. The samples have been simulated using the POWHEG2 box generator [177] at NLO, and PYTHIA8 generator [180] with Tune CP5 to model the hadronisation and decay modes. For 2016, as already mentioned for the MC backgrounds and data, the samples have been split in two different campaigns: one with the APV reconstruction and the other with the classic one.

For the gluon fusion production mechanism, the values of the Higgs boson masses, widths, and the Yukawa couplings have been computed automatically using the POWHEG-BOX/gg_H_MSSM [182], for the scalar and pseudoscalar Higgs production in gluon fusion in the MSSM theory, based on the four flavour NLO QCD calculations [183, 184]. On the other hand, for the b-associated production mechanism, the POWHEG-BOX/bbH [185] SM Higgs production in association with bottom quarks has been used, correcting the Higgs mass and width, for each $(m_A, \tan \beta)$ point, using FeynHiggs [186] by the LHC Higgs Cross Section Working Group [26, 187], and using the five flavour next-to NLO (NNLO) QCD calculations [188].

The generator cards [189] and central productions of the signal MC samples [190–193], have been handled by myself during my doctoral experience as Monte Carlo contact for the CMS Higgs HExtended Physics Analysis Group.

5.5 OBJECT RECONSTRUCTION

The experimental signature of the MSSM Higgs boson ϕ considered in this analysis is a pair of opposite-charged, isolated muon tracks with high transverse momentum. The invariant mass of the muon pair corresponds directly to the mass of the ϕ boson, within the experimental resolution. Moreover the event is characterised by a small transverse missing energy. If the ϕ boson is produced in association with a $b\bar{b}$ pair, the additional presence of at least one b-tagged jet is expected. Before the off-line selection, the events are required to fire the single muon High Level Trigger (HLT). The Particle Flow (PF) technique (explained in Section 3.6), that combines the information from all the subdetectors, is used to reconstruct the full event. The events are required to fulfil the following general criteria:

- the two muon tracks must be associated to the same primary vertex (PV);
- each muon track must fulfil track quality and isolation criteria;
- the missing transverse energy (E_T^{miss}) has to be small;
- jet b tagging is applied to discriminate between GF and BA production.

m_A (GeV)	$\tan \beta$ values	Num. of events 2016
Process: ggA, ggH, bbA, bbH		
130	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
150	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
170	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
200	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
250	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
300	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
350	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
400	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
450	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
500	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
600	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
700	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
800	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
1000	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
1200	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
1500	2,5,10,15,20,25,30,40,50,60	4600 + 5400 (APV)
Process: ggh, bbh		
125	30	4600 + 5400 (APV)

Table 5.4: Grid of m_A and $\tan \beta$ points used to simulate the signal samples. For each point, the pseudoscalar A and the H bosons produced via gluon fusion (gg) and in association to b-quarks (bb) have been generated.

A description of the above-mentioned objects explored in this analysis is given in the following subsections [10, 11], while the event selection based on such objects will be explored in more detail on Chapter 6.

5.5.1 Trigger requirements

The trigger, described in more detail in Section 3.5, is designed and optimised to reach a signal efficiency close to 100%. The offline events are collected with some dedicated collection of HLT triggers which are designed specifically to select particle candidates with some specific requirement: in particular, the events are required to pass an unprescaled single-muon trigger. At least one HLT muon candidate must have a p_T value above a given threshold in the pseudorapidity range $|\eta| < 2.4$. The trigger path used in this analysis is HLT_IsoMu24, where the number in the path name indicates the corresponding p_T threshold [194]. This is among the trigger algorithms with the lowest p_T threshold whose output is not artificially reduced to limit the event rate and that cover the entire η acceptance of the muon detector.

5.5.2 Muons

The analysis described in this thesis, but in general the entire CMS physics program largely relies on muons, involved in many of the signatures studied by the experiment concerning precision measurements of the SM and searches for new physics up to the TeV scale. Given that muons are the only charged particles able to reach the outermost layers of the detector, they provide a clear signature, easy to trigger on and to reconstruct with high efficiency and purity.

The first step in the reconstruction of muons with the CMS detector comes from the local reconstruction with the information coming from the hits in the individual DT and CSC chambers (explored in detail in Section 3.4). Such hits are then used to form *segments*, approximable as a straight lines since the magnetic field is almost totally confined in the steel return yoke of the magnet. The DT track segments are reconstructed from the hits originating in the local DT electronics. The measurement of the radial position and the bending angle is provided by exploiting two different views: the hit reconstruction in a DT cell specifies the transverse distance between the wire and the intersection of the muon trajectory with the plane containing the wires in the layer. The CSC track segments, instead, are obtained by combining the cathode and anode hits (one per each layer of each station) in the transverse plane.

Later on, to build the tracks, an iterative approach based on the Kalman Filter technique [195] is performed, taking into account the effect of the magnetic field and the energy loss/multiple scattering in the steel. From this segments, the position and direction vectors are used to give an initial estimate of the muon transverse momentum p_T . Such information is then used to generate *seeds* that are propagated to each station layer, looking for the most compatible segment based on the χ^2 of the fit. At this stage, hit clusters reconstructed in the RPC system (explored in Section 3.4.3) are also included in the tracks.

This steps gives as output a collection of reconstructed track objects, referred to as *standalone muons*. To improve the momentum resolution, the beam spot position in the transverse plane can be also used in the fit. To achieve that, the full muon tracks given by the local reconstruction in the muon spectrometer get further combined with tracks reconstructed in the inner detector. A *global muon* (GLB) is obtained by matching standalone muon tracks and inner tracks, using the above-mentioned Kalman Filter technique. In such global reconstruction, as in the standalone one, the minimum requirement for a track is to have hits or segments in two station layers with at least one from DTs or CSCs. On the contrary, a *tracker muon* (TRK) is reconstructed by extrapolating an inner track inside-out to all the DT and CSC station layers and looking for compatible DT and CSC segments. The lower p_T reach of tracker muons is improved with respect to global muons reaching. Every track with $p > 2.5$ GeV and $p_T > 0.5$ GeV, matched with at least one segment in the muon spectrometer, is labelled as tracker muon.

The majority of muons produced in pp collisions ($\approx 99\%$) with a sufficient high momentum is reconstructed by both algorithms. The information from all muon track types

is finally combined in a single object, providing a coherent view of a muon candidate and to ensure an efficient final reconstruction. At this point in the computation, energy deposits in the calorimeters can also be included [196].

Muon identification criteria. Since the collection of reconstructed muons also contains misidentified charged hadrons or muons coming from in-flight decays, additional quality requirements must be applied to have as pure as possible samples of muon candidates. Therefore, a set of variables and selection criteria are defined, e.g. the type and number of missing hits, the track χ^2 and the compatibility between tracker and standalone muons tracks in a global muon. The main identification (ID) criteria for muons adopted in the CMS physics analyses are:

- A *loose muon* which is a tracker or global muon selected by the PF algorithm. The loose ID is oriented to identify prompt muons originating at the primary vertex as well as muons from light and heavy flavour decays, avoiding the increase in the misidentification rate of charged hadrons as muons.
- A *medium muon* which is optimised for prompt muons and for muons from heavy flavour decay. A medium muon is defined as a loose muon reconstructed from a tracker track with many hits in the inner tracker layers. If a muon is only reconstructed as a tracker muon, the *muon segment compatibility*³ is required to be larger than 0.451. Instead, if the muon is reconstructed as tracker and global muon, a less stringent cut on the muon segment compatibility is complemented by some quality requirements on the χ^2 of the track [197].
- A *tight muon* which is a loose muon reconstructed from a tracker track with at least one pixel hit and at least six hits in the inner tracker. It must be reconstructed as both a tracker muon and a global muon and must be compatible with the primary vertex. The tracker muon must also have segment matching in at least two of the muon stations, consistently with the logic of the muon trigger. The global muon fit must include at least one hit from the muon system. The idea behind such criteria is the suppression of the so called *fake muons*, or muons coming from in-flight decays and/or hadronic punch-through.
- A *soft muon* which is a tracker muon with a high purity track, using hits from at least six layers of the inner tracker and including at least one pixel hit. The tracker track is matched with at least one muon segment in any station in both x and y coordinates. This criteria is mainly optimised for analyses with low p_T muons objects.
- A *highPt* muon which is a global and tracker muon and not PF. This allows the reduction of the bad muon effect due to electromagnetic showers developed in the calorimeters. The muon candidate has to fulfil the same criteria of the Tight ID with the exception of the requirement on the χ^2 of the global fit, which is replaced by a cut

³The muon segment compatibility is a variable, ranging from 0 to 1, computed on the basis of the distance between the inner track, extrapolated to all the muon stations, and the closest DT and CSC segments from the same stations.

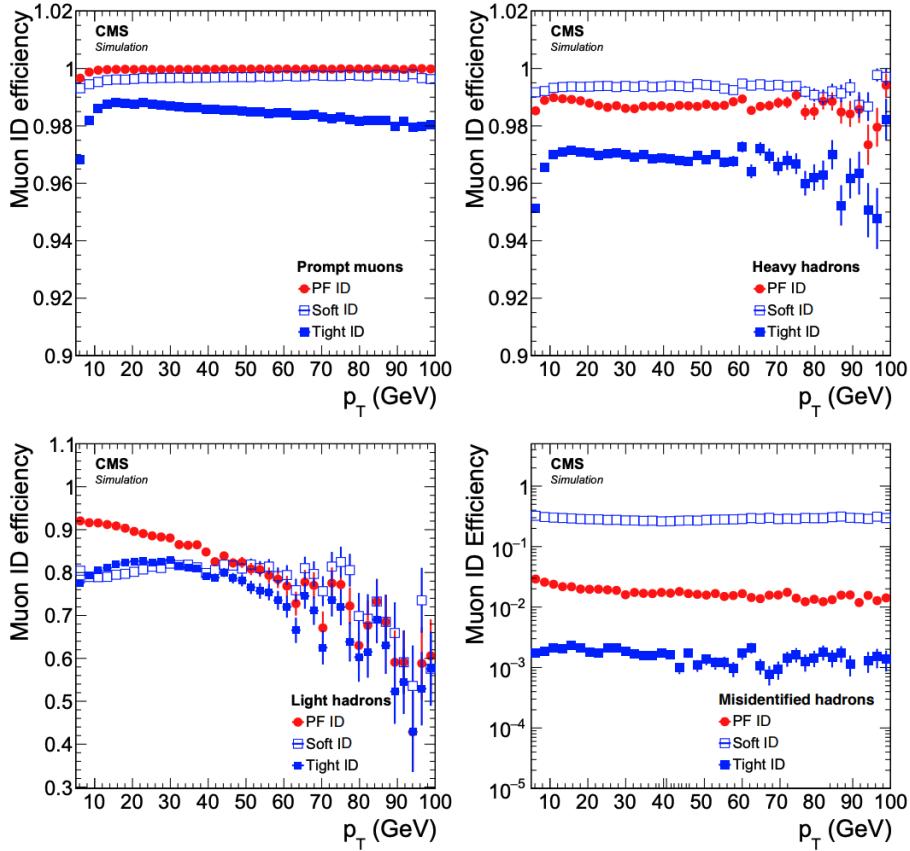


Figure 5.6: Efficiency for different algorithms (PF/loose, soft, and tight) to identify a simulated muon track that has been reconstructed as a tracker muon, as a function of the p_T of the reconstructed track. From top left to bottom right, the efficiency of the three identification algorithms is shown for: prompt muons, muons from heavy-flavour decays, muons from light-flavour decays, and misidentified hadrons. Reproduced from [115].

on the relative p_T error, $\delta p_T / p_T < 0.3$. Moreover, the tracker muon must be matched to segments in at least one muon station if the station belongs to the first layer of the muon system. However, if the matching occurs in the first muon layer, an additional match in another layer or with more than two RPC layers is also required.

A performance plot, comparing some of the ID criteria, is shown in Figure 5.6.

Muon Isolation. Another important criterion used to select leptons originating from the W and Z decays is the isolation. The isolation distinguishes prompt leptons, such as the ones originating from W and Z boson decays, from leptons produced in hadron decays that usually are embedded in a jet. Using all the PF candidates, the isolation is usually defined as:

$$\text{ISO}_\mu^{\text{PF}} = \frac{1}{p_T^\mu} \left(\sum_{\text{charged}} p_T + \max \left[0, \sum_{\text{neutral}} p_T + \sum_\gamma p_T - 0.5 p_T^{\text{PU}} \right] \right), \quad (5.1)$$

where the sum runs on the PF candidates contained in a cone of given $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$, centred around the lepton direction, and p_T^{PU} is a correction that takes into account the pileup conditions.

5.5.3 Jets

In CMS, jets are reconstructed from particle flow objects using the anti- k_T clustering algorithm [198], implemented in the FASTJET package [199, 200]. The algorithm belongs to a class of sequential recombination algorithms, which includes also the k_T and Cambridge-Aachen algorithms [201, 202]. These three methods are prevalent in high-energy physics nowadays and are adopted depending on the different analysis strategies. All the algorithms of this family are *infrared* and *collinear safe*: infrared safeness means that the results of the jet clustering are not altered if an arbitrary number of extra particles with momentum that tends to zero is included in the jet clustering, while collinear safeness means that the clustering is not sensitive to a splitting of a particle into two collinear ones each taking a fraction of the momentum.

In such algorithms, for each pair of objects to be clustered, e.g. the PF objects for CMS, the distance d_{ij} is computed:

$$d_{ij} = \min(p_{T,i}^{2n}, p_{T,j}^{2n}) \cdot \frac{\Delta R_{ij}^2}{\Delta R^2}, \quad (5.2)$$

where $\Delta R_{ij} = \sqrt{(\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2}$ and ΔR is a fixed distance parameter which defines the cone amplitude. The pair that minimises the distance d_{ij} is merged to form a new object. The distances are then computed with the new set of objects. At each step the pair that minimises d_{ij} is merged, until a stopping condition is reached. In the k_T algorithm $n = 1$, in the Cambridge-Aachen $n = 0$, while in the anti- k_T $n = -1$. In the anti- k_T algorithm, therefore, the radiation is clustered around high- p_T objects first, so that conical jets of radius equal to the distance parameter ΔR are typically produced, unless multiple hard objects are clustered. At CMS, the standard clustering distance is $\Delta R = 0.4$. At the same time, larger jets with a ΔR of 0.8 are used when looking for boosted heavy particles decaying into hadrons.

5.5.4 Identification of b jets

Concerning the b-associated signal production of the Higgs boson, another fundamental quantity to identify are the jets coming from the b quark. Such jets have a similar behaviour to the other jets and do not need a special treatment in their reconstruction. However, due to the hadronisation of the b quark in a B hadron, the reconstruction of its decay products is fundamental to tag the b jets. Jets containing B hadrons can be distinguished thanks to the long B hadron lifetime $\tau_B = (1.638 \pm 0.004) \cdot 10^{-12}$ s, corresponding to an average distance of $c\tau \approx 500 \mu\text{m}$ [4]. Such long lifetime, compared to other quarks, is caused by the need for b quarks to decay weakly into lighter quarks, as shown in Figure 5.7a: the top quark final state would be favoured, but it is not kinematically possible due to the mass

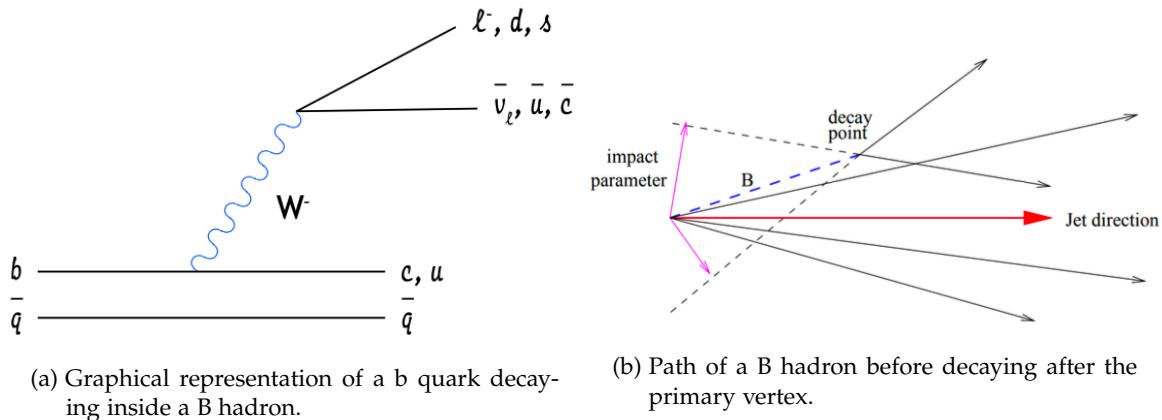


Figure 5.7: Hadronisation of a b quark in a B hadron, and relative decay products used for b tagging.

of ≈ 170 GeV of the top quark. The transition to lighter quarks, belonging to the second or the first family, comes with a sizeable suppression factor⁴ thus resulting in a longer lifetime. The long lifetime results then in a higher impact parameter of the decay products w.r.t. the primary vertex, which are reconstructed as tracks (as shown in Figure 5.7b).

CMS b tagging algorithms. The identification of a jet from the b quark (also known as *b tagging*), is one of the areas where Machine Learning is fundamental in order to obtain an optimal performance (more details in Section 4.8.2). The CMS standard algorithms, optimised with Machine Learning, rely both on secondary vertices and tracks for the discrimination. At first, these algorithms were created by using single observables, or a few of them, without using learning techniques.

The Track Counting (TC) algorithm, for example, sorts tracks in a jet by decreasing values of the IP significance. The first track has little discriminating power: however, the probability to have several tracks with high positive values is low for light-flavour and gluon jets. Therefore, there are two versions of the algorithm that use the IP significance of the second and third ranked track as the discriminator value. The two versions are called Track Counting High Efficiency (TCHE) and Track Counting High Purity (TCHP) algorithm, respectively. Another one, based instead on the secondary vertex observable, is called Simple Secondary Vertex (SSV) algorithm which uses the significance of the secondary vertex flight distance as discriminating variable. If several vertices are present the one with the highest significance is used. As for the Track Counting algorithms, two SSV versions optimised for different purity exist: the High Efficiency version (SSVHE) uses vertices with at least two associated tracks, while for the High Purity version (SSVHP) at least three tracks are required.

Such algorithms were already existing before the beginning of the LHC data taking. During Run 1, the first discriminators based on Machine Learning were employed: the Combined Secondary Vertex (CSV) algorithm, which combines secondary vertices and

⁴ Due to the mixing of weak interaction eigenstates and mass eigenstates in the electroweak interaction, described in Section 1.1.2, the decay into a different family fermion is highly reduced.

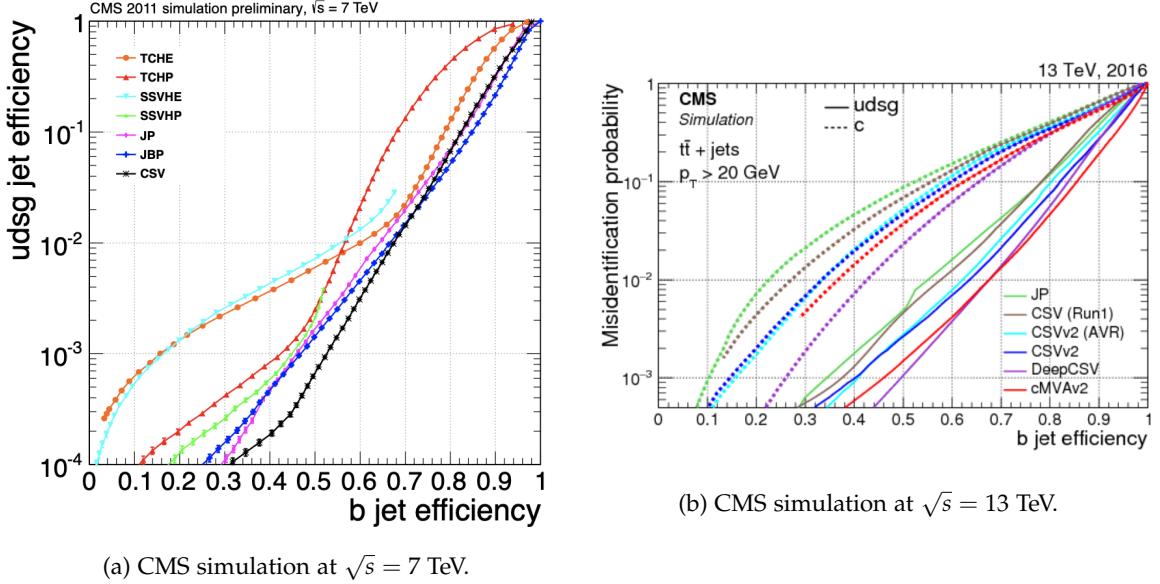


Figure 5.8: Misidentification efficiency for c and light-flavour jets vs b jet identification efficiency (ROC curve) for various tagging algorithms in $t\bar{t}$ events [203, 204].

tracks via a likelihood ratio was first developed. The performance of these algorithms is shown in Figure 5.8a. The ROC curve (explained in Section 4.4.4), comparing the efficiency of b jets and the mistag of light-flavour jets for the $t\bar{t}$ 7 TeV simulation, is shown. The CSV algorithm outperforms all the older algorithms. Afterwards, even better versions of the CSV algorithms were developed. The upgraded CSVv2 algorithm [203], in fact, requires at least 2 tracks per jet compatible with the primary vertex. Additionally, any combination of two tracks compatible with the mass of the K_s^0 meson is rejected. The training of the algorithm is then performed in three independent vertex categories: the first category contains jets with at least one associated reconstructed secondary vertex; the second, called *pseudovertex*, contains jets whose tracks with an IP significance larger than 2, allowing for the computation of a subset of SV observables. Otherwise, a "no vertex" category with track-based variables only is defined.

Finally, during Run 2 also Deep Learning algorithms were introduced for b tagging. The DeepCSV algorithm has been developed using a deep feed-forward neural network, explained in Section 4.6. The DeepCSV algorithm exploit the same information as the CSVv2 one, but the training is performed using more events and a more flexible algorithm. This allows to solve the entire b tagging problem in one step, i.e. a training including all categories and all jet flavours, and allows improved performances. Figure 5.8b shows the performance of b tagging CMS algorithms at 13 TeV.

Currently, in the CMS Collaboration, a new innovation concerning the jet b tagging using deep models has been made, processing a large number of particles and features, called DeepJet [205]. The jet representation used by DeepJet, fully captures the description of the jet as produced by the PF algorithm. Up to 25 charged PF candidates and up to 16 neutral PF candidates per jet are used as input to a deep Neural Network. In addition, also the Inclusive Vertex Finder (IVF) reconstruction of the secondary vertices is exploited: up to 4

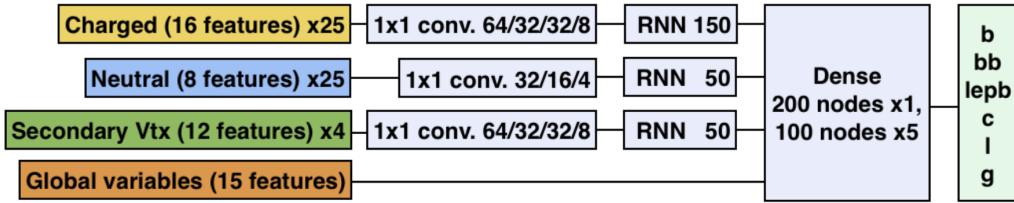


Figure 5.9: Deep NN architecture for the DeepJet network. The different particle sequences are shown in different colors as well as the dimensionality of the inputs. In total, about 265000 trainable parameters are considered during training.

vertices matched geometrically to the jet are given as input to the Neural Network. Other variables used are: the high level features, providing a global description of the jet suitable for b tagging, the so called *tagging variables* used also in the previous CSV algorithm, as well as the jet kinematics (p_T, η). The sequences of particles include respectively 16, 8 and 12 input features for the charged PF candidates, the neutral PF candidates and the secondary vertices. The global variables are 15 in total. The architecture of DeepJet is quite complex and it is shown briefly in Figure 5.9: in a nutshell, after some feature rescaling on the first epoch, each sequence of particle is processed with convolutional 1×1 filters [206], with a specific topology based on the candidate, followed by a Long Short Term Memory (LSTM) recurrent Neural Network [207] with different output nodes, also chosen based on the particle candidate. The outputs from the LSTM's are finally merged with the global jet properties, and fed through one dense layer (see Section 4.6) with 200 nodes with 7 subsequent hidden layers with 100 nodes each.

The target is the jet flavour with 6 flavour categories employed in the training. The categories are: gluon, light-flavour quark, charm quark and bottom quark jets, with b quark jets further split into 3 categories: b with no leptons, b with leptonic decays and bb, which are then merged for evaluation purposes. The loss function is the categorical cross-entropy (see Equation 4.18). The NN is implemented using the KERAS package with TENSORFLOW backend (see Section 4.7). The optimisation is performed using stochastic gradient descent (SGD) with the Adam optimiser [208], which is based on the adaptive estimation of the first and second order moments in the gradient computation.

The DeepJet algorithm has been trained on a total sample of 100M jets. The jets come from $t\bar{t}$ production and QCD multi-jet production simulated events. All flavours are included. The performance of DeepJet represents an important step forward in the jet b tagging performances. Compared to the CSV algorithm, DeepJet was found to improve on the standard algorithm both in $t\bar{t}$ and QCD multi-jet simulation, against both charm quark, light-flavour quark and gluon jets. In Figure 5.10, the ROC curve describing the performance of the b jet algorithm, is shown. For the loose, medium and tight working points, the data-to-simulation scale factors have been applied (see Section 6.3.1) and are represented by the triangles with error bars. Circular markers, instead, represent the performance of the respective working point in simulated samples [11].

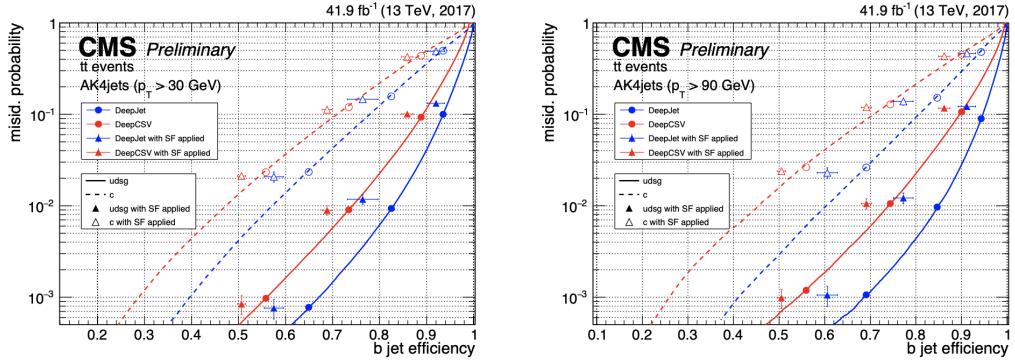


Figure 5.10: Performance of the CMS DeepJet and DeepCSV algorithms. The ROC curves show the b tagging efficiency as a function of the probability of misidentifying non b-jets as b jets. On the left plot, jets with $|\eta| < 2.5$ and $p_T > 30$ GeV are considered; on the right plot, jets with $|\eta| < 2.5$ and $p_T > 90$ GeV are considered. Such results are obtained from simulated top pair events. For the loose, medium and tight working points, the data-to-simulation scale factors have been applied, represented by the triangles with error bars. Circular markers, instead, represent the performance of the respective working point in simulated data.

5.5.5 Missing energy in the transverse plane

The PF technique reconstructs the full event by identifying all the particles: muons, electrons, photons, charged and neutral hadrons. The reconstructed particles are given as input to the jets and E_T^{miss} (or MET) reconstruction algorithms, defined as the magnitude of the negative vector p_T sum of all the PF objects. Jets reconstructed using PF have a good angular resolution ($\delta\eta$, $\delta\phi$ in the 0.3-0.01 range for $p_T < 100$ GeV, ≈ 0.01 for $p_T > 100$ GeV), while the energy needs to be corrected in multiple steps. Such corrections are called "jet energy corrections" (JEC), applied to the 4-momentum magnitude. The solution to the problem of JEC in CMS is factorised, where each level of correction takes care of a different effect.

The corrections are summarised as:

- Pileup offset corrections, computed in simulation. First, tracks coming from pileup vertices are removed (charged hadron subtraction, CHS). Then an offset correction is applied to account for residual contamination, determined from the per-event median energy density ρ computed with the k_T clustering algorithm. Such corrections are parametrised in terms of η , p_T and the jet area.
- Simulated response corrections, determined in simulation as a function of η and p_T by evaluating the response with respect to generator level jets.
- Residual corrections, determined in data and compared with simulations. The correction is derived in two main steps: first, the jet energy response is corrected as a function of η relatively to the better calibrated barrel region ($|\eta| < 1.3$), these being usually called "relative" jet corrections; then the jet momentum is scaled in order to match a reference object within jet $\eta < 1.3$: these are usually called "absolute" corrections.

The residual data/MC corrections are measured in dedicated topologies: the relative corrections are measured using dijet events, while absolute corrections are calibrated using the balancing of Z/ γ +jet events. Analogous topologies are used to measure the difference in resolution between jets and simulation as a function of η and p_T and correct for it (JER correction).

Jet energy corrections are propagated to the missing transverse energy in the event, forming the *Type I* corrections. Event filters recommended by the JetMET POG are applied to the event in order to veto events where problems in the reconstruction raise to unphysical events. These filters include the BadMuon and BadChargedHadron filters, noise filters in HBHE, EcalDeadCellTriggerPrimitive filter, good vertices filter, EE bad super cluster filter, and beam halo filter.

6

ANALYSIS STRATEGY

In the previous Chapter, the physics objects used to perform the analysis have been listed and described. Now it is possible to present how such objects are combined, to select and categorise events in the MSSM Higgs to $\mu^+\mu^-$ analysis. The definition and optimisation of the signal region selection will be presented, followed by the properties and Machine Learning techniques exploited to discriminate the H/A boson signals against the relevant backgrounds and to classify the events depending on the specific topology of each H/A boson production mode. In order to study the bosons production mechanisms and improve the discrimination, the selected events are further classified into mutually exclusive categories. The categorisation scheme that targets all the main production mechanisms is going to be described in detail, because it represents a fundamental point to understand the multivariate analysis strategy and performance presented below, as well as the final results presented in the next Chapter.

6.1 EVENT SELECTION

The events selected for the MSSM Higgs to $\mu^+\mu^-$ analysis are preselected by the trigger system, described in Section 5.5.1, requiring a muon candidate within the pseudorapidity range of $|\eta| < 2.4$, and satisfying the following criteria: $p_T > 24$ GeV with isolation requirements (IsoMu24).

The search for an Higgs boson signal is performed on a wide mass range, with a corresponding muon p_T varying from tens to hundreds of GeV. Therefore, to select a good quality muon the *Tight ID*, described in Section 5.5.2, has been used. Events with a pair of opposite-charge muons, coming from the PV, are selected requiring both muons to satisfy the same ID. In addition, a few additional quality criteria must be applied:

- The χ^2/ndf of the muon track fit must be smaller than 10, to guarantee a good measurement of the track momentum;
- The transverse and longitudinal impact parameter of each muon track must be $|d_{xy}| < 0.02$ cm and $|d_z| < 0.1$ cm respectively, to discard muons from cosmic rays.

At least one of the two muon candidates has to match (in η and ϕ) the muon that triggered the event. Offline reconstructed muons with $|\eta| < 2.4$ are considered. Their offline p_T is required to be higher than 26 GeV, to be compatible with the muon that triggered the event.

Muon selection	muon <i>Tight ID</i>
Online selection: Single muon	$ \eta < 2.4$ $p_T > 24 \text{ GeV}$ Online iso
Offline selection: Two opposite-charged muons	$ \eta < 2.4$ $p_T > 26 \text{ GeV}$ Offline iso < 0.25

Table 6.1: Summary of the muon selection criteria.

A cut on the relative isolation is also applied: muons with a relative isolation value above the loose working point are rejected, as recommended by the Muon POG. For the *Tight ID*, the PF isolation in a cone of aperture 0.4 is used, where the isolation variable is defined in Eq. 5.1 and it is required to be $\text{ISO}_\mu^{\text{PF}} < 0.25$.

The *Rochester* correction [209] is applied to the selected muons to improve their momentum resolution. This correction is performed to take into account alignment effects, like shifts and deformations in the transverse plane, and magnetic field effect, that affects the muon momentum resolution. The correction is applied on both data and simulated events.

Energy deposits in the calorimeters are not included, since electromagnetic showers can develop from photons radiated by a high- p_T muon. The invariant mass of the Higgs boson candidate is then reconstructed from the two highest p_T opposite-charge muon candidates in the event. The dimuon selection criteria are summarised in Table 6.1.

Jets are reconstructed using the anti- k_T clustering algorithm, described in Section 5.5.3, with a distance parameter of 0.4, as implemented in the FASTJET package. The missing transverse energy E_T^{miss} , described in Section 5.5.5, is defined as the magnitude of the negative vector p sum of all the PF objects (charged and neutral) in the event, and is modified by corrections to the energy scale of reconstructed jets. The events must have at least one well reconstructed PV, i.e. the number of degrees of freedom of the vertex should be $ndf_{\text{PV}} > 4$, to ensure a good quality of the PV determination. The reconstructed vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex. The distance of the reconstructed vertex along the z axis from the nominal centre of the detector has to be $|z_{\text{PV}}| < 24 \text{ cm}$.

The DeepJet discriminator algorithm, described in Section 5.5.4, is used to b tag the jets. The discriminating variable, b_{disc} , is built out of the tracks and their impact parameter significance within each jet is aimed to distinguish b-quark from c and light-flavour jets. A jet is considered b-tagged with an associated $b_{\text{disc}} > 0.2489$ for 2016 with no-HIPM, or $b_{\text{disc}} > 0.2598$, for 2016 with HIPM, which corresponds to the medium working point recommended by the BTV POG [204]. The transverse momentum of the b-tagged jets has to be larger than 20 GeV, in the pseudorapidity range $|\eta| < 2.4$.

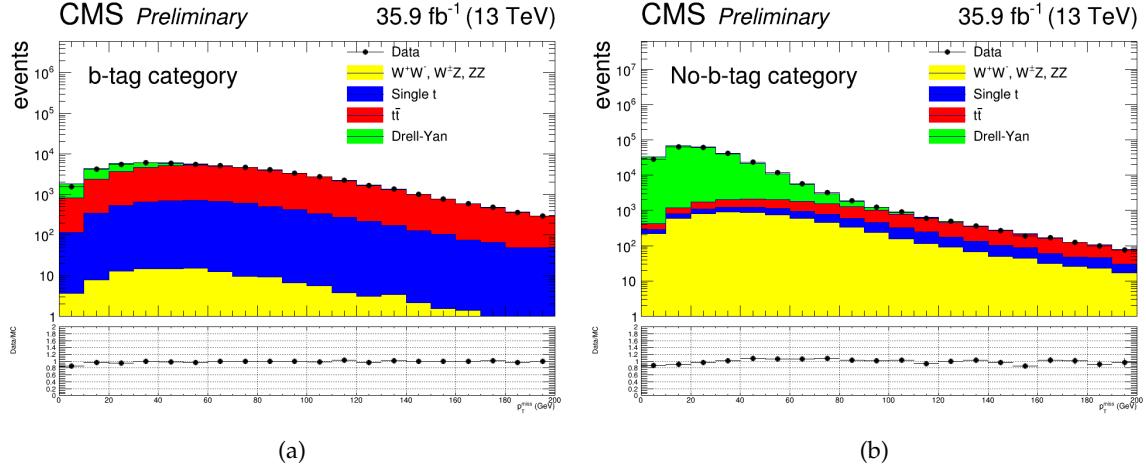


Figure 6.1: On the left, the transverse missing energy in the event for b-tagged events (CAT1), while on the right the same quantity for non b-tagged events (CAT2). The quantities refer to events with dimuon invariant mass larger than 130 GeV. The MC background samples are normalized to the integrated luminosity.

To summarise, the event are selected if they have at least two opposite-charge *Tight* muons:

- $p_T > 26$ GeV and $|\eta| < 2.4$;
- relative PF isolation < 0.25 ;
- at least one muon firing the IsoMu24 HLT trigger path.

6.2 EVENT CATEGORISATION

Events containing b-jet candidates provide the highest sensitivity for the b-associated production channel, and events that do not contain b-tagged jets provide the best sensitivity for the gluon fusion production channel. For this reason, the events have been split into two exclusive categories: the b tag category (CAT1), containing events with at least one b-jet, and the no b tag category (CAT2), containing events without b-tagged jets.

Signal events are characterised by a rather small p_T^{miss} . However, the background content is quite different for the two categories. The background from $t\bar{t}$ events, characterised by a relatively large p_T^{miss} from W boson decays, is much more relevant for the b tag category, as shown in Figure 6.1a. For the no b tag category, instead, the dominant background is the Drell-Yan production, whose events are characterised by a p_T^{miss} distribution that is similar to that of the signal, as shown in Figure 6.1b.

6.3 MULTIVARIATE ANALYSIS

The analysis can be summarised in two main steps. The first step is the event selection and categorisation, mentioned in the previous Sections, while the second step consists of a multivariate analysis used to discriminate between signal and background samples. In this

second step, the most discriminating variables are chosen and combined into a Machine Learning discriminator able to maximise the signal sensitivity, relying on a precise modelling of the background. The backgrounds in question are modelled using Monte Carlo simulated samples, described in more detail in Section 5.4.2. But first, in order to verify the reliability of the background model in the signal region, control regions for the simulated backgrounds have been identified in data.

6.3.1 Control regions

Although the MC background events are not used to determine the Higgs exclusion limits, they are adopted for the multivariate model training, and therefore it is important to show how well the simulation models the data. A mismodelling of the simulation in describing the quantities used to select the events can introduce incorrect determination of the efficiency to detect signal events. In order to account for differences between data and simulation, the scale factors (SF) recommended by the CMS POGs are used to re-weight the MC events. The SFs are applied as an unique global event weight.

The Muon POG provided the SF related to the muon identification, isolation, and trigger selection [210]. The SFs are applied as an event weight according to the the p_T and η of each of the two selected muons. An exception is made for the trigger SF, because the trigger can be fired by one or both muons.

The efficiency to trigger an event is calculated as the *logical-OR* of the single muon-trigger efficiency with $\epsilon_{\text{trg}} = 1 - \prod_{i=1}^N (1 - \epsilon_{\mu_i})$, where N is the total number of muons, with $p_T > 26 \text{ GeV}$, that fired the trigger in the event. The corresponding SF is therefore:

$$\text{SF}_{\text{trg}} = \frac{\epsilon_{\text{trg}}^{\text{data}}}{\epsilon_{\text{trg}}^{\text{MC}}}. \quad (6.1)$$

The event reconstruction and isolation efficiencies are calculated as the *logical-AND* of the single muon efficiencies with $\epsilon_{\text{id/iso}} = \prod_{i=1}^N \epsilon_{\mu,i}$, yielding to the event SF as the product of the muons SFs:

$$\text{SF}_{\text{id/iso}} = \frac{\epsilon_{\text{id/iso}}^{\text{data}}}{\epsilon_{\text{id/iso}}^{\text{MC}}}. \quad (6.2)$$

The BTV POG provided the data/MC SF (`DeepJet_106XUL16postVFPsf_v2.csv` for 2016 with no-HIPM and `DeepJet_106XUL16preVFPsf_v1.csv` with HIPM) for the b tagging identification and mis-identification [211]. It accounts for all the jets in the selection, distinguishing between light jet (udsg) mistag rate or heavy jet (b/c) efficiency using MC truth information. To apply the b tag SF corrections, the `correctionlib` library has been used, within the CMS NanoAOD central tools [212, 213].

Given the different reconstruction modes among the different run periods, different SFs are provided: run BCDEF, with HIPM reconstruction, and run GH with the classical

track reconstruction (see Section 5.4.1). Event SFs are averaged using as weights the corresponding integrated luminosity:

$$w_{\text{BF}} = \frac{\mathcal{L}_{\text{BF}}}{\mathcal{L}_{\text{tot}}}, \quad (6.3)$$

$$w_{\text{GH}} = \frac{\mathcal{L}_{\text{GH}}}{\mathcal{L}_{\text{tot}}}, \quad (6.4)$$

$$\text{SF} = w_{\text{BF}} \cdot \left(\text{SF}_{\text{trg}}^{\text{BF}} \cdot \text{SF}_{\text{id}}^{\text{BF}} \cdot \text{SF}_{\text{iso}}^{\text{BF}} \cdot \text{SF}_{\text{btag}}^{\text{BF}} \right) + w_{\text{GH}} \cdot \left(\text{SF}_{\text{trg}}^{\text{GH}} \cdot \text{SF}_{\text{id}}^{\text{GH}} \cdot \text{SF}_{\text{iso}}^{\text{GH}} \cdot \text{SF}_{\text{btag}}^{\text{GH}} \right). \quad (6.5)$$

Equation 6.5 represents the total muon efficiency correction applied to simulation.

Figures 6.2 and 6.3 show the data-to-simulation comparison for some muon and dimuon related quantities, for events with dimuon invariant mass larger than 130 GeV, belonging to CAT1 and CAT2 events respectively. The MC background samples are normalized to the integrated luminosity.

Figure 6.4 shows the data-to-simulation comparison for two quantities relevant for the b tag: the number of b-tagged jets in the event and the transverse momentum of the b-tagged jet with the highest transverse momentum in the event, for events with dimuon invariant mass larger than 130 GeV, belonging only to CAT1 since CAT2 does not contain b-jet events by assumption. The MC background samples are normalized to the integrated luminosity. Figure 6.5 shows the dimuon invariant mass distribution for the events belonging to the two categories, with the data-to-simulation comparison.

The muon momentum measurement is crucial for the reconstruction of the Higgs boson mass peak, which is dominated by the detector resolution. Improving the dimuon mass resolution in data is an important factor in the sensitivity of the analysis. Moreover, the mean and the resolution of the dimuon mass peaks in Monte Carlo must match in data in order to set limits accurately. In order to address these issues, the Kalman filter muon correction has been applied to both data and MC signal events. The Kalman correction is valid for muons of p_T smaller than 100 GeV and it is meant to provide consistent measurements of the Z peak in the different ϕ and η regions of the detector, improving the net resolution in data. The correction is also applied to the simulated muons to align the scale and resolution to those measured in the data.

6.3.2 Parameterised Neural Network

In Section 4.6, network-based learning techniques have been explained in detail. Such networks, in the current days, have been applied to a wide variety of problems in high-energy physics [214, 215], from the event classification [147, 216] to object reconstruction [217, 218] and triggering [219, 220]. These networks, however, are applied to solve a specific isolated problem, even when this problem is part of a set of closely related problems.

In the MSSM Higgs to $\mu^+ \mu^-$ analysis, the different m_A hypothesis are distinct classification tasks, but related: current approaches would require the training of a set of isolated networks [221, 222], each of which is ignorant of the larger context and with a lack in the ability to smoothly interpolate, or to use a single signal sample in training [223],

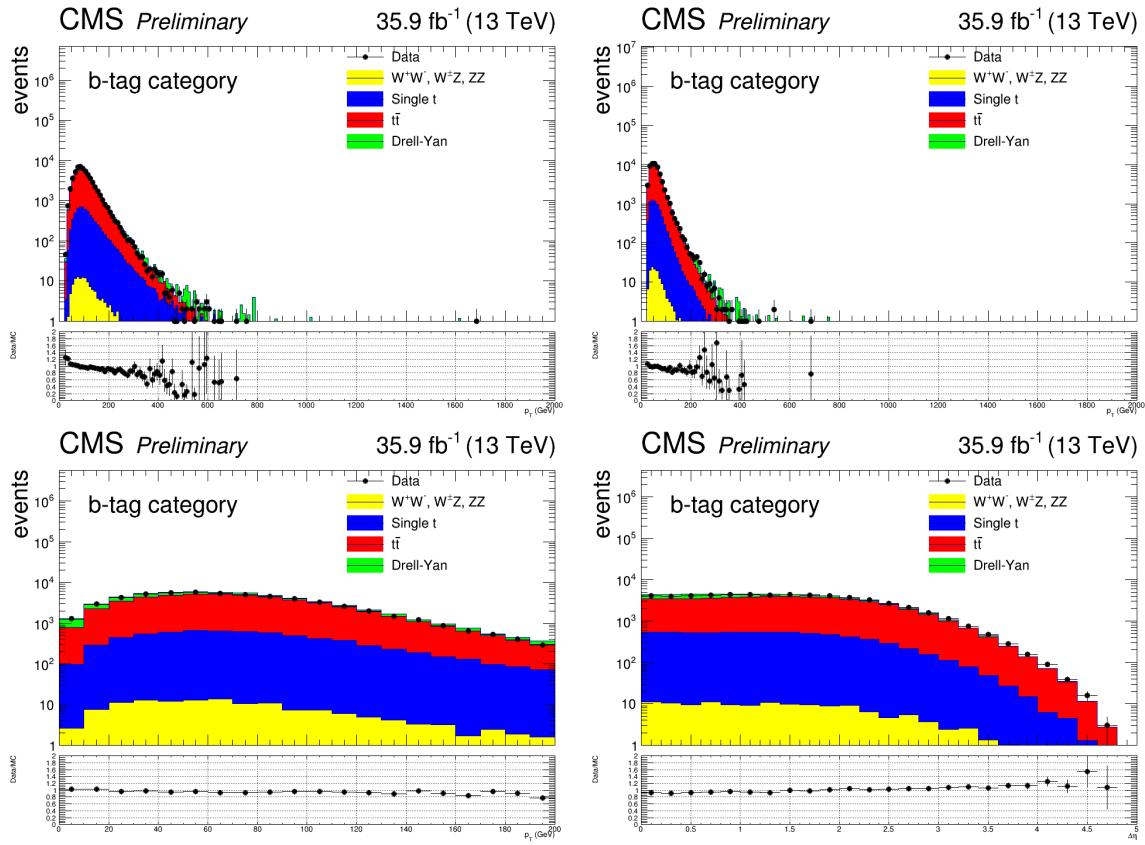


Figure 6.2: From left to right, and from top to bottom: p_T of the leading muon, p_T of the 2nd leading muon, the transverse momentum of the dimuon system, and the pseudorapidity of the dimuon system, for events with dimuon invariant mass larger than 130 GeV and belonging to CAT1. The MC background samples are normalized to the integrated luminosity.

sacrificing performances at other values. For this reason, a new network approach has been adopted, the *parameterised Neural Network* (pNN), in which a single network tackles the full set of related tasks [224]. This particular approach is relatively new in CMS, and it is already used by some published analyses, e.g. [225, 226].

This is done by simply extending the list of input features to include not only the traditional set of event-level features but also one or more *parameters* that describe the larger scope of the problem, for the MSSM analysis being the mass of the pseudoscalar boson A . With a single parameterised network, in fact, it is possible to replace a set of individual networks trained for specific cases, as well as to interpolate cases where it has not been trained. In the case of the search for a MSSM neutral Higgs boson, this greatly simplifies the task – by requiring only one network – as well as making the results more powerful by allowing them to be interpolated between specific values. In addition, they may outperform isolated networks by generalizing from the full parameter-dependent dataset.

