A deep learning based search for a heavy CP-even Higgs boson in dileptonic $H \rightarrow WW$ decays with the CMS experiment

Vorgelegt der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Master of Science

von

B.Sc.

Peter Fackeldey

Advisor: Univ.-Prof. Dr. rer. nat. Achim Stahl
Second advisor: Apl.-Prof. Dr. Wolfgang Lohmann

September 26, 2018

Contents

Figures

Tables

Chapter 1

Introduction

This thesis presents a search for additional Higgs bosons in WW final states using modern deep learning techniques. The analysis is based on events recorded by the CMS experiment at the Large Hadron Collider (LHC) with a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ in the year 2016. The recorded data of pp collisions corresponds to an integrated luminosity of $35.9\,\text{fb}^{-1}$.
With the discovery of the Higgs boson in 2012 particle physicists could solve a very fundamental problem of the standard model (SM) of particle physics, namely how fundamental particles acquire mass. But there are more questions, that cannot be answered by the SM, such as dark matter, the disappearance of anti-matter in the universe or the hierarchy problem. These may be explained with beyond the SM (BSM) theories. One of these theories is the minimal supersymmetric extension of the SM (MSSM), which is a constrained theory of the more general supersymmetry (SUSY). Especially the Higgs sector of BSM theories is of great interest, since multiple Higgs bosons are predicted. Furthermore the precision in the measurement of couplings of the observed Higgs boson is not yet sufficient to determine, whether it is the SM Higgs boson or one from BSM theories. This analysis focusses on the search for an additional $\mathcal{CP}$-even Higgs boson, which decays in two W bosons. In the MSSM the most sensitive decay channel is $\text{H} \to \tau\tau$, due to an enhanced coupling to down-type fermions. Thus this channel can exclude the largest region in the parameter space of the MSSM, spanned up by the two parameters m_A and $\tan\beta$. Nevertheless there is a non-negligible unexcluded region in this parameter space at low values of m_A and $\tan\beta$, which can currently not be covered by the $\text{H} \to \tau\tau$ analysis. For this region the most sensitive Higgs decay channel is the one in two W bosons.
Since W bosons are unstable, they decay further leptonically or hadronically. This thesis analyses the dileptonic final state, where one W boson decays into an electron and a neutrino and the other W boson into a muon and a neutrino. The decay channel of the Higgs boson to the two leptons and two neutrinos via two W bosons is very sensitive to low values of m_A and $\tan\beta$. In this region a lot of SM backgrounds are still present. Therefore a dedicated multi-class classification algorithm using deep learning is developed, exploiting inner representations and correlations of the data, in order to suppress and control these backgrounds.
The final statistical analysis is performed in a model-independent search for a heavy $\mathcal{CP}$-even Higgs scalar in addition to the SM Higgs boson, a more general two Higgs doublet model (2HDM) of type 2 and two MSSM scenarios. In the region, where the aforementioned SM backgrounds are most prominent, an improvement in expected limits has been achieved compared to standard analysis techniques. The analysis sensitivity in the MSSM interpretation is improved over the full parameter space. There is large improvement for the signal interpretation in the 2HDM type 2 model.

Chapter 2

Theoretical foundation

2.1. The standard model of particle physics

Four fundamental forces are known up to now: the electromagnetic force, the weak force, the strong force and the gravitational force. The first three forces are well described within the standard model of particle physics [1–3], hereafter referred to as SM, while the latter can be described by general relativity. For all these interactions, with exception of the gravitational force, a unique quantum field theory exists. The interactions are mediated by gauge bosons. The SM in its full magnificence is represented by gauge groups of the form $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $SU(3)_C$ gauge group describes the quantum chromodynamics (QCD) and its force (strong force) couples to three colour charges (red, green or blue). Demanding local gauge symmetry within the QCD Lagrangian 8 gauge bosons are predicted for the strong force, namely gluons with different colour charge. The weak and electromagnetic force can be unified to the electroweak force, which is described by the gauge group $SU(2)_L \times U(1)_Y$. The physical gauge bosons of this unification appear via mixing of the neutral abstract fields (B) and (W^3) with a rotation matrix parametrised by the weak mixing angle θ_W and by combining the charged abstract fields W^1 and W^2 to be the photon γ, the Z^0 boson and the two charged $W^{\pm}$ bosons. All gauge bosons are listed in the following table.