While a typical network takes as input a vector of features \mathbf{x} , giving as output a function of those features $f(\mathbf{x})$ after training, a parameterised network is part of a larger context, described by one or more parameters θ . It is then straightforward to construct

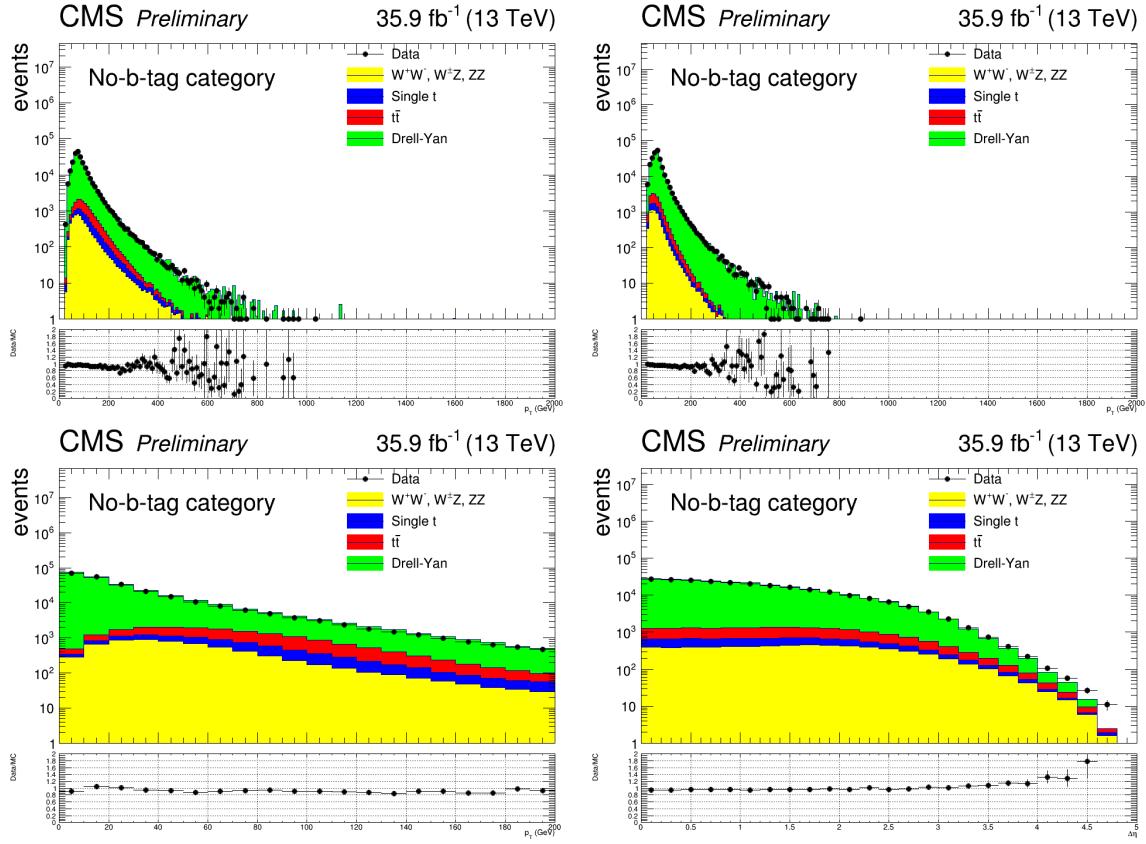


Figure 6.3: From left to right, and from top to bottom: p_T of the leading muon, p_T of the 2nd leading muon, the transverse momentum of the dimuon system, and the pseudorapidity of the dimuon system, for events with dimuon invariant mass larger than 130 GeV and belonging to CAT2. The MC background samples are normalized to the integrated luminosity.

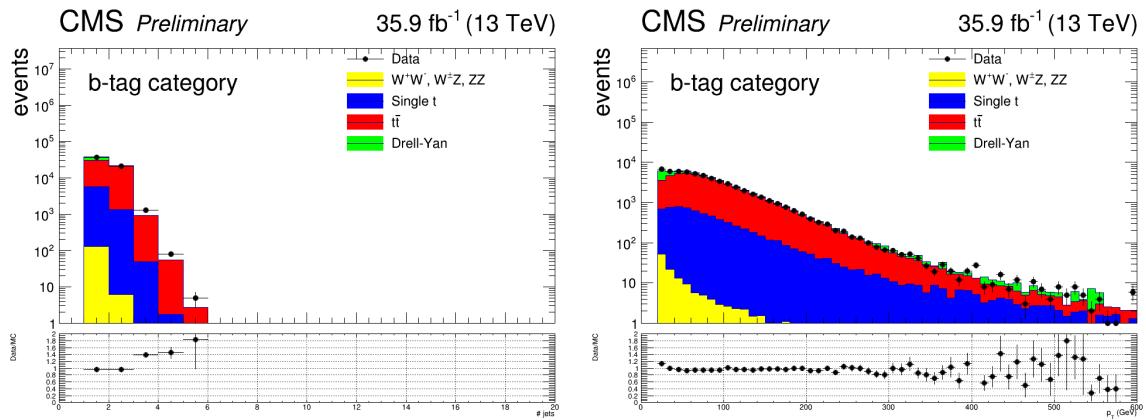


Figure 6.4: From left to right: the number of b-tagged jets and the p_T of the leading b-jet. The quantities refer to events with dimuon invariant mass larger than 130 GeV and belonging to CAT1. The MC background samples are normalized to the integrated luminosity.

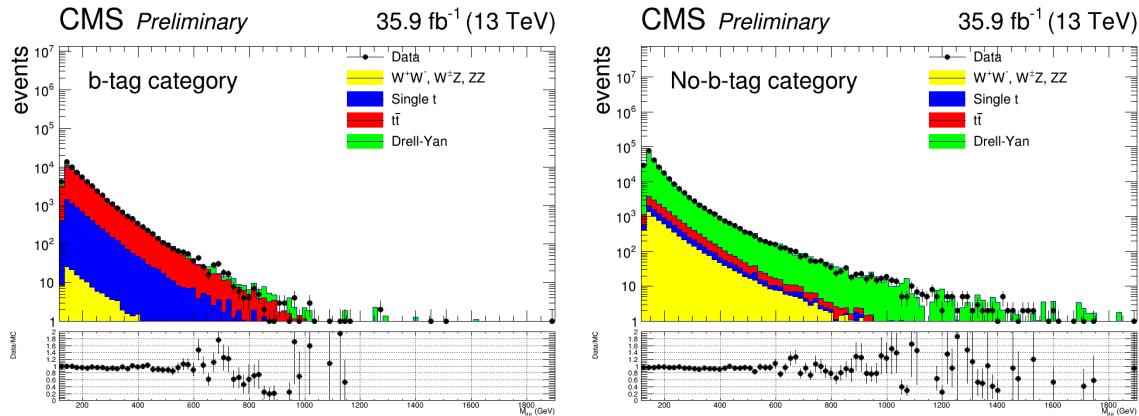


Figure 6.5: On the left, the dimuon invariant mass for events that belong to CAT1, while on the right the dimuon invariant mass for events that belong to CAT2, for dimuon invariant masses larger than 130 GeV. The MC background samples are normalized to the integrated luminosity.

a different network that uses both sets of input, \mathbf{x} and $\boldsymbol{\theta}$, and is a function of both: $f(\mathbf{x}, \boldsymbol{\theta})$. Given a set of inputs \mathbf{x}_0 , a classical network is evaluated as a real number $f(\mathbf{x}_0)$. A parameterised network, however, as a result that is also parameterised in terms of $\boldsymbol{\theta}$: $f(\mathbf{x}_0, \boldsymbol{\theta})$, yielding different output values for different choices of the parameters $\boldsymbol{\theta}$, as shown in Figure 6.6.

Training data for the parameterised network can be written in the form $(\mathbf{x}, \boldsymbol{\theta}, y)_i$, where y is the label for the target class (the model training is explained in more detail in Section 4.3.1). Adding $\boldsymbol{\theta}$ introduces additional considerations in the training procedure: while traditionally the training only requires the conditional distribution of \mathbf{x} given $\boldsymbol{\theta}$ (which is predicted by the theory and detector simulation), now the training data has some implicit prior distribution over $\boldsymbol{\theta}$ as well (which is arbitrary). When the network is used, in practice it will be to predict y conditional on both \mathbf{x} and $\boldsymbol{\theta}$, so the distribution of $\boldsymbol{\theta}$ used for training is only relevant in how it affects the quality of the resulting parameterised network, without implying a Bayesian inference.

An issue regarding pNN is about the possibility that some or all of the components of $\boldsymbol{\theta}$ may not be meaningful for a particular target class. In the MSSM analysis, for instance, the m_A parameter is not particularly meaningful for the background training examples (which obviously do not have a mass). To solve this issue, a viable solution is the assignment of the m_A parameter in the background samples, with a series of rules described in the next Section.

6.3.3 Feature description

Depending on the event category (defined in Section 6.2), several features have been selected for the model training.

If the event belongs to the no-bjet category (CAT2) the features are:

- ΔR of the dimuon pair (*dimuon_deltar*), where the dimuon object has been selected adopting the rules described in Section 6.1;

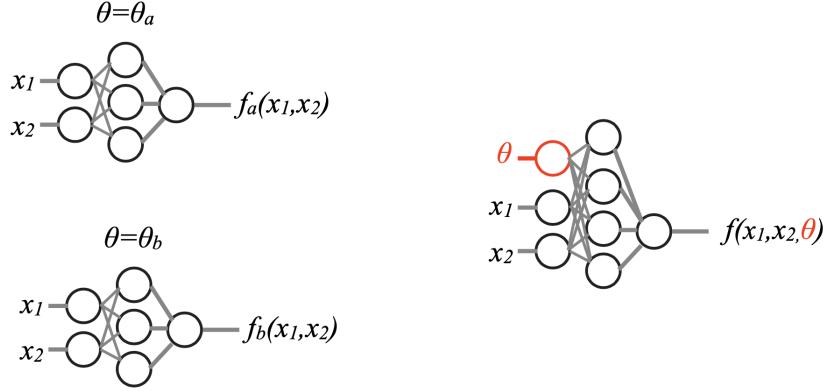


Figure 6.6: On the left, some exemplified individual networks with input features (x_1, x_2) , each one trained with a single value of some parameter $\theta = \theta_a, \theta_b$. The individual networks are purely a function of the input features. On the right, a single network trained with the same input features (x_1, x_2) as well as the input parameter θ ; such a network is trained with examples at several values of the parameter θ .

- $\Delta\eta$ of the dimuon pair (*dimuon_deltaeta*);
- $\Delta\phi$ of the dimuon pair (*dimuon_deltaphi*);
- E_{miss}^T or MET (*met_pt*), defined in Section 5.5.5, of the event;
- p_T of the highest- p_T light quark jet (*ljet_1_pt*);
- η of the highest- p_T light quark jet (*ljet_1_eta*);
- number of light quark jets in the event (*ljet_n*).

In addition, for the b-jet category (CAT1), the following features have been included during training:

- ΔR between the dimuon pair and the highest- p_T b-jet (*deltar_bjet1_dimuon*), where a b quark jet is defined via the DeepJet discriminator described in Section 5.5.4;
- $\Delta\eta$ between the dimuon pair and the highest- p_T b-jet (*deltaeta_bjet1_dimuon*);
- $\Delta\phi$ between the dimuon pair and the highest- p_T b-jet (*deltaphi_bjet1_dimuon*);
- Δp_T between the dimuon pair and the highest- p_T b-jet (*deltapt_bjet1_dimuon*);
- p_T of the highest- p_T b quark jet (*ljet_1_pt*);
- η of the highest- p_T b quark jet (*ljet_1_eta*);
- number of b quark jets in the event (*bjet_n*).

Since the network approach is parameterised, an additional feature has been added, acting as *parameter θ* of the pNN:

- m_A , or the mass hypothesis of the pseudoscalar boson A of the 2HDM model (explained in Section 2.2.1). For the signal, it is extrapolated from the MC at generator level (in particular from the powheg cards used for the production [190–193]) and has a unique discrete value, as shown in Figures 6.7 and 6.8, while for the background it is assigned to a specific m_A depending on a specific range in the dimuon invariant mass $m_{\mu^+\mu^-}$, as listed below:

- $m_{\mu^+\mu^-} = (115, 180)$ GeV $\rightarrow m_A = 130$ GeV;
- $m_{\mu^+\mu^-} = (115, 200)$ GeV $\rightarrow m_A = 150$ GeV;
- $m_{\mu^+\mu^-} = (120, 220)$ GeV $\rightarrow m_A = 170$ GeV;
- $m_{\mu^+\mu^-} = (150, 250)$ GeV $\rightarrow m_A = 200$ GeV;
- $m_{\mu^+\mu^-} = (200, 300)$ GeV $\rightarrow m_A = 250$ GeV;
- $m_{\mu^+\mu^-} = (225, 375)$ GeV $\rightarrow m_A = 300$ GeV;
- $m_{\mu^+\mu^-} = (275, 425)$ GeV $\rightarrow m_A = 350$ GeV;
- $m_{\mu^+\mu^-} = (300, 500)$ GeV $\rightarrow m_A = 400$ GeV;
- $m_{\mu^+\mu^-} = (350, 550)$ GeV $\rightarrow m_A = 450$ GeV;
- $m_{\mu^+\mu^-} = (350, 650)$ GeV $\rightarrow m_A = 500$ GeV;
- $m_{\mu^+\mu^-} = (400, 800)$ GeV $\rightarrow m_A = 600$ GeV;
- $m_{\mu^+\mu^-} = (500, 900)$ GeV $\rightarrow m_A = 700$ GeV;
- $m_{\mu^+\mu^-} = (600, 1000)$ GeV $\rightarrow m_A = 800$ GeV;
- $m_{\mu^+\mu^-} = (700, 1800)$ GeV $\rightarrow m_A = 1000$ GeV;
- $m_{\mu^+\mu^-} = (700, 1800)$ GeV $\rightarrow m_A = 1200$ GeV;
- $m_{\mu^+\mu^-} = (700, 1800)$ GeV $\rightarrow m_A = 1500$ GeV.

The size of the interval increases at higher values of m_A : this is caused by a limited statistics of the background samples available at high invariant mass as well as to include a much wider signal distribution at higher m_A values. In order to partially compensate the background lack of statistics at high mass values, a few additional samples have been included for the pNN model training, containing MC events at specific binned mass values both for DY and $t\bar{t}$ processes (listed in detail in Table 5.3), visible also in the bumps of Figures 6.9 and 6.10.

The distributions of the above-mentioned features, for both signal and background, are shown in Figure 6.11 for CAT1 events and Figure 6.12 for CAT2 events.

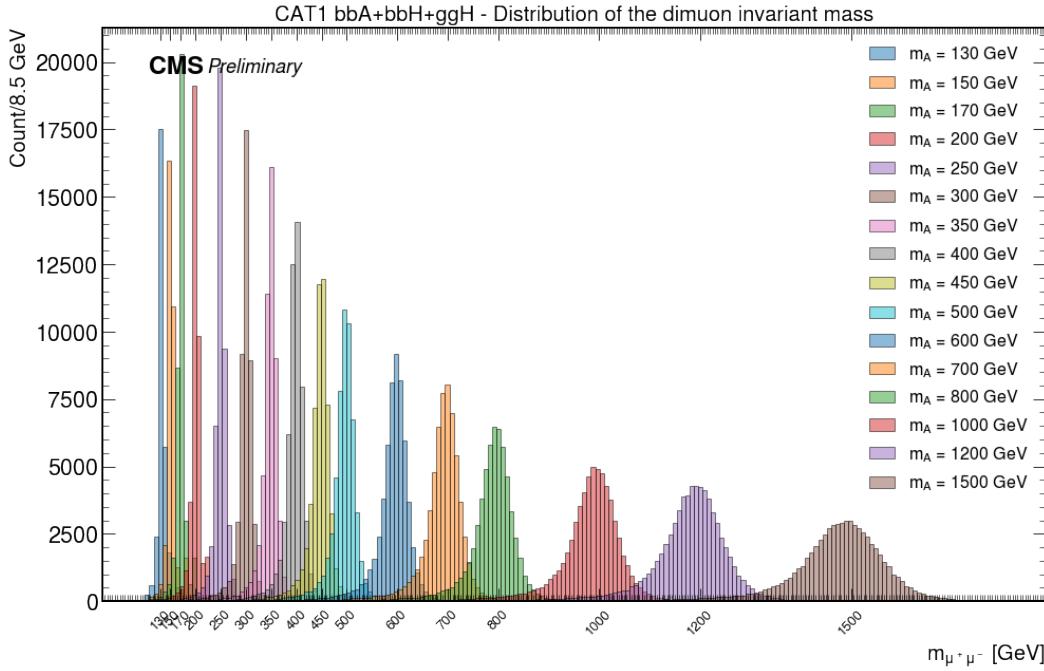


Figure 6.7: Distribution of the dimuon invariant mass $m_{\mu^+\mu^-}$, for the b-associated and gluon fusion signal production of the H and A bosons, for the CAT1 event category. The different colours, in the legend, show the exact m_A hypothesis parameter assigned for the pNN training.

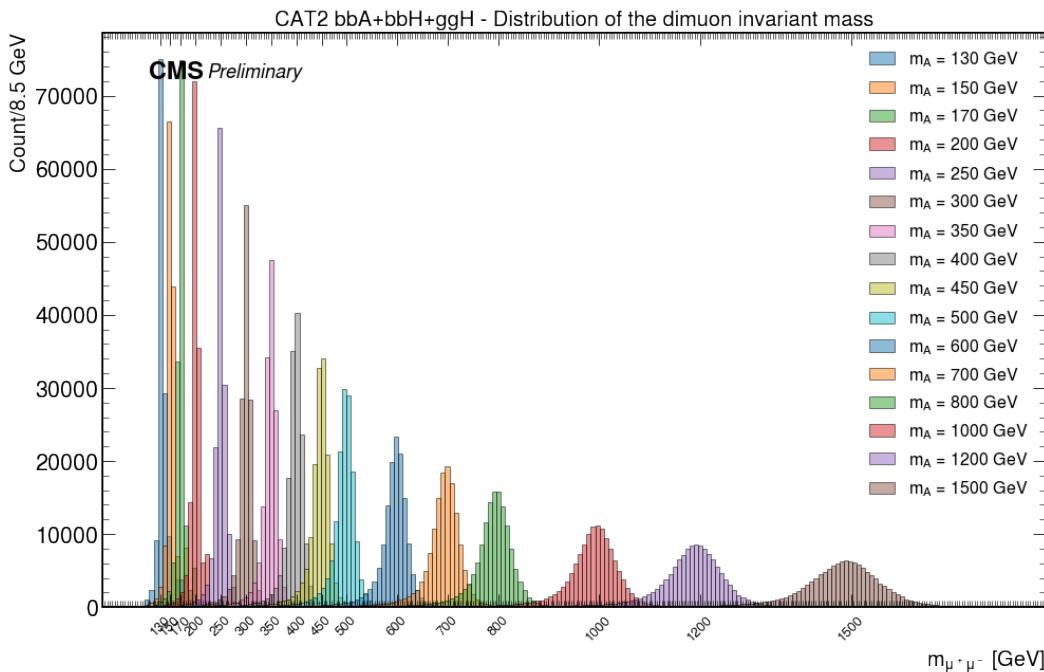


Figure 6.8: Distribution of the dimuon invariant mass $m_{\mu^+\mu^-}$, for the b-associated and gluon fusion signal production of the H and A bosons, for the CAT2 event category. The different colours, in the legend, show the exact m_A hypothesis parameter assigned for the pNN training.

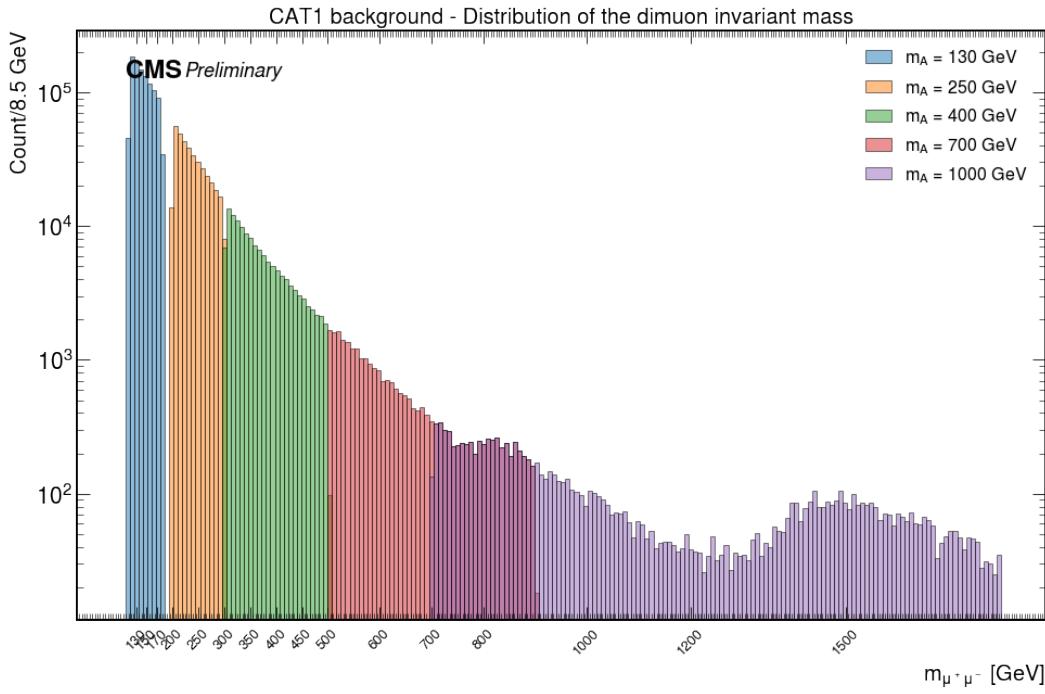


Figure 6.9: Distribution of the dimuon invariant mass $m_{\mu^+\mu^-}$, for the total amount of background samples, for the CAT1 event category. The different colours, in the legend, show the exact m_A hypothesis parameter assigned for the pNN training (only in a few m_A examples, for the sake of clarity). The y-axis is shown in logarithmic scale due to the few samples available at very high masses: the bumps visible are caused by the additional statistics provided by the $t\bar{t}$ and DY binned samples.

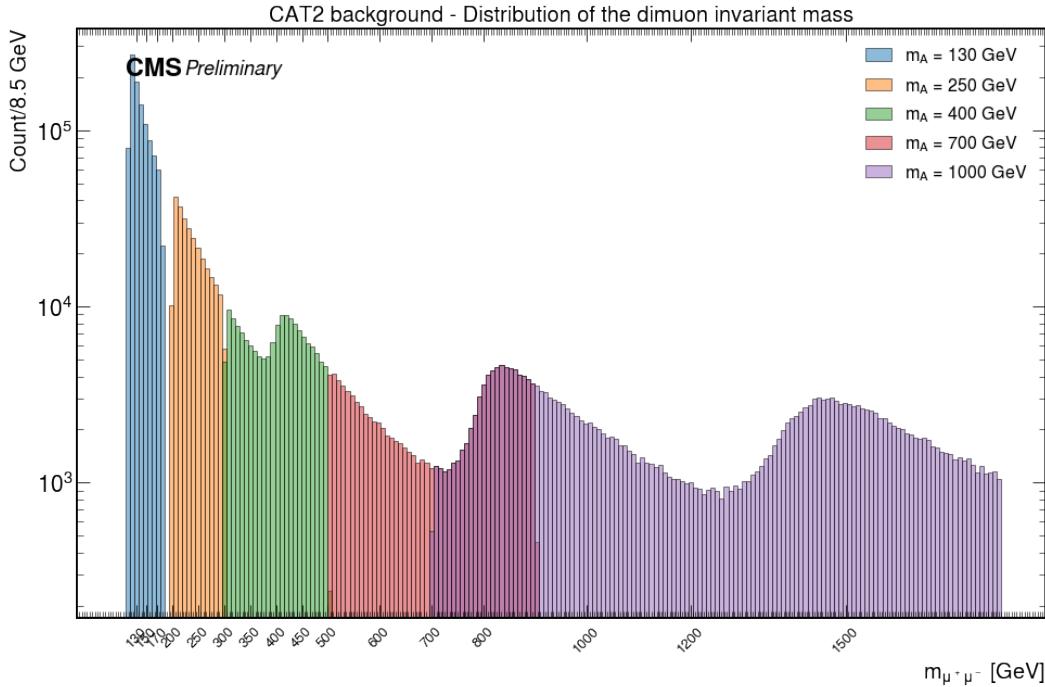


Figure 6.10: Distribution of the dimuon invariant mass $m_{\mu^+\mu^-}$, for the total amount of background samples, for the CAT2 event category. The different colours, in the legend, show the exact m_A hypothesis parameter assigned for the pNN training (only in a few m_A examples, for the sake of clarity). The y-axis is shown in logarithmic scale due to the few samples available at very high masses: the bumps visible are caused by the additional statistics provided by the $t\bar{t}$ and DY binned samples.

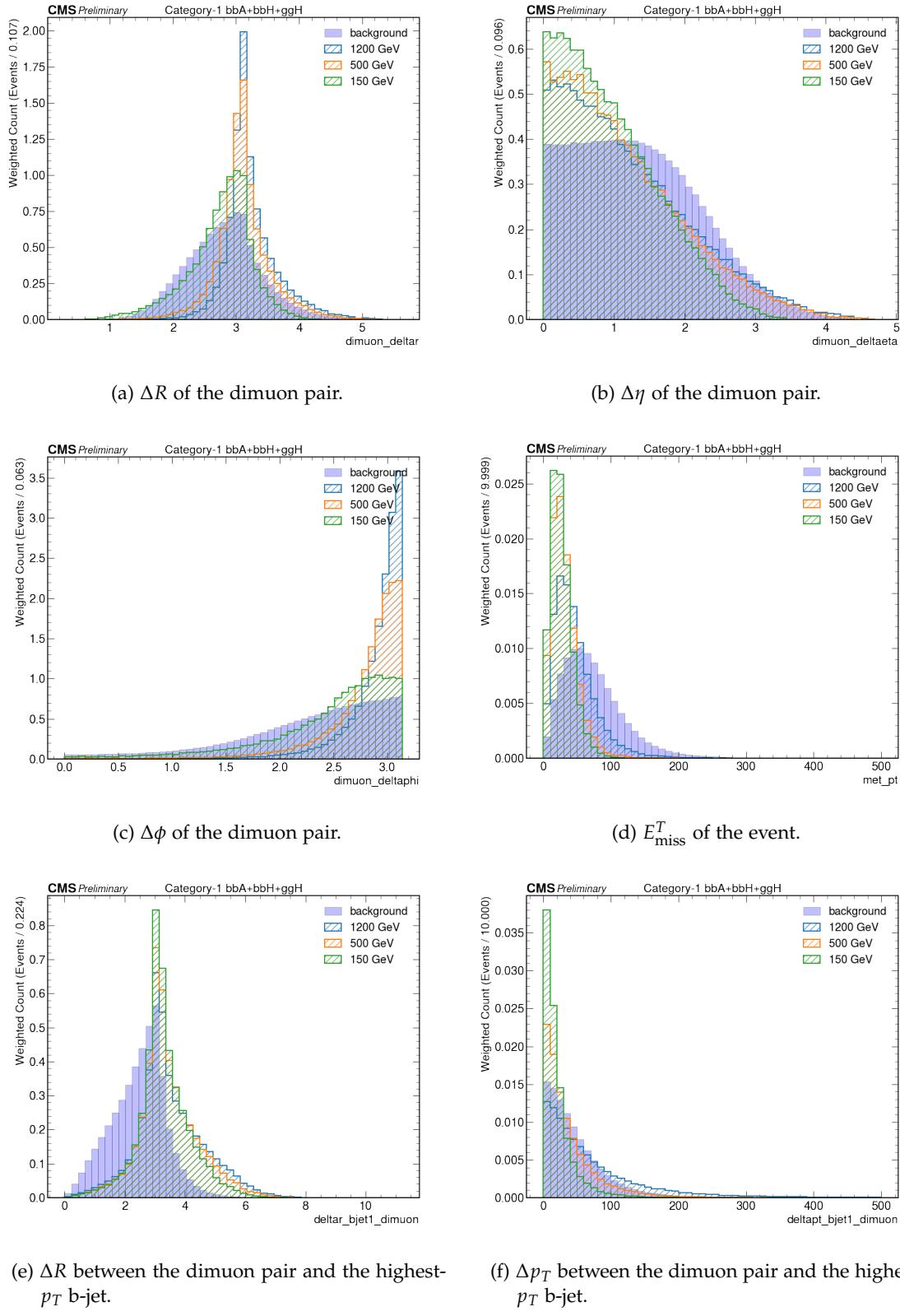


Figure 6.11: Distribution of the features used for the pNN training, for CAT1. Concerning the signal samples, only a few values of m_A are shown: in a low (150 GeV in green), middle (500 GeV in orange) and high (1200 GeV in blue) mass value. [Continues next page]

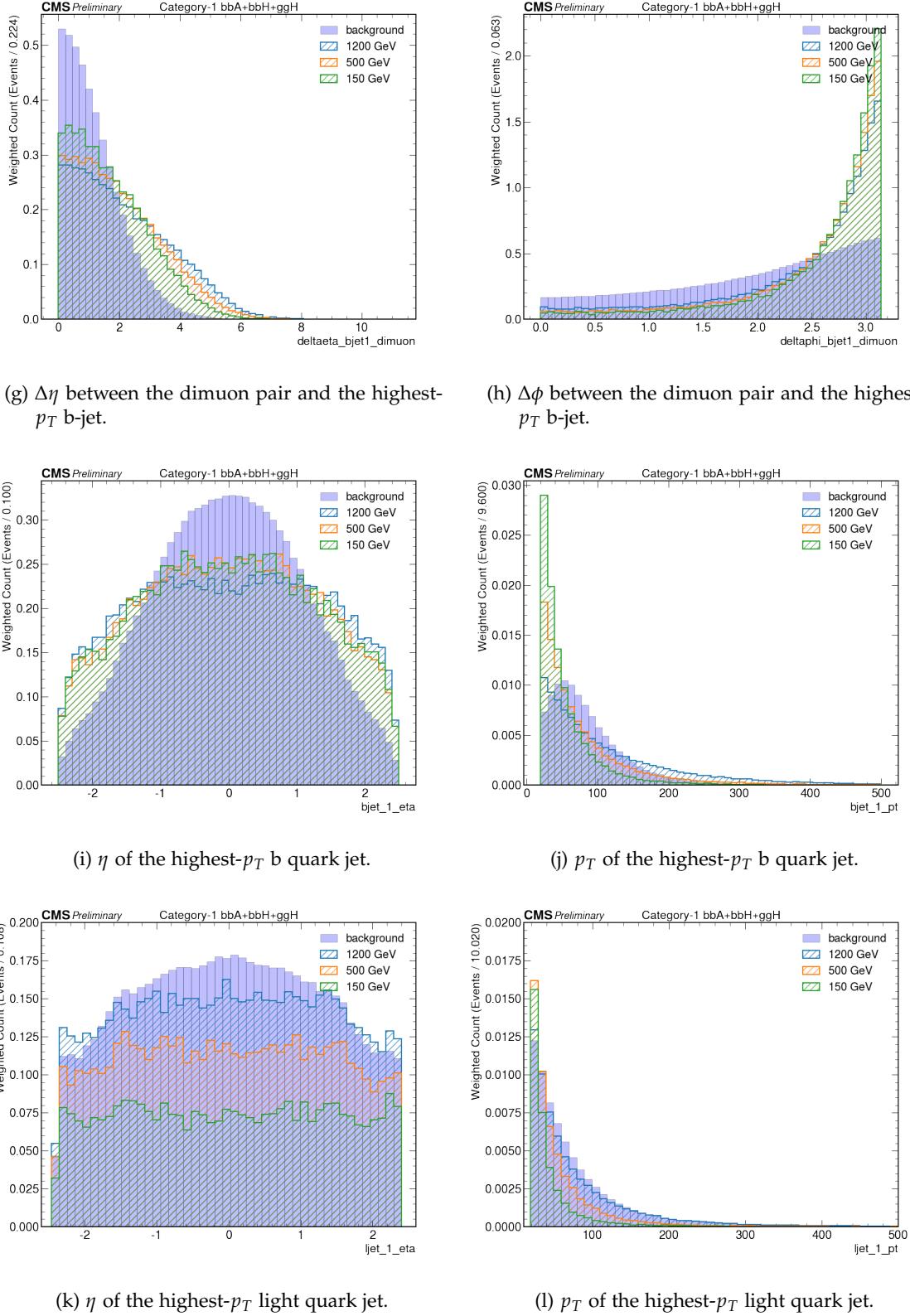


Figure 6.11: Distribution of the features used for the pNN training, for CAT1. Concerning the signal samples, only a few values of m_A are shown: in a low (150 GeV in green), middle (500 GeV in orange) and high (1200 GeV in blue) mass value. [Continues next page]

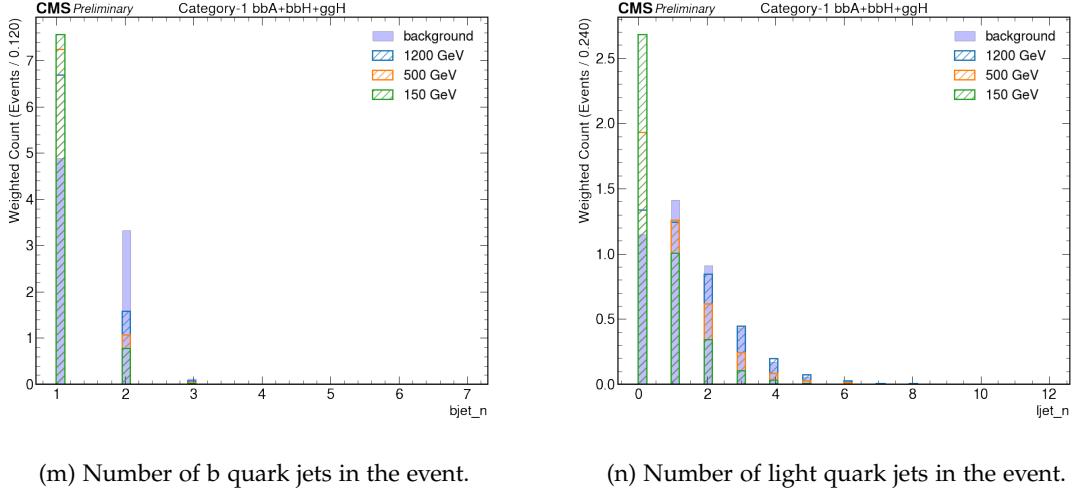


Figure 6.11: Distribution of the features used for the pNN training, for CAT1. Concerning the signal samples, only a few values of m_A are shown: in a low (150 GeV in green), middle (500 GeV in orange) and high (1200 GeV in blue) mass value.

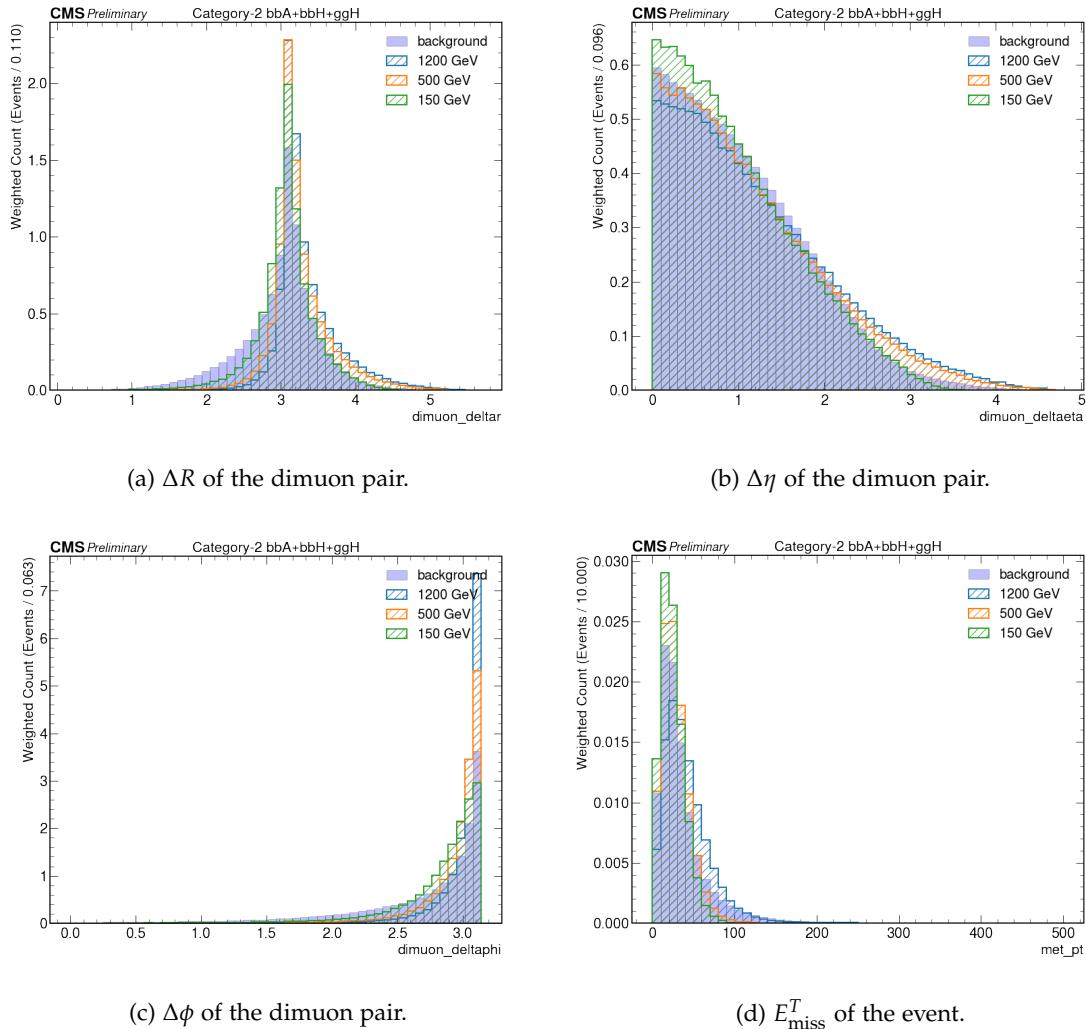


Figure 6.12: Distribution of the features used for the pNN training, for CAT2. Concerning the signal samples, only a few values of m_A are shown: in a low (150 GeV in green), middle (500 GeV in orange) and high (1200 GeV in blue) mass value. [Continues next page]

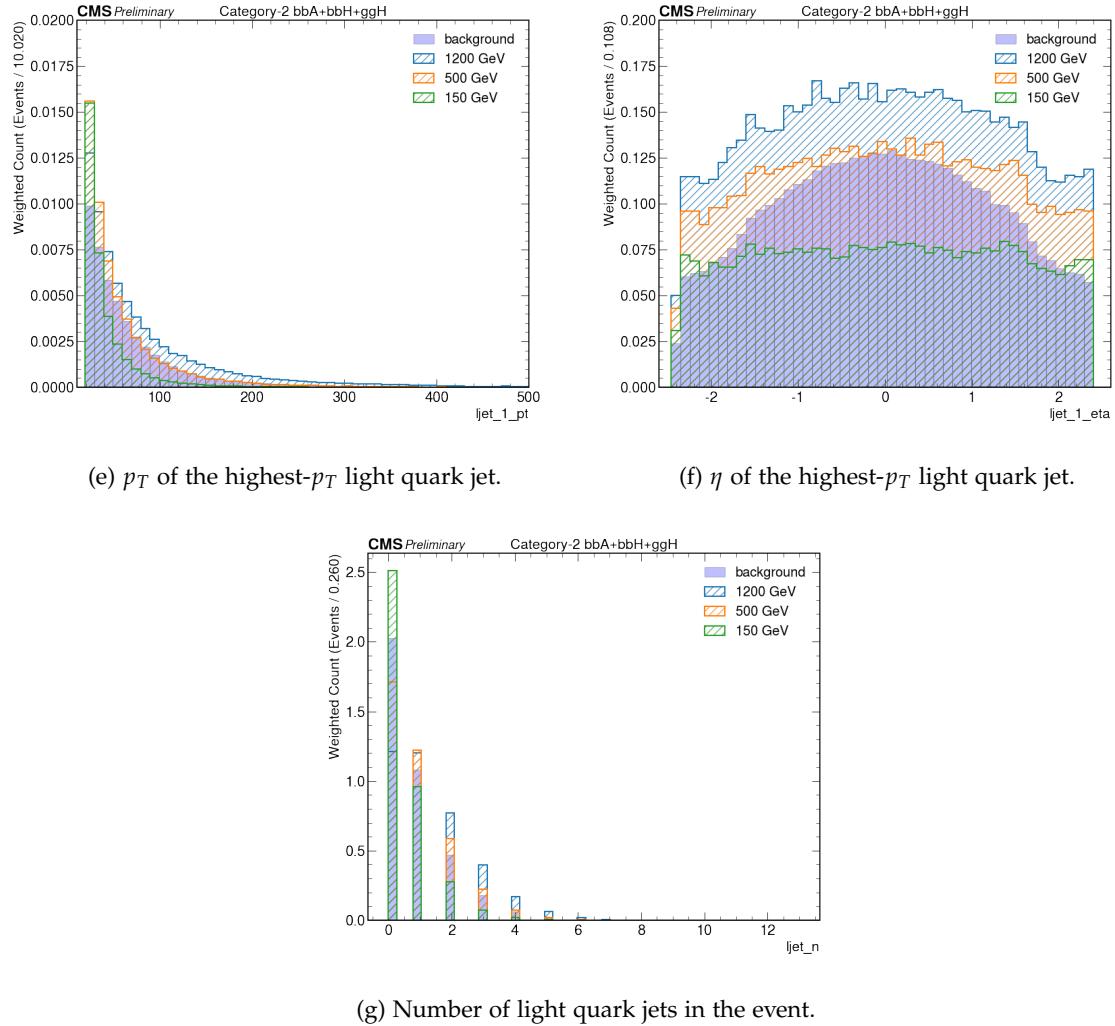


Figure 6.12: Distribution of the features used for the pNN training, for CAT2. Concerning the signal samples, only a few values of m_A are shown: in a low (150 GeV in green), middle (500 GeV in orange) and high (1200 GeV in blue) mass value.

6.3.4 Feature correlation

In the parameterised Neural Network literature [224], the choice for the parameter θ in the background class is randomly assigned according to the same distributions of the signal class. For the purpose of this analysis, however, this particular choice of assignment would be possible only if the features, other than being physically relevant for the discrimination between signal and background events, are uncorrelated with the $\mu^+\mu^-$ invariant mass. Any correlation, in fact, will provide a strong discrimination due to the highly different distribution between the signal (shaped as a Breit-Wigner distribution [227]) and the background (shaped as a exponentially decaying distribution) dimuon invariant mass, biasing the network to learn primarily only from this particular variable and not from the ensemble of the training features. This training bias will result later in a poor network generalisation capability, using new and unseen data, with an overall bad performance during the model inference. In order to check this kind of correlation, some profile histograms has been made: in these histograms, instead of the entire distribution, only the mean value of the y-axis (representing the feature x_i) and relative error is shown, for each bin of the x-axis (representing the dimuon invariant mass). In case of correlations, a non flat behaviour is expected. Figures 6.13 and 6.14 show each feature profile histogram with respect to the dimuon invariant mass, for CAT1 and CAT2 events respectively: from these histograms, it is possible to observe a slight correlation in some of the features used for training the network.

For this reason a different choice for the background m_A parameter initialisation has been made, compared to the original prescription: instead of a random sampling, the value has been assigned according to a specific dimuon invariant mass interval, as previously mentioned in Section 6.3.3.

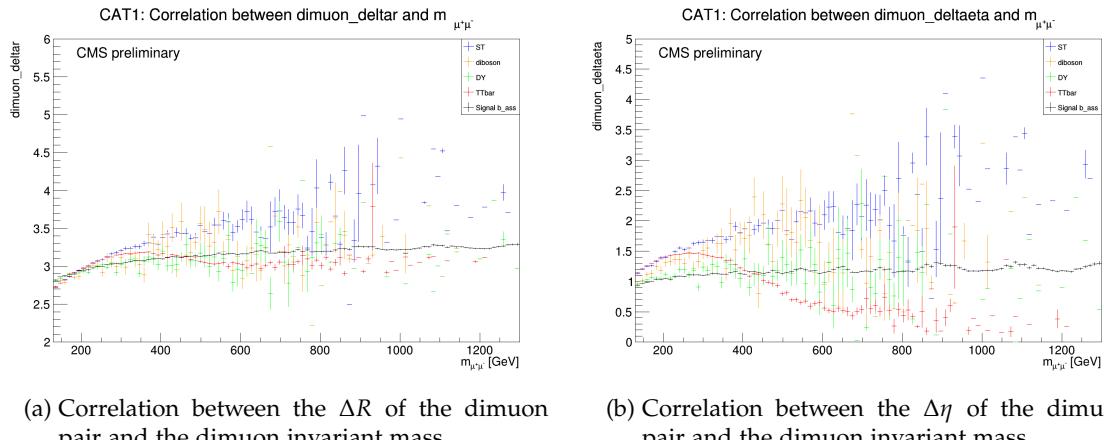


Figure 6.13: Profile histogram showing the correlation between the features of CAT1 (shown in the y-axis) and the dimuon invariant mass (shown in the x-axis). For each bin in the x-axis, the mean value and its error of the y-axis is shown. The signal is shown in black (for both the H and A bosons), while the background events are shown in different colours. [Continues next page]

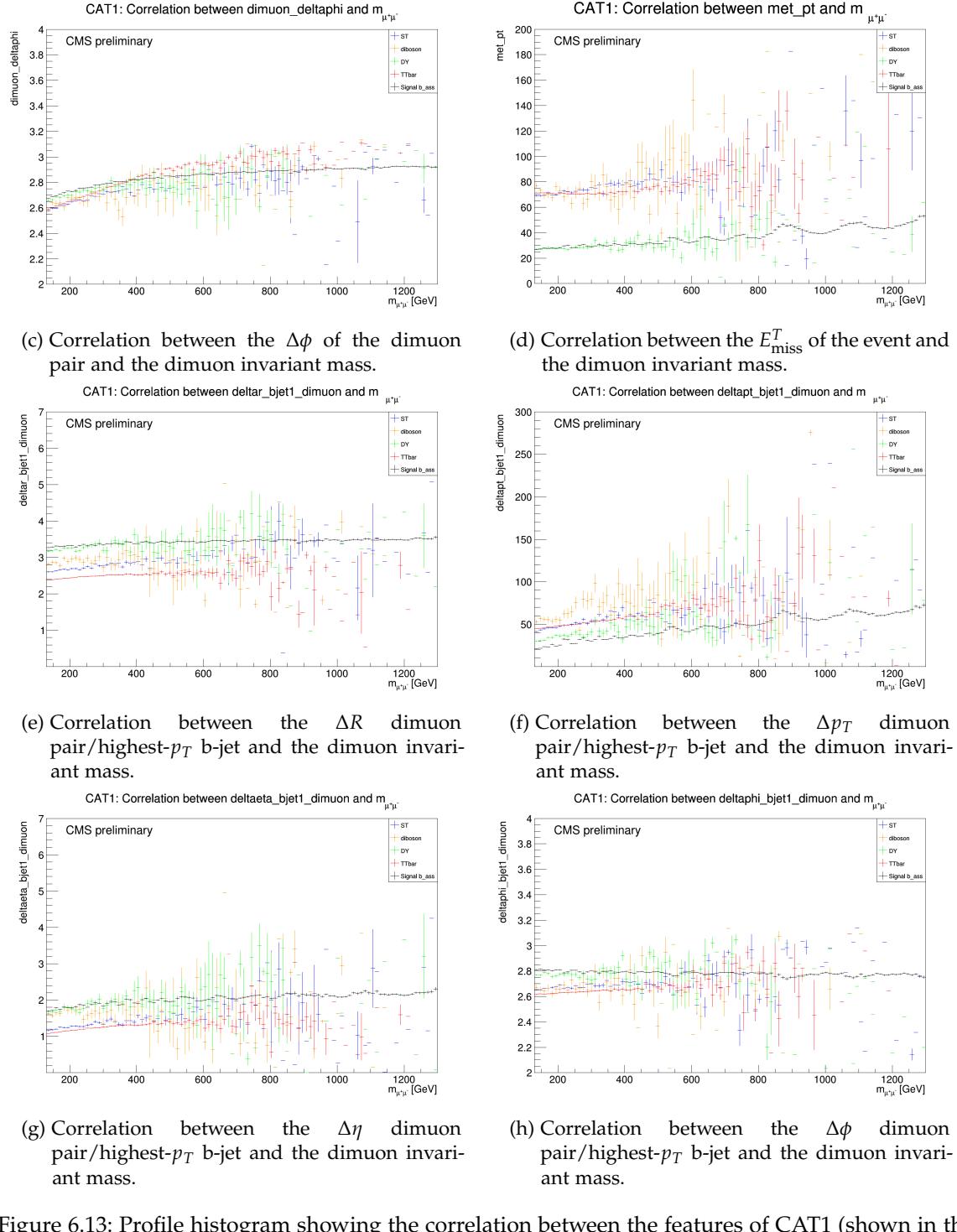


Figure 6.13: Profile histogram showing the correlation between the features of CAT1 (shown in the y-axis) and the dimuon invariant mass (shown in the x-axis). For each bin in the x-axis, the mean value and its error of the y-axis is shown. The signal is shown in black (for both the H and A bosons), while the background events are shown in different colours. [Continues next page]

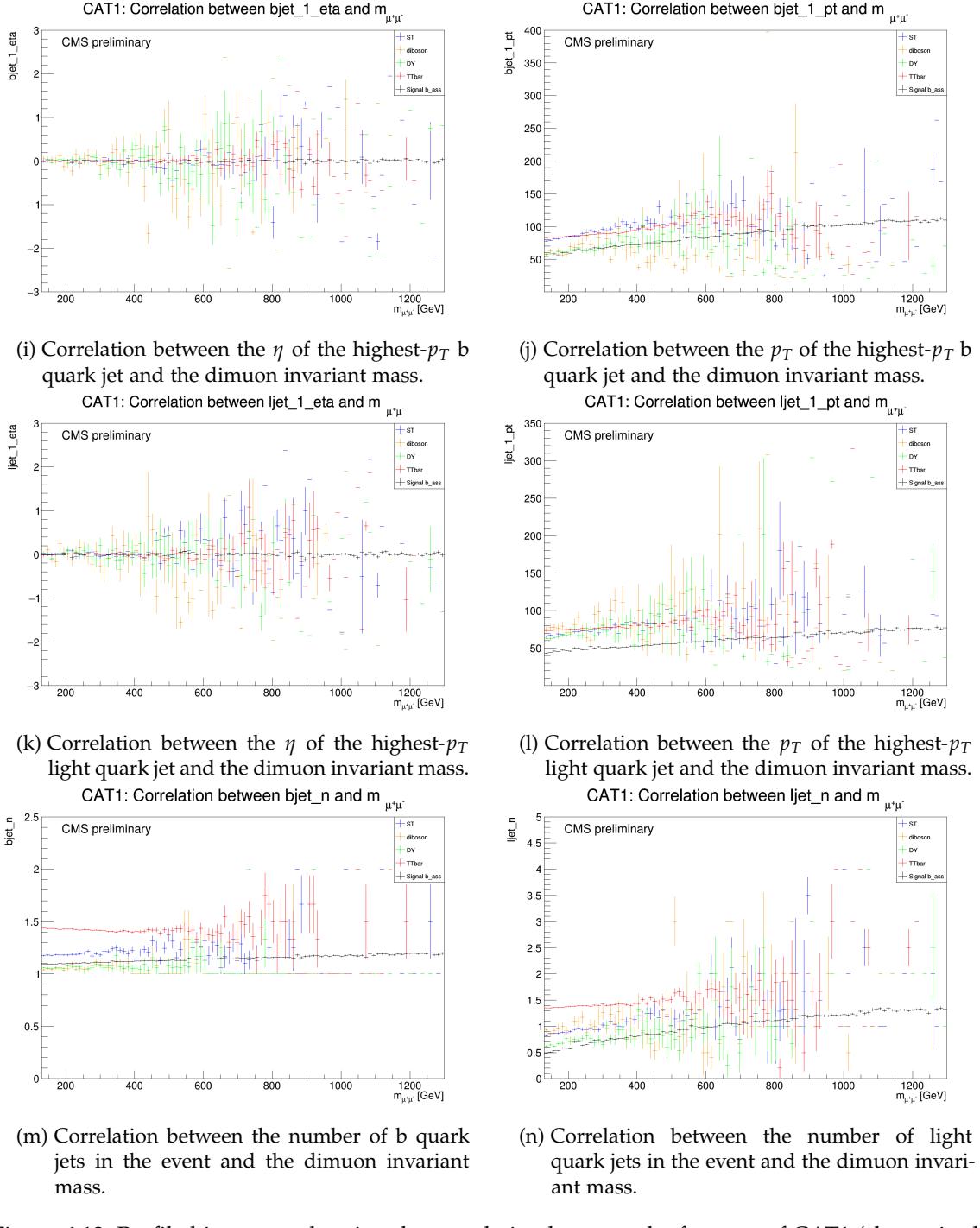


Figure 6.13: Profile histogram showing the correlation between the features of CAT1 (shown in the y-axis) and the dimuon invariant mass (shown in the x-axis). For each bin in the x-axis, the mean value and its error of the y-axis is shown. The signal is shown in black (for both the H and A bosons), while the background events are shown in different colours.

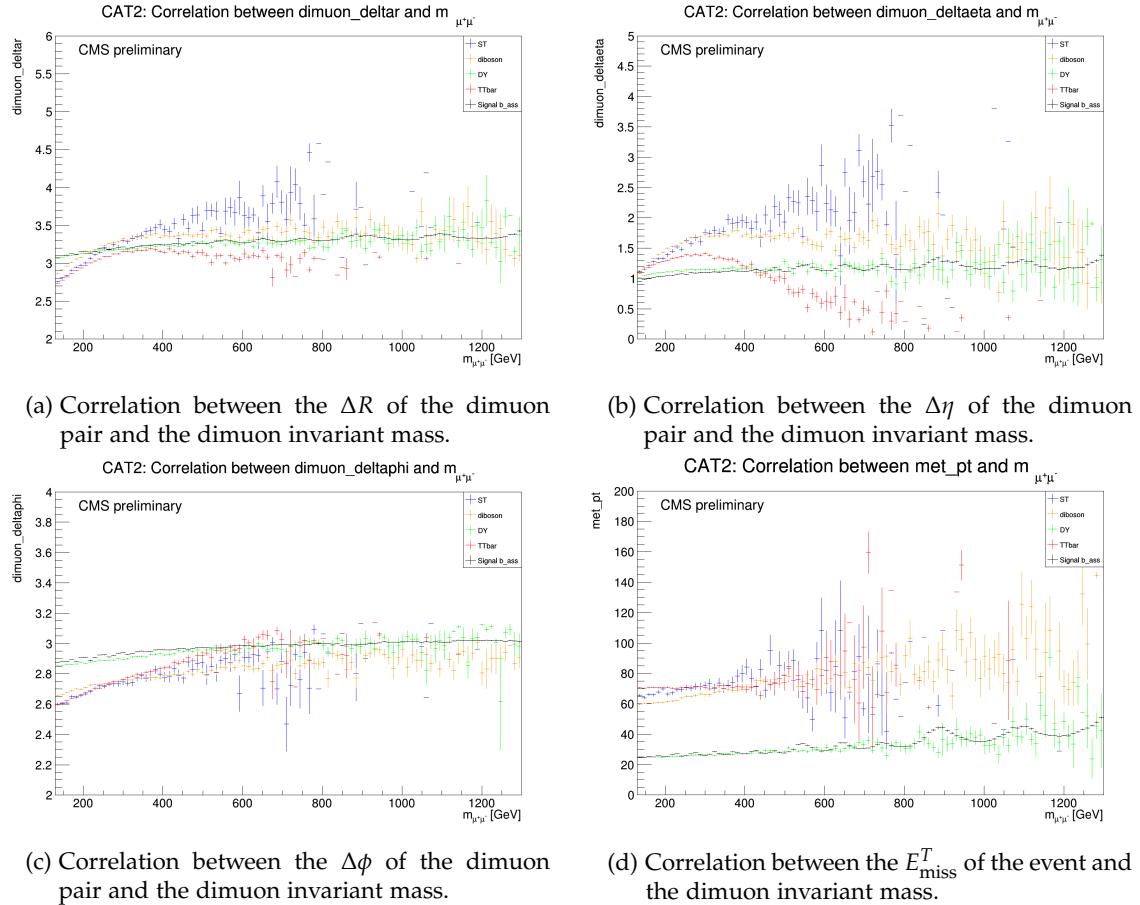


Figure 6.14: Profile histogram showing the correlation between the features of CAT2 (shown in the y-axis) and the dimuon invariant mass (shown in the x-axis). For each bin in the x-axis, the mean value and its error of the y-axis is shown. The signal is shown in black (for both the H and A bosons), while the background events are shown in different colours. [Continues next page]

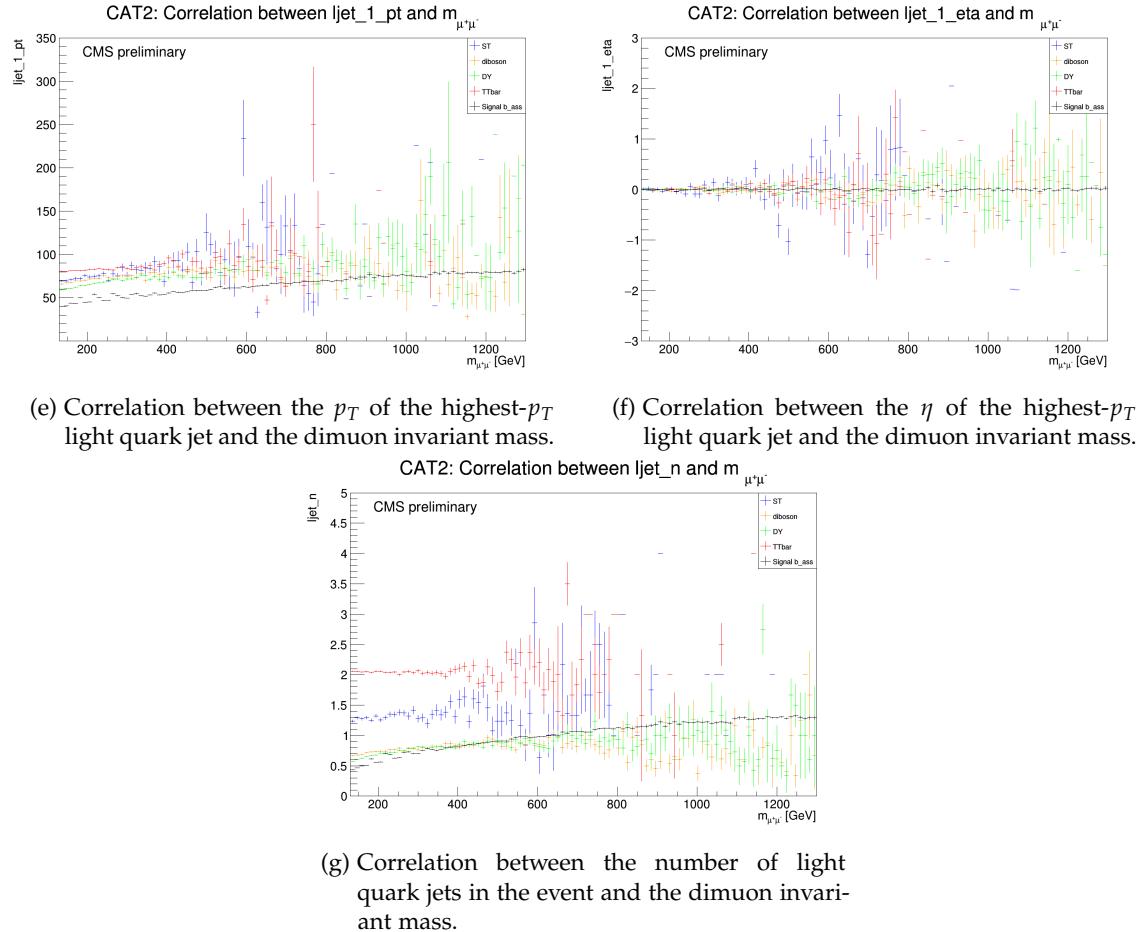


Figure 6.14: Profile histogram showing the correlation between the features of CAT2 (shown in the y-axis) and the dimuon invariant mass (shown in the x-axis). For each bin in the x-axis, the mean value and its error of the y-axis is shown. The signal is shown in black (for both the H and A bosons), while the background events are shown in different colours.

6.3.5 Data preprocessing

Before training the model, the Monte Carlo simulations for background and signal, as well as the Run 2 datasets, must undergo some preprocessing operations.

First, a skim on the NanoAOD ROOT files is performed, using the CMS NanoAOD central Tools [212]. Two different skims are performed:

- The first one is a data quality skim based on the certification made by the CMS Data Quality Monitoring group. This skim only affects Run 2 data (not the MC simulations) and it is encoded in a .json mask file that indicates which runs are good and which runs must be discarded (in Table 5.1, the complete file name of the .json mask is shown);
- The second one is a skim mainly performed as a pre-selection cut on the muon objects: only the events with two muons - one with $p_T > 20$ GeV and the other with $p_T > 10$ GeV - are selected. This is done on both data and MC, and mainly for computing efficiency: since an high amount of samples is used for the analysis, the resulting total file size is difficult to handle and to be stored; after the skim, $\approx 70\%$ of the events are discarded, thus resulting in more light and easily manageable datasets.

After the skim procedure, the .root data and MC datasets are converted in a plain .csv (comma separated values) file, extracting only the features required by the pNN during training and test. This is accomplished via a python script based on *uproot* [228], a library for reading and writing ROOT files in pure Python and NumPy [229]. Unlike the standard ROOT implementation in C++, uproot is purely an I/O library, primarily intended to stream data, coming from complex ntuples, into machine learning libraries written in Python. This has multiple advantages, since a lot of tools and frameworks (explained in more detail in Section 4.7) widely used in Data Science applications, only relies on standalone Python APIs, which are not included natively in the ROOT framework¹. The resulting .csv files are then manipulated, including information like the class label (0 for background and 1 for signal) and the physical weights of the event, given as the ratio between the initial dataset event number and its relative production cross section (shown in Tables 5.2, 5.3, 5.4).

Finally, the output .csv files are stored on disk and used for the model training, explored in more detail in the next Section.

6.3.6 Model training

The pNN is trained using half of the simulated sample, so that the remaining half is used for the test (see Section 6.3.7), establishing the best cuts for the different discriminators, at each m_A value. In the training set, an additional 25% has been split (12.5% of the

¹ In the last few years, the CERN ROOT team has put a lot of effort in upgrading its multivariate analysis toolkit TMVA, by adding the support with the most commonly used learning frameworks like Keras and PyTorch. However, for the time being, the original frameworks have been preferred for the prosecution of this analysis work.

entire dataset) as a *validation* set, to perform sanity checks and avoid unwanted effect i.e. overtraining (more information regarding overtraining and split techniques are provided in Sections 4.3.2 and 4.3.4, respectively). Two separate network models have been created and trained, one per each event category: CAT1 and CAT2: the architecture is identical, changing only the features used for training.

Two different classes are used at training time: the background class, including all the backgrounds together, and the signal class, including all the m_A samples and merging together every $\tan \beta$ for each m_A value. All the input variables are standardised, subtracting the mean and dividing by the standard deviation of the original distributions, and an additional clipping on the input features has been performed to avoid uninteresting few outlier events. The training is performed in a balanced scenario, where the signal and background samples are taken in the same quantity: this is performed on the network mini-batch level, sampling each time the events randomly but in a balanced fashion, having per each m_A hypothesis the same amount of signal and background events (the background m_A assignment is defined in Section 6.3.3). Concerning the different physical background composition, also another balancing is performed, levelling the number of events per each process (diboson, $t\bar{t}$, single top and Drell-Yan).

The parameterised Neural Network architecture and tuning, has been performed using the KERAS package of the TENSORFLOW version 2 framework, described in more detail in Section 4.7. The model architecture is shown in more detail in Figure 6.15: the input features x are clipped and standardised (using the StandardScaler method of sklearn [230]), and concatenated with the mass parameter m (after a division layer for normalisation). After the concatenation, a regular feed-forward Neural Network is built, with 4 hidden dense layers of [300,150,100,50] nodes for each layer. After each dense layer, a Dropout layer is added (explained in more detail in Section 4.6.2), with a pruning percentage of 25% the layer's neurons. The loss function adopted is the *binary crossentropy*, described in Equation 4.18, with the addition of a L2 regularisation term (also explored in more detail in Section 4.6.2) with a regularisation rate of 10^{-4} for the weights and 10^{-5} for the bias terms.

Each hidden layer has a ReLU activation function (see Section 4.6), with the exception of the output layer which has a sigmoid activation function, in order to provide the output discriminating value in the [0,1] range. The stochastic gradient descent optimisation method adopted (see Section 4.6.1) is Adam, with a learning rate initialisation of $3 \cdot 10^{-4}$. The validation set, at each epoch of training, keeps under control a possible overtraining by computing the same metrics of the training set, and stopping the training if a specific metric do not satisfy a certain condition: for this parameterised model, the metric under control is the *validation AUC* (metric explored in detail on Section 4.4.4) and the stopping condition is the maximum value of such metric within a certain number of epochs (fixed to 50). The training loss trend for each epoch as well as the AUC metric trend are shown in Figure 6.16. Once the training is complete, only the model that has achieved the best performance is kept and the relative matrix weights is saved on file.

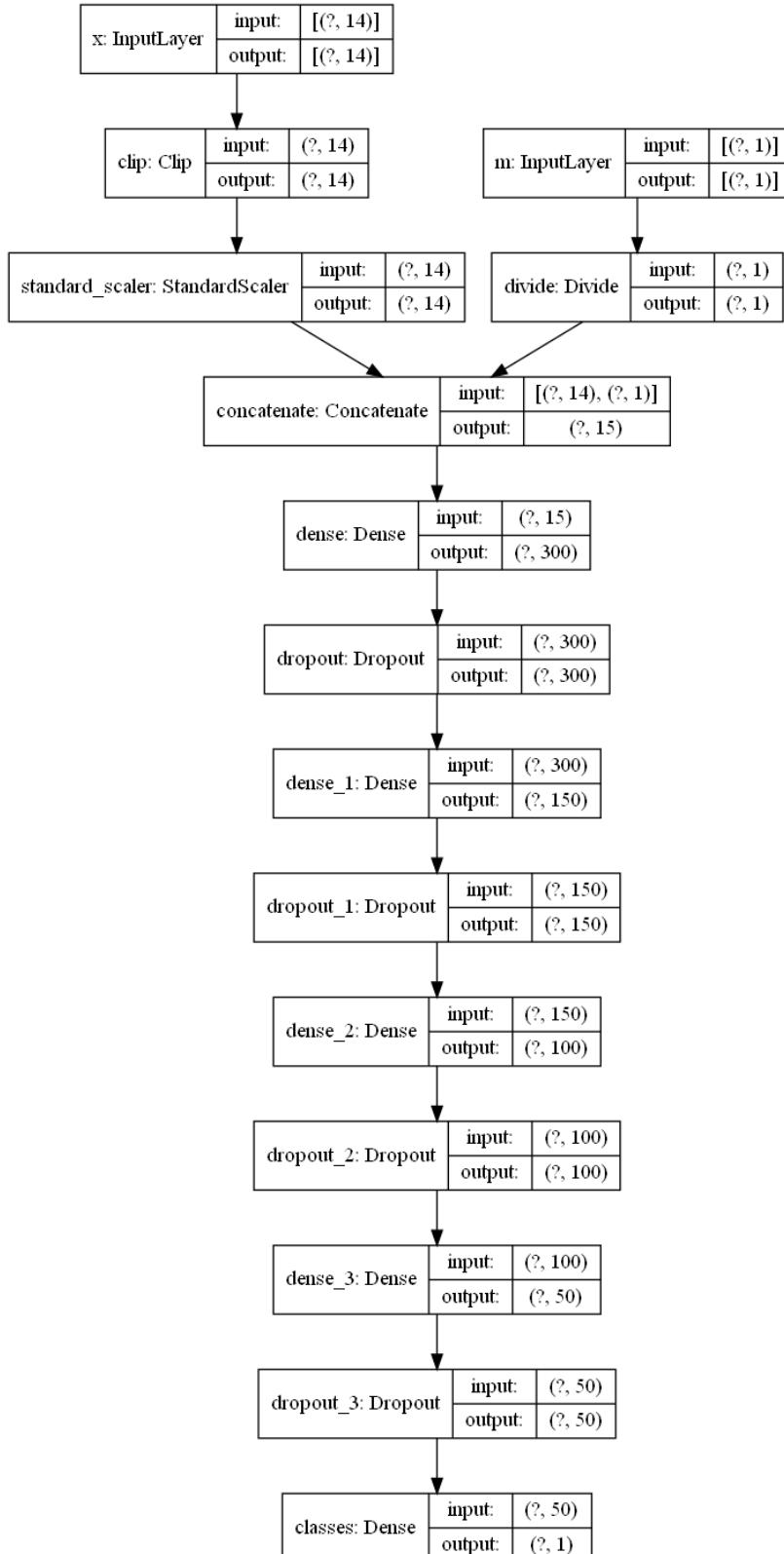


Figure 6.15: Model architecture of the parameterised Neural Network used to discriminate signal and background, for CAT1 features. The input features x are concatenated with the mass parameter m before the feed forward neural network structure: 4 hidden layers, each one followed by a dropout layer. Finally, the output layer produces the output discriminator in the $[0,1]$ interval.

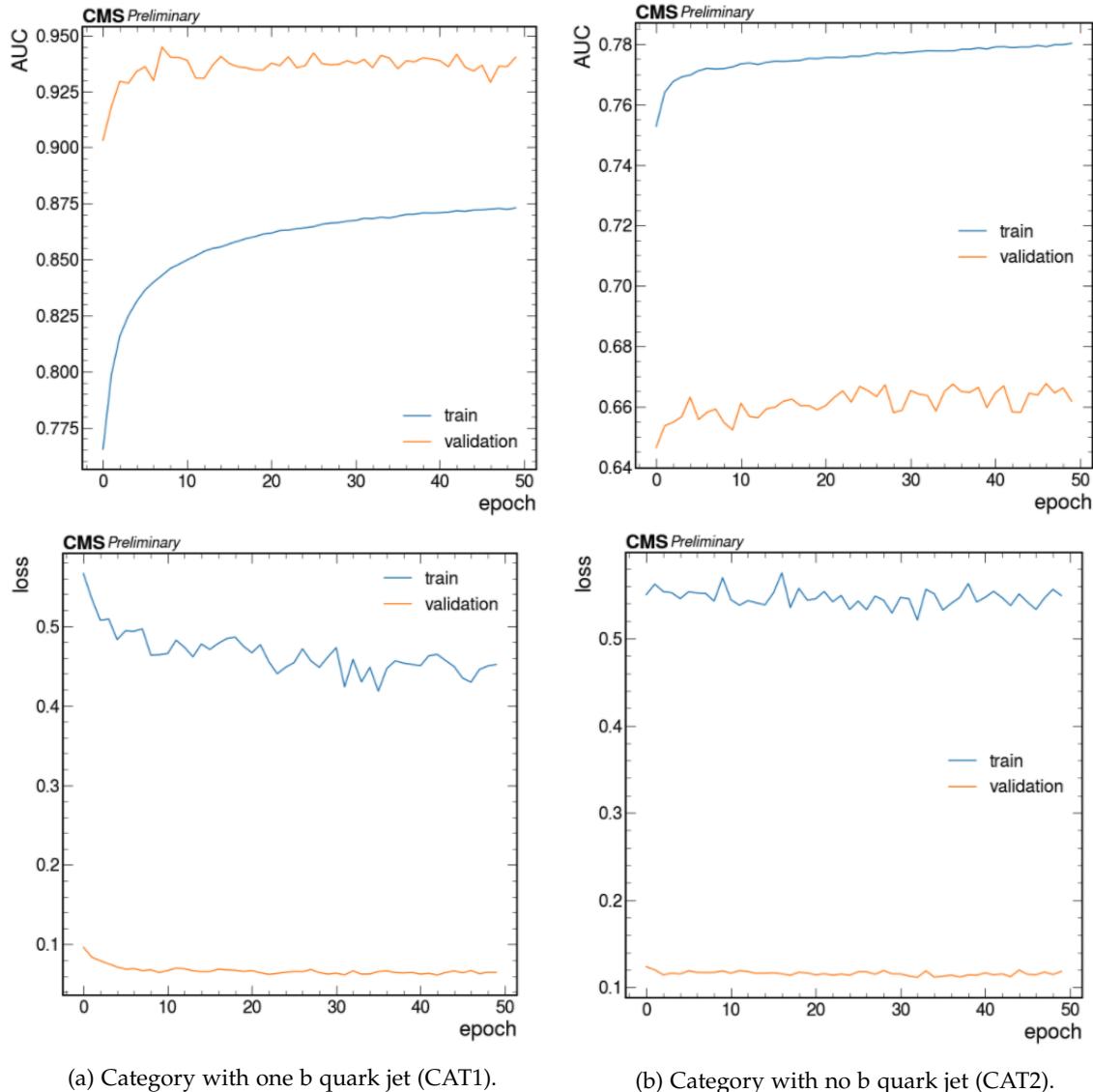


Figure 6.16: Training behaviour for each epoch, in the a) CAT1 and b) CAT2 event categorisation. The top plots show the AUC metric vs the training epoch, for the training set (in blue) and the validation set (in orange); the bottom plots show the loss vs the training epoch, with the same colour convention.

6.3.7 Testing the model for the best cut selection

After the pNN training, the final step is the test of such models in a separate and independent dataset. The KERAS model, defined and trained using the network architecture described in the previous Section, is now loaded with the best configuration and the resulting weights matrix is applied to the new unseen samples, to obtain a class prediction based on the training process. As previously stated, the test set consists of half of the simulated signal and background samples, which have been created and processed in the same way as the training sample.

The output of the parameterised Neural Network is given in a continuous range between 0 and 1, coming from the application of the sigmoid activation function, implemented in the last layer of the network. An event with an output closer to 0 will be assigned more likely as a background event, while an event with an output closer to 1 will be assigned more likely as a signal one. However, to assign uniquely an event as signal or background, a decision cut on this output must be performed. In order to choose the best cut, a metric has to be defined: usually, in a standard multivariate approach for a physical analysis, the significance S from Equation 4.10 is used. For this parameterised approach, however, the usage of such metric is not particularly meaningful, since its numeric value is not comparable between different m_A hypothesis, due to the different amount of signal and background events falling in each m_A hypothesis value. For this reason, a slightly different metric has been selected, derived from the significance: the *significance ratio*, defined as:

$$\text{SR} = \frac{S}{S_{\max}} = \frac{N_s \cdot \sqrt{N_{s_{\max}}}}{N_{s_{\max}} \cdot \sqrt{N_s + N_b}}, \quad (6.6)$$

where $S = N_s / \sqrt{N_s + N_b}$ is the standard significance, defined with a certain cut selecting N_s signal events and N_b background events, while $S_{\max} = N_{s_{\max}} / \sqrt{N_{s_{\max}}}$ is the maximum significance, defined in the ideal case of a perfect classifier which has a background rejection efficiency of 100%.

The pNN output distribution plots, for each m_A hypothesis value in the signal space (bbA, bbH and ggH) together with the relative backgrounds using the m_A assignment described in Section 6.3.3, divided for each event category (CAT1 and CAT2) are shown in Appendix B.1. For the sake of clarity, the same plots for $m_A = 500$ GeV are shown in Figure 6.17 and 6.18: the grey histogram, filled with diagonal lines, represents the network's output distribution for the signal simulated events; the coloured histograms, instead, represent the same output for the background simulated events, each colour showing a different background process weighted for its relative production cross section. The green line, superimposed, represents the computation of the significance ratio, defined in Equation 6.6, for each bin with a dashed vertical line corresponding to its maximum value, used for the network output's best cut decision.

The first category, including the events with at least one b-tagged jet, has a background mainly consisting of $t\bar{t}$ and single top events, as visible from Figure 6.17. It is also possible to notice a good discrimination between signal and background, thanks to the different

m_A (GeV)	Best cut value					
	bbA (CAT1)	bbH (CAT1)	ggH (CAT 1)	bbA (CAT2)	bbH (CAT2)	ggH (CAT2)
130	0.49	0.44	0.34	0.48	0.47	0.06
150	0.49	0.49	0.33	0.44	0.48	0.06
170	0.49	0.52	0.41	0.21	0.06	0.08
200	0.48	0.51	0.43	0.46	0.49	0.07
250	0.48	0.49	0.27	0.40	0.17	0.09
300	0.49	0.47	0.26	0.42	0.48	0.13
350	0.52	0.51	0.35	0.48	0.13	0.13
400	0.51	0.52	0.25	0.15	0.14	0.09
450	0.47	0.53	0.26	0.15	0.15	0.15
500	0.56	0.51	0.36	0.11	0.15	0.10
600	0.55	0.58	0.42	0.16	0.17	0.13
700	0.62	0.61	0.35	0.10	0.10	0.10
800	0.61	0.63	0.48	0.16	0.11	0.18
1000	0.65	0.66	0.55	0.22	0.21	0.23
1200	0.63	0.63	0.42	0.41	0.41	0.16
1500	0.67	0.67	0.52	0.40	0.31	0.31

Table 6.2: Value for the best cut, for each m_A and signal process, extracted from the evaluation of the pNN on the test set. For each of these cuts, the maximum significance ratio is achieved, as shown in the individual plots of Appendix B.1 (with the same binning configuration). In bold, the two processes selected to represent the entire event category for the overall signal efficiency.

b-tagged jets kinematics: in fact, more than half of the signal events from b-associated production are characterised by b jets emitted at large η , out of the acceptance of the tracking detector and failing the b tag requirements, whereas b jets from $t\bar{t}$ are preferentially emitted in the central η region. The second category, including the events with no b-tagged jets, has instead a background mainly consisting of Drell-Yan events, as visible from Figure 6.18, which is an irreducible background with a p_T^{miss} distribution that is similar to that of the signal. In fact, it is possible to notice a worst discriminating power compared to the first category network's output.

The list of best cuts, for each m_A and signal process is shown in Table 6.2. Despite being computed for each process (b-associated and gluon fusion) and event category, to obtain the final signal efficiencies (described in Chapter 7) only a single array of cut values has been selected per category. For CAT1 events, since the best cut values are quite similar between processes (for any given m_A value), only the bbA cuts are chosen; for CAT2 events, instead, since the ggH process has the worst cut values, it has been chosen to ensure the most conservative selection possible.

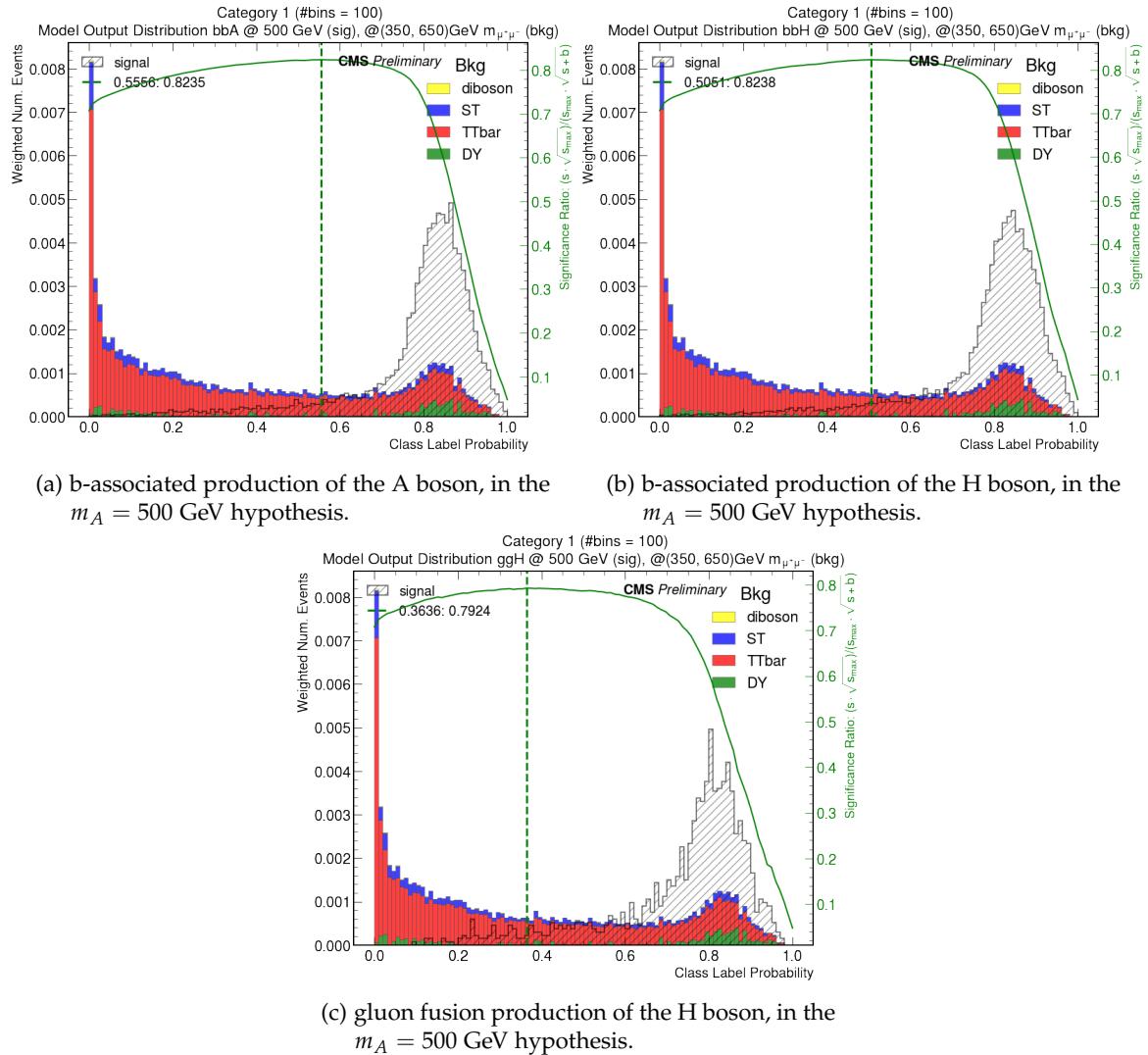


Figure 6.17: pNN output distribution histogram for the $m_A = 500$ GeV hypothesis, for CAT1 events. The grey histogram, filled with diagonal lines, represents the output network distribution for the signal simulated events; the coloured histograms, instead, represent the same output for the background simulated events, with the same m_A , each colour showing a different background process weighted for its relative production cross section. The green line, superimposed, represents the computation of the significance ratio for each bin, with a dashed vertical line corresponding to the maximum value, used for the output's best cut decision. The numerical values for the cut are also shown in the plot's left legend, with the notation: (best cut: maximum significance ratio).

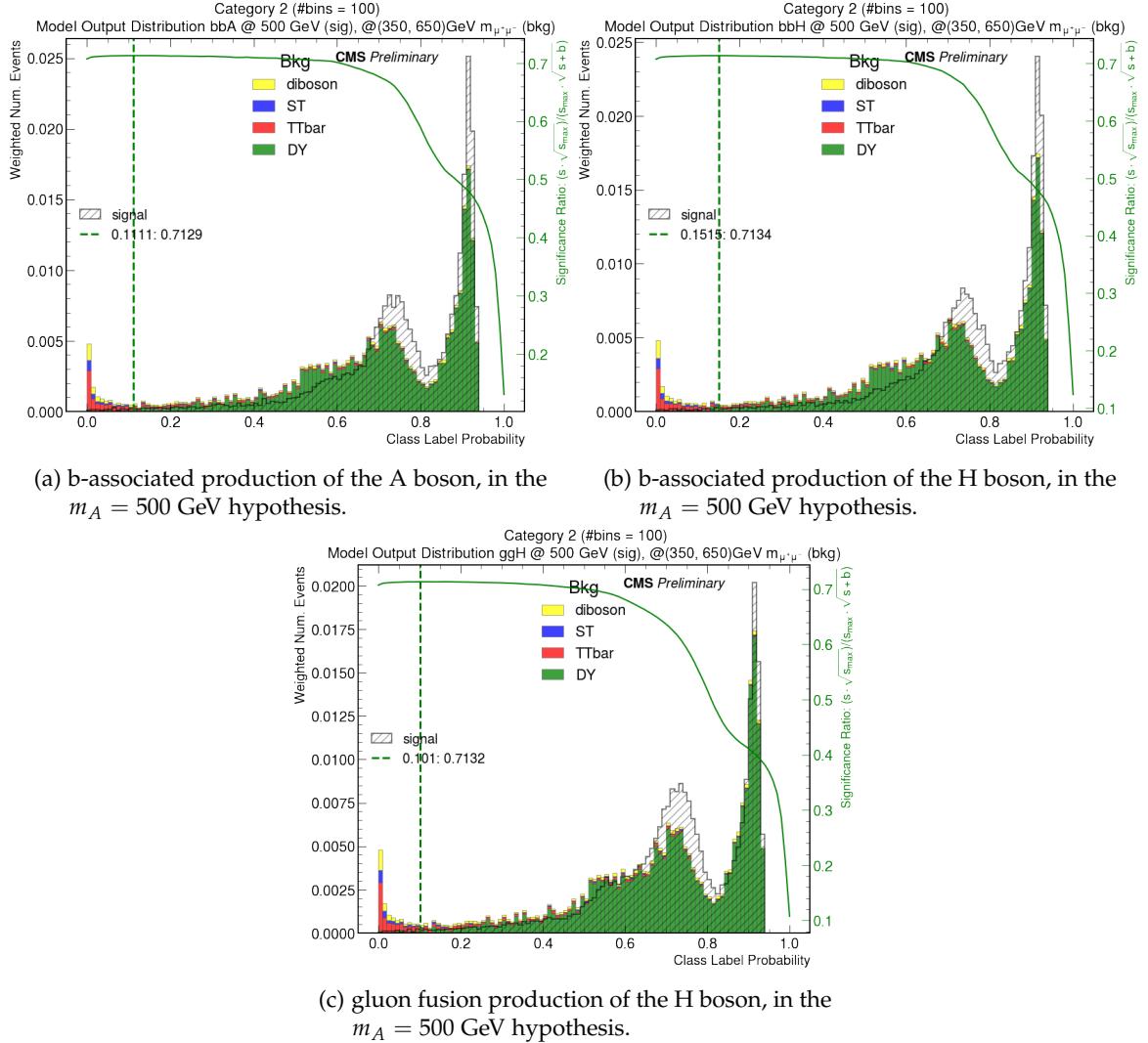


Figure 6.18: pNN output distribution histogram for the $m_A = 500$ GeV hypothesis, for CAT2 events. The grey histogram, filled with diagonal lines, represents the output network distribution for the signal simulated events; the coloured histograms, instead, represent the same output for the background simulated events, with the same m_A , each colour showing a different background process weighted for its relative production cross section. The green line, superimposed, represents the computation of the significance ratio for each bin, with a dashed vertical line corresponding to the maximum value, used for the output's best cut decision. The numerical values for the cut are also shown in the plot's left legend, with the notation: (best cut: maximum significance ratio).

6.3.8 Model performances

In Section 4.4, a description of the most commonly used metrics for checking the model quality has been described in detail. To quantify the performance of the parameterised Neural Networks, two different plots have been considered:

- The ROC curve, described in Section 4.4.4, describing the network’s signal efficiency versus the network’s background efficiency, with relative AUC;
- The PR curve, or *Precision-Recall* curve, combining the information coming from both the precision (described in Section 4.4.2) and recall (described in Section 4.4.3) metrics. This curve, in high-energy physics terms, can be also interpreted as the network’s *signal efficiency* versus the network’s *signal purity* curve.

The ROC curves, for each m_A hypothesis in the signal space (bbA, bbH and ggH) together with the relative backgrounds using the m_A assignment described in Section 6.3.3, divided for the two event categories (CAT1 and CAT2) are shown in Appendix B.2. For the sake of clarity, two plots showing three overlapped ROC curves at different m_A hypothesis (150 GeV, 500 GeV and 1200 GeV) are also shown, for the two event categories, in Figures 6.19 and 6.20. In these plots, the ROC curves of the Neural Network are compared with the performances of the cut-based selection criteria adopted in the previous analysis work [171]: the working point for each ROC curve, corresponding to the best cut selection described in the previous Section, is represented with a dot along the curve and it is possible to observe that the network’s signal efficiency is always higher, with a fixed background efficiency, compared to the cut-based working point, represented as a single dot².

The PR curves, instead, for each m_A hypothesis in the signal space (bbA, bbH and ggH) together with the relative backgrounds using the m_A assignment described in Section 6.3.3, divided for the two event categories (CAT1 and CAT2) are shown in Appendix B.3. For the sake of clarity, two plots showing three overlapped PR curves at different m_A hypothesis (150 GeV, 500 GeV and 1200 GeV) are also shown, for the two event categories, in Figures 6.21 and 6.22. Similarly to the ROC curves, also the PR curves of the Neural Network are compared with the performances of the previous cut-based selection: the working point for each PR curve, obtained with the best cut selection described in the previous Section, is represented with a dot along the curve and it is possible to observe that the network’s signal efficiency is higher, with a fixed signal purity, compared to the cut-based working point, represented as a single dot.

The combined information from the ROC and PR curves show that, in the majority of the m_A hypothesis considered, the performances of the networks are higher than a regular cut-based selection. In particular, such improvement is higher in the first category model, where the discrimination between signal and background is better, and smaller in

² The cut-based approach is described single point by definition, since it is based on a unique cut in a single variable. The multivariate approach, instead, has an infinite amount of possible cuts, based on the network output and it is plotted as a continuous ROC curve. The best cut choice, therefore, allows to fix a working point in such curve.

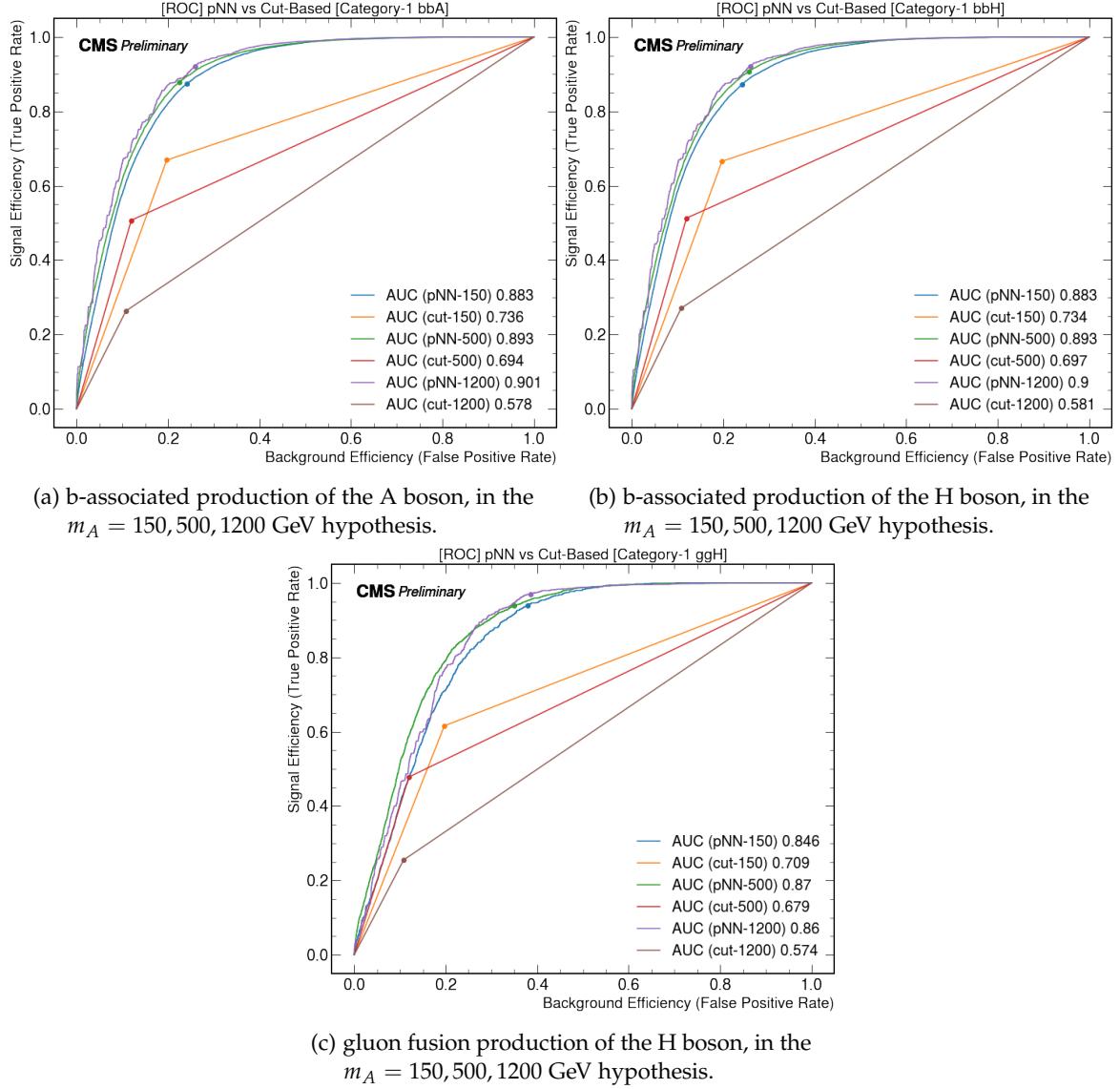


Figure 6.19: The blue, green and purple curves are the ROC for the $m_A = 150, 500, 1200$ GeV hypothesis, respectively, for CAT1 events. The AUCs of the ROC curves are shown in the plot's legend and the working points, corresponding to the best cut selections, are represented with a dot of the same colour. The orange, red and brown broken lines represent the signal efficiency versus the background efficiency for the cut-based selection method adopted in the previous analysis work, for the same m_A hypothesis: only the dots with the cut-based unique working point, have to be considered (the connecting lines are shown only for illustration purposes).

the second category model, where the irreducible Drell-Yan background causes a worst discrimination.

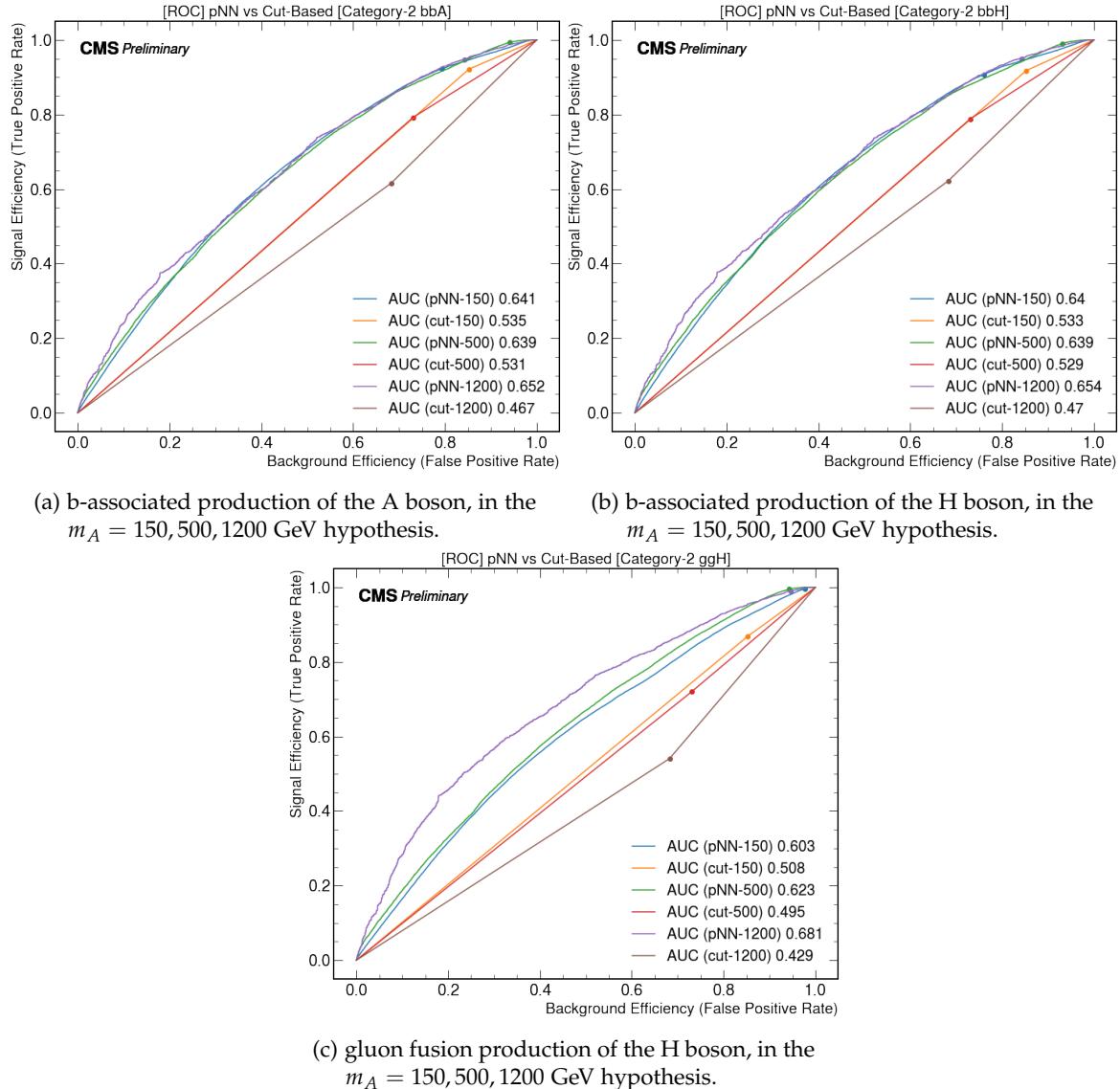


Figure 6.20: The blue, green and purple curves are the ROC for the $m_A = 150, 500, 1200$ GeV hypothesis, respectively, for CAT2 events. The AUCs of the ROC curves are shown in the plot's legend and the working points, corresponding to the best cut selections, are represented with a dot of the same colour. The orange, red and brown broken lines represent the signal efficiency versus the background efficiency for the cut-based selection method adopted in the previous analysis work, for the same m_A hypothesis: only the dots with the cut-based unique working point, have to be considered (the connecting lines are shown only for illustration purposes).

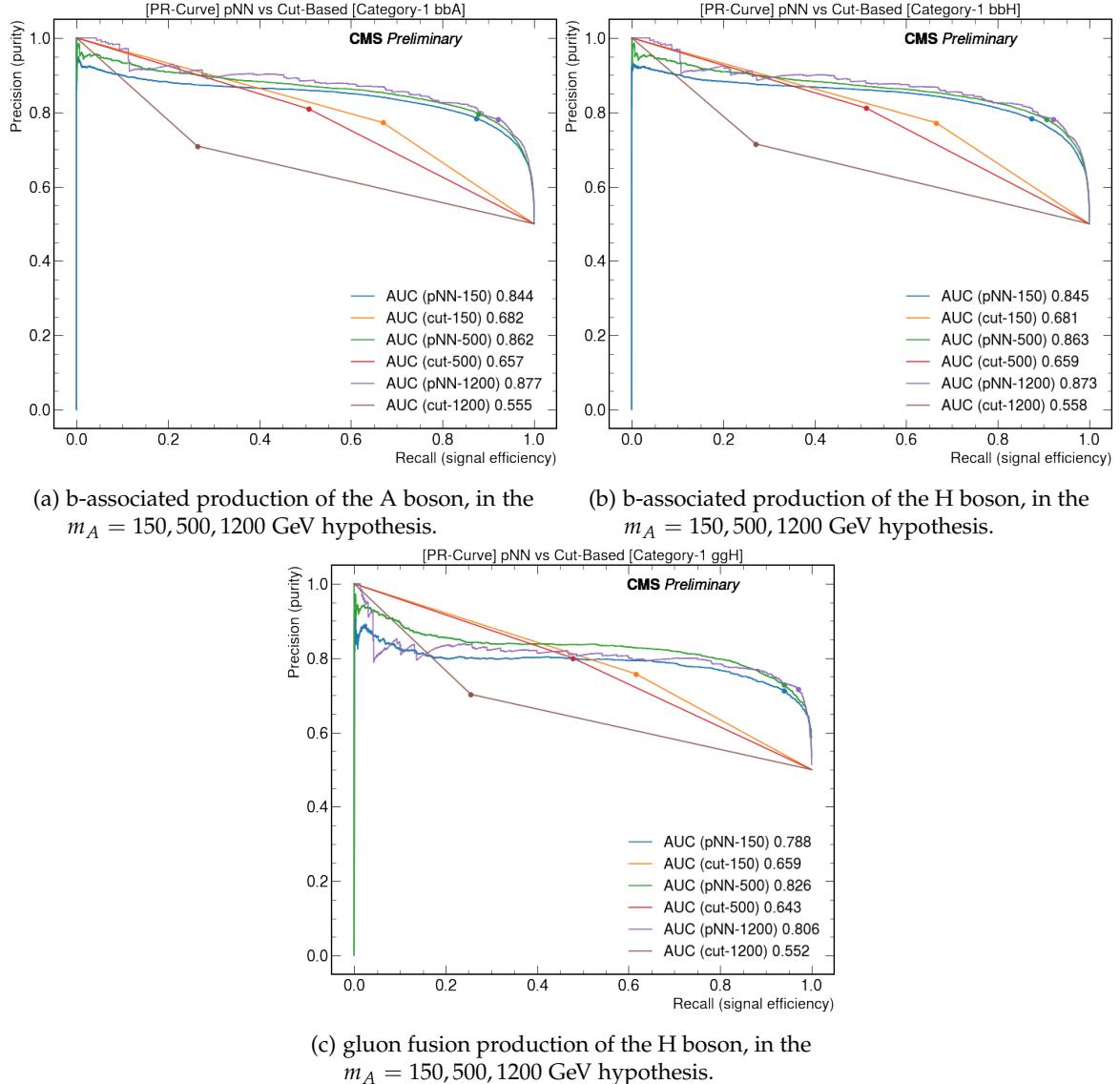


Figure 6.21: The blue, green and purple curves are the PR curves for the $m_A = 150, 500, 1200$ GeV hypothesis, respectively, for CAT1 events. The AUCs of the PR curves are shown in the plot's legend and the working points, corresponding to the best cut selections, are represented with a dot of the same colour. The orange, red and brown broken lines represent the signal efficiency versus the signal purity for the cut-based selection method adopted in the previous analysis work, for the same m_A hypothesis: only the dots with the cut-based unique working point, have to be considered (the connecting lines are shown only for illustration purposes).

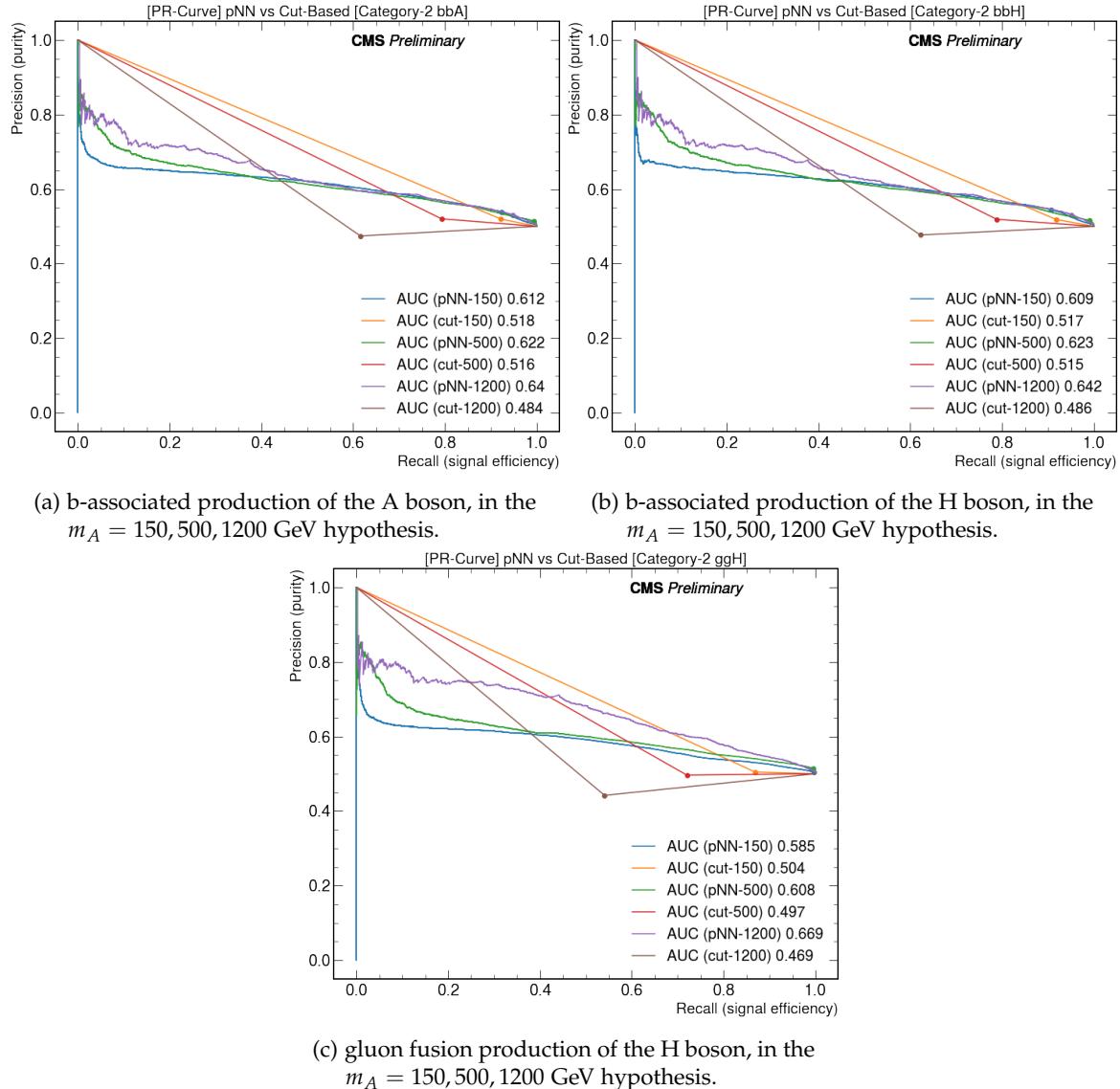


Figure 6.22: The blue, green and purple curves are the PR curves for the $m_A = 150, 500, 1200$ GeV hypothesis, respectively, for CAT2 events. The AUCs of the PR curves are shown in the plot's legend and the working points, corresponding to the best cut selections, are represented with a dot of the same colour. The orange, red and brown broken lines represent the signal efficiency versus the signal purity for the cut-based selection method adopted in the previous analysis work, for the same m_A hypothesis: only the dots with the cut-based unique working point, have to be considered (the connecting lines are shown only for illustration purposes).

6.3.9 Apply the network discriminator output to the ROOT files

After the creation of the two parameterised Neural Network models (one per category) and their test to choose the best cuts, the resulting weight matrices have been saved and later used for the computation of the network's output discriminator. This operation has been performed for every sample used in the analysis, both MC and data. For the signal MC samples, a single variable with the pNN output distribution has been created, corresponding to the sample's specific m_A hypothesis. Instead, for the background MC and data samples, multiple variables with the pNN output distributions have been written, each one corresponding to a single m_A hypothesis assigned with the interval rules described in Section 6.3.3.

A custom ROOT macro has been written, capable of writing the discriminating variables back to the original ROOT files used for the analysis. This procedure allows the prosecution of the analysis work within the ROOT framework, which is naturally suited with all the high-energy physics tools and methods for the computation of the analysis' results, like the signal selection efficiencies and the 95% CL exclusion limits for the cross section times branching ratio $\sigma \cdot B(\phi \rightarrow \mu^+ \mu^-)$ of a MSSM neutral Higgs boson ϕ , in the model independent scenario, which are shown in the next Chapter.

RESULTS

The dimuon mass distributions, obtained analysing data and simulated signal samples, are used for a statistical treatment of results. This procedure is aimed at testing the presence of a signal. A signal search is performed by testing the so-called *null hypothesis* (H_0), defined only with the background, against the *alternative hypothesis* (H_1), which includes both signal and background. In the case of setting limits, instead, the model with signal plus background becomes H_0 , tested against the background-only hypothesis, H_1 .

Searching for a possible signal, therefore, consists in the quantification of the agreement between the observed data and the null hypothesis H_0 , with a quantity called *p-value* usually expressed in its equivalent Gaussian significance. The p-value is the probability, under the assumption of the null hypothesis H_0 , of finding data of equal or greater incompatibility with the prediction of H_0 . In particle physics, the p-value is usually converted in an equivalent significance, Z , defined as the number of standard deviations that corresponds to an upper-tail probability of p , under a Gaussian distribution:

$$Z = \phi^{-1}(1 - p), \quad (7.1)$$

where ϕ^{-1} is the quantile function¹.

The discovery of a new signal process, e.g. the Higgs boson, requires the rejection of the background-only hypothesis with a significance of at least $Z = 5$, usually indicated as 5σ , where σ is the standard deviation of the Gaussian distribution. This corresponds to $p = 2.87 \cdot 10^{-7}$. For purposes of excluding a signal hypothesis, instead, a threshold p-value of 0.05 (i.e. 95% confidence level) is often used, which corresponds to $Z = 1.64$.

As introduced in Chapter 5, the analysis work described in this thesis is performed in a *model independent* scenario: compared to the MSSM benchmark scenarios described in Section 2.3, in a model independent search no assumption is made on the cross section, mass, and width of the ϕ boson, which is sought as a single resonance with mass m_ϕ . Therefore, to compute the signal efficiency for a generic ϕ boson decaying to a muon pair, the b-associated A boson (bbA) and gluon-fusion H boson (ggH) processes have been considered as a template. Also, in order to be model independent, this analysis has been performed in a narrow width approximation, for which the intrinsic width of the signal has to be smaller compared to the experimental resolution. Since the signal intrinsic width increases with $\tan\beta$, only the low value of $\tan\beta = 5$ has been used.

Finally, the upper limits on the cross section times branching ratio are shown: for the

¹ Inverse of the cumulative distribution $\phi(x) = \int_{-\infty}^x f(t)dt$ of the standardised Gaussian $f(x)$.

purpose of this thesis, which is the evaluation of the parameterised Network approach for the signal/background discrimination, only the 95% CL expected exclusion limits are shown (representing the background-only hypothesis) and compared with the same limits computed using the cut-based approach adopted in the previous MSSM analysis [171].

7.1 SIGNAL EFFICIENCY

For each value of m_A , the signal efficiency for each ϕ boson is defined as the fraction of generated events that fulfil the selection criteria. As explained in Section 5.4.3, the Pythia generator used to simulate the signals incorrectly describes the production mechanism of H and A bosons for high values of m_A and $\tan\beta$. In order to properly compute the signal efficiency and its shape, only those events with a generated mass value within three times the intrinsic signal width from the nominal mass value are considered.

The efficiency is calculated for the various steps of the event-selection, and it is separately shown for the A and H bosons, produced respectively via b-associated and gluon fusion production mechanisms. It is shown as a function of m_A , in the $\tan\beta = 5$ hypothesis which implies a narrow width Higgs signal, for the model independent scenario.

The acceptance is defined as the fraction of events with at least two muon candidates satisfying the kinematic cuts of the analysis. The acceptance as a function of m_A is shown in Figure 7.1. As expected, the acceptance increases as a function of m_A , since the average p_T of the Higgs decay products also increases with the Higgs mass.

The muon identification and isolation efficiencies as a function of m_A , are shown in Figure 7.2 and 7.3. Such efficiencies are defined as the fraction of events, that passed the previous acceptance selection, which also have at least two isolated and well reconstructed muons, according to the identification and isolation criteria defined in Section 6.1.

Once the dimuon pair selection is performed, the events are categorised according to the presence or absence of b-tagged jets, as shown in Section 6.2. The efficiency to detect at least one b-tagged jet (CAT1), is shown in Figure 7.4, and it is calculated on top of the muon selection. If the event does not contain b-tagged jets, it belongs to no b tag category (CAT2), whose efficiency is shown in Figure 7.5. The efficiency in the case of b tag category is small even for the b-associated production mechanism. This is because the b jets are preferentially emitted in the forward region of the detector, out or near the edge of the tracker-detector acceptance. Furthermore, the b-quarks of such jets have a rather soft p_T , whereas the b tagging algorithm is applied to jet with $p_T > 20$ GeV.

The last step is the efficiency of the multivariate signal selection. The pNN cut values, different for the two categories, has been already discussed in Section 6.3.7. Also in this case, the efficiency of the selection has been calculated on top of all the above-mentioned steps. The resulting efficiencies are shown in Figure 7.6 and 7.7, for the two different categories. These efficiencies must be compared with the performances of the cut-based signal selection, adopted in the previous analysis [171] and shown, for the respective categories, in Figure 7.8 and 7.9: for CAT1, it is possible to notice a significant improvement, especially in the high mass region; for CAT2, due to the irreducible Drell-Yan background, discrim-

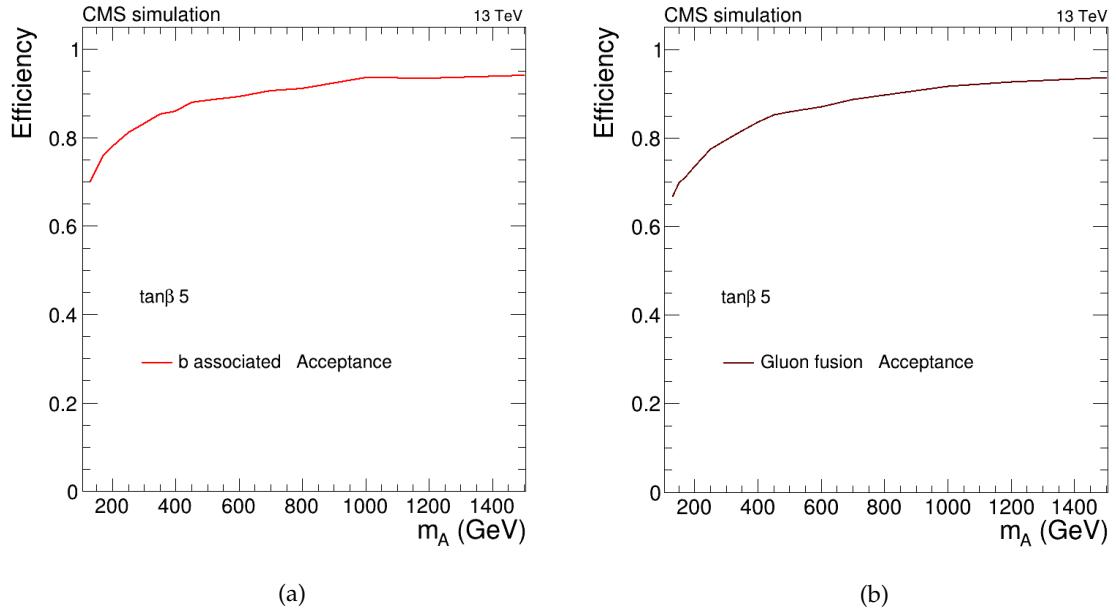


Figure 7.1: Acceptance as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

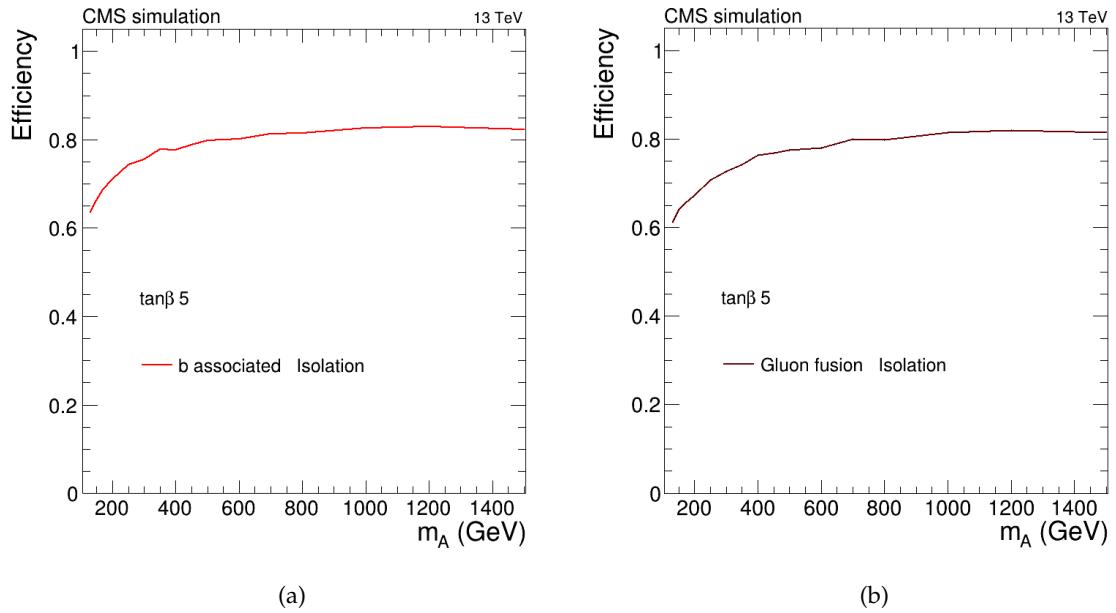


Figure 7.2: Muon isolation as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

ination is a bit harder and the efficiency is similar (only a few % higher in the high mass region) compared to the previous cut-based one. This difference is also visible from the pNN output distributions shown in Appendix B.1.

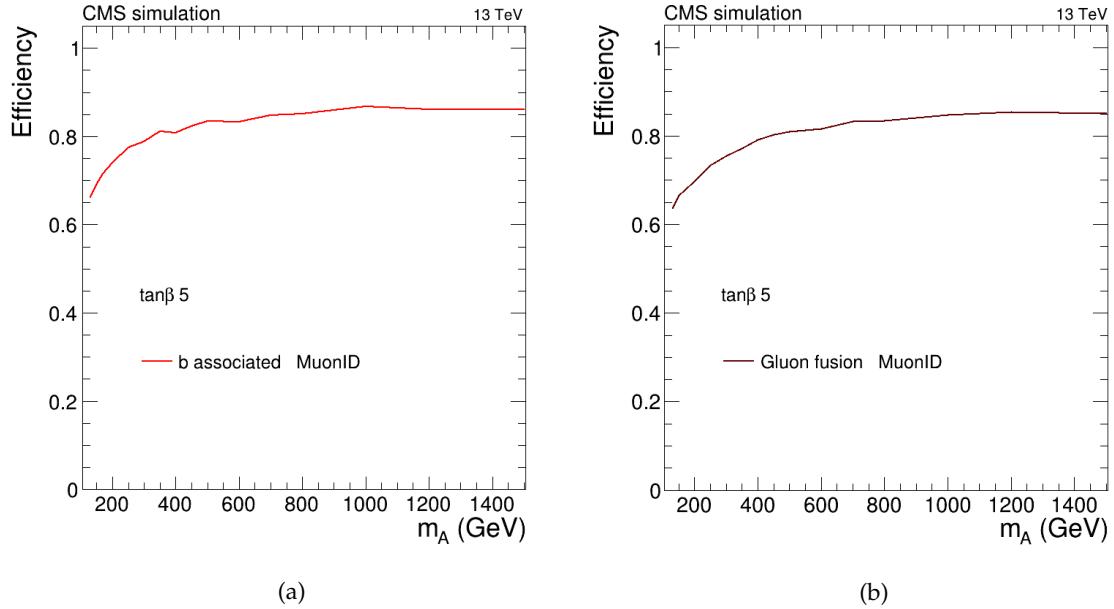


Figure 7.3: Muon identification as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

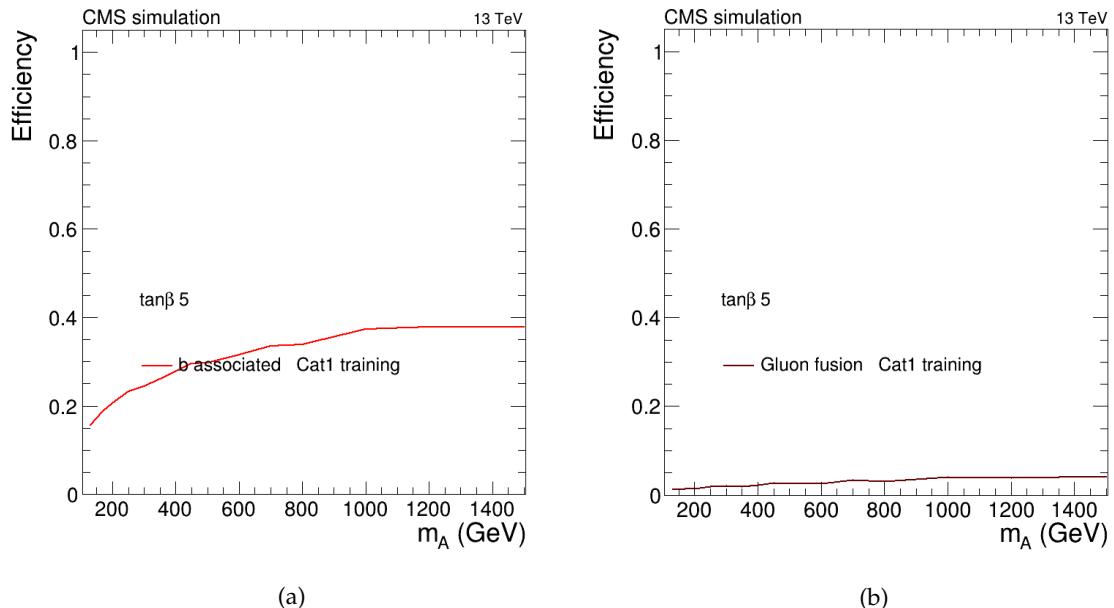


Figure 7.4: Efficiency of detecting at least one b-tagged jet in the event (CAT1), before the pNN cut, as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

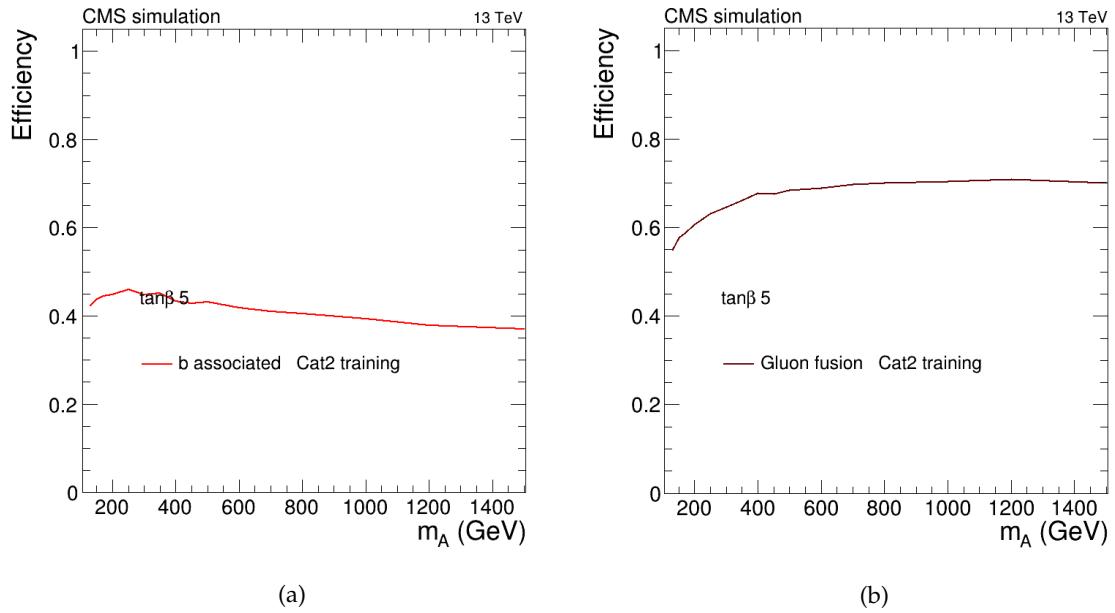


Figure 7.5: Efficiency of detecting no b-tagged jet in the event (CAT2), before the pNN cut, as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

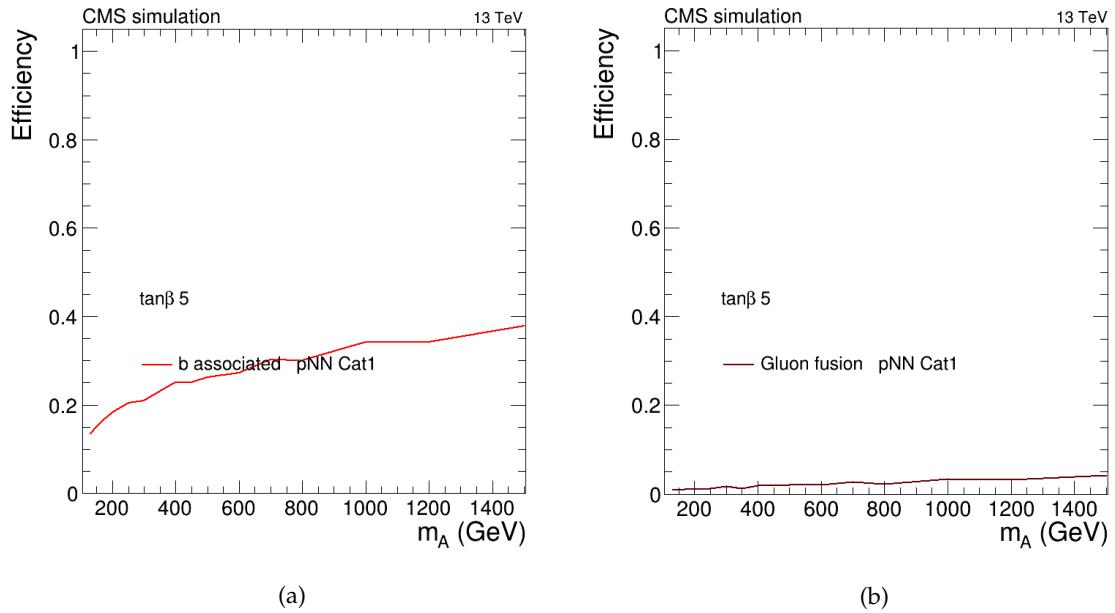


Figure 7.6: Efficiency of the pNN selection for events that belong to the b tag category (CAT1), as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

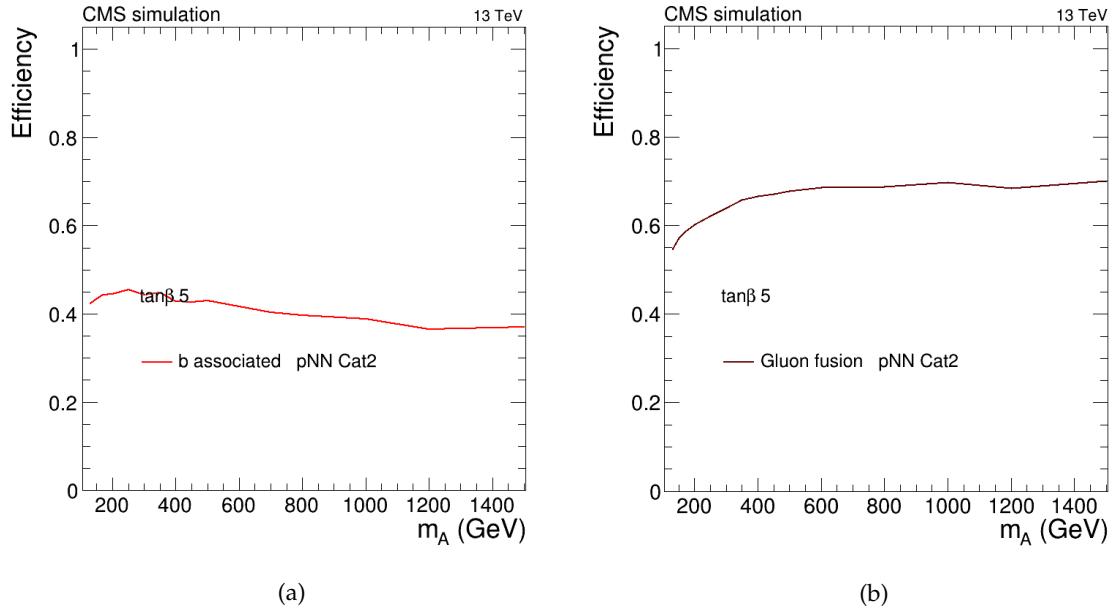


Figure 7.7: Efficiency of the pNN selection for events that belong to the no b tag category (CAT2), as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

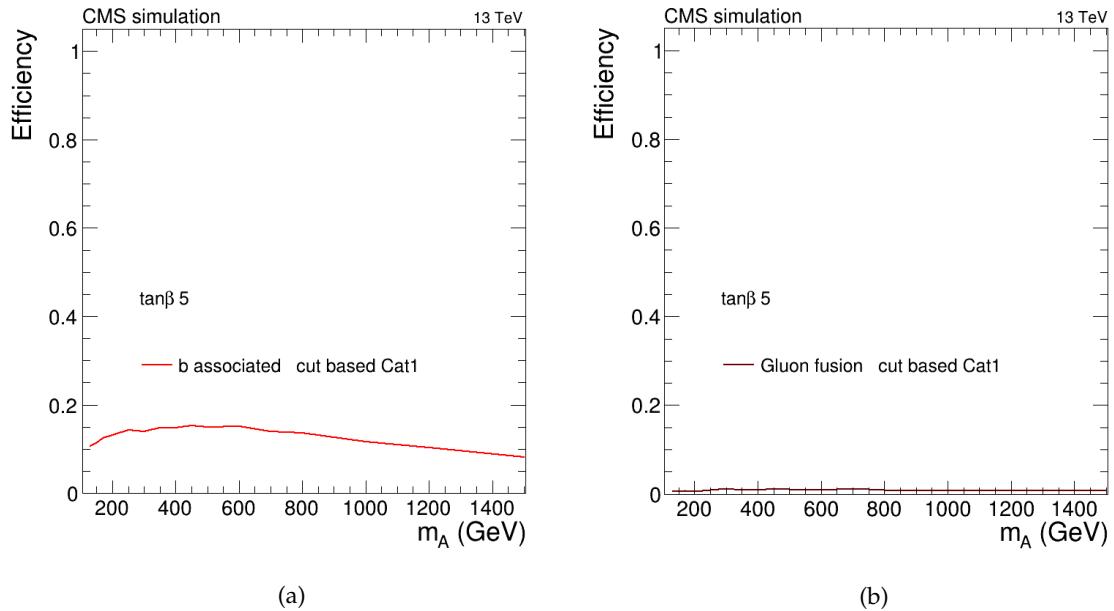


Figure 7.8: Efficiency of the cut-based selection for events that belong to the b tag category (CAT1), as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

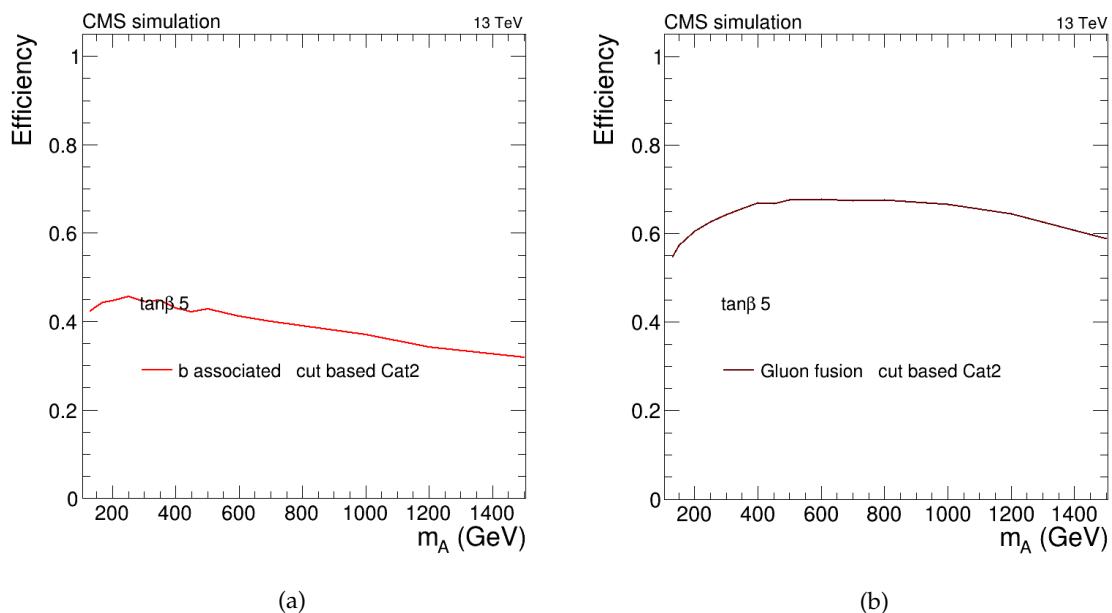


Figure 7.9: Efficiency of the cut-based selection for events that belong to the no b tag category (CAT2), as a function of m_A for bbA (left) and ggH (right) processes. Only the $\tan\beta = 5$ value has been considered.

7.2 FITTING PROCEDURE

This analysis, as already mentioned in the previous Chapter, adopts a data driven approach, in which the background contribution is estimated by a fit to the data. The invariant mass distributions of the events that belong to CAT1 and CAT2 are used to determine the background shape. The signal shape, for each m_A listed in Table 5.4 and with $\tan \beta = 5$, is instead estimated by a fit to MC samples, for a generic ϕ boson in the model independent scenario. These invariant mass distributions are properly weighted to account for data-simulation discrepancies, as described in Section 6.3.1.

7.2.1 Signal shape modelling

The invariant mass spectrum of the signal events that pass the event selection is used to determine the signal yield. The mass spectrum of the ϕ boson is fitted by a function F_{sig} , separately for the two categories, for any m_A with $\tan \beta = 5$.

The function F_{sig} , defined as:

$$F_{sig} = w \cdot F, \quad (7.2)$$

describes the signal. The function F , describing the mass shape of the expected signal, is a convolution of a Breit-Wigner (BW), to describe the signal resonance, with a Gaussian to account for the detector resolution. The quantity w , instead, is the number of events expected for the corresponding boson. In the fit, the mass and width parameters of the BW, and the sigma of the Gaussian, are left free to vary.

7.2.2 Background shape modelling

In the range of Higgs masses tested in this analysis, the background is mainly composed by Drell-Yan and top pair production. In particular, for the b tag category the latter is dominant, whereas dimuon pairs from Drell-Yan events are the main background for non b-tagged events.

This analysis does not use background determination from simulation, because of the little statistical power of the MC samples with respect to the data in the region of interests, as well as because of the large uncertainties due to extra terms in QCD or EW expansions, re-summation effects, parton density function and scale uncertainties, and unknown modelling of the long range correlation uncertainties. Therefore, given the smooth dependence of the background shape as a function of the dimuon invariant mass, the strategy chosen is to estimate it from data, by assuming proper functional form to describe its dependence as a function of the reconstructed dimuon invariant mass x .

One of such functional forms is F_{bkg} , defined as:

$$F_{bkg} = e^{\lambda x} \left[\frac{f}{N_1} \cdot \frac{1}{(x - m_Z)^2 + \frac{\Gamma_Z^2}{4}} + \frac{(1-f)}{N_2} \cdot \frac{1}{x^2} \right]. \quad (7.3)$$

The quantity $e^{\lambda x}$ parameterises the exponential part of the parton density function, and f represents the weight of the BW term with respect to the DY photon-exchange. These two quantities are free parameters of the fit. N_1 and N_2 are normalisation quantities, corresponding to the integral of each term in F_{bkg} . The parameters Γ_Z and m_Z are separately determined for the two event categories by fitting the dimuon mass distribution close to the Z boson mass. The fit provides the effective values of such quantities, which include detector and resolution effects. Their values are then kept constant when using F_{bkg} in the final fit. This function, has been used in the previous cut-based analysis with 2016 Run 2 data [171] and also in the 7-8 TeV analysis [172].

7.2.3 Signal plus background parameterisation

A linear combination of the two functions described in Section 7.2.1 and 7.2.2 is then used to perform an unbinned fit to the data:

$$F_{fit} = (1 - f_{bkg}) \cdot F_{sig} + f_{bkg} \cdot F_{bkg}. \quad (7.4)$$

The fit is performed for each m_A and with $\tan\beta = 5$ hypothesis, as the yield of the signal events and the shape of F_{sig} depend on these quantities. The parameters that describe the signal are determined by fitting the simulated samples that pass the event selection with Equation 7.2, for each m_A and $\tan\beta = 5$ pair, as explained above. Subsequently they are assigned as constant terms in F_{fit} . The quantity f_{bkg} is a free parameter in the fit, and the fraction of signal events is defined as $f_{sig} = (1 - f_{bkg})$. The overall normalization is also a free parameter and is profiled in the fit.

For each m_A assumption, the function F_{fit} is used to fit the data over an $m_{\mu^+\mu^-}$ range centred on m_A . The range has to be large enough to account for the signal width, including the experimental resolution, and it is ± 50 GeV for $m_A < 300$ GeV, ± 75 GeV for $300 \leq m_A < 400$ GeV, ± 100 GeV for $400 \leq m_A < 500$ GeV, ± 150 GeV for $m_A = 500$ GeV, and ± 200 GeV for $500 < m_A \leq 800$ GeV. For values of $m_A \leq 150$ GeV the lower bound of the mass window is set to 115 GeV. For $m_A > 800$ GeV, the range from 700 to 1800 GeV is used.

7.3 SYSTEMATIC UNCERTAINTIES

The main systematic uncertainties considered are briefly described below.

7.3.1 Efficiency uncertainties

The limited number of events of the simulated signal samples introduces an uncertainty on the signal efficiency that is between 1.5% and 2.5%, depending on the number of generated events.

In order to account for the differences between data and simulation in the muon trigger efficiency, the muon identification and isolation efficiency, the recommended scale factors

determined by the CMS muon POG are used to re-weight the MC distributions, as described in Section 6.3.1. A similar procedure is used to account for data-to-simulation discrepancies in the b tag efficiency [211]. Efficiency corrections are applied to the simulation as scale factors (SFs), and corrected by applying a weight to that event. The uncertainty associated with each SF is propagated to the analysis and the impact is assigned as a systematic uncertainty.

An event-by-event weight is also applied to account for the modelling of the pileup in the simulation. The uncertainty in the knowledge of the pileup multiplicity is evaluated by varying the total inelastic cross section [231, 232] by $\pm 5\%$, which translates into an uncertainty smaller than 1% in the signal efficiency.

The uncertainty of the jet energy scale (see Section 5.5.5) [233] is estimated with the rescale of the jet momentum by a factor that depends on the p_T and η of each jet. This variation is also propagated to the p_T^{miss} determination. Its effect on the signal selection efficiency is about 1.6 (0.4)% for the CAT1 (CAT2) category.

Systematic uncertainties in the unclustered energy are also propagated to the p_T^{miss} determination. The effect on the signal efficiency is 4.1% for CAT1, and 0.3% for CAT2.

The systematic uncertainty of the b tagging algorithm (see Section 5.5.4) affects the signal yield and the category migration with an impact on the signal efficiency of 2% for CAT1 and 0.6% for CAT2. The uncertainty in the total integrated luminosity is 2.5% [234] and affects the signal yield.

A conservative uncertainty of <0.1% is used to account for the Parton Density Functions (PDFs), used to generate the MSSM Higgs signals at NLO with the NNPDF3.1 set. Also, an uncertainty of 0.2 (0.7) has been taken from the choice of the renormalisation and factorisation scales for the CAT1 (CAT2) category.

Table 7.1 shows the systematic uncertainties in the signal efficiency for the two categories: the b tag efficiency is applied only to the b-associated signal process.

7.3.2 Uncertainty on the background parameterisation

The uncertainty introduced by the choice of the analytical function used to parametrize the background affects the determination of the exclusion limit. A method similar to that described in [24] is used to address this aspect. The method aims to determine the number of spurious signal events that are introduced by the choice of the background function F_{bkg} , when the background is fit by the function F_{fit} .

The invariant mass spectrum is fitted with the function F_{bkg}^a , chosen among various forms: one is shown in Eq. 7.3 while others include similar forms like a BW plus exponentials, and sum of exponentials. All these functional forms adequately describe the background distribution observed in data. The fit is performed in the proper mass range centred around the assumed value of m_A , and the parameters of F_{bkg}^a are determined.

After that, thousands of MC pseudo-experiments are generated, each one with the same number of events observed in data, distributed according the function F_{bkg}^a . For each pseudo-experiment, the invariant mass distribution is then fit with the function F_{fit} of

Source	Systematic uncertainty (%)	
	CAT1 (b tag category)	CAT2 (no b tag category)
MC statistical uncertainty	1.8-2.5	1.5-1.8
Trigger efficiency	0.9	0.9
Muon reconstruction	2	2
Muon isolation	1	2
Pileup	0.8	0.9
Jet energy scale	1.6	0.4
Unclustered energy	4.1	0.3
PDF	< 0.1	< 0.1
Scale	0.2	0.7
b tag (only for b-associated)	2	0.6

Table 7.1: Systematic uncertainties in the signal efficiency for the two event categories. The systematic uncertainties hold for both Higgs boson production processes except for b tag efficiency, which applies to the b-associated production process only. For those sources of systematics that depend on m_A the range of uncertainty is quoted.

Eq. 7.4, once using F_{bkg}^a , and then using a different function F_{bkg}^b , given by Eq. 7.3. For each pseudo-experiment, the spurious signal yield, expressed by the number of events N_{bias}^a and N_{bias}^b , is determined. The quantity N_{bias}^a is on average consistent with zero within statistical fluctuations. The quantity N_{bias}^b represents, instead, the number of spurious signal events that are found in the signal yield if the function F_{bkg}^b is used to describe the background, when the background itself is actually distributed according to F_{bkg}^a .

The median of the distribution of the difference $N_{bias}^a - N_{bias}^b$ obtained from the pseudo-experiments is defined as the bias introduced by using the function F_{bkg}^b , relative to the tested mass m_A . This procedure is repeated for each function F_{bkg}^a among the functional forms mentioned above, and the largest bias is taken as the systematic uncertainty in the number of signal events obtained from the maximum likelihood fit, due to choosing Eq. 7.3 for the background parameterisation.

7.4 LIMITS CALCULATION

As mentioned in this Chapter's introduction, to search for the MSSM Higgs bosons in the dimuon decay channel, a comparison is done between what it is expected and what is observed. The expected value is constituted by the background-only hypothesis, while the observed comes out after the fit to the data with the F_{fit} function. This function represents the signal plus background hypothesis, and its f_{bkg} parameter returns the relative amount of background. The significance of possible deviations of f_{bkg} from unity is evaluated performing a statistical test using the Asymptotic CL algorithm [235]. The outcome of this test is expressed in terms of Confidence Level (CL). Where the presence of background only is confirmed at 95% of probability, the signal is automatically excluded. Therefore, in

absence of signal in data, the following results give an exclusion limit for MSSM neutral Higgs boson.

7.4.1 Model independent limits

Limits on the production cross section times decay branching ratio $\sigma \cdot B(\phi \rightarrow \mu^+ \mu^-)$ for a generic single neutral boson ϕ , in the model independent scenario, are determined. In the model independent interpretation, the ϕ boson is searched as a single resonance with mass m_ϕ , using as simulated signal the bbA and ggH samples at the $\tan\beta = 5$ case (with small intrinsic width, dominated by the detector resolution) as a template to compute the detection efficiency of a generic ϕ boson decaying to a muon pair (from now called bb ϕ and gg ϕ).

In this analysis the ϕ boson is assumed to be produced entirely either via the b-associated production or the gluon fusion process, therefore the exclusion limit at 95% of CL on $\sigma \cdot B(\phi \rightarrow \mu^+ \mu^-)$ is determined as a function of m_ϕ for the two production mechanisms separately.

Figure 7.10 shows the expected 95% CL model independent expected limits on the cross section times the decay branching ratio for b-associated and gluon fusion productions, as a function of the ϕ mass hypothesis, merging the two categories. The 1 and 2 σ uncertainties are represented by the coloured green and yellow error bands, respectively. The black line with dots represents the same limits (showing only the central values), for the cut-based approach adopted in the previous analysis [171]: it is noticeable, in the b-associated signal production, a much more stringent upper limit using the pNN discriminator, reaching an improvement of 50% in the high mass region. In the gluon fusion signal production, instead, the improvement in the exclusion limits is smaller. These results are in agreement with the previous consideration on the parameterised Network's performances and the relative signal efficiencies.

For the b-associated production, Figure 7.11 shows the 95% CL expected model independent exclusion limits, shown separately for each category before the merging. Since, for the limits on the b-associated production the main contribution comes from the b-tagged category (CAT1), the dominant improvement is led in the bb ϕ CAT1 limit with only a minor contribution from CAT2.

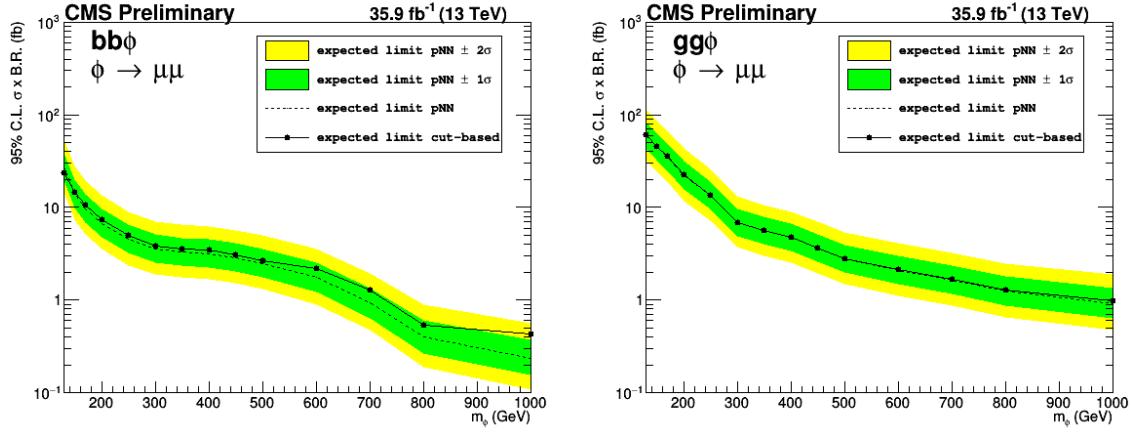


Figure 7.10: The 95% CL expected model independent upper limits, including the 68 and 95% CL bands, on the production cross section times branching ratio of a generic ϕ boson decaying to a dimuon pair, in the case of b-associated (left) and gluon fusion (right) production, after using the pNN. The results are obtained using a signal template with an intrinsic narrow width. The black line with dots represents the central value of the expected upper limits for the previous cut-based analysis [171], as a comparison.

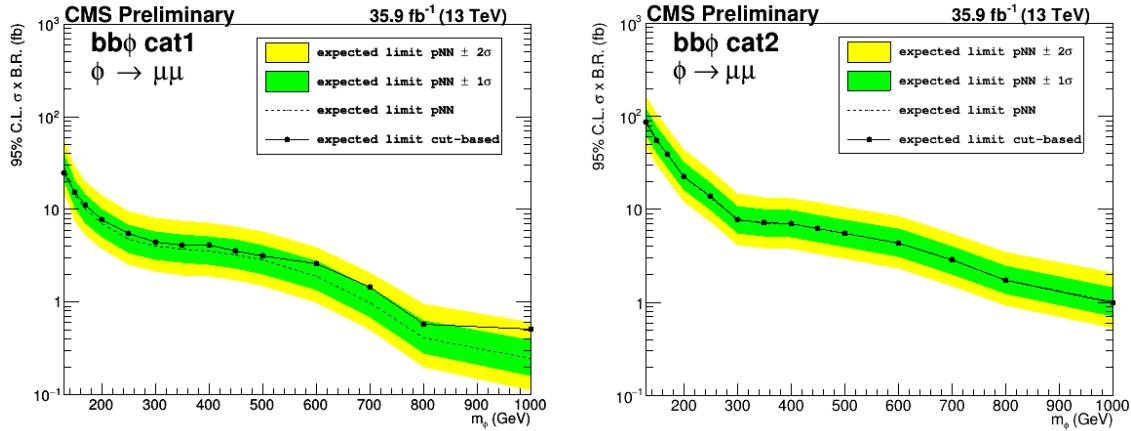


Figure 7.11: The 95% CL expected model independent upper limits, including the 68 and 95% CL bands, on the production cross section times branching ratio of a generic ϕ boson decaying to a dimuon pair, in the b-associated production for the b tag category (left) and the no b tag category (right), after using the pNN. The results are obtained using a signal template with an intrinsic narrow width. The black line with dots represents the central value of the expected upper limits for the previous cut-based analysis [171], as a comparison.

8

CONCLUSIONS AND OUTLOOK

This PhD thesis described a model independent search for a narrow neutral BSM Higgs boson, decaying to $\mu^+\mu^-$ with the CMS experiment, where the Higgs boson is produced either via gluon fusion process or in association with a $b\bar{b}$ pair. These processes, in fact, have the highest cross section for a neutral MSSM Higgs boson at the LHC. The search is performed on the 2016 data collected at CMS in proton-proton collisions at the LHC, with a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 35.9 fb^{-1} .

The dimuon final state, despite having a very low branching ratio, has been chosen for its very clean signature, easy to detect thanks to the CMS excellent muon identification capabilities. Therefore, to enhance the signal sensitivity for the b -associated production and gluon fusion, an event categorisation based on the presence or the absence of a jet coming from a b quark has been made.

For the signal/background discrimination, a multivariate approach is then performed, adopting a parameterised Neural Network (pNN) approach, in which an entire set of networks per each m_A hypothesis has been replaced by a single network, trained including the m_A values as an input parameter, together with the physical features of interest. The signal events used for training are simulated in the MSSM parameter space ($m_A, \tan\beta$), scanning from 2 to 60 in $\tan\beta$, and from 130 to 1500 GeV in m_A . The background events used for the model training are also simulated, consisting in dimuon pairs produced in Drell-Yan process and $t\bar{t}$ decays as well as some minor contribution from diboson and single top events, making sure that the simulation models the data correctly. The model training and testing have been performed separately for each category, and a series of best cuts are obtained maximising the significance ratio for each value of m_A and signal process.

The expected exclusion 95% CL upper limits for a model independent scenario have been computed, using simulated samples only for the signal and estimating the background directly from the data. The statistical treatment of the data has been performed separately for each category, making no assumptions on the cross section, mass and width of a generic ϕ neutral Higgs boson, and using the b -associated A boson (bbA) and gluon-fusion H boson (ggH) processes as a template to compute the signal shape fit, with only $\tan\beta = 5$ to have the smallest signal intrinsic width possible compared to the experimental detector resolution. To fully evaluate the potential improvement of the pNN approach, compared to the cut-based signal selection adopted in the previous analysis work, only the expected limits on the background-only hypothesis are compared between the two methods,

showing a good improvement in the b-associated production (up to 50% on the high mass region) and showing a similar behaviour with a slight improvement on the gluon fusion production.

FUTURE OUTLOOK

This thesis work is inserted in an active CMS analysis group, focused on the prosecution of the analysis work described in this thesis in view of a future publication in a year time range. Some aspects of the analysis are still not finished and will require an additional effort:

- Simulated gluon fusion events of the H Higgs boson are still missing, due to several delays in the CMS central production system. Unable to be used during the PhD time scale, they will be implemented in the near future;
- The parameterised Network approach, used to discriminate between signal and background events, is a relatively new architecture used only on a limited amount of analyses in the CMS collaboration with a specific use-case. For the purpose of this thesis, such approach has been currently tested only on 2016 data, to perform sanity checks with the results obtained in the previous 2016 analysis [171], with the final goal of including also both the 2017 and 2018 datasets, for the full Run 2 data analysis paper;
- In Section 6.3.7, Eq. 6.6 shows the metric used for the best cut selection: the significance ratio. This metric has been chosen to compare its numeric value among different m_A hypotheses, having each one a different amount of signal and background events. However, the most common s/\sqrt{b} significance metric has been investigated as well. In Figure 8.1, the comparison between such two metrics is shown for CAT1 events: in particular, the s/\sqrt{b} metric selects a very narrow region characterised by an almost empty background. Such study shows a promising new sub-categorisation, with a greatly reduced background and, therefore, resulting in a high purity. For CAT2, instead, almost no difference between metrics has been observed. This possibility will be explored with full Run 2 statistics, where the lower statistical error might result in an efficient split and an increased overall signal efficiency.
- The model independent 95% CL upper limits on the cross section times branching ratio only show the expected limits on the background-only hypothesis, for a comparison between the pNN and the cut-based approach. The following step would be the inclusion of the observed limits, from the signal plus background hypothesis, to be able to put actual exclusion limits on the neutral Higgs signal processes.

In addition, further developments on the pNN are also possible and currently under investigation: from a better hyperparameter tuning and a more optimised feature pre-processing, to improved model architectures with more complex mechanisms of concatenation between features and mass parameter, currently under study.

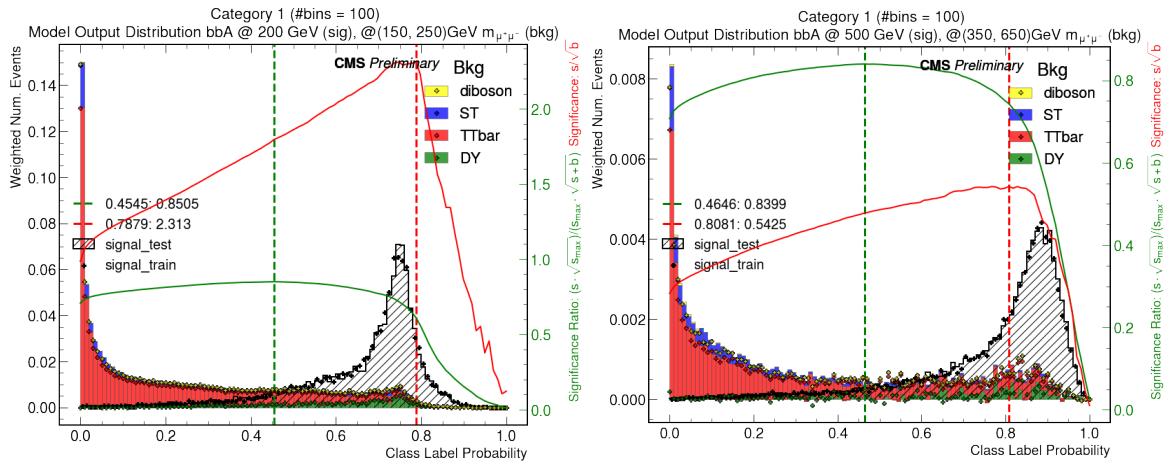


Figure 8.1: pNN output distribution histograms for the $m_A = 200$ GeV hypothesis (on the left) and $m_A = 500$ GeV (on the right), for CAT1 events. The black histogram, filled with diagonal lines, represents the output network distribution for the signal simulated events; the coloured histograms, instead, represent the same output for the background simulated events, with the same m_A , each colour showing a different background process weighted for its relative production cross section. Superimposed, the dots represent the output distributions for the training set, for an overtraining check. The green curve shows the computation of the significance ratio for each bin, with a dashed green vertical line corresponding to its maximum value. The red line shows the computation of the classical s/\sqrt{b} significance for each bin, with a dashed red vertical line corresponding to its maximum value.

A

THE HIGH-LUMINOSITY LHC PROJECT

Since 2010, the LHC has been producing proton-proton collisions with a 7 and then 8 TeV centre-of-mass energy. After a two year long shutdown in spring 2015, the machine has been restarted delivering collisions between protons at a new record of 13 TeV centre-of-mass energy. In order to increase its discovery potential by mid-2020, the LHC would need an upgrade to increase the total number of collisions by a factor of 10. A more powerful LHC will allow even rarer events to be detected and will increase our understanding of the energy frontier.

How this should be done is at the heart of the *High-Luminosity LHC project* (HL-LHC) [236], which design study came to a close on October 2015 with the publication of a technical design report and a budget of 950 million CHF over a period of 10 years. The HL-LHC relies on a number of new technologies, including 11-12 Tesla superconducting magnets, compact superconducting crab cavities with a precise phase control for beam rotation, new beam collimators etc... raising the instantaneous luminosity from $\approx 2.01 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ up to $7.5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

Figure A.1 shows the timeline for HL-LHC, currently scheduled to start its data taking in 2026 and to operate through 2030s.

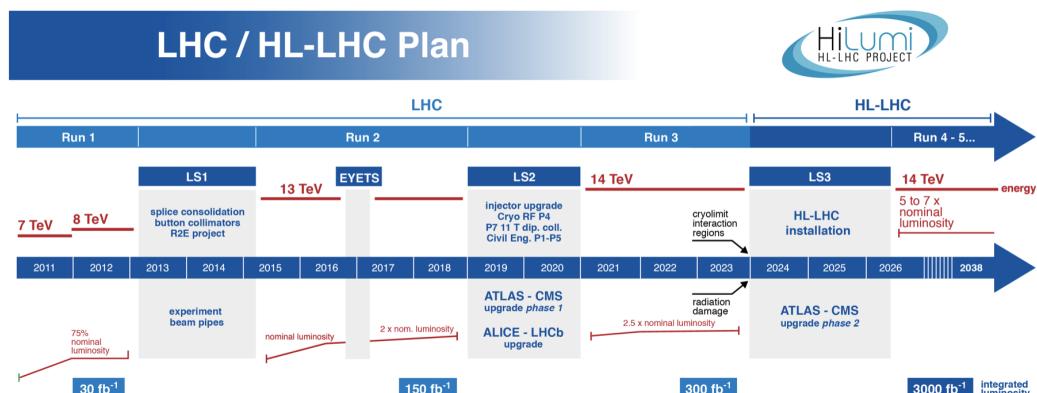


Figure A.1: Timeline of operations at the LHC for the next years. After a Long Shutdown time period (LS3), the HL-LHC is supposed to start its life cycle.

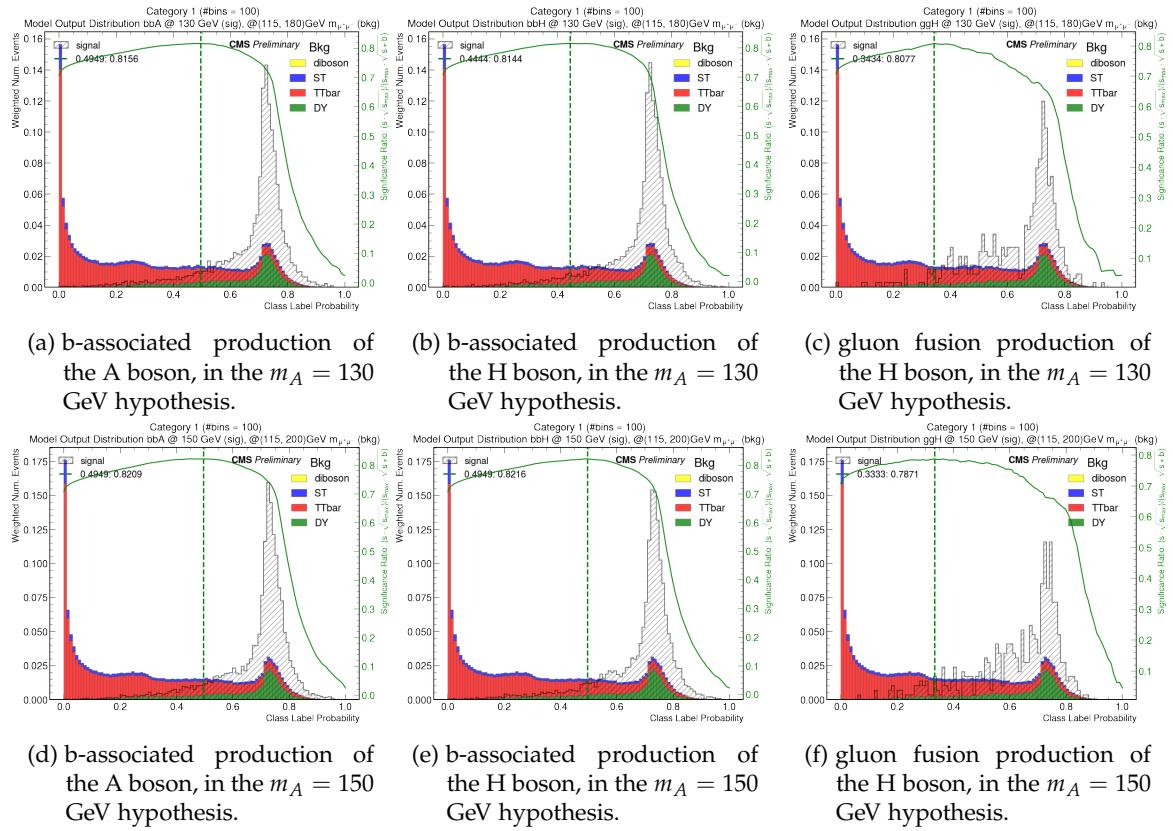
B

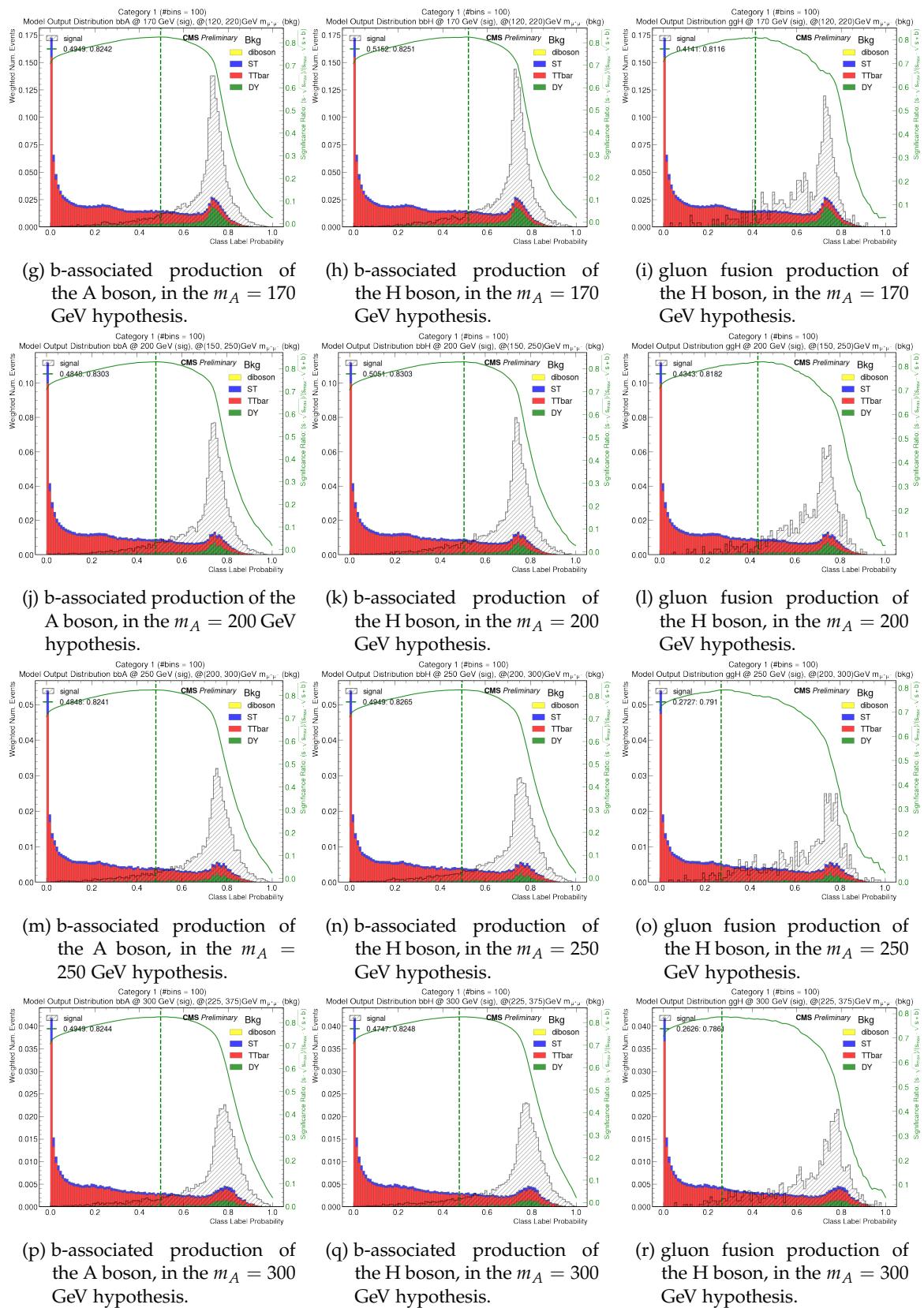
PARAMETRIC NEURAL NETWORK ADDITIONAL PLOTS

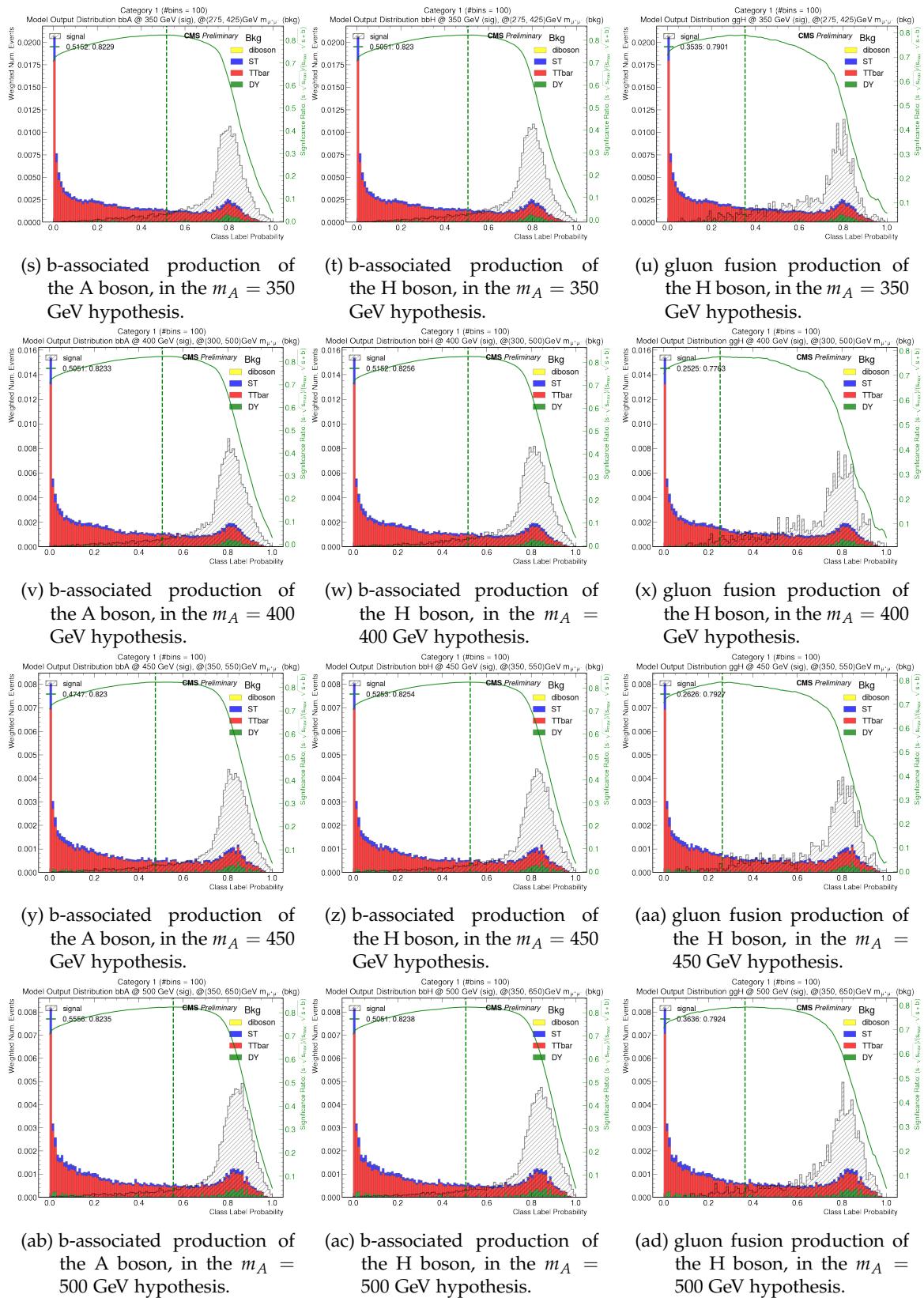
In this Appendix, all the pNN plots described in Chapter 6 are shown.

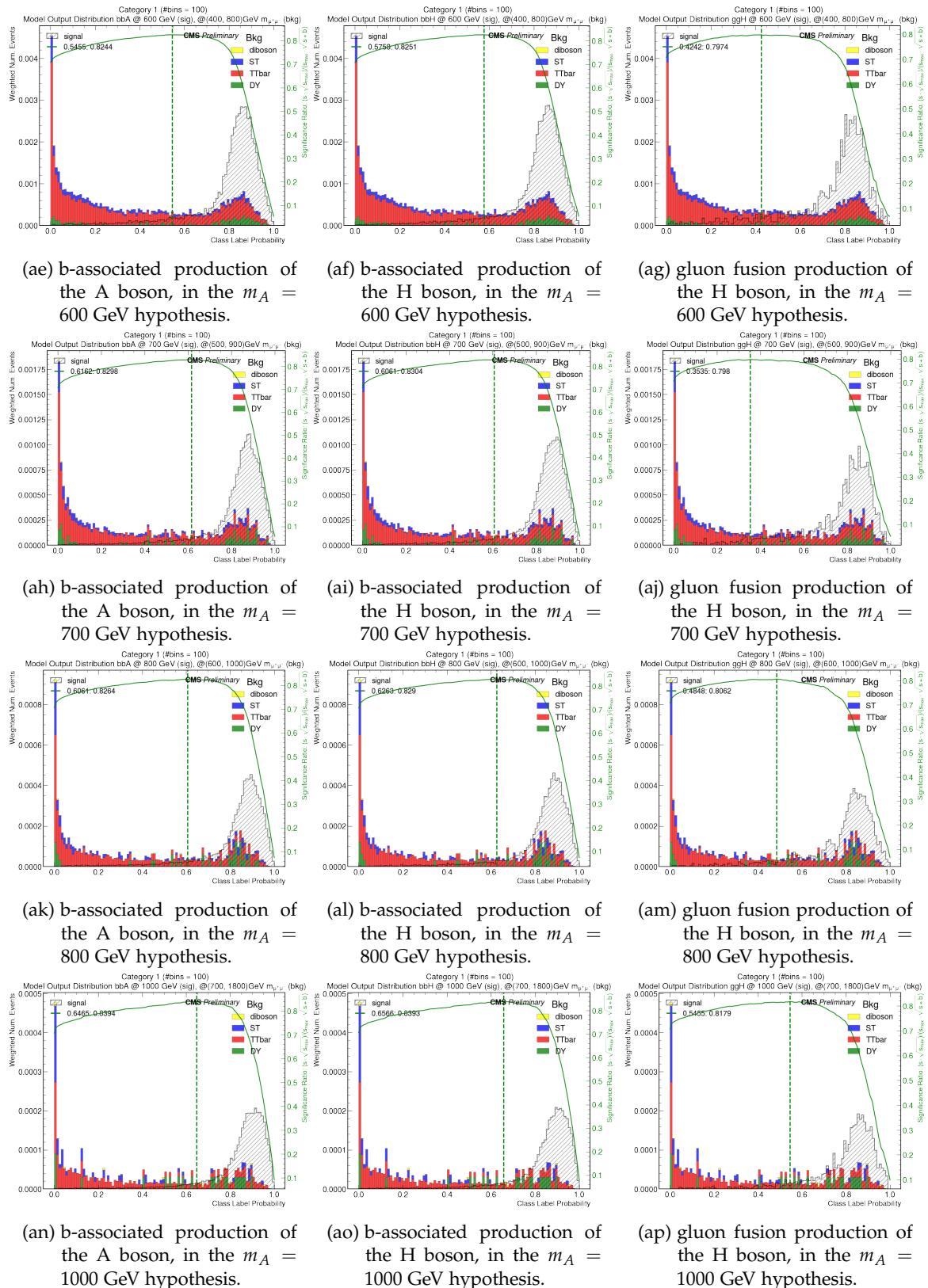
b.1 PNN OUTPUT DISTRIBUTIONS

These are all the plots, for each m_A hypothesis, discussed in Section 6.3.7, and including also Figure 6.17 and 6.18, which are added in the main text for the sake of clarity.









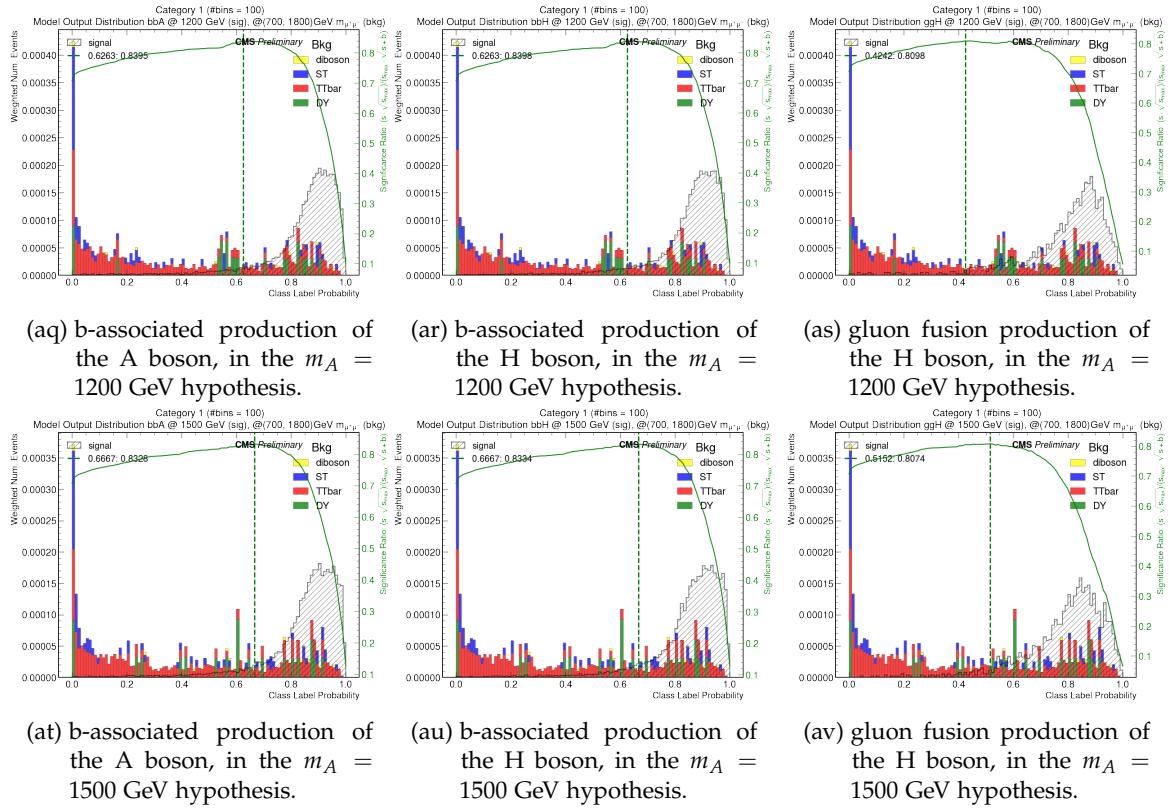
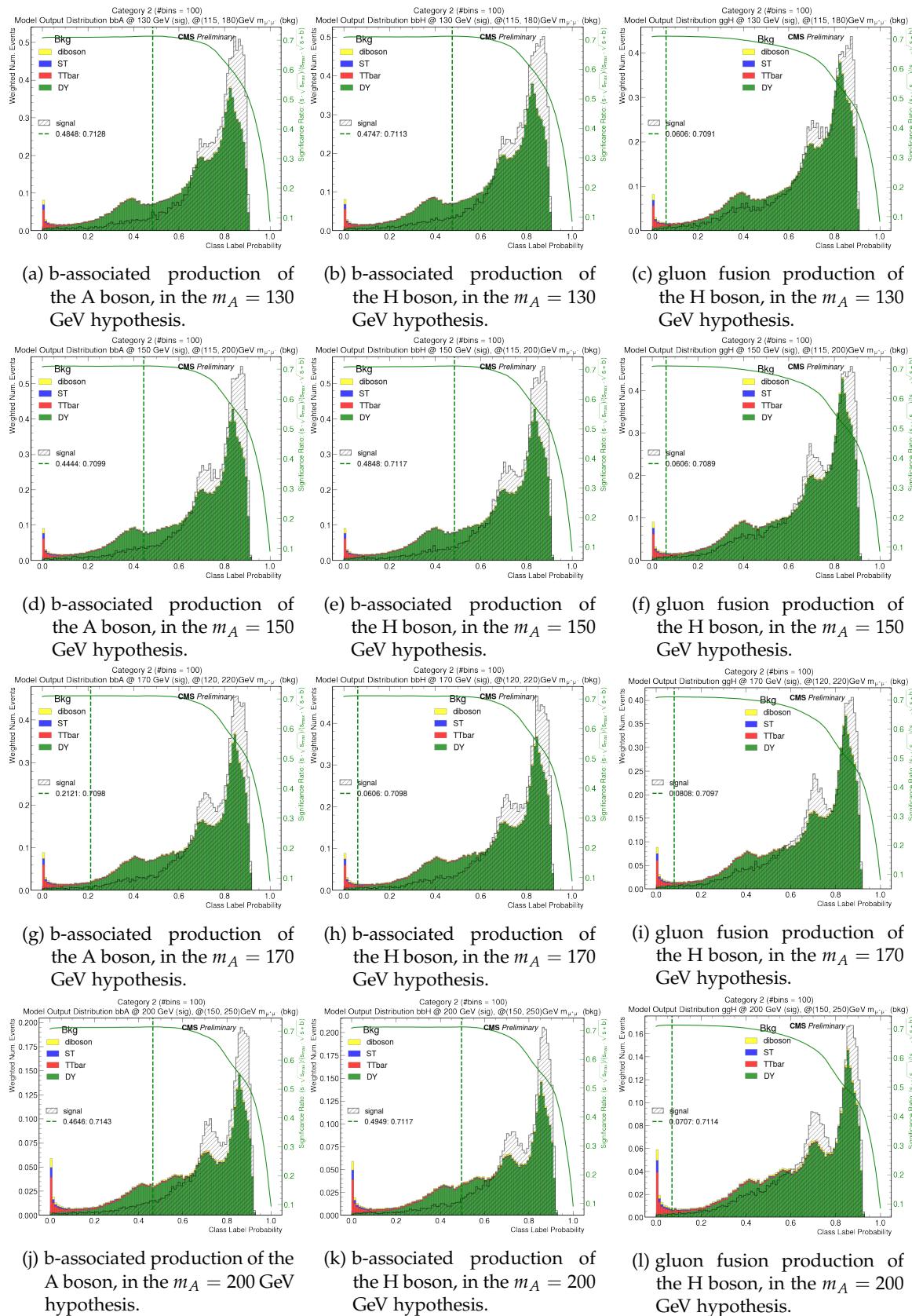
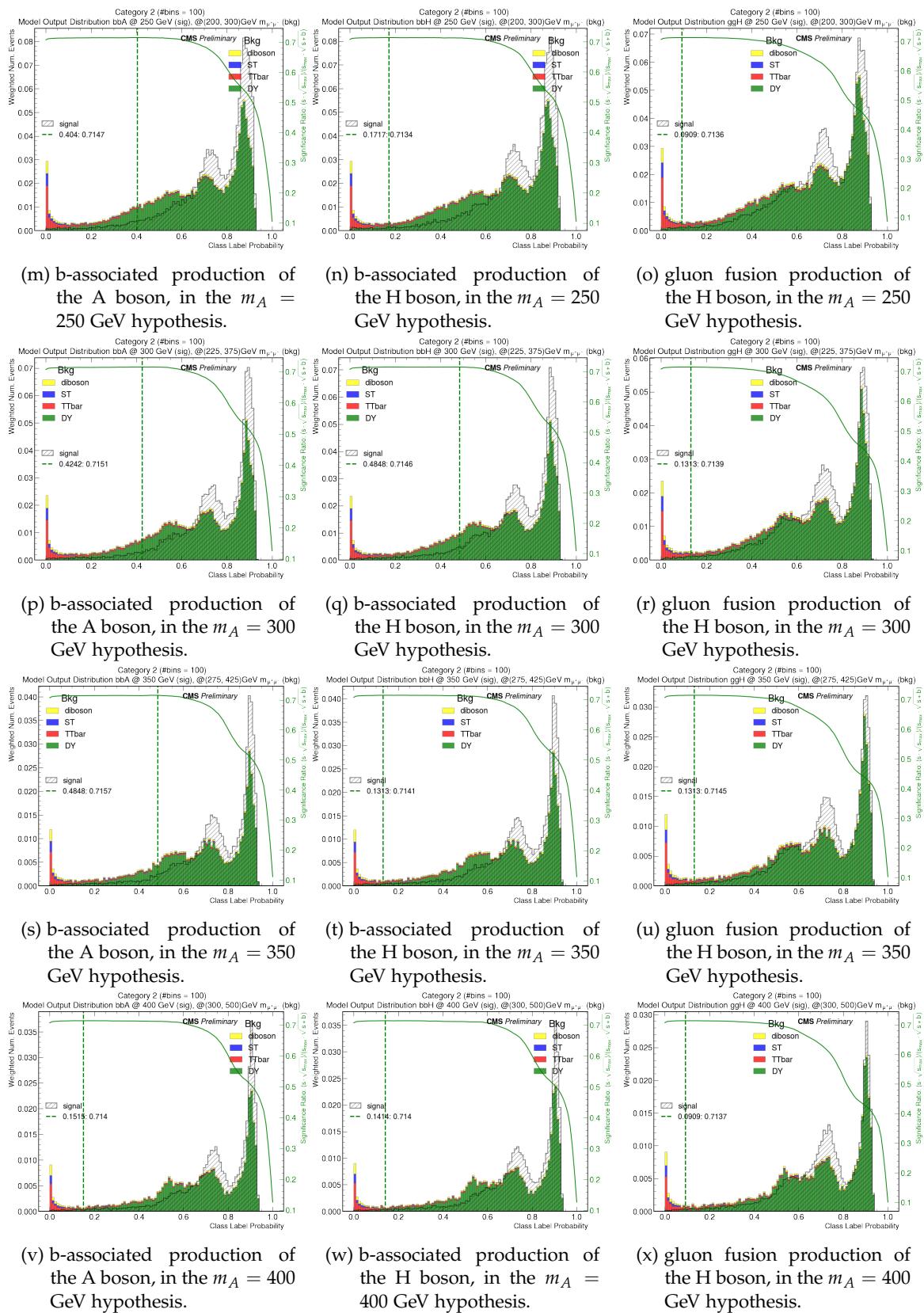
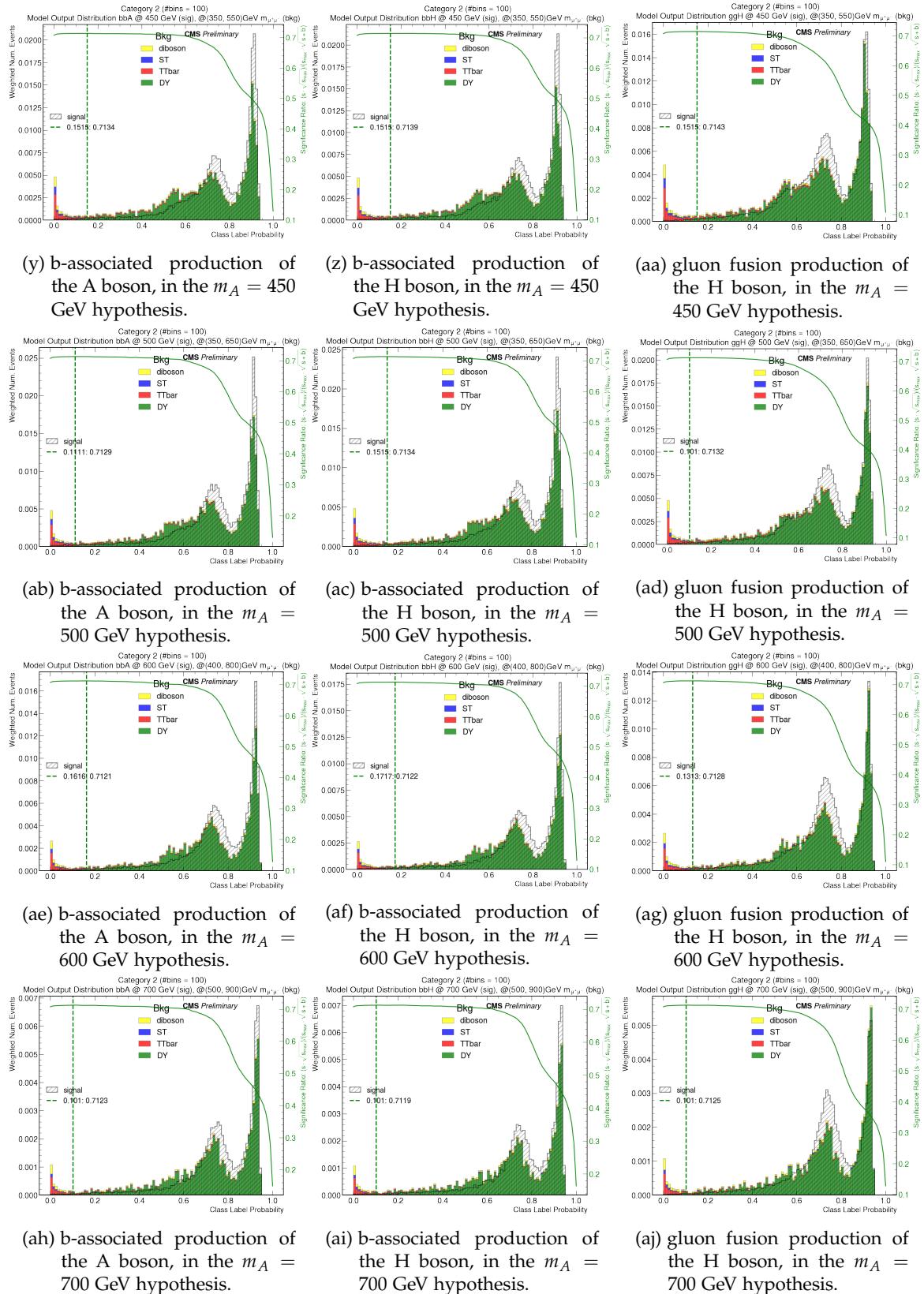


Figure B.1: pNN output distribution histograms for any m_A hypothesis, for CAT1 events. The grey histogram, filled with diagonal lines, represents the output network distribution for the signal simulated events; the coloured histograms, instead, represent the same output for the background simulated events, each color showing a different background process weighted for its relative production cross section. The green line, superimposed, represents the computation of the significance ratio for each bin, with a dashed vertical line corresponding to the maximum value, used for the output's best cut decision. The numerical values for the cut are also shown in the plot's left legend, with the notation: (best cut: maximum significance ratio).







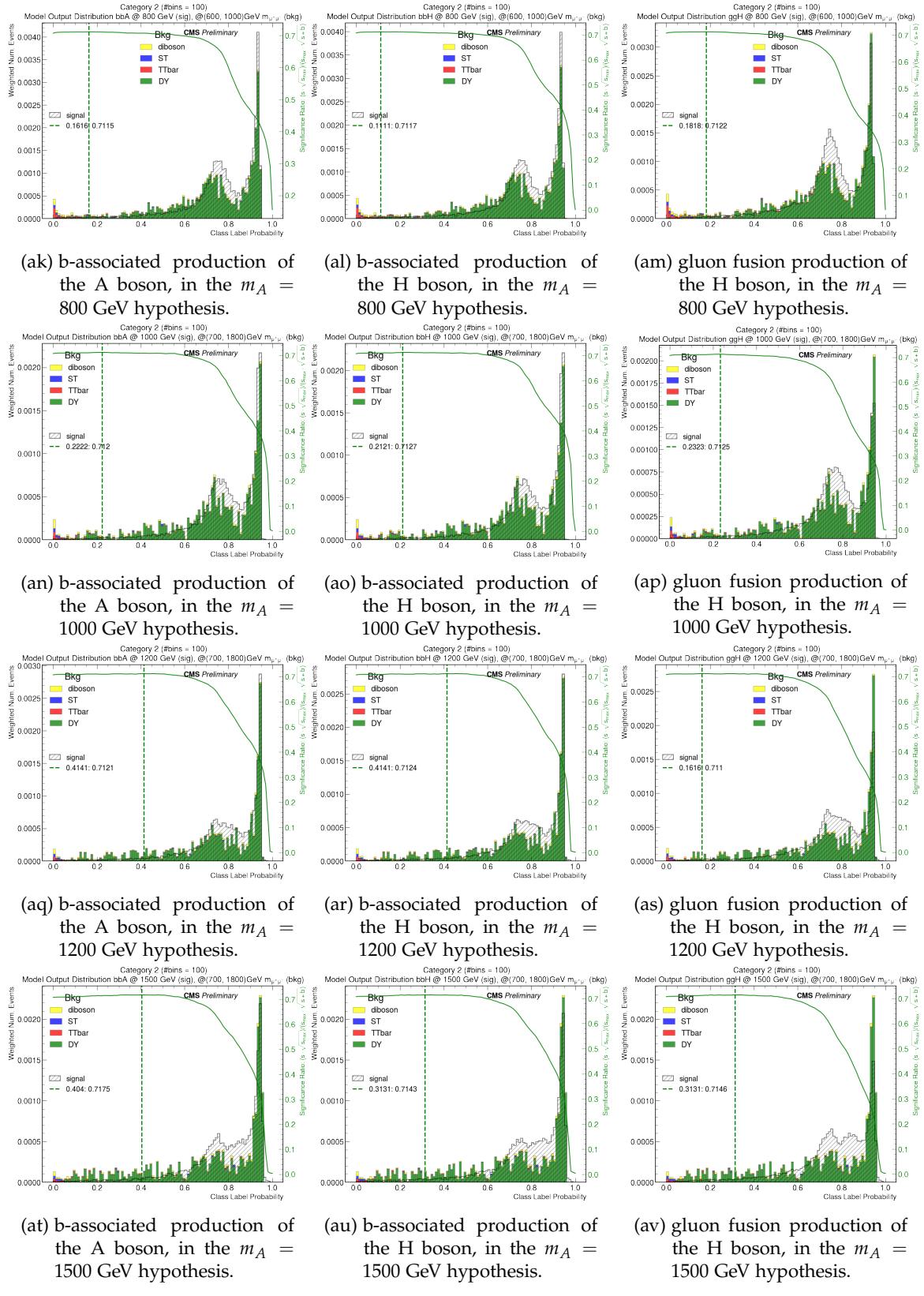
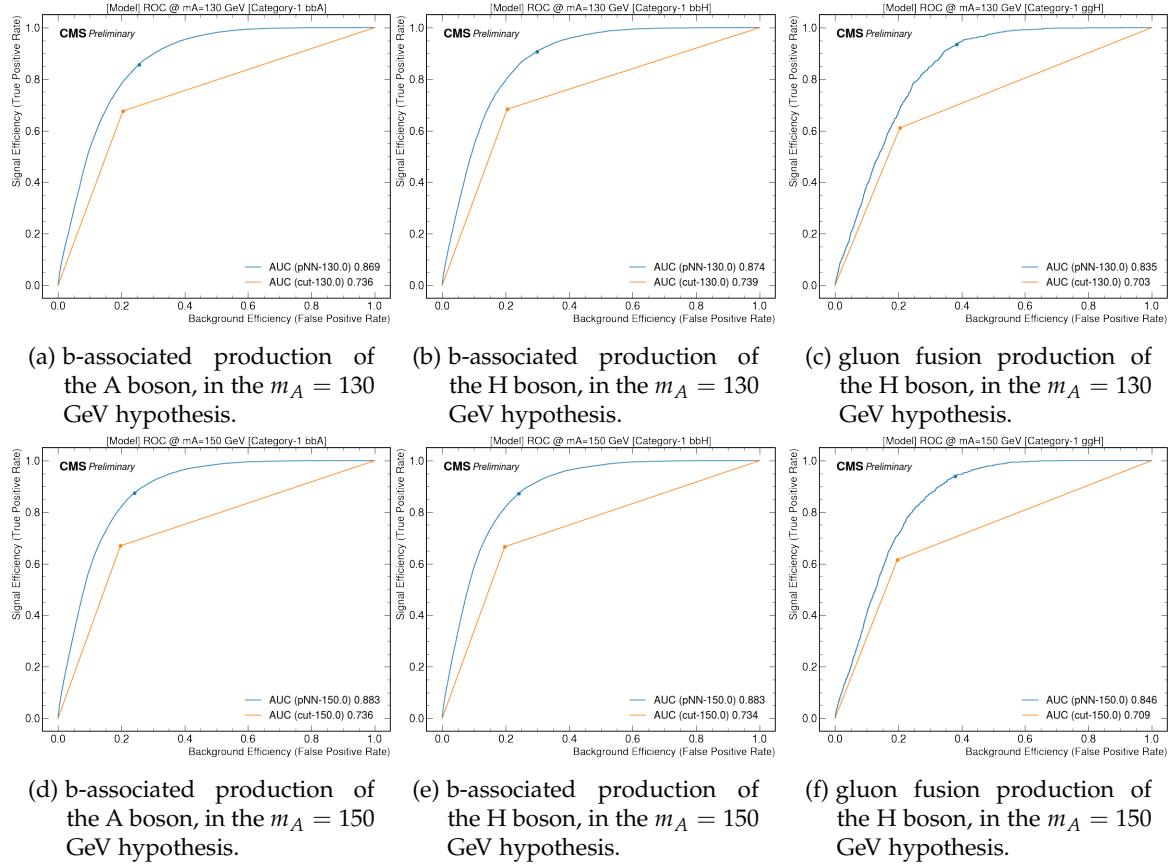
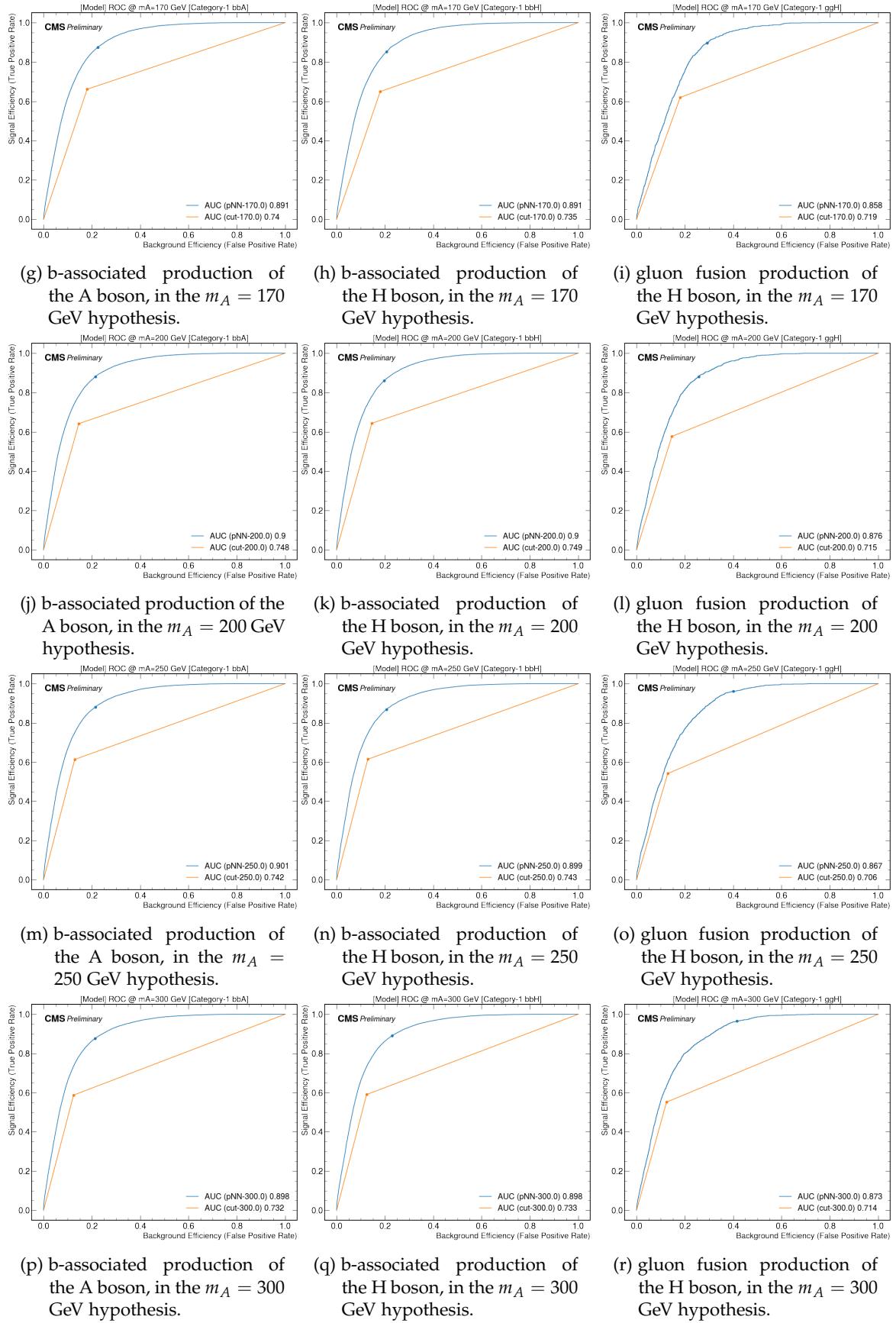


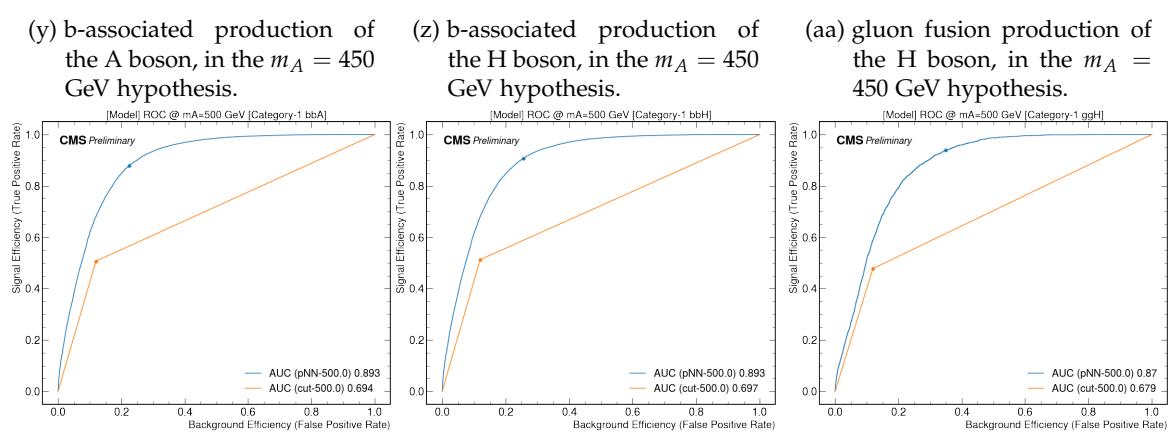
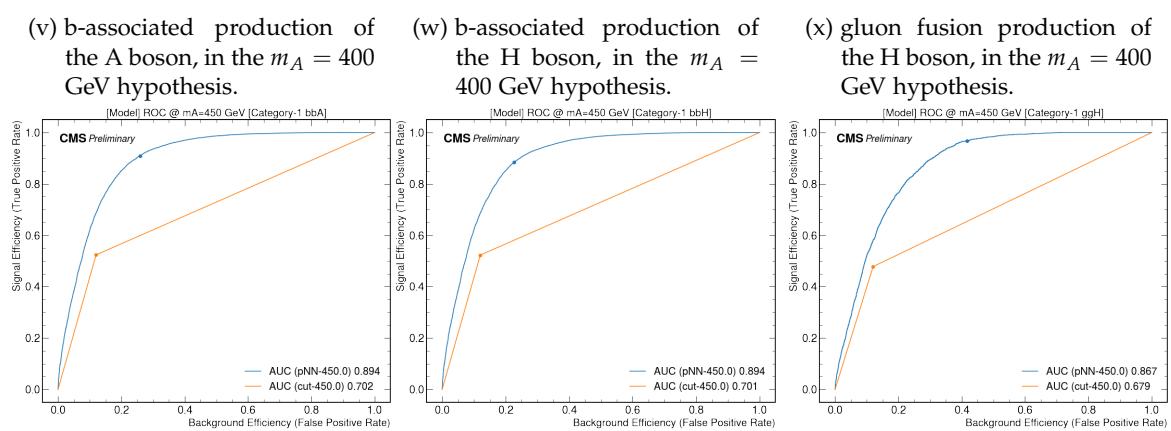
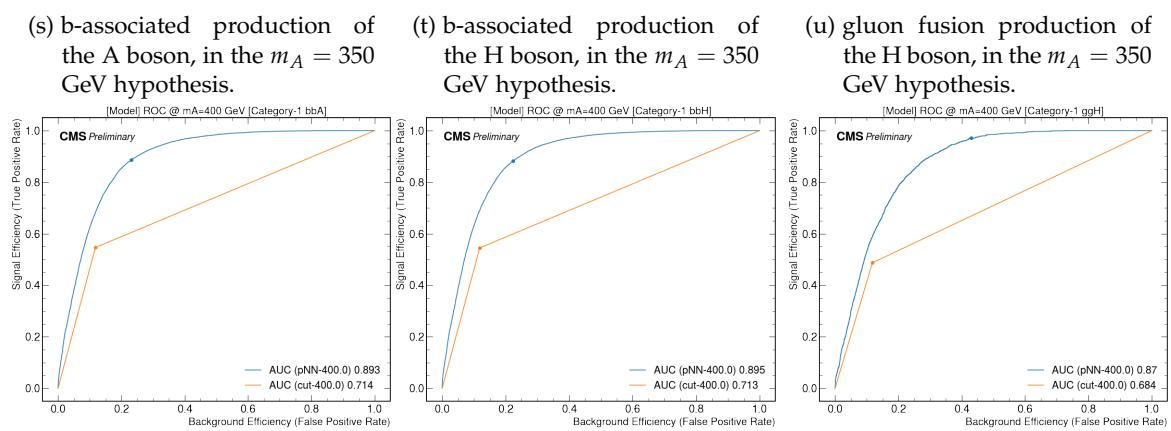
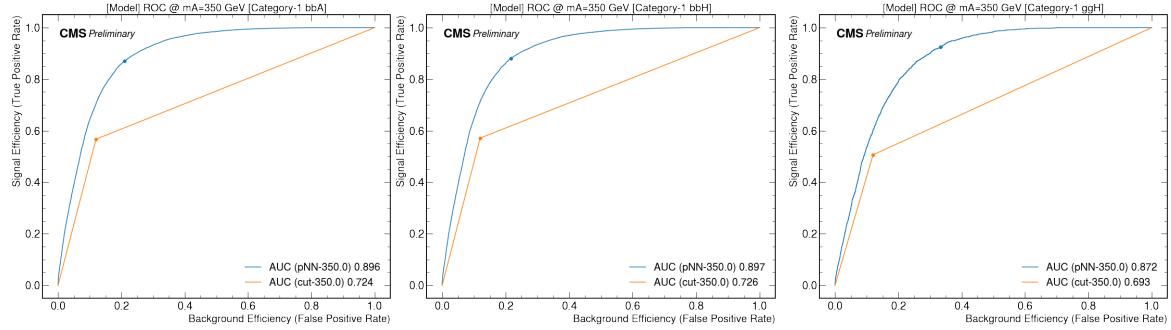
Figure B.2: pNN output distribution histograms for any m_A hypothesis, for CAT2 events. The grey histogram, filled with diagonal lines, represents the output network distribution for the signal simulated events; the coloured histograms, instead, represent the same output for the background simulated events, each color showing a different background process weighted for its relative production cross section. The green line, superimposed, represents the computation of the significance ratio for each bin, with a dashed vertical line corresponding to the maximum value, used for the output's best cut decision. The numerical values for the cut are also shown in the plot's left legend, with the notation: (best cut: maximum significance ratio).

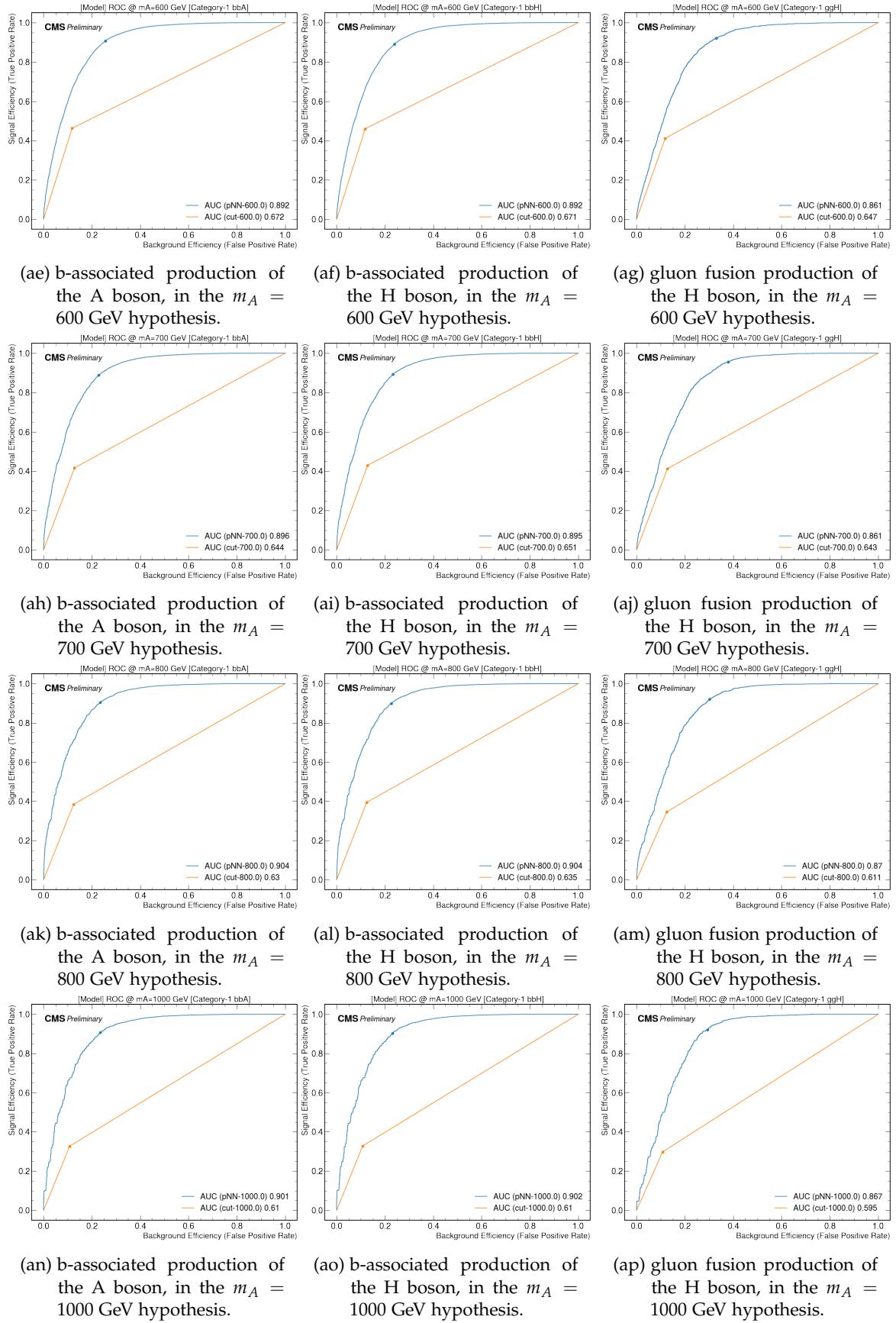
B.2 PNN ROC CURVES

These are all the plots, for each m_A hypothesis, discussed in Section 6.3.8, and including also Figure 6.19 and 6.20, which are added in the main text for the sake of clarity.









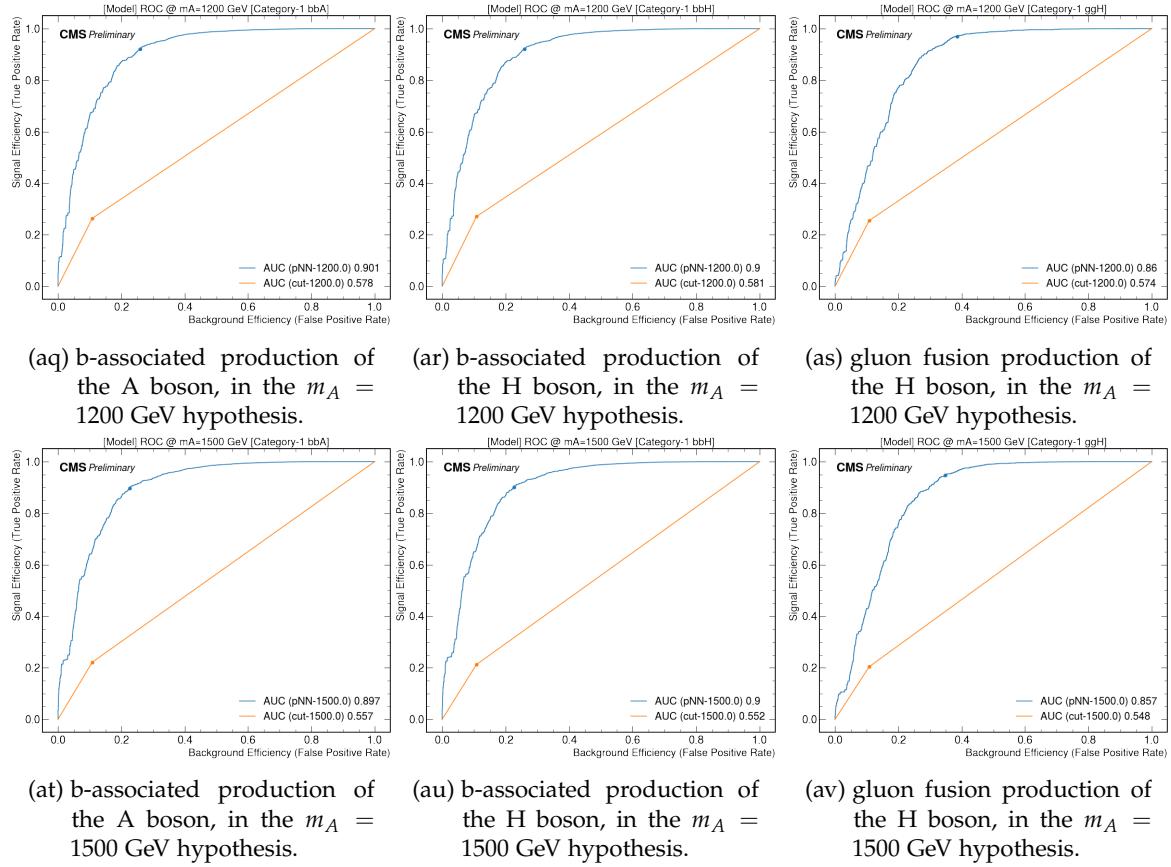
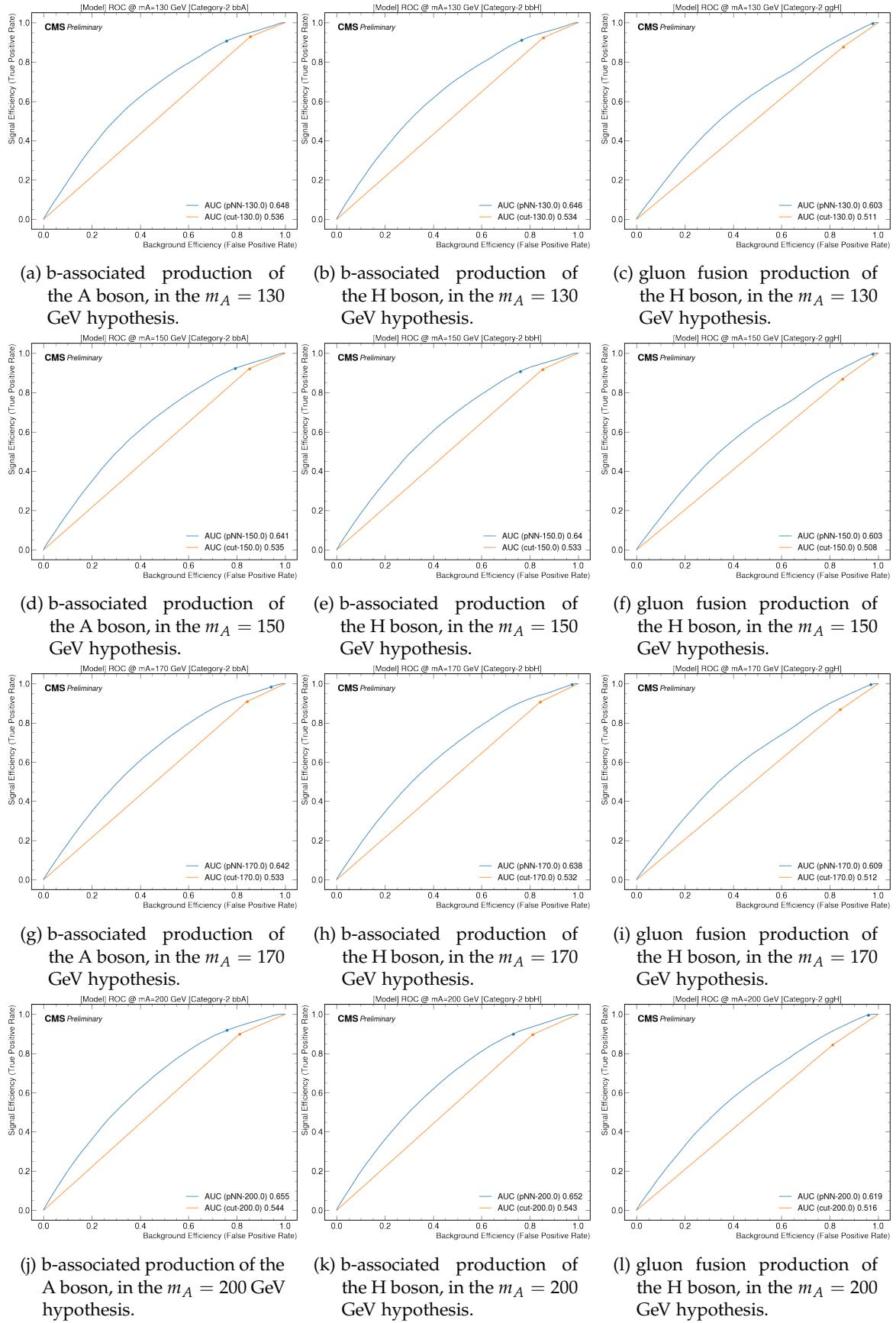
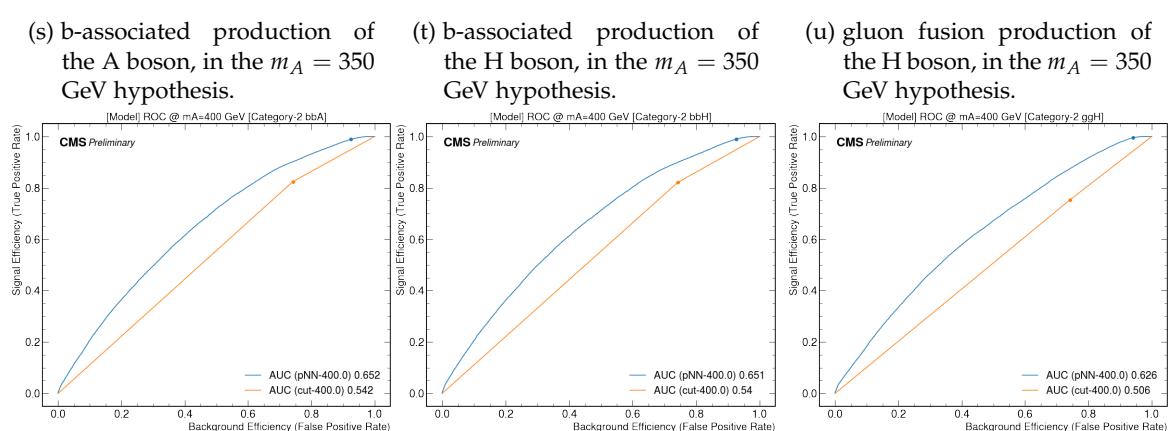
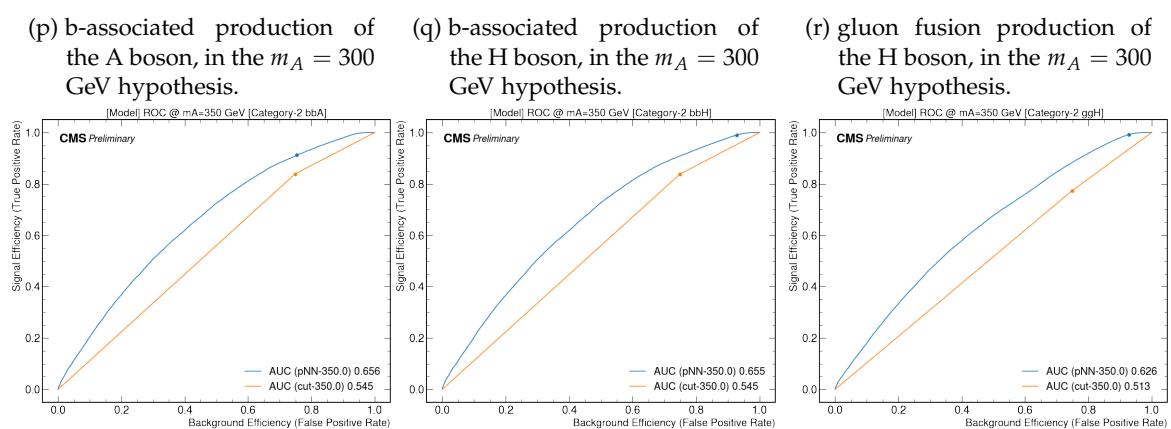
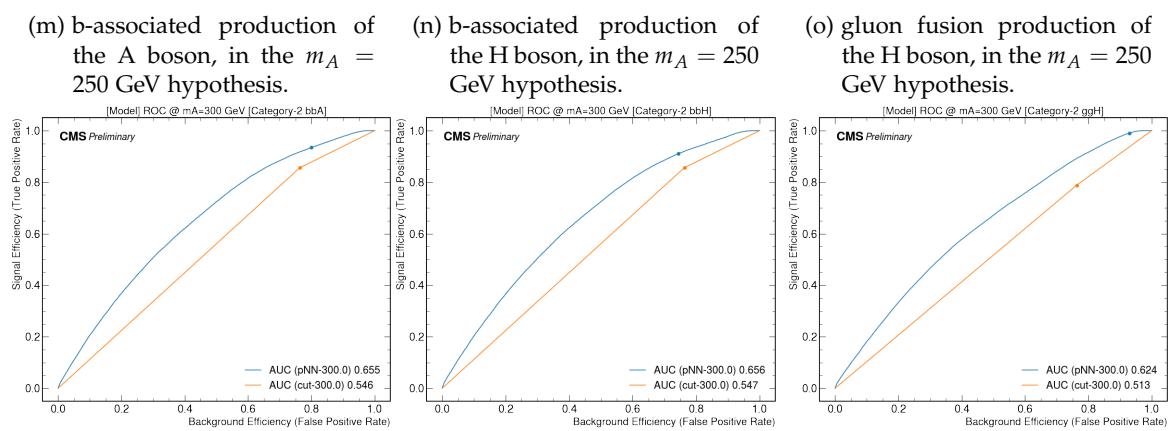
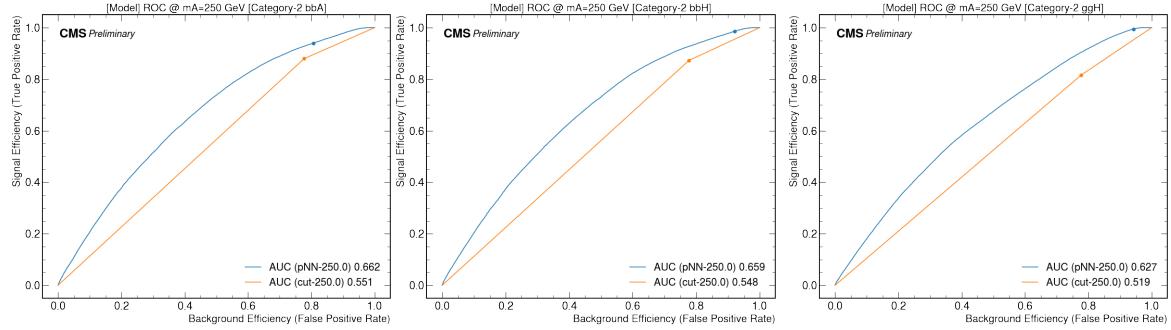
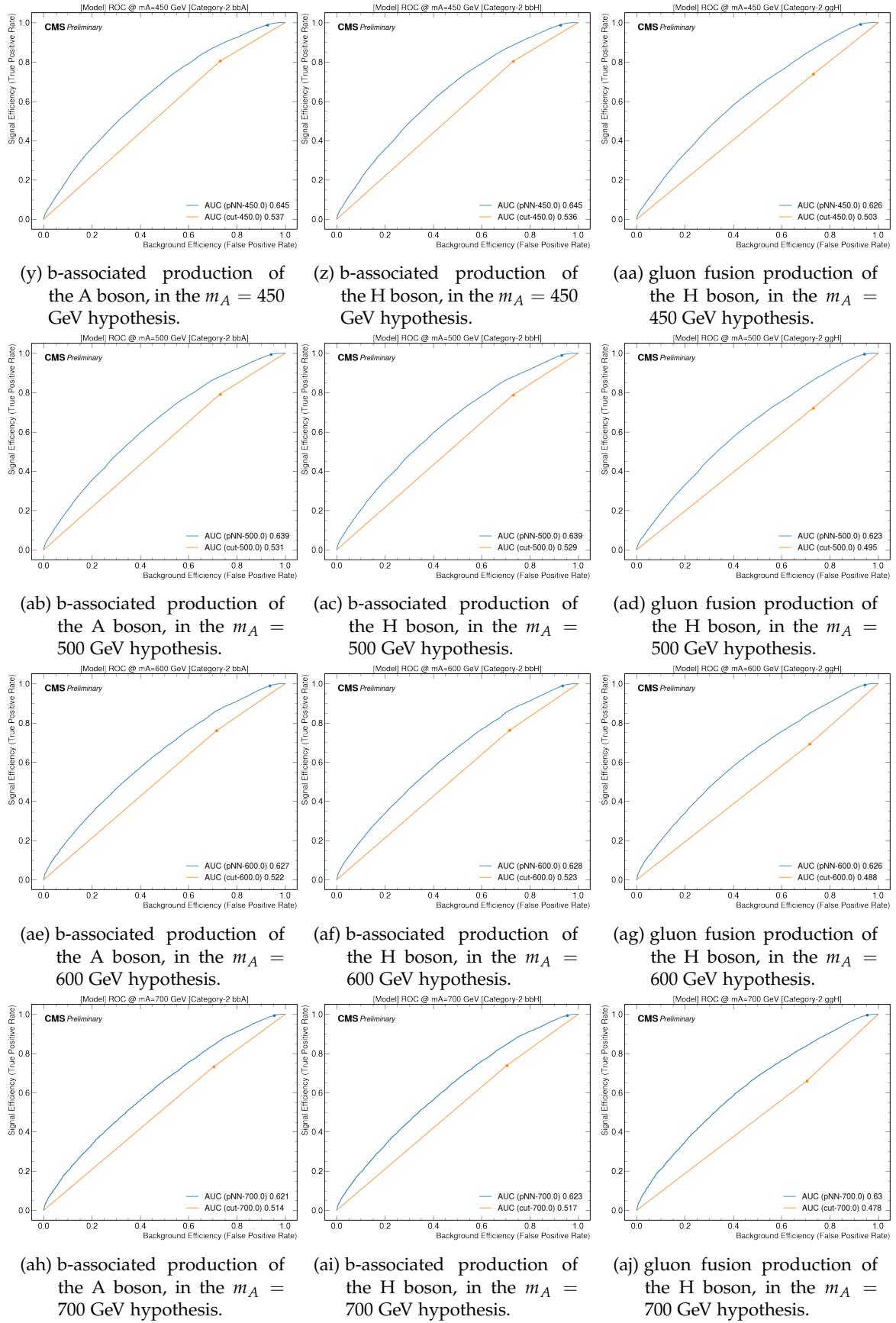


Figure B.3: In blue, the ROC curves for any m_A hypothesis, for CAT1 events. The AUC of the ROC curve is shown in the plot's legend and the working point, corresponding to the best cut selection, is represented with a blue dot. The orange broken line represents the signal efficiency versus the background efficiency for the cut-based selection method adopted in the previous analysis work: only the orange dot with the cut-based working point, have to be considered (the connecting orange lines are shown only for illustration purposes).







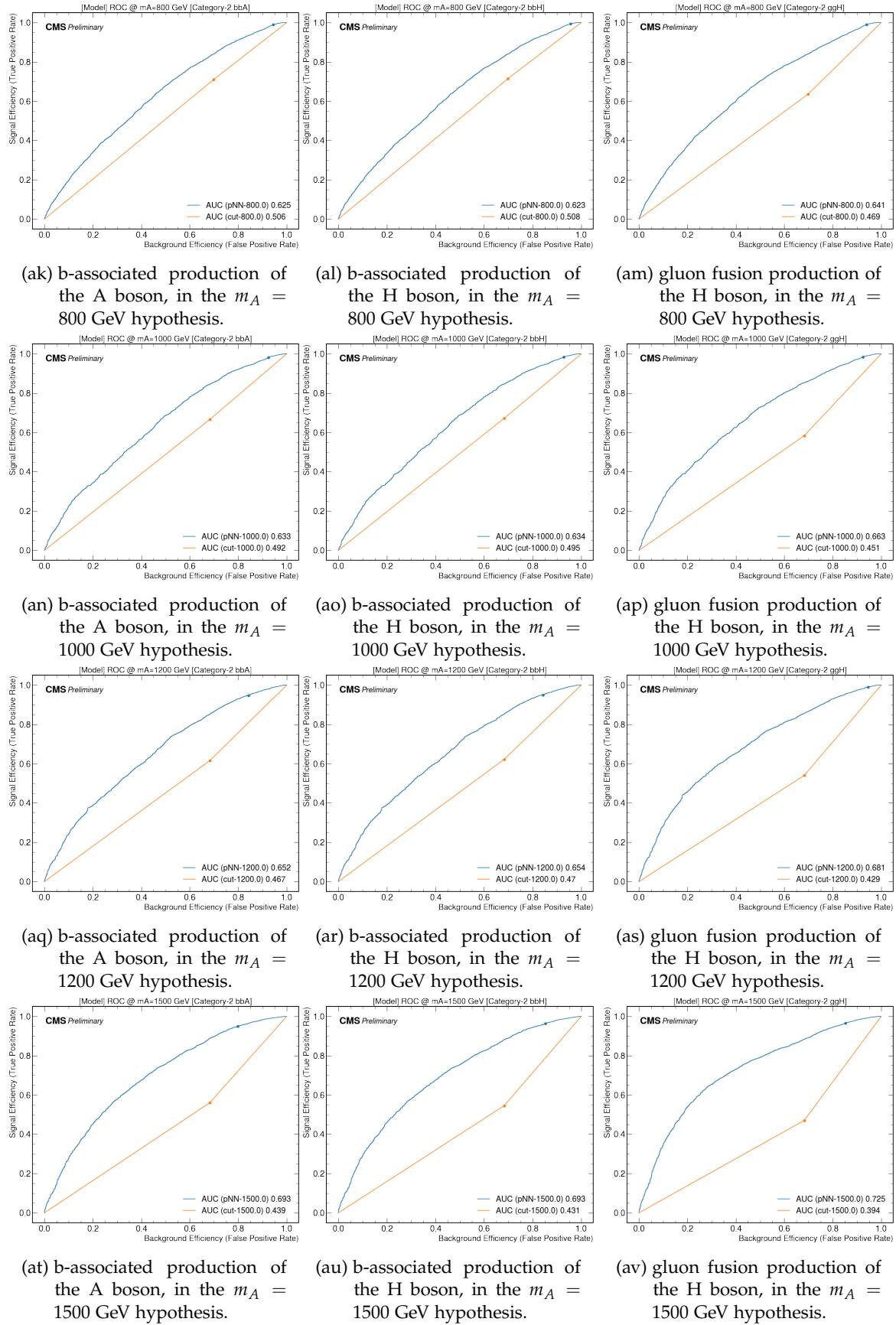
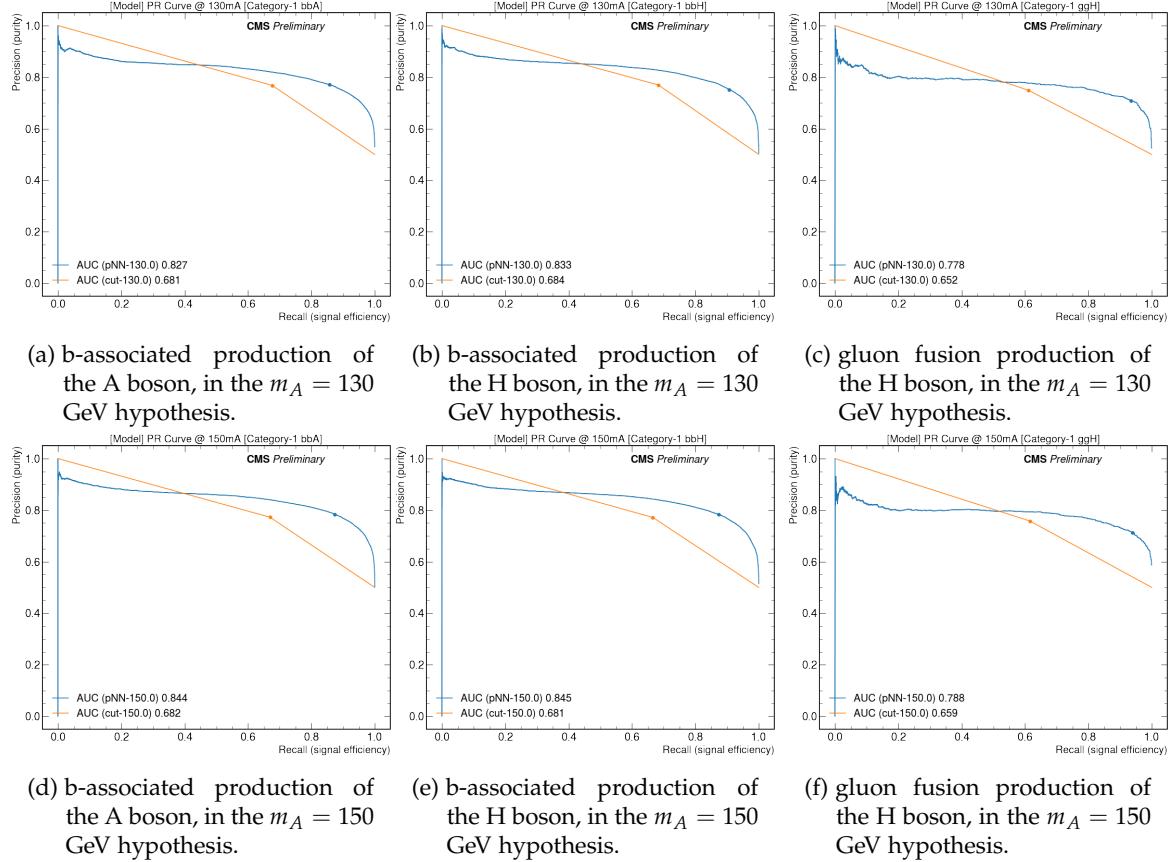
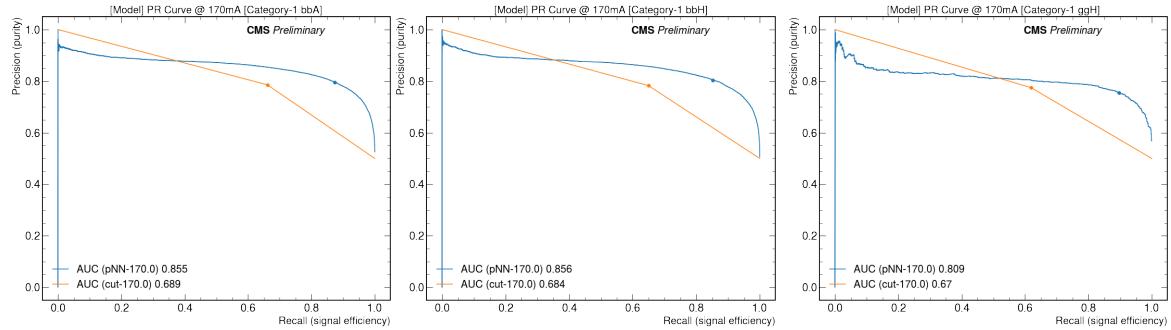
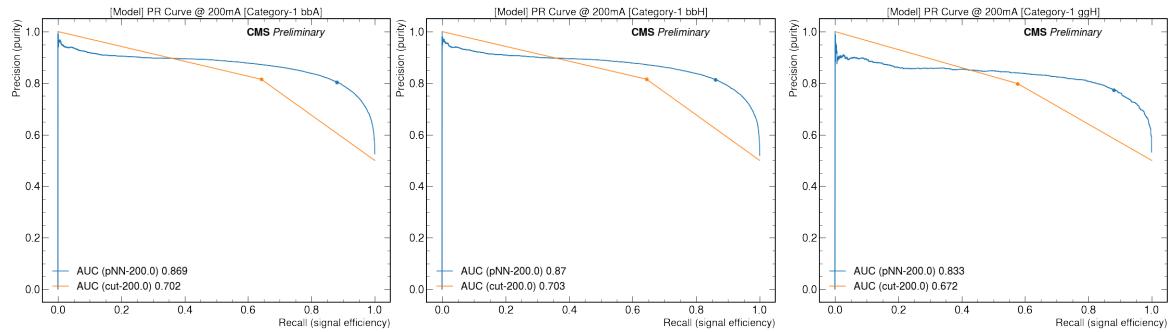
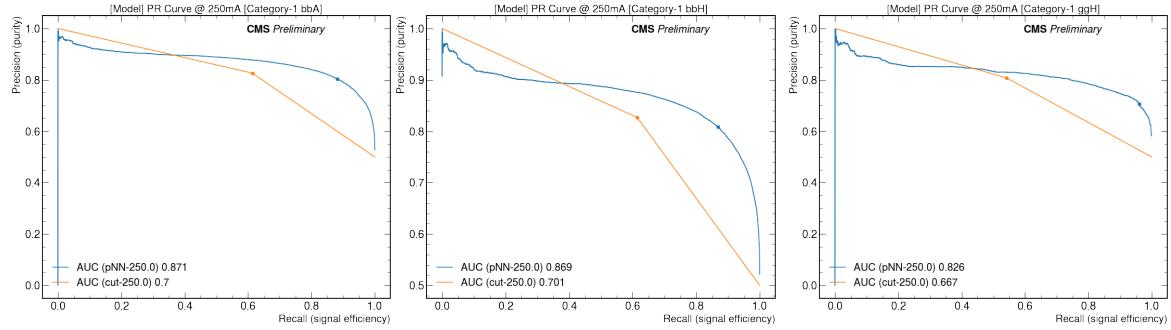
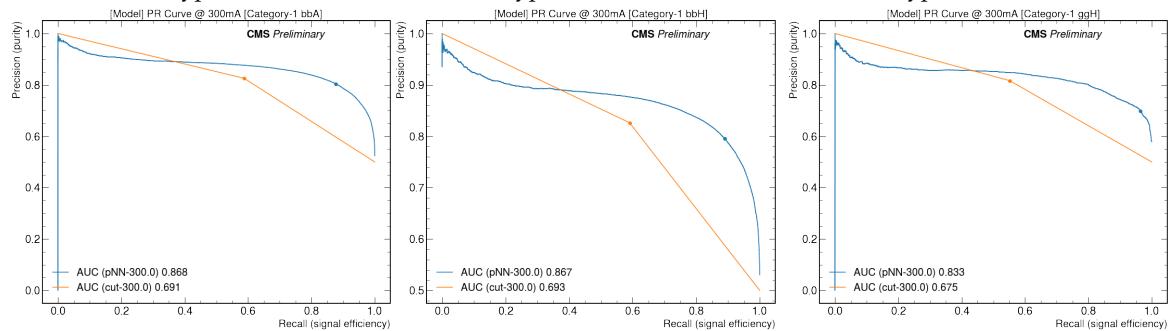


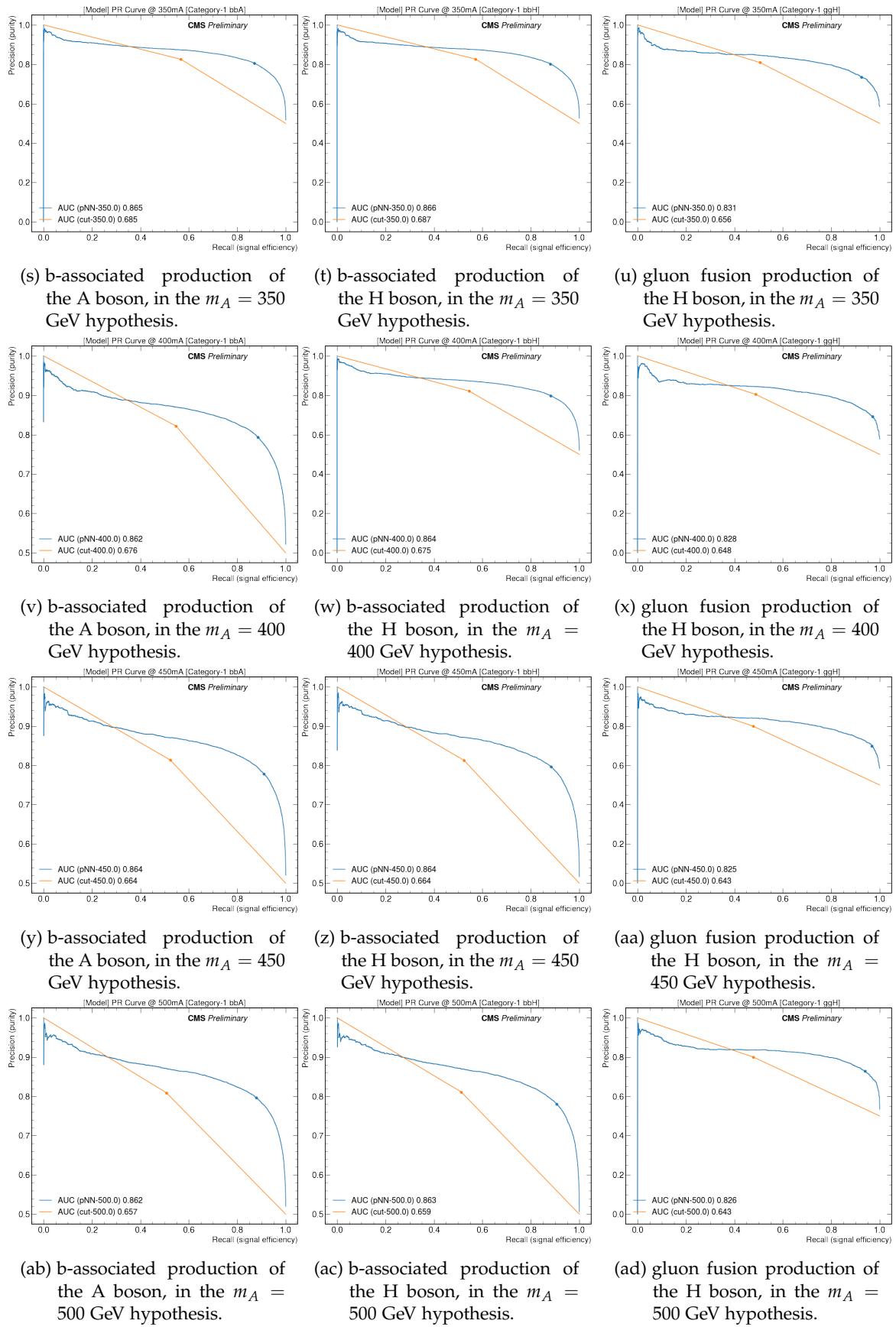
Figure B.4: In blue, the ROC curves for any m_A hypothesis, for CAT2 events. The AUC of the ROC curve is shown in the plot's legend and the working point, corresponding to the best cut selection, is represented with a blue dot. The orange broken line represents the signal efficiency versus the background efficiency for the cut-based selection method adopted in the previous analysis work: only the orange dot with the cut-based working point, have to be considered (the connecting orange lines are shown only for illustration purposes).

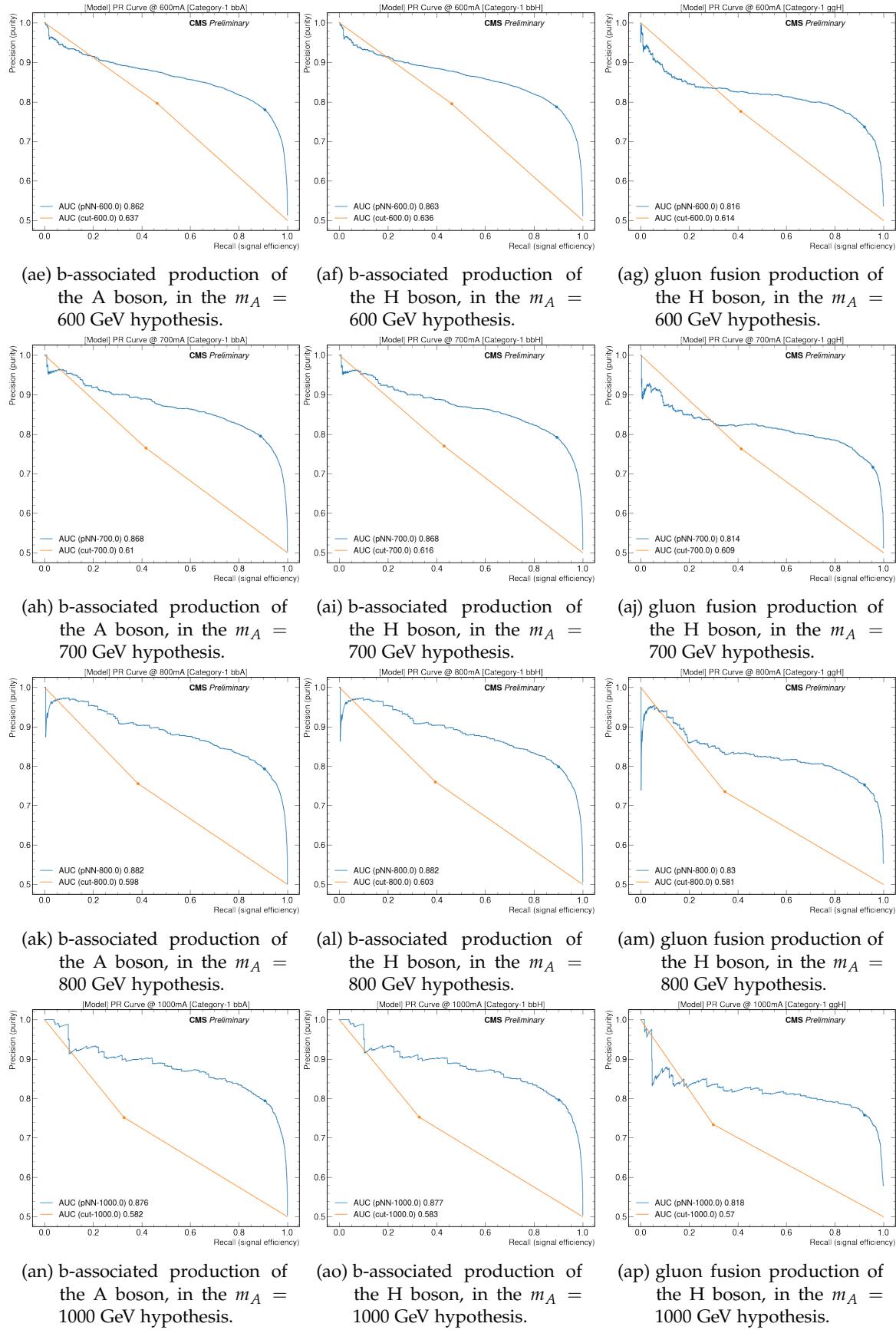
B.3 PNN PR CURVES

These are all the PR plots, for each m_A hypothesis, discussed in Section 6.3.8, and including also Figure 6.21 and 6.22, which are added in the main text for the sake of clarity.



(g) b-associated production of the A boson, in the $m_A = 170$ GeV hypothesis.(h) b-associated production of the H boson, in the $m_A = 170$ GeV hypothesis.(i) gluon fusion production of the H boson, in the $m_A = 170$ GeV hypothesis.(j) b-associated production of the A boson, in the $m_A = 200$ GeV hypothesis.(k) b-associated production of the H boson, in the $m_A = 200$ GeV hypothesis.(l) gluon fusion production of the H boson, in the $m_A = 200$ GeV hypothesis.(m) b-associated production of the A boson, in the $m_A = 250$ GeV hypothesis.(n) b-associated production of the H boson, in the $m_A = 250$ GeV hypothesis.(o) gluon fusion production of the H boson, in the $m_A = 250$ GeV hypothesis.(p) b-associated production of the A boson, in the $m_A = 300$ GeV hypothesis.(q) b-associated production of the H boson, in the $m_A = 300$ GeV hypothesis.(r) gluon fusion production of the H boson, in the $m_A = 300$ GeV hypothesis.





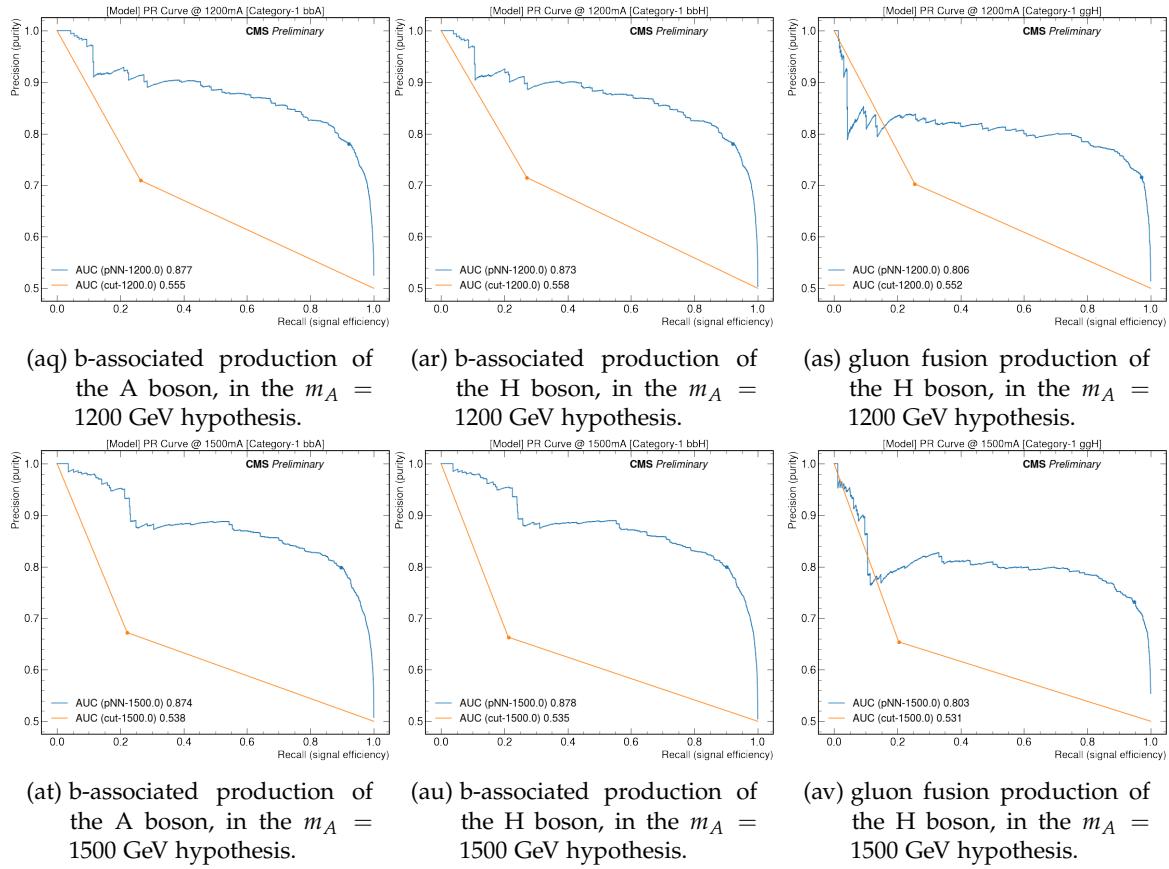
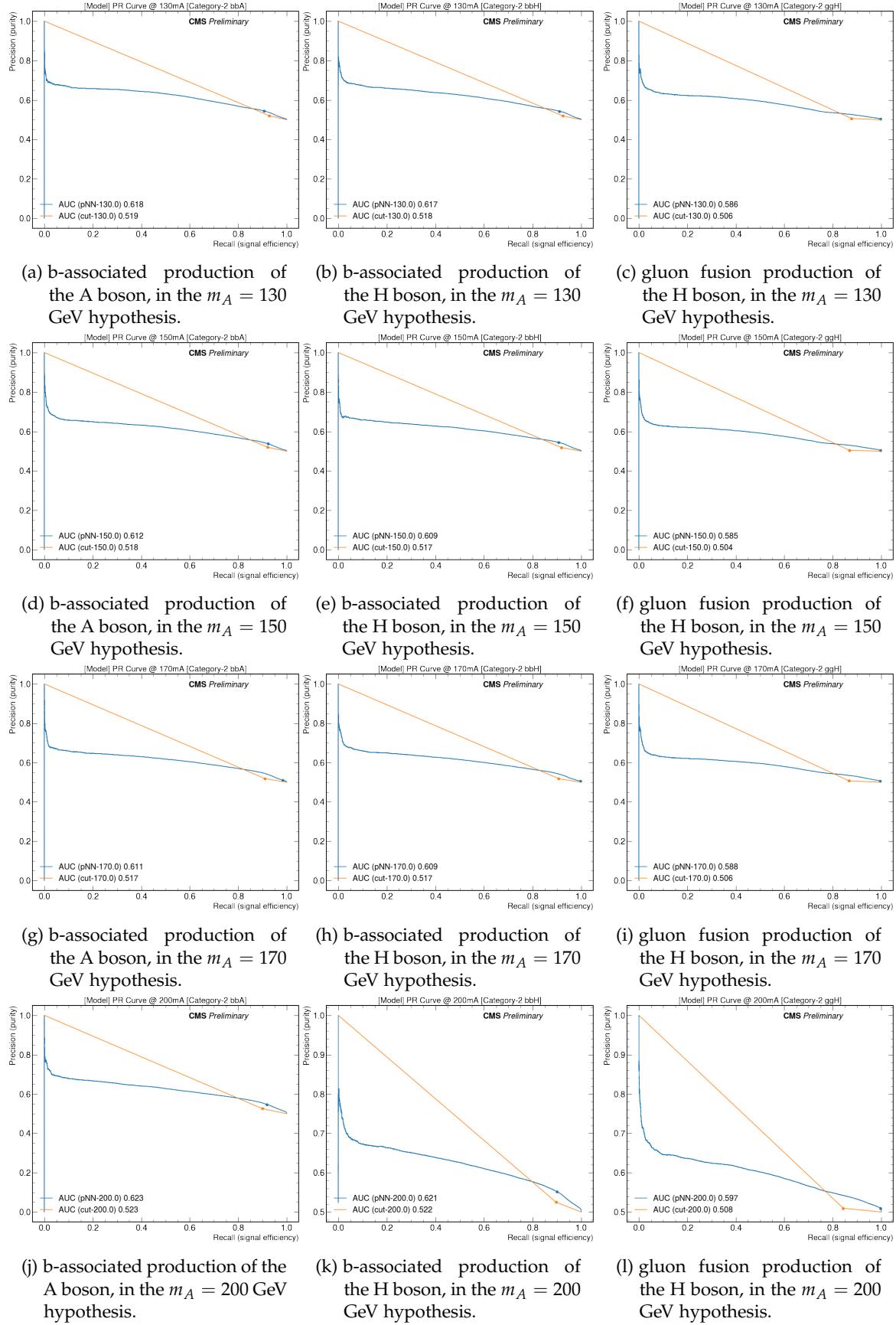
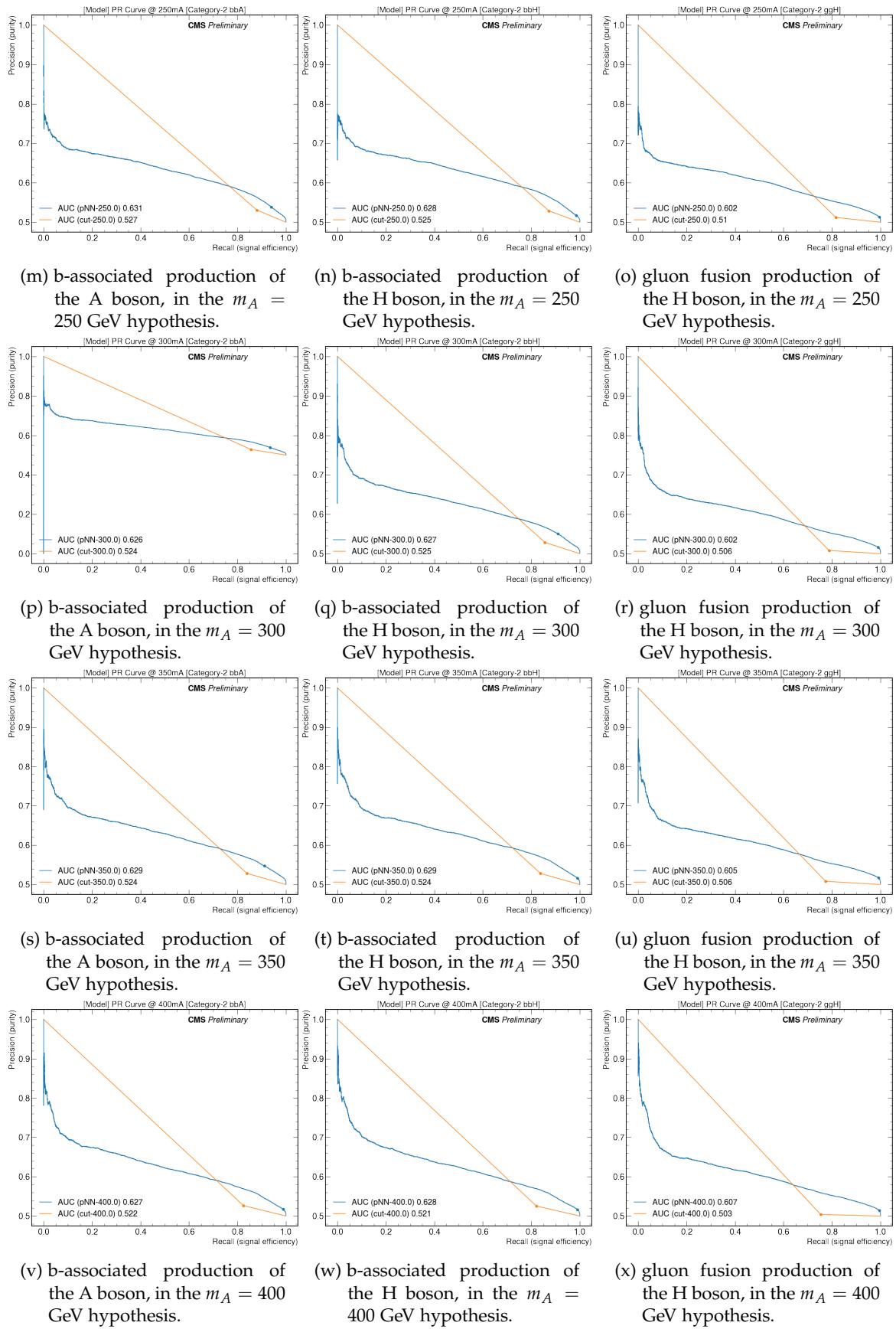
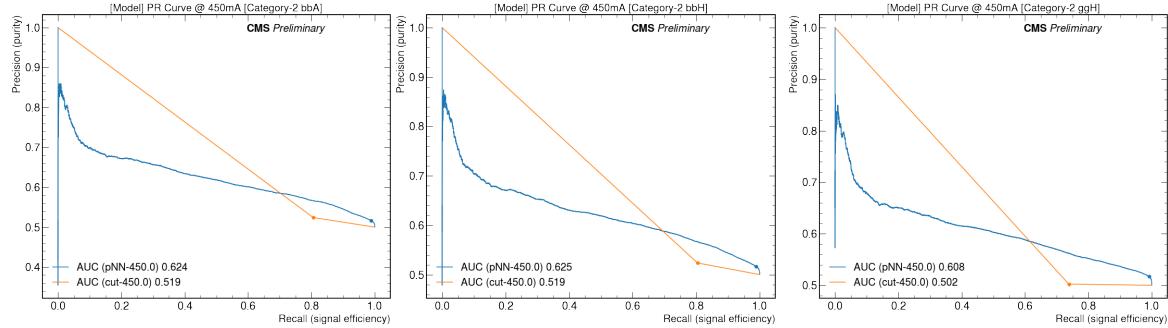
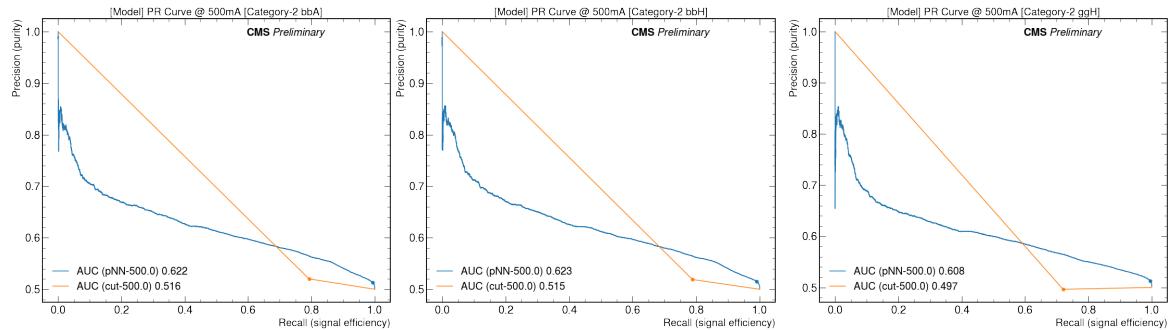
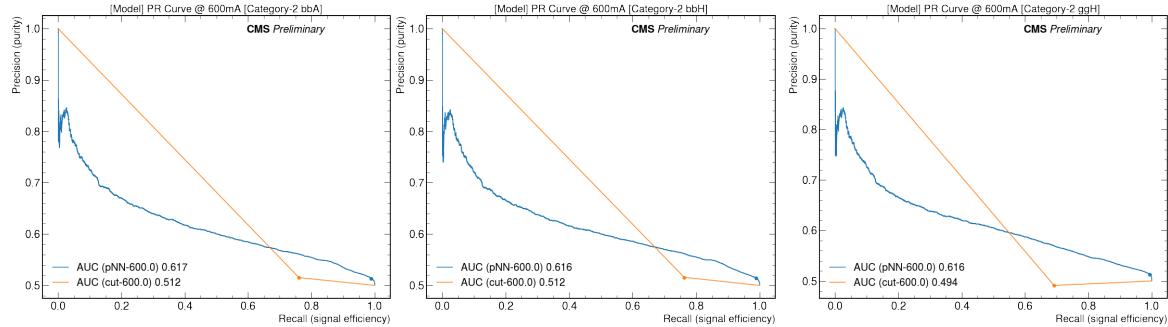
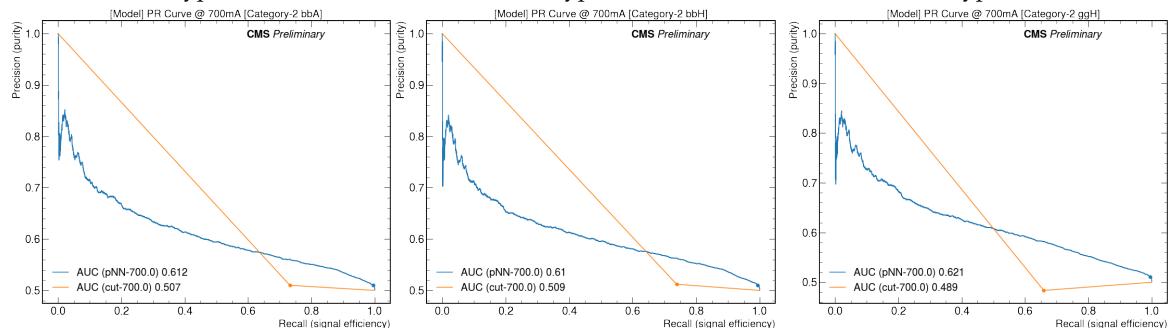


Figure B.5: In blue, the PR curves for any m_A hypothesis, for CAT1 events. The AUC of the PR curve is shown in the plot's legend and the working point, corresponding to the best cut selection, is represented with a blue dot. The orange broken line represents the signal efficiency versus the signal purity for the cut-based selection method adopted in the previous analysis work: only the orange dot with the cut-based working point, have to be considered (the connecting orange lines are shown only for illustration purposes).





(y) b-associated production of the A boson, in the $m_A = 450$ GeV hypothesis.(z) b-associated production of the H boson, in the $m_A = 450$ GeV hypothesis.(aa) gluon fusion production of the H boson, in the $m_A = 450$ GeV hypothesis.(ab) b-associated production of the A boson, in the $m_A = 500$ GeV hypothesis.(ac) b-associated production of the H boson, in the $m_A = 500$ GeV hypothesis.(ad) gluon fusion production of the H boson, in the $m_A = 500$ GeV hypothesis.(ae) b-associated production of the A boson, in the $m_A = 600$ GeV hypothesis.(af) b-associated production of the H boson, in the $m_A = 600$ GeV hypothesis.(ag) gluon fusion production of the H boson, in the $m_A = 600$ GeV hypothesis.(ah) b-associated production of the A boson, in the $m_A = 700$ GeV hypothesis.(ai) b-associated production of the H boson, in the $m_A = 700$ GeV hypothesis.(aj) gluon fusion production of the H boson, in the $m_A = 700$ GeV hypothesis.

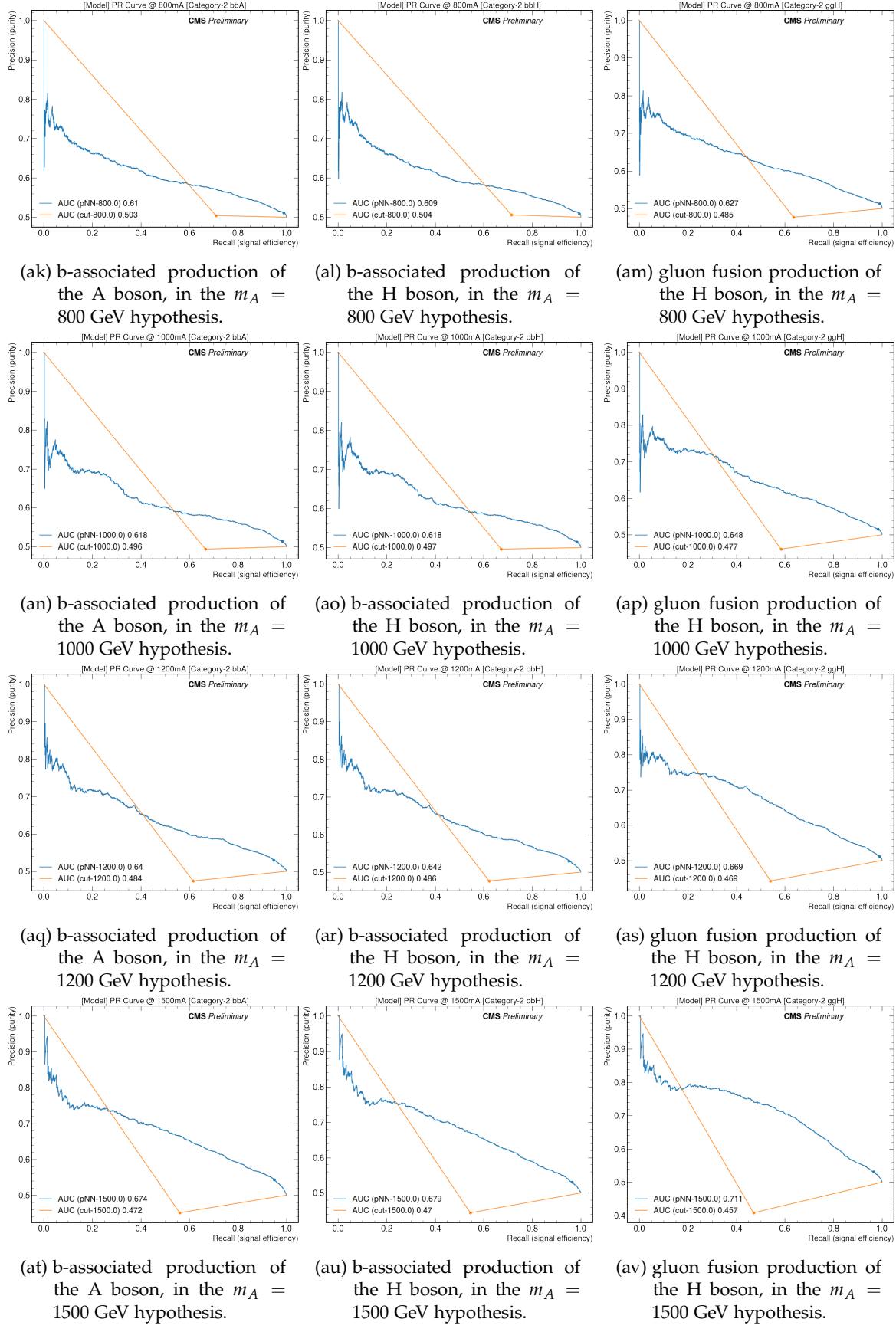


Figure B.6: In blue, the PR curves for any m_A hypothesis, for CAT2 events. The AUC of the PR curve is shown in the plot's legend and the working point, corresponding to the best cut selection, is represented with a blue dot. The orange broken line represents the signal efficiency versus the signal purity for the cut-based selection method adopted in the previous analysis work: only the orange dot with the cut-based working point, have to be considered (the connecting orange lines are shown only for illustration purposes).

BIBLIOGRAPHY

- [1] ‘Standard model particles’, [Online]. Available: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [2] S. H. Neddermeyer and C. D. Anderson, ‘Note on the Nature of Cosmic Ray Particles’, *Phys. Rev.*, vol. 51, pp. 884–886, 1937. doi: [10.1103/PhysRev.51.884](https://doi.org/10.1103/PhysRev.51.884).
- [3] M. L. Perl *et al.*, ‘Evidence for Anomalous Lepton Production in e^+e^- Annihilation’, *Phys. Rev. Lett.*, vol. 35, pp. 1489–1492, 1975. doi: [10.1103/PhysRevLett.35.1489](https://doi.org/10.1103/PhysRevLett.35.1489).
- [4] M. Tanabashi *et al.* [Particle Data Group Collaboration], ‘Review of Particle Physics’, *Phys. Rev. D*, vol. 98, no. 3, p. 030001, 2018. doi: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001).
- [5] R. Aaij *et al.* [LHCb Collaboration], ‘Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays’, *Phys. Rev. Lett.*, vol. 115, p. 072001, 2015. doi: [10.1103/PhysRevLett.115.072001](https://doi.org/10.1103/PhysRevLett.115.072001). arXiv: [1507.03414 \[hep-ex\]](https://arxiv.org/abs/1507.03414).
- [6] G. Cotugno, R. Faccini, A. D. Polosa and C. Sabelli, ‘Charmed Baryonium’, *Phys. Rev. Lett.*, vol. 104, p. 132005, 2010. doi: [10.1103/PhysRevLett.104.132005](https://doi.org/10.1103/PhysRevLett.104.132005). arXiv: [0911.2178 \[hep-ph\]](https://arxiv.org/abs/0911.2178).
- [7] S. L. Glashow, ‘Partial Symmetries of Weak Interactions’, *Nucl. Phys.*, vol. 22, pp. 579–588, 1961. doi: [10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [8] S. Weinberg, ‘A Model of Leptons’, *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967. doi: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [9] A. Salam, ‘Weak and Electromagnetic Interactions’, *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968. doi: [10.1142/9789812795915_0034](https://doi.org/10.1142/9789812795915_0034).
- [10] E. Fontanesi, ‘Precision measurements of the Higgs boson properties: from the $H \rightarrow ZZ^* \rightarrow 4l$ analysis with CMS at the LHC to the future large lepton colliders’, 2021. [Online]. Available: <http://amsdottorato.unibo.it/9661/>.
- [11] L. Giannini, ‘Deep Learning techniques for the observation of the Higgs boson decay to bottom quarks with the CMS experiment’, Presented 28 Jul 2020. [Online]. Available: <https://cds.cern.ch/record/2730094>.
- [12] F. Englert and R. Brout, ‘Broken Symmetry and the Mass of Gauge Vector Mesons’, *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 321–323, 1964. doi: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [13] P. W. Higgs, ‘Broken symmetries, massless particles and gauge fields’, *Phys. Lett.*, vol. 12, pp. 132–133, 1964. doi: [10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9).
- [14] P. W. Higgs, ‘Broken Symmetries and the Masses of Gauge Bosons’, *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 508–509, 1964. doi: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [15] P. W. Higgs, ‘Spontaneous Symmetry Breakdown without Massless Bosons’, *Phys. Rev.*, vol. 145, pp. 1156–1163, 1966. doi: [10.1103/PhysRev.145.1156](https://doi.org/10.1103/PhysRev.145.1156).
- [16] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, ‘Global Conservation Laws and Massless Particles’, *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 585–587, 1964. doi: [10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585).
- [17] T. W. B. Kibble, ‘Symmetry breaking in nonAbelian gauge theories’, *Phys. Rev.*, vol. 155, J. C. Taylor, Ed., pp. 1554–1561, 1967. doi: [10.1103/PhysRev.155.1554](https://doi.org/10.1103/PhysRev.155.1554).
- [18] Y. Nambu and G. Jona-Lasinio, ‘Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1.’, *Phys. Rev.*, vol. 122, T. Eguchi, Ed., pp. 345–358, 1961. doi: [10.1103/PhysRev.122.345](https://doi.org/10.1103/PhysRev.122.345).
- [19] M. Gell-Mann and M. Levy, ‘The axial vector current in beta decay’, *Nuovo Cim.*, vol. 16, p. 705, 1960. doi: [10.1007/BF02859738](https://doi.org/10.1007/BF02859738).
- [20] J. Ellis, M. K. Gaillard and D. V. Nanopoulos, ‘A Historical Profile of the Higgs Boson’, in *The standard theory of particle physics: Essays to celebrate CERN’s 60th anniversary*, L. Maiani and L. Rolandi, Eds. 2016, pp. 255–274. doi: [10.1142/9789814733519_0014](https://doi.org/10.1142/9789814733519_0014). arXiv: [1504.07217 \[hep-ph\]](https://arxiv.org/abs/1504.07217).
- [21] J. Alcaraz *et al.* [ALEPH, DELPHI, L3, OPAL, LEP Electroweak Working Group Collaboration], ‘A Combination of preliminary electroweak measurements and constraints on the Standard Model’, Dec. 2006. arXiv: [hep-ex/0612034](https://arxiv.org/abs/hep-ex/0612034).

- [22] G. Bernardi, V. Buescher, A. Khanov, J.-F. Grivaz, B. Heinemann, S. Mrenna, J. Nachtman and S. M. Wang [D0, CDF Collaboration], ‘Combined D0 and CDF upper limits on standard-model Higgs-boson production’, in *33rd International Conference on High Energy Physics*, Dec. 2006. arXiv: [hep-ex/0612044](#).
- [23] G. Aad *et al.* [ATLAS Collaboration], ‘Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC’, *Phys. Lett. B*, vol. 716, pp. 1–29, 2012. doi: [10.1016/j.physletb.2012.08.020](#). arXiv: [1207.7214 \[hep-ex\]](#).
- [24] S. Chatrchyan *et al.* [CMS Collaboration], ‘Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC’, *Phys. Lett. B*, vol. 716, pp. 30–61, 2012. doi: [10.1016/j.physletb.2012.08.021](#). arXiv: [1207.7235 \[hep-ex\]](#).
- [25] S. Chatrchyan *et al.* [CMS Collaboration], ‘Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV’, *JHEP*, vol. 06, p. 081, 2013. doi: [10.1007/JHEP06\(2013\)081](#). arXiv: [1303.4571 \[hep-ex\]](#).
- [26] D. de Florian *et al.* [LHC Higgs Cross Section Working Group Collaboration], ‘Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector’, vol. 2/2017, Oct. 2016. doi: [10.23731/CYRM-2017-002](#). arXiv: [1610.07922 \[hep-ph\]](#).
- [27] G. Aad *et al.* [ATLAS, CMS Collaboration], ‘Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments’, *Phys. Rev. Lett.*, vol. 114, p. 191803, 2015. doi: [10.1103/PhysRevLett.114.191803](#). arXiv: [1503.07589 \[hep-ex\]](#).
- [28] G. Aad *et al.* [ATLAS, CMS Collaboration], ‘Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV’, *JHEP*, vol. 08, p. 045, 2016. doi: [10.1007/JHEP08\(2016\)045](#). arXiv: [1606.02266 \[hep-ex\]](#).
- [29] M. Aaboud *et al.* [ATLAS Collaboration], ‘Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector’, *Phys. Rev. D*, vol. 99, p. 072001, 2019. doi: [10.1103/PhysRevD.99.072001](#). arXiv: [1811.08856 \[hep-ex\]](#).
- [30] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector’, *Phys. Lett. B*, vol. 779, pp. 283–316, 2018. doi: [10.1016/j.physletb.2018.02.004](#). arXiv: [1708.00373 \[hep-ex\]](#).
- [31] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Observation of $t\bar{t}H$ production’, *Phys. Rev. Lett.*, vol. 120, no. 23, p. 231801, 2018. doi: [10.1103/PhysRevLett.120.231801](#). arXiv: [1804.02610 \[hep-ex\]](#).
- [32] M. Aaboud *et al.* [ATLAS Collaboration], ‘Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector’, *Phys. Lett. B*, vol. 784, pp. 173–191, 2018. doi: [10.1016/j.physletb.2018.07.035](#). arXiv: [1806.00425 \[hep-ex\]](#).
- [33] M. Aaboud *et al.* [ATLAS Collaboration], ‘Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector’, *Phys. Lett. B*, vol. 786, pp. 59–86, 2018. doi: [10.1016/j.physletb.2018.09.013](#). arXiv: [1808.08238 \[hep-ex\]](#).
- [34] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Observation of Higgs boson decay to bottom quarks’, *Phys. Rev. Lett.*, vol. 121, no. 12, p. 121801, 2018. doi: [10.1103/PhysRevLett.121.121801](#). arXiv: [1808.08242 \[hep-ex\]](#).
- [35] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Evidence for Higgs boson decay to a pair of muons’, *JHEP*, vol. 01, p. 148, 2021. doi: [10.1007/JHEP01\(2021\)148](#). arXiv: [2009.04363 \[hep-ex\]](#).
- [36] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV’, *Eur. Phys. J. C*, vol. 79, no. 5, p. 421, 2019. doi: [10.1140/epjc/s10052-019-6909-y](#). arXiv: [1809.10733 \[hep-ex\]](#).
- [37] S. M. Bilenky and S. T. Petcov, ‘Massive Neutrinos and Neutrino Oscillations’, *Rev. Mod. Phys.*, vol. 59, p. 671, 1987, [Erratum: *Rev.Mod.Phys.* 61, 169 (1989), Erratum: *Rev.Mod.Phys.* 60, 575–575 (1988)]. doi: [10.1103/RevModPhys.59.671](#).
- [38] A. Gando *et al.* [KamLAND-Zen Collaboration], ‘Limit on Neutrinoless $\beta\beta$ Decay of ^{136}Xe from the First Phase of KamLAND-Zen and Comparison with the Positive Claim in ^{76}Ge ’, *Phys. Rev. Lett.*, vol. 110, no. 6, p. 062502, 2013. doi: [10.1103/PhysRevLett.110.062502](#). arXiv: [1211.3863 \[hep-ex\]](#).
- [39] Y. Nagashima, *Beyond the Standard Model of elementary particle physics*. Weinheim, USA: Wiley-VCH, 2014, ISBN: 978-3-527-41177-1, 978-3-527-66505-1.
- [40] A. J. Buras, J. R. Ellis, M. K. Gaillard and D. V. Nanopoulos, ‘Aspects of the Grand Unification of Strong, Weak and Electromagnetic Interactions’, *Nucl. Phys. B*, vol. 135, pp. 66–92, 1978. doi: [10.1016/0550-3213\(78\)90214-6](#).

- [41] D. I. Kazakov, ‘Beyond the Standard Model: In search of supersymmetry’, in *2000 European School of High-Energy Physics*, Aug. 2000, pp. 125–199. arXiv: [hep-ph/0012288](#).
- [42] S. P. Martin, ‘A Supersymmetry primer’, *Adv. Ser. Direct. High Energy Phys.*, vol. 18, G. L. Kane, Ed., pp. 1–98, 1998. doi: [10.1142/9789812839657_0001](#). arXiv: [hep-ph/9709356](#).
- [43] Y. Nagashima, *Elementary particle physics. Vol. 1: Quantum field theory and particles*. 2010.
- [44] G. R. Farrar and P. Fayet, ‘Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry’, *Phys. Lett. B*, vol. 76, pp. 575–579, 1978. doi: [10.1016/0370-2693\(78\)90858-4](#).
- [45] P. Fayet, ‘Relations Between the Masses of the Superpartners of Leptons and Quarks, the Goldstino Couplings and the Neutral Currents’, *Phys. Lett. B*, vol. 84, p. 416, 1979. doi: [10.1016/0370-2693\(79\)91229-2](#).
- [46] P. Fayet, ‘Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions’, *Phys. Lett. B*, vol. 69, p. 489, 1977. doi: [10.1016/0370-2693\(77\)90852-8](#).
- [47] S. Dimopoulos and H. Georgi, ‘Softly Broken Supersymmetry and SU(5)’, *Nucl. Phys. B*, vol. 193, pp. 150–162, 1981. doi: [10.1016/0550-3213\(81\)90522-8](#).
- [48] P. Fayet, ‘Supersymmetry and Weak, Electromagnetic and Strong Interactions’, *Phys. Lett. B*, vol. 64, p. 159, 1976. doi: [10.1016/0370-2693\(76\)90319-1](#).
- [49] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher and J. P. Silva, ‘Theory and phenomenology of two-Higgs-doublet models’, *Phys. Rept.*, vol. 516, pp. 1–102, 2012. doi: [10.1016/j.physrep.2012.02.002](#). arXiv: [1106.0034 \[hep-ph\]](#).
- [50] V. D. Barger, J. L. Hewett and R. J. N. Phillips, ‘New Constraints on the Charged Higgs Sector in Two Higgs Doublet Models’, *Phys. Rev. D*, vol. 41, pp. 3421–3441, 1990. doi: [10.1103/PhysRevD.41.3421](#).
- [51] T. N. Dao, ‘On the Phenomenology of Charged Higgs Bosons in the complex MSSM’, 2012.
- [52] A. Jelinskas, ‘Search for heavy neutral MSSM Higgs bosons in the di-tau final state with the ATLAS detector’, Presented 12 Apr 2019, Jan. 2019. [Online]. Available: <http://cds.cern.ch/record/2672453>.
- [53] D. Alvarez Piqueras, ‘Measurement of the Higgs boson coupling with tau leptons and search for an additional neutral MSSM Higgs boson with the ATLAS detector’, Presented 07 May 2018, Apr. 2018. [Online]. Available: <http://cds.cern.ch/record/2622452>.
- [54] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for top squark pair production in pp collisions at $\sqrt{s} = 13$ TeV using single lepton events’, *JHEP*, vol. 10, p. 019, 2017. doi: [10.1007/JHEP10\(2017\)019](#). arXiv: [1706.04402 \[hep-ex\]](#).
- [55] M. Aaboud *et al.* [ATLAS Collaboration], ‘Search for supersymmetry in events with b -tagged jets and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector’, *JHEP*, vol. 11, p. 195, 2017. doi: [10.1007/JHEP11\(2017\)195](#). arXiv: [1708.09266 \[hep-ex\]](#).
- [56] M. Carena, S. Heinemeyer, O. Stål, C. E. M. Wagner and G. Weiglein, ‘MSSM Higgs Boson Searches at the LHC: Benchmark Scenarios after the Discovery of a Higgs-like Particle’, *Eur. Phys. J. C*, vol. 73, no. 9, p. 2552, 2013. doi: [10.1140/epjc/s10052-013-2552-1](#). arXiv: [1302.7033 \[hep-ph\]](#).
- [57] A. Djouadi, L. Maiani, G. Moreau, A. Polosa, J. Quevillon and V. Riquer, ‘The post-Higgs MSSM scenario: Habemus MSSM?’, *Eur. Phys. J. C*, vol. 73, p. 2650, 2013. doi: [10.1140/epjc/s10052-013-2650-0](#). arXiv: [1307.5205 \[hep-ph\]](#).
- [58] M. Carena, S. Heinemeyer, C. E. M. Wagner and G. Weiglein, ‘Suggestions for Benchmark Scenarios for MSSM Higgs Boson Searches at Hadron Colliders’, *Eur. Phys. J. C*, vol. 26, P. Aurenche, G. Belanger, F. Boudjema, J. P. Guillet and E. Pilon, Eds., pp. 601–607, 2003. doi: [10.1140/epjc/s2002-01084-3](#). arXiv: [hep-ph/0202167](#).
- [59] E. Bagnaschi *et al.*, ‘Benchmark scenarios for low $\tan\beta$ in the MSSM’, Aug. 2015. [Online]. Available: <http://cds.cern.ch/record/2039911>.
- [60] A. Bailey [ATLAS, CMS Collaboration], ‘Searches for BSM Higgs bosons at ATLAS and CMS’, *PoS*, vol. LHCP2020, B. Mansoulie, G. Marchiori, R. Salern and T. Bos, Eds., p. 011, 2021. doi: [10.22323/1.382.0011](#).
- [61] G. Aad *et al.* [ATLAS Collaboration], ‘Search for the neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector’, *JHEP*, vol. 02, p. 095, 2013. doi: [10.1007/JHEP02\(2013\)095](#). arXiv: [1211.6956 \[hep-ex\]](#).
- [62] G. Aad *et al.* [ATLAS Collaboration], ‘Search for neutral Higgs bosons of the minimal supersymmetric standard model in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector’, *JHEP*, vol. 11, p. 056, 2014. doi: [10.1007/JHEP11\(2014\)056](#). arXiv: [1409.6064 \[hep-ex\]](#).

- [63] M. Aaboud *et al.* [ATLAS Collaboration], ‘Search for Minimal Supersymmetric Standard Model Higgs bosons H/A and for a Z' boson in the $\tau\tau$ final state produced in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector’, *Eur. Phys. J. C*, vol. 76, no. 11, p. 585, 2016. doi: [10.1140/epjc/s10052-016-4400-6](https://doi.org/10.1140/epjc/s10052-016-4400-6). arXiv: [1608.00890 \[hep-ex\]](https://arxiv.org/abs/1608.00890).
- [64] M. Aaboud *et al.* [ATLAS Collaboration], ‘Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector’, *JHEP*, vol. 01, p. 055, 2018. doi: [10.1007/JHEP01\(2018\)055](https://doi.org/10.1007/JHEP01(2018)055). arXiv: [1709.07242 \[hep-ex\]](https://arxiv.org/abs/1709.07242).
- [65] S. Chatrchyan *et al.* [CMS Collaboration], ‘Search for Neutral MSSM Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV’, *Phys. Rev. Lett.*, vol. 106, p. 231801, 2011. doi: [10.1103/PhysRevLett.106.231801](https://doi.org/10.1103/PhysRevLett.106.231801). arXiv: [1104.1619 \[hep-ex\]](https://arxiv.org/abs/1104.1619).
- [66] S. Chatrchyan *et al.* [CMS Collaboration], ‘Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV’, *Phys. Lett. B*, vol. 713, pp. 68–90, 2012. doi: [10.1016/j.physletb.2012.05.028](https://doi.org/10.1016/j.physletb.2012.05.028). arXiv: [1202.4083 \[hep-ex\]](https://arxiv.org/abs/1202.4083).
- [67] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV’, *JHEP*, vol. 09, p. 007, 2018. doi: [10.1007/JHEP09\(2018\)007](https://doi.org/10.1007/JHEP09(2018)007). arXiv: [1803.06553 \[hep-ex\]](https://arxiv.org/abs/1803.06553).
- [68] S. Chatrchyan *et al.* [CMS Collaboration], ‘Search for a Higgs Boson Decaying into a b Quark Pair and Produced in Association with b Quarks in Proton-Proton Collisions at 7 TeV’, *Phys. Lett. B*, vol. 722, pp. 207–232, 2013. doi: [10.1016/j.physletb.2013.04.017](https://doi.org/10.1016/j.physletb.2013.04.017). arXiv: [1302.2892 \[hep-ex\]](https://arxiv.org/abs/1302.2892).
- [69] V. Khachatryan *et al.* [CMS Collaboration], ‘Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks’, *JHEP*, vol. 11, p. 071, 2015. doi: [10.1007/JHEP11\(2015\)071](https://doi.org/10.1007/JHEP11(2015)071). arXiv: [1506.08329 \[hep-ex\]](https://arxiv.org/abs/1506.08329).
- [70] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for beyond the standard model Higgs bosons decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV’, *JHEP*, vol. 08, p. 113, 2018. doi: [10.1007/JHEP08\(2018\)113](https://doi.org/10.1007/JHEP08(2018)113). arXiv: [1805.12191 \[hep-ex\]](https://arxiv.org/abs/1805.12191).
- [71] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for a heavy Higgs boson decaying to a pair of W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV’, *JHEP*, vol. 03, p. 034, 2020. doi: [10.1007/JHEP03\(2020\)034](https://doi.org/10.1007/JHEP03(2020)034). arXiv: [1912.01594 \[hep-ex\]](https://arxiv.org/abs/1912.01594).
- [72] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV’, *Phys. Lett. B*, vol. 778, pp. 101–127, 2018. doi: [10.1016/j.physletb.2018.01.001](https://doi.org/10.1016/j.physletb.2018.01.001). arXiv: [1707.02909 \[hep-ex\]](https://arxiv.org/abs/1707.02909).
- [73] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for a heavy pseudoscalar Higgs boson decaying into a 125 GeV Higgs boson and a Z boson in final states with two tau and two light leptons at $\sqrt{s} = 13$ TeV’, *JHEP*, vol. 03, p. 065, 2020. doi: [10.1007/JHEP03\(2020\)065](https://doi.org/10.1007/JHEP03(2020)065). arXiv: [1910.11634 \[hep-ex\]](https://arxiv.org/abs/1910.11634).
- [74] S. Schael *et al.* [ALEPH, DELPHI, L3, OPAL, LEP Working Group for Higgs Boson Searches Collaboration], ‘Search for neutral MSSM Higgs bosons at LEP’, *Eur. Phys. J. C*, vol. 47, pp. 547–587, 2006. doi: [10.1140/epjc/s2006-02569-7](https://doi.org/10.1140/epjc/s2006-02569-7). arXiv: [hep-ex/0602042](https://arxiv.org/abs/hep-ex/0602042).
- [75] T. Aaltonen *et al.* [CDF Collaboration], ‘Search for Higgs bosons predicted in two-Higgs-doublet models via decays to tau lepton pairs in 1.96-TeV p anti-p collisions’, *Phys. Rev. Lett.*, vol. 103, p. 201801, 2009. doi: [10.1103/PhysRevLett.103.201801](https://doi.org/10.1103/PhysRevLett.103.201801). arXiv: [0906.1014 \[hep-ex\]](https://arxiv.org/abs/0906.1014).
- [76] T. Aaltonen *et al.* [CDF Collaboration], ‘Search for Higgs Bosons Produced in Association with b-quarks’, *Phys. Rev. D*, vol. 85, p. 032005, 2012. doi: [10.1103/PhysRevD.85.032005](https://doi.org/10.1103/PhysRevD.85.032005). arXiv: [1106.4782 \[hep-ex\]](https://arxiv.org/abs/1106.4782).
- [77] V. M. Abazov *et al.* [D0 Collaboration], ‘Search for Neutral Higgs Bosons in the Multi-b-Jet Topology in 5.2fb^{-1} of $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV’, *Phys. Lett. B*, vol. 698, pp. 97–104, 2011. doi: [10.1016/j.physletb.2011.02.062](https://doi.org/10.1016/j.physletb.2011.02.062). arXiv: [1011.1931 \[hep-ex\]](https://arxiv.org/abs/1011.1931).
- [78] V. M. Abazov *et al.* [D0 Collaboration], ‘Search for Higgs bosons decaying to $\tau\tau$ pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV’, *Phys. Lett. B*, vol. 707, pp. 323–329, 2012. doi: [10.1016/j.physletb.2011.12.050](https://doi.org/10.1016/j.physletb.2011.12.050). arXiv: [1106.4555 \[hep-ex\]](https://arxiv.org/abs/1106.4555).
- [79] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic and J. P. P. Proudlock, *LHC Design Report*. 2012, vol. 1, 2, 3. [Online]. Available: <http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html>.
- [80] L. Evans and P. Bryant, ‘LHC Machine’, *Journal of Instrumentation*, vol. 3, no. 08, IOPscience, Ed., S08001, 2008. [Online]. Available: <http://iopscience.iop.org/1748-0221/3/08/S08001>.
- [81] ‘CERN webpage’, [Online]. Available: <http://www.cern.ch>.

- [82] ‘LHC Vacuum’, [Online]. Available: <http://home.web.cern.ch/about/engineering/vacuum-empty-interstellar-space>.
- [83] ‘LHC Superconducting Electromagnets’, [Online]. Available: <http://home.web.cern.ch/about/engineering/pulling-together-superconducting-electromagnets>.
- [84] J.-L. Caron, ‘Cross section of LHC dipole. Dipole LHC: coupe transversale.’, AC Collection. Legacy of AC. Pictures from 1992 to 2002., May 1998. [Online]. Available: <https://cds.cern.ch/record/841539>.
- [85] ‘LHC Radiofrequency cavities’, [Online]. Available: <http://home.web.cern.ch/about/engineering/radiofrequency-cavities>.
- [86] ‘The ALICE experiment’, [Online]. Available: <http://home.web.cern.ch/about/experiments/alice>.
- [87] K. Aamodt *et al.* [ALICE Collaboration], ‘The ALICE experiment at the CERN LHC’, *JINST*, vol. 3, S08002, 2008. doi: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [88] ‘The ATLAS Experiment’, [Online]. Available: <http://home.web.cern.ch/about/experiments/atlas>.
- [89] G. Aad *et al.* [ATLAS Collaboration], ‘The ATLAS Experiment at the CERN Large Hadron Collider’, *JINST*, vol. 3, S08003, 2008. doi: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [90] ‘The CMS Experiment’, [Online]. Available: <http://home.web.cern.ch/about/experiments/cms>.
- [91] S. Chatrchyan *et al.* [CMS Collaboration], ‘The CMS Experiment at the CERN LHC’, *JINST*, vol. 3, S08004, 2008. doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [92] ‘The LHCb Experiment’, [Online]. Available: <http://home.web.cern.ch/about/experiments/lhcb>.
- [93] A. A. Alves Jr. *et al.* [LHCb Collaboration], ‘The LHCb Detector at the LHC’, *JINST*, vol. 3, S08005, 2008. doi: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [94] ‘The LHCf experiment’, [Online]. Available: <http://home.web.cern.ch/about/experiments/lhcfc>.
- [95] O. Adriani *et al.* [LHCf Collaboration], ‘The LHCf detector at the CERN Large Hadron Collider’, *JINST*, vol. 3, S08006, 2008. doi: [10.1088/1748-0221/3/08/S08006](https://doi.org/10.1088/1748-0221/3/08/S08006).
- [96] ‘The TOTEM Experiment’, [Online]. Available: <http://home.web.cern.ch/about/experiments/totem>.
- [97] G. Anelli *et al.* [TOTEM Collaboration], ‘The TOTEM experiment at the CERN Large Hadron Collider’, *JINST*, vol. 3, S08007, 2008. doi: [10.1088/1748-0221/3/08/S08007](https://doi.org/10.1088/1748-0221/3/08/S08007).
- [98] ‘CMS Luminosity public plots’, [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [99] M. Della Negra *et al.* [CMS Collaboration], ‘CMS: The Compact Muon Solenoid: Letter of intent for a general purpose detector at the LHC’, Oct. 1992.
- [100] ‘CMS, the Compact Muon Solenoid: Technical proposal’, Dec. 1994. [Online]. Available: <http://cds.cern.ch/record/290969>.
- [101] S. Chatrchyan *et al.* [CMS Collaboration], ‘The CMS Experiment at the CERN LHC’, *JINST*, vol. 3, S08004, 2008. doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [102] V. Karimäki [CMS Collaboration], ‘The CMS tracker system project: Technical Design Report’, M. Mannelli, P. Siegrist, H. Breuker, A. Caner, R. Castaldi, K. Freudenreich, G. Hall, R. Horisberger, M. Huhtinen and A. Cattai, Eds., 1997.
- [103] ‘The CMS tracker: addendum to the Technical Design Report’, 2000. [Online]. Available: <https://cds.cern.ch/record/490194>.
- [104] ‘CMS Technical Design Report for the Pixel Detector Upgrade’, D. A. Matzner Dominguez *et al.*, Eds., Sep. 2012. doi: [10.2172/1151650](https://doi.org/10.2172/1151650).
- [105] ‘The CMS electromagnetic calorimeter project: Technical Design Report’, 1997. [Online]. Available: <https://cds.cern.ch/record/349375>.
- [106] S. Chatrchyan *et al.* [CMS Collaboration], ‘Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV’, *JINST*, vol. 8, P09009, 2013. doi: [10.1088/1748-0221/8/09/P09009](https://doi.org/10.1088/1748-0221/8/09/P09009). arXiv: [1306.2016 \[hep-ex\]](https://arxiv.org/abs/1306.2016).
- [107] ‘CMS: The hadron calorimeter technical design report’, Jun. 1997. [Online]. Available: <https://cds.cern.ch/record/357153>.
- [108] S. Abdullin *et al.* [USCMS, ECAL/HCAL Collaboration], ‘The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c’, *Eur. Phys. J. C*, vol. 60, pp. 359–373, 2009, [Erratum: *Eur.Phys.J.C* 61, 353–356 (2009)]. doi: [10.1140/epjc/s10052-009-0959-5](https://doi.org/10.1140/epjc/s10052-009-0959-5).
- [109] ‘The CMS muon project: Technical Design Report’, 1997. [Online]. Available: <https://cds.cern.ch/record/343814>.

- [110] M. Aguilar-Benitez *et al.*, ‘Construction and test of the final CMS Barrel Drift Tube Muon Chamber prototype’, *Nucl. Instrum. Meth.*, vol. A480, pp. 658–669, 2002. doi: [10.1016/S0168-9002\(01\)01227-X](https://doi.org/10.1016/S0168-9002(01)01227-X).
- [111] J. Hauser [CMS Collaboration], ‘Cathode strip chambers for the CMS endcap muon system’, *Nucl. Instrum. Meth.*, vol. A384, pp. 207–210, 1996. doi: [10.1016/S0168-9002\(96\)00905-9](https://doi.org/10.1016/S0168-9002(96)00905-9).
- [112] G. Wrochna, ‘The RPC system for the CMS experiment at LHC’, in *Resistive plate chambers and related detectors. Proceedings, 3rd International Workshop, Pavia, Italy, October 11-12, 1995*, 1995, pp. 63–77.
- [113] A. Tapper and D. Acosta [CMS Collaboration], ‘CMS Technical Design Report for the Level-1 Trigger Upgrade’, 2013. [Online]. Available: <https://cds.cern.ch/record/1556311>.
- [114] P. Sphicas [CMS Collaboration], ‘CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger’, 2002. [Online]. Available: <https://cds.cern.ch/record/578006>.
- [115] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Particle-flow reconstruction and global event description with the CMS detector’, *JINST*, vol. 12, no. 10, P10003, 2017. doi: [10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003). arXiv: [1706.04965 \[physics.ins-det\]](https://arxiv.org/abs/1706.04965).
- [116] A. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, N. Y. Springer, Ed., pp. 335–365. doi: [10.1007/978-1-4613-8716-9_14](https://doi.org/10.1007/978-1-4613-8716-9_14).
- [117] T. Diotalevi, D. Bonacorsi, C. Battilana and L. Guiducci, ‘Development of Machine Learning based muon trigger algorithms for the Phase2 upgrade of the CMS detector’, *PoS*, vol. LHCP2018, p. 092, 2018. doi: [10.22323/1.321.0092](https://doi.org/10.22323/1.321.0092).
- [118] ‘Machine learning tutorial’, [Online]. Available: <https://class.coursera.org/ml-005/lecture>.
- [119] S. Kaplan, ‘Deep generative models for synthetic retinal image generation’, Jul. 2017. [Online]. Available: https://www.researchgate.net/publication/319093376_DEEP_GENERATIVE_MODELS_FOR_SYNTHETIC_RETINAL_IMAGE_GENERATION.
- [120] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018. [Online]. Available: <https://mitpress.mit.edu/books/reinforcement-learning-second-edition>.
- [121] ‘Machine learning crash course - google developers’, [Online]. Available: <https://developers.google.com/machine-learning/crash-course>.
- [122] A. Artemov, ‘Intro into machine learning’, [Online]. Available: <https://indico.cern.ch/event/687473/sessions/259653/>.
- [123] ‘Overfitting in machine learning: What it is and how to prevent it’, [Online]. Available: <https://elitedatascience.com/overfitting-in-machine-learning>.
- [124] C. Adam-Boudarios, G. Cowan, C. Germain-Renaud, I. Guyon, B. Kégl and D. Rousseau, ‘The Higgs Machine Learning Challenge’, *J. Phys. Conf. Ser.*, vol. 664, no. 7, p. 072015, 2015. doi: [10.1088/1742-6596/664/7/072015](https://doi.org/10.1088/1742-6596/664/7/072015).
- [125] N. Kazeev, ‘Machine Learning for particle identification in the LHCb detector.’, Presented 21 Oct 2020, 2020. [Online]. Available: <https://cds.cern.ch/record/2744601>.
- [126] B. Schölkopf, J. Platt and T. Hofmann, *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference* (Bradford Books). The MIT Press, 2007, ISBN: 0262195682.
- [127] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006, ISBN: 0387310738.
- [128] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [129] T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009. [Online]. Available: <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>.
- [130] *Statistical Decision Series*, ser. Harvard Business Review reprint series pt. 4. HBR, 1959. [Online]. Available: <https://books.google.ru/books?id=8WwPAQAAQAAJ>.
- [131] L. Cam and J. Neyman, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Held at the Statistical Laboratory, University of California, June 21-July 18, 1965 and December 27, 1965-January 7, 1966*, ser. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability v. 1. University of California Press, 1967. [Online]. Available: <https://books.google.ru/books?id=QsBQCBgTx8gc>.
- [132] A. Moodley, ‘Language identification with decision trees: Identification of individual words in the south african languages’, Ph.D. dissertation, Jan. 2016. doi: [10.13140/RG.2.2.25539.81445](https://doi.org/10.13140/RG.2.2.25539.81445).

- [133] J. H. Friedman, 'Greedy function approximation: A gradient boosting machine.', *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. doi: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451). [Online]. Available: <https://doi.org/10.1214/aos/1013203451>.
- [134] T. Chen and C. Guestrin, 'Xgboost', *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016. doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785). [Online]. Available: <http://dx.doi.org/10.1145/2939672.2939785>.
- [135] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y. Liu, 'Lightgbm: A highly efficient gradient boosting decision tree', in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>.
- [136] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush and A. Gulin, 'Catboost: Unbiased boosting with categorical features', in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, Eds., vol. 31, Curran Associates, Inc., 2018. [Online]. Available: <https://proceedings.neurips.cc/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf>.
- [137] S. Samarasinghe, *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*. CRC Press, 2016, ISBN: 9781420013061. [Online]. Available: <https://books.google.it/books?id=EyFeUiJibooC>.
- [138] A. L. Maas, 'Rectifier nonlinearities improve neural network acoustic models', 2013. [Online]. Available: https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.
- [139] P. Ramachandran, B. Zoph and Q. V. Le, *Searching for activation functions*, 2017. arXiv: [1710.05941 \[cs.NE\]](https://arxiv.org/abs/1710.05941).
- [140] J. Kiefer and J. Wolfowitz, 'Stochastic Estimation of the Maximum of a Regression Function', *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 462–466, 1952. doi: [10.1214/aoms/1177729392](https://doi.org/10.1214/aoms/1177729392). [Online]. Available: <https://doi.org/10.1214/aoms/1177729392>.
- [141] 'Google brain team', [Online]. Available: <https://research.google.com/teams/brain>.
- [142] 'Tensorflow documentation', [Online]. Available: <https://www.tensorflow.org>.
- [143] 'Scikit-learn documentation', [Online]. Available: <http://scikit-learn.org>.
- [144] 'Keras documentation', [Online]. Available: <https://keras.io>.
- [145] D. Guest, K. Cranmer and D. Whiteson, 'Deep Learning and its Application to LHC Physics', *Ann. Rev. Nucl. Part. Sci.*, vol. 68, pp. 161–181, 2018. doi: [10.1146/annurev-nucl-101917-021019](https://doi.org/10.1146/annurev-nucl-101917-021019). arXiv: [1806.11484 \[hep-ex\]](https://arxiv.org/abs/1806.11484).
- [146] D. Bourilkov, 'Machine and Deep Learning Applications in Particle Physics', *Int. J. Mod. Phys. A*, vol. 34, no. 35, p. 1930019, 2020. doi: [10.1142/S0217751X19300199](https://doi.org/10.1142/S0217751X19300199). arXiv: [1912.08245 \[physics.data-an\]](https://arxiv.org/abs/1912.08245).
- [147] P. Abreu *et al.* [DELPHI Collaboration], 'Classification of the hadronic decays of the Z0 into b and c quark pairs using a neural network', *Phys. Lett. B*, vol. 295, pp. 383–395, 1992. doi: [10.1016/0370-2693\(92\)91580-3](https://doi.org/10.1016/0370-2693(92)91580-3).
- [148] J. K. Kohne *et al.*, 'Realization of a second level neural network trigger for the H1 experiment at HERA', *Nucl. Instrum. Meth. A*, vol. 389, M. Werlen and D. Perret-Gallix, Eds., pp. 128–133, 1997. doi: [10.1016/S0168-9002\(97\)00062-4](https://doi.org/10.1016/S0168-9002(97)00062-4).
- [149] V. M. Abazov *et al.* [D0 Collaboration], 'Search for Single Top Quark Production at D0 Using Neural Networks', *Phys. Lett. B*, vol. 517, pp. 282–294, 2001. doi: [10.1016/S0370-2693\(01\)01009-7](https://doi.org/10.1016/S0370-2693(01)01009-7). arXiv: [hep-ex/0106059](https://arxiv.org/abs/hep-ex/0106059).
- [150] T. Aaltonen *et al.* [CDF Collaboration], 'A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and b-quark Jets in p anti-p Collisions at s***(1/2) = 1.96-TeV', *Phys. Rev. Lett.*, vol. 104, p. 141801, 2010. doi: [10.1103/PhysRevLett.104.141801](https://doi.org/10.1103/PhysRevLett.104.141801). arXiv: [0911.3935 \[hep-ex\]](https://arxiv.org/abs/0911.3935).
- [151] V. Khachatryan *et al.* [CMS Collaboration], 'Observation of the Diphoton Decay of the Higgs Boson and Measurement of Its Properties', *Eur. Phys. J. C*, vol. 74, no. 10, p. 3076, 2014. doi: [10.1140/epjc/s10052-014-3076-z](https://doi.org/10.1140/epjc/s10052-014-3076-z). arXiv: [1407.0558 \[hep-ex\]](https://arxiv.org/abs/1407.0558).
- [152] P. Baldi, P. Sadowski and D. Whiteson, 'Searching for Exotic Particles in High-Energy Physics with Deep Learning', *Nature Commun.*, vol. 5, p. 4308, 2014. doi: [10.1038/ncomms5308](https://doi.org/10.1038/ncomms5308). arXiv: [1402.4735 \[hep-ph\]](https://arxiv.org/abs/1402.4735).
- [153] M. Aaboud *et al.* [ATLAS Collaboration], 'Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2', *Eur. Phys. J. C*, vol. 77, no. 10, p. 673, 2017. doi: [10.1140/epjc/s10052-017-5225-7](https://doi.org/10.1140/epjc/s10052-017-5225-7). arXiv: [1704.07983 \[hep-ex\]](https://arxiv.org/abs/1704.07983).

- [154] A. Salzburger, ‘Optimisation of the ATLAS Track Reconstruction Software for Run-2’, CERN, Geneva, Tech. Rep., May 2015. doi: [10.1088/1742-6596/664/7/072042](https://doi.org/10.1088/1742-6596/664/7/072042). [Online]. Available: <https://cds.cern.ch/record/2018442>.
- [155] J. Machado Miguens [ATLAS Collaboration], ‘The ATLAS Run-2 Trigger: Design, Menu, Performance and Operational Aspects’, *PoS*, vol. ICHEP2016, p. 244, 2016. doi: [10.22323/1.282.0244](https://doi.org/10.22323/1.282.0244).
- [156] J. Cogan, M. Kagan, E. Strauss and A. Schwarzman, ‘Jet-images: Computer vision inspired techniques for jet tagging’, *Journal of High Energy Physics*, vol. 2015, no. 2, Feb. 2015, ISSN: 1029-8479. doi: [10.1007/jhep02\(2015\)118](https://doi.org/10.1007/jhep02(2015)118). [Online]. Available: [http://dx.doi.org/10.1007/JHEP02\(2015\)118](http://dx.doi.org/10.1007/JHEP02(2015)118).
- [157] ‘Performance of heavy flavour identification algorithms in proton-proton collisions at 13 TeV at the CMS experiment’, May 2017. [Online]. Available: <https://cds.cern.ch/record/2263801>.
- [158] J. Duarte *et al.*, ‘Fast inference of deep neural networks in fpgas for particle physics’, *Journal of Instrumentation*, vol. 13, no. 07, Jul. 2018, ISSN: 1748-0221. doi: [10.1088/1748-0221/13/07/p07027](https://doi.org/10.1088/1748-0221/13/07/p07027). [Online]. Available: <http://dx.doi.org/10.1088/1748-0221/13/07/P07027>.
- [159] T. Diotalevi, M. Lorusso, R. Travaglini, C. Battilana and D. Bonacorsi [CMS Collaboration], ‘Deep Learning fast inference on FPGA for CMS Muon Level-1 Trigger studies’, *PoS*, vol. ISGC2021, p. 005, 2021. doi: [10.22323/1.378.0005](https://doi.org/10.22323/1.378.0005).
- [160] G. Aad *et al.*, ‘Artificial Neural Networks on FPGAs for Real-Time Energy Reconstruction of the ATLAS LAr Calorimeters’, *Comput. Softw. Big Sci.*, vol. 5, no. 1, p. 19, 2021. doi: [10.1007/s41781-021-00066-y](https://doi.org/10.1007/s41781-021-00066-y).
- [161] ‘High-pt muon refit with ml’, [Online]. Available: <https://indico.cern.ch/event/973558/>.
- [162] M. Paganini, L. de Oliveira and B. Nachman, ‘Accelerating science with generative adversarial networks: An application to 3d particle showers in multilayer calorimeters’, *Physical Review Letters*, vol. 120, no. 4, Jan. 2018, ISSN: 1079-7114. doi: [10.1103/physrevlett.120.042003](https://doi.org/10.1103/physrevlett.120.042003). [Online]. Available: <http://dx.doi.org/10.1103/PhysRevLett.120.042003>.
- [163] L. de Oliveira, M. Paganini and B. Nachman, ‘Learning particle physics by example: Location-aware generative adversarial networks for physics synthesis’, *Computing and Software for Big Science*, vol. 1, no. 1, Sep. 2017, ISSN: 2510-2044. doi: [10.1007/s41781-017-0004-6](https://doi.org/10.1007/s41781-017-0004-6). [Online]. Available: <http://dx.doi.org/10.1007/s41781-017-0004-6>.
- [164] T. Diotalevi, A. Falabella, B. Martelli, D. Michelotto, L. Morganti, D. Bonacorsi, L. Giommi and S. Rossi Tisbeni, ‘Collection and harmonization of system logs and prototypal Analytics services with the Elastic (ELK) suite at the INFN-CNAF computing centre’, *PoS*, vol. ISGC2019, p. 027, 2019. doi: [10.22323/1.351.0027](https://doi.org/10.22323/1.351.0027). arXiv: [2106.02612 \[cs.DC\]](https://arxiv.org/abs/2106.02612).
- [165] L. Giommi, D. Bonacorsi, T. Diotalevi, L. Rinaldi, L. Morganti, A. Falabella, E. Ronchieri, A. Ceccanti, B. Martelli and S. Tisbeni, ‘Towards Predictive Maintenance with Machine Learning at the INFN-CNAF computing centre’, *PoS*, vol. ISGC2019, p. 003, 2019. doi: [10.22323/1.351.0003](https://doi.org/10.22323/1.351.0003).
- [166] ‘Machine learning applications for data certification’, [Online]. Available: <https://agenda.infn.it/event/19755/contributions/101002/>.
- [167] P. Bechtle, S. Heinemeyer, O. Stål, T. Stefaniak and G. Weiglein, ‘Probing the Standard Model with Higgs signal rates from the Tevatron, the LHC and a future ILC’, *JHEP*, vol. 11, p. 039, 2014. doi: [10.1007/JHEP11\(2014\)039](https://doi.org/10.1007/JHEP11(2014)039). arXiv: [1403.1582 \[hep-ph\]](https://arxiv.org/abs/1403.1582).
- [168] M. Carena, H. E. Haber, I. Low, N. R. Shah and C. E. M. Wagner, ‘Complementarity between Non-standard Higgs Boson Searches and Precision Higgs Boson Measurements in the MSSM’, *Phys. Rev. D*, vol. 91, no. 3, p. 035 003, 2015. doi: [10.1103/PhysRevD.91.035003](https://doi.org/10.1103/PhysRevD.91.035003). arXiv: [1410.4969 \[hep-ph\]](https://arxiv.org/abs/1410.4969).
- [169] L. Maiani, A. D. Polosa and V. Riquer, ‘Bounds to the Higgs Sector Masses in Minimal Supersymmetry from LHC Data’, *Phys. Lett. B*, vol. 724, pp. 274–277, 2013. doi: [10.1016/j.physletb.2013.06.026](https://doi.org/10.1016/j.physletb.2013.06.026). arXiv: [1305.2172 \[hep-ph\]](https://arxiv.org/abs/1305.2172).
- [170] A. Djouadi, L. Maiani, A. Polosa, J. Quevillon and V. Riquer, ‘Fully covering the MSSM Higgs sector at the LHC’, *JHEP*, vol. 06, p. 168, 2015. doi: [10.1007/JHEP06\(2015\)168](https://doi.org/10.1007/JHEP06(2015)168). arXiv: [1502.05653 \[hep-ph\]](https://arxiv.org/abs/1502.05653).
- [171] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for MSSM Higgs bosons decaying to $\mu^+ \mu^-$ in proton-proton collisions at $s=13\text{TeV}$ ’, *Phys. Lett. B*, vol. 798, p. 134 992, 2019. doi: [10.1016/j.physletb.2019.134992](https://doi.org/10.1016/j.physletb.2019.134992). arXiv: [1907.03152 \[hep-ex\]](https://arxiv.org/abs/1907.03152).
- [172] V. Khachatryan *et al.* [CMS Collaboration], ‘Search for neutral MSSM Higgs bosons decaying to $\mu^+ \mu^-$ in pp collisions at $\sqrt{s} = 7$ and 8 TeV’, *Phys. Lett. B*, vol. 752, pp. 221–246, 2016. doi: [10.1016/j.physletb.2015.11.042](https://doi.org/10.1016/j.physletb.2015.11.042). arXiv: [1508.01437 \[hep-ex\]](https://arxiv.org/abs/1508.01437).
- [173] V. Khachatryan *et al.* [CMS Collaboration], ‘Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions’, *JHEP*, vol. 10, p. 160, 2014. doi: [10.1007/JHEP10\(2014\)160](https://doi.org/10.1007/JHEP10(2014)160). arXiv: [1408.3316 \[hep-ex\]](https://arxiv.org/abs/1408.3316).

- [174] T. A. Aaltonen *et al.* [CDF Collaboration], ‘Search for s -Channel Single-Top-Quark Production in Events with Missing Energy Plus Jets in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV’, *Phys. Rev. Lett.*, vol. 112, no. 23, p. 231805, 2014. doi: [10.1103/PhysRevLett.112.231805](https://doi.org/10.1103/PhysRevLett.112.231805). arXiv: [1402.3756 \[hep-ex\]](https://arxiv.org/abs/1402.3756).
- [175] ‘UL2016 Datasets for Analysis’, [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMS/PdmVDatasetsUL2016>.
- [176] M. Peruzzi, G. Petrucciani and A. Rizzi [CMS Collaboration], ‘The NanoAOD event data format in CMS’, *J. Phys. Conf. Ser.*, vol. 1525, no. 1, p. 012038, 2020. doi: [10.1088/1742-6596/1525/1/012038](https://doi.org/10.1088/1742-6596/1525/1/012038).
- [177] P. Nason, ‘A New method for combining NLO QCD with shower Monte Carlo algorithms’, *JHEP*, vol. 11, p. 040, 2004. doi: [10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040). arXiv: [hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [178] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, ‘MadGraph 5 : Going Beyond’, *JHEP*, vol. 06, p. 128, 2011. doi: [10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128). arXiv: [1106.0522 \[hep-ph\]](https://arxiv.org/abs/1106.0522).
- [179] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. ... Shao, T. Stelzer, P. Torrielli and M. Zaro, ‘The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations’, *JHEP*, vol. 07, p. 079, 2014. doi: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079). arXiv: [1405.0301 \[hep-ph\]](https://arxiv.org/abs/1405.0301).
- [180] T. Sjostrand, S. Mrenna and P. Z. Skands, ‘A Brief Introduction to PYTHIA 8.1’, *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008. doi: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036). arXiv: [0710.3820 \[hep-ph\]](https://arxiv.org/abs/0710.3820).
- [181] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements’, *Eur. Phys. J. C*, vol. 80, no. 1, p. 4, 2020. doi: [10.1140/epjc/s10052-019-7499-4](https://doi.org/10.1140/epjc/s10052-019-7499-4). arXiv: [1903.12179 \[hep-ex\]](https://arxiv.org/abs/1903.12179).
- [182] E. Bagnaschi, G. Degrassi, P. Slavich and A. Vicini, ‘Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM’, *JHEP*, vol. 02, p. 088, 2012. doi: [10.1007/JHEP02\(2012\)088](https://doi.org/10.1007/JHEP02(2012)088). arXiv: [1111.2854 \[hep-ph\]](https://arxiv.org/abs/1111.2854).
- [183] S. Dittmaier, M. Krämer and M. Spira, ‘Higgs radiation off bottom quarks at the Tevatron and the CERN LHC’, *Phys. Rev. D*, vol. 70, p. 074010, 2004. doi: [10.1103/PhysRevD.70.074010](https://doi.org/10.1103/PhysRevD.70.074010). arXiv: [hep-ph/0309204](https://arxiv.org/abs/hep-ph/0309204).
- [184] S. Dawson, C. B. Jackson, L. Reina and D. Wackerlo, ‘Exclusive Higgs boson production with bottom quarks at hadron colliders’, *Phys. Rev. D*, vol. 69, p. 074027, 2004. doi: [10.1103/PhysRevD.69.074027](https://doi.org/10.1103/PhysRevD.69.074027). arXiv: [hep-ph/0311067](https://arxiv.org/abs/hep-ph/0311067).
- [185] B. Jager, L. Reina and D. Wackerlo, ‘Higgs boson production in association with b jets in the POWHEG BOX’, *Phys. Rev. D*, vol. 93, no. 1, p. 014030, 2016. doi: [10.1103/PhysRevD.93.014030](https://doi.org/10.1103/PhysRevD.93.014030). arXiv: [1509.05843 \[hep-ph\]](https://arxiv.org/abs/1509.05843).
- [186] T. Hahn, S. Heinemeyer, W. Hollik, H. Rzehak and G. Weiglein, ‘FeynHiggs: A program for the calculation of MSSM Higgs-boson observables - Version 2.6.5’, *Comput. Phys. Commun.*, vol. 180, pp. 1426–1427, 2009. doi: [10.1016/j.cpc.2009.02.014](https://doi.org/10.1016/j.cpc.2009.02.014).
- [187] J. R. Andersen *et al.* [LHC Higgs Cross Section Working Group Collaboration], ‘Handbook of LHC Higgs Cross Sections: 3. Higgs Properties’, S. Heinemeyer, C. Mariotti, G. Passarino and R. Tanaka, Eds., Jul. 2013. doi: [10.5170/CERN-2013-004](https://doi.org/10.5170/CERN-2013-004). arXiv: [1307.1347 \[hep-ph\]](https://arxiv.org/abs/1307.1347).
- [188] R. V. Harlander and W. B. Kilgore, ‘Higgs boson production in bottom quark fusion at next-to-next-to leading order’, *Phys. Rev. D*, vol. 68, p. 013001, 2003. doi: [10.1103/PhysRevD.68.013001](https://doi.org/10.1103/PhysRevD.68.013001). arXiv: [hep-ph/0304035](https://arxiv.org/abs/hep-ph/0304035).
- [189] ‘POWHEG cards for the MSSM Higgs to $\mu^+\mu^-$ analysis’, [Online]. Available: <https://github.com/cms-sw/genproductions/tree/master/bin/Powheg/production/2017/13TeV/Higgs/MSSM>.
- [190] ‘Monte Carlo central production for the 2016 signal samples’, [Online]. Available: https://cms-pdmv.cern.ch/mcm/requests?page=0&dataset_name=MSSMGluGluTo*ToMuMu_MA*&member_of_campaign=RunIISummer20UL16wmLHEGEN.
- [191] ‘Monte Carlo central production for the 2016APV signal samples’, [Online]. Available: https://cms-pdmv.cern.ch/mcm/requests?page=0&dataset_name=MSSMGluGluTo*ToMuMu_MA*&member_of_campaign=RunIISummer20UL16wmLHEGENAPV.
- [192] ‘Monte Carlo central production for the 2017 signal samples’, [Online]. Available: https://cms-pdmv.cern.ch/mcm/requests?page=0&dataset_name=MSSMGluGluTo*ToMuMu_MA*&member_of_campaign=RunIISummer20UL17wmLHEGEN.
- [193] ‘Monte Carlo central production for the 2018 signal samples’, [Online]. Available: https://cms-pdmv.cern.ch/mcm/requests?page=0&dataset_name=MSSMGluGluTo*ToMuMu_MA*&member_of_campaign=RunIISummer20UL18wmLHEGEN.

- [194] ‘Muon HLT paths’, [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonHLT>.
- [195] W. Adam, B. Mangano, T. Speer and T. Todorov, ‘Track reconstruction in the CMS tracker’, Dec. 2005.
- [196] S. Chatrchyan *et al.* [CMS Collaboration], ‘Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV’, *JINST*, vol. 7, P10002, 2012. doi: [10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002). arXiv: [1206.4071 \[physics.ins-det\]](https://arxiv.org/abs/1206.4071).
- [197] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV’, *JINST*, vol. 13, no. 06, P06015, 2018. doi: [10.1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015). arXiv: [1804.04528 \[physics.ins-det\]](https://arxiv.org/abs/1804.04528).
- [198] M. Cacciari, G. P. Salam and G. Soyez, ‘The anti- k_t jet clustering algorithm’, *JHEP*, vol. 04, p. 063, 2008. doi: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv: [0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189).
- [199] M. Cacciari and G. P. Salam, ‘Dispelling the N^3 myth for the k_t jet-finder’, *Phys. Lett. B*, vol. 641, pp. 57–61, 2006. doi: [10.1016/j.physletb.2006.08.037](https://doi.org/10.1016/j.physletb.2006.08.037). arXiv: [hep-ph/0512210](https://arxiv.org/abs/hep-ph/0512210).
- [200] M. Cacciari, G. P. Salam and G. Soyez, ‘FastJet User Manual’, *Eur. Phys. J. C*, vol. 72, p. 1896, 2012. doi: [10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2). arXiv: [1111.6097 \[hep-ph\]](https://arxiv.org/abs/1111.6097).
- [201] S. D. Ellis and D. E. Soper, ‘Successive combination jet algorithm for hadron collisions’, *Phys. Rev. D*, vol. 48, pp. 3160–3166, 1993. doi: [10.1103/PhysRevD.48.3160](https://doi.org/10.1103/PhysRevD.48.3160). arXiv: [hep-ph/9305266](https://arxiv.org/abs/hep-ph/9305266).
- [202] M. Wobisch and T. Wengler, ‘Hadronization corrections to jet cross-sections in deep inelastic scattering’, in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, Apr. 1998, pp. 270–279. arXiv: [hep-ph/9907280](https://arxiv.org/abs/hep-ph/9907280).
- [203] S. Chatrchyan *et al.* [CMS Collaboration], ‘Identification of b-Quark Jets with the CMS Experiment’, *JINST*, vol. 8, P04013, 2013. doi: [10.1088/1748-0221/8/04/P04013](https://doi.org/10.1088/1748-0221/8/04/P04013). arXiv: [1211.4462 \[hep-ex\]](https://arxiv.org/abs/1211.4462).
- [204] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV’, *JINST*, vol. 13, no. 05, P05011, 2018. doi: [10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011). arXiv: [1712.07158 \[physics.ins-det\]](https://arxiv.org/abs/1712.07158).
- [205] E. Bols, J. Kieseler, M. Verzetti, M. Stoye and A. Stakia, ‘Jet Flavour Classification Using DeepJet’, *JINST*, vol. 15, no. 12, P12012, 2020. doi: [10.1088/1748-0221/15/12/P12012](https://doi.org/10.1088/1748-0221/15/12/P12012). arXiv: [2008.10519 \[hep-ex\]](https://arxiv.org/abs/2008.10519).
- [206] A. J. Larkoski, I. Moult and B. Nachman, ‘Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning’, *Phys. Rept.*, vol. 841, pp. 1–63, 2020. doi: [10.1016/j.physrep.2019.11.001](https://doi.org/10.1016/j.physrep.2019.11.001). arXiv: [1709.04464 \[hep-ph\]](https://arxiv.org/abs/1709.04464).
- [207] S. Hochreiter and J. Schmidhuber, ‘Long Short-Term Memory’, *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [208] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: [1412.6980 \[cs.LG\]](https://arxiv.org/abs/1412.6980).
- [209] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto and A. Strelnikov, ‘Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments’, *Eur. Phys. J. C*, vol. 72, p. 2194, 2012. doi: [10.1140/epjc/s10052-012-2194-8](https://doi.org/10.1140/epjc/s10052-012-2194-8). arXiv: [1208.3710 \[hep-ex\]](https://arxiv.org/abs/1208.3710).
- [210] ‘Reference muon id, isolation and trigger efficiencies for Run II’, [Online]. Available: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffsRun2>.
- [211] ‘Usage of b/c Tag Objects for 13 TeV Data’, [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco>.
- [212] ‘CMS NanoAOD Tools GitHub repository’, [Online]. Available: <https://github.com/cms-nanoAOD/nanoAOD-tools>.
- [213] ‘NanoAOD Tools GitHub repository’, [Online]. Available: <https://github.com/cms-nanoAOD/correctionlib>.
- [214] B. H. Denby, ‘Neural Networks and Cellular Automata in Experimental High-energy Physics’, *Comput. Phys. Commun.*, vol. 49, pp. 429–448, 1988. doi: [10.1016/0010-4655\(88\)90004-5](https://doi.org/10.1016/0010-4655(88)90004-5).
- [215] C. Peterson, T. Rognvaldsson and L. Lonnblad, ‘JETNET 3.0: A Versatile artificial neural network package’, *Comput. Phys. Commun.*, vol. 81, pp. 185–220, 1994. doi: [10.1016/0010-4655\(94\)90120-1](https://doi.org/10.1016/0010-4655(94)90120-1).
- [216] H. Kolanoski, ‘Application of artificial neural networks in particle physics’, *Nucl. Instrum. Meth. A*, vol. 367, W. Bartl, M. Krammer, G. Neuhofer, M. Regler and A. Taurok, Eds., pp. 14–20, 1995. doi: [10.1016/0168-9002\(95\)00743-1](https://doi.org/10.1016/0168-9002(95)00743-1).
- [217] C. Peterson, ‘Track Finding With Neural Networks’, *Nucl. Instrum. Meth. A*, vol. 279, p. 537, 1989. doi: [10.1016/0168-9002\(89\)91300-4](https://doi.org/10.1016/0168-9002(89)91300-4).

- [218] G. Aad *et al.* [ATLAS Collaboration], ‘A Neural Network clustering algorithm for the ATLAS silicon pixel detector’, *Journal of Instrumentation*, vol. 9, no. 09, Sep. 2014, ISSN: 1748-0221. doi: [10.1088/1748-0221/9/09/p09009](https://doi.org/10.1088/1748-0221/9/09/p09009). [Online]. Available: <http://dx.doi.org/10.1088/1748-0221/9/09/P09009>.
- [219] L. Lonnblad, C. Peterson and T. Rognvaldsson, ‘Finding Gluon Jets With a Neural Trigger’, *Phys. Rev. Lett.*, vol. 65, pp. 1321–1324, 1990. doi: [10.1103/PhysRevLett.65.1321](https://doi.org/10.1103/PhysRevLett.65.1321).
- [220] B. H. Denby, M. Campbell, F. Bedeschi, N. Chriss, C. Bowers and F. Nesti, ‘Neural Networks for Triggering’, *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 248–254, 1990. doi: [10.1109/23.106627](https://doi.org/10.1109/23.106627).
- [221] T. Aaltonen *et al.* [CDF, D0 Collaboration], ‘Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron’, *Phys. Rev. Lett.*, vol. 109, p. 071804, 2012. doi: [10.1103/PhysRevLett.109.071804](https://doi.org/10.1103/PhysRevLett.109.071804). arXiv: [1207.6436](https://arxiv.org/abs/1207.6436) [hep-ex].
- [222] S. Chatrchyan *et al.* [CMS Collaboration], ‘Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV’, *Phys. Lett. B*, vol. 710, pp. 26–48, 2012. doi: [10.1016/j.physletb.2012.02.064](https://doi.org/10.1016/j.physletb.2012.02.064). arXiv: [1202.1488](https://arxiv.org/abs/1202.1488) [hep-ex].
- [223] G. Aad *et al.* [ATLAS Collaboration], ‘Search for $W' \rightarrow t\bar{b}$ in the lepton plus jets final state in proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV with the ATLAS detector’, *Phys. Lett. B*, vol. 743, pp. 235–255, 2015. doi: [10.1016/j.physletb.2015.02.051](https://doi.org/10.1016/j.physletb.2015.02.051). arXiv: [1410.4103](https://arxiv.org/abs/1410.4103) [hep-ex].
- [224] P. Baldi, K. Cranmer, T. Fauchet, P. Sadowski and D. Whiteson, ‘Parameterized neural networks for high-energy physics’, *Eur. Phys. J. C*, vol. 76, no. 5, p. 235, 2016. doi: [10.1140/epjc/s10052-016-4099-4](https://doi.org/10.1140/epjc/s10052-016-4099-4). arXiv: [1601.07913](https://arxiv.org/abs/1601.07913) [hep-ex].
- [225] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell^+\ell^-$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV’, *JHEP*, vol. 01, p. 054, 2018. doi: [10.1007/JHEP01\(2018\)054](https://doi.org/10.1007/JHEP01(2018)054). arXiv: [1708.04188](https://arxiv.org/abs/1708.04188) [hep-ex].
- [226] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Search for a charged Higgs boson decaying into top and bottom quarks in events with electrons or muons in proton-proton collisions at $\sqrt{s} = 13$ TeV’, *JHEP*, vol. 01, p. 096, 2020. doi: [10.1007/JHEP01\(2020\)096](https://doi.org/10.1007/JHEP01(2020)096). arXiv: [1908.09206](https://arxiv.org/abs/1908.09206) [hep-ex].
- [227] A. R. Bohm and Y. Sato, ‘Relativistic resonances: Their masses, widths, lifetimes, superposition, and causal evolution’, *Physical Review D*, vol. 71, no. 8, Apr. 2005, ISSN: 1550-2368. doi: [10.1103/physrevd.71.085018](https://doi.org/10.1103/physrevd.71.085018). [Online]. Available: <http://dx.doi.org/10.1103/PhysRevD.71.085018>.
- [228] J. Pivarski, P. Elmer and D. Lange, ‘Awkward arrays in python, c++, and numba’, *EPJ Web of Conferences*, vol. 245, C. Doglioni, D. Kim, G. Stewart, L. Silvestris, P. Jackson and W. Kamleh, Eds., p. 05023, 2020, ISSN: 2100-014X. doi: [10.1051/epjconf/202024505023](https://doi.org/10.1051/epjconf/202024505023). [Online]. Available: <http://dx.doi.org/10.1051/epjconf/202024505023>.
- [229] C. R. Harris *et al.*, ‘Array programming with NumPy’, *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). [Online]. Available: <https://doi.org/10.1038/s41586-020-2649-2>.
- [230] F. Pedregosa *et al.*, ‘Scikit-learn: Machine learning in Python’, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [231] A. M. Sirunyan *et al.* [CMS Collaboration], ‘Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV’, *JHEP*, vol. 07, p. 161, 2018. doi: [10.1007/JHEP07\(2018\)161](https://doi.org/10.1007/JHEP07(2018)161). arXiv: [1802.02613](https://arxiv.org/abs/1802.02613) [hep-ex].
- [232] M. Aaboud *et al.* [ATLAS Collaboration], ‘Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC’, *Phys. Rev. Lett.*, vol. 117, no. 18, p. 182002, 2016. doi: [10.1103/PhysRevLett.117.182002](https://doi.org/10.1103/PhysRevLett.117.182002). arXiv: [1606.02625](https://arxiv.org/abs/1606.02625) [hep-ex].
- [233] S. Chatrchyan *et al.* [CMS Collaboration], ‘Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS’, *JINST*, vol. 6, P11002, 2011. doi: [10.1088/1748-0221/6/11/P11002](https://doi.org/10.1088/1748-0221/6/11/P11002). arXiv: [1107.4277](https://arxiv.org/abs/1107.4277) [physics.ins-det].
- [234] ‘CMS Luminosity Measurements for the 2016 Data Taking Period’, 2017. [Online]. Available: <https://cds.cern.ch/record/2257069>.
- [235] G. Cowan, K. Cranmer, E. Gross and O. Vitells, ‘Asymptotic formulae for likelihood-based tests of new physics’, *Eur. Phys. J. C*, vol. 71, p. 1554, 2011, [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. doi: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). arXiv: [1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an].
- [236] ‘The HL-LHC project’, [Online]. Available: <http://hilumilhc.web.cern.ch>.