Table 2.1.: Gauge bosons

Interaction	**Gauge Boson**	**Mass [GeV]**
electromagnetic	Photon γ	0
weak	Z^0	91.18
	$W^{\pm}$	80.40
strong	8 Gluons g	0

The remaining fundamental particles are fermions. They are subdivided into leptons and quarks. These are further divided into three generations due to their different masses. A full list of fermions (left-handed doublets) is given in table 2.2.

Table 2.2.: Left-handed elementary fermions form weak isospin doublets. Right-handed elementary fermions do not carry any weak isospin, but carry weak hypercharge. Neutrinos appear only in left-handed doublets.

Fermions	**Generation**			**Charge**			
	1	2	3	El. Charge	Weak Isospin	Weak Hypercharge	Colour
Leptons	ν_e	ν_μ	ν_τ	0	$+\frac{1}{2}$	-1	0
	e	μ	τ	$-e$	$-\frac{1}{2}$		
Quarks	u	c	t	$+\frac{2}{3}e$	$+\frac{1}{2}$	$+\frac{1}{3}$	red, green, blue
	d	s	b	$-\frac{1}{3}e$	$-\frac{1}{2}$		

2.1.1. Gauge invariance and interactions

A particle's kinematic behaviour and interactions can be described by a Lagrangian. The QED Lagrangian for a spin-1/2 field with electromagnetic interactions in natural units is shown in equation 2.1:

$$\mathcal{L} = -m\overline{\psi}\psi + \mathrm{i}\,\overline{\psi}\gamma^\mu\partial_\mu\psi + q\overline{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.1}$$

Interactions of particles are described demanding local gauge invariance of an already global gauge invariant Lagrangian. By introducing gauge fields into the Lagrangian, local gauge invariance can be achieved and thus leads to the known force mediator, listed in table 2.1. An example of this effect is shown for QED.
Global gauge invariance of a free spin-1/2 particle in QED:

$$\mathcal{L} \to \mathcal{L}' \qquad \text{with} \qquad \psi \to \psi' = e^{\mathrm{i}\,\alpha}\psi \tag{2.2}$$

$$\begin{aligned}
\mathcal{L}' &= -m\overline{\psi'}\psi' + \mathrm{i}\,\overline{\psi'}\gamma^\mu\partial_\mu\psi' && (2.3)\\
&= -m\overline{\psi}e^{-i\alpha}e^{i\alpha}\psi + \mathrm{i}\,\overline{\psi}e^{-\mathrm{i}\,\alpha}\gamma^\mu\partial_\mu e^{\mathrm{i}\,\alpha}\psi && (2.4)\\
&= \mathcal{L} && (2.5)
\end{aligned}$$

Local gauge invariance of a free spin-1/2 particle in QED:

$$\mathcal{L} \to \mathcal{L}' \qquad \text{with} \qquad \psi \to \psi' = e^{i\alpha(x)}\psi \tag{2.6}$$

$$\begin{aligned}
\mathcal{L}' &= -m\overline{\psi'}\psi' + \mathrm{i}\,\overline{\psi'}\gamma^\mu\partial_\mu\psi' && (2.7)\\
&= -m\overline{\psi}e^{-\mathrm{i}\,\alpha(x)}e^{\mathrm{i}\,\alpha(x)}\psi + \mathrm{i}\,\overline{\psi}e^{-\mathrm{i}\,\alpha(x)}\gamma^\mu\partial_\mu e^{\mathrm{i}\,\alpha(x)}\psi && (2.8)\\
&= -m\overline{\psi}\psi + \mathrm{i}\,\overline{\psi}\gamma^\mu\partial_\mu\psi - \overline{\psi}\gamma^\mu\psi\partial_\mu\alpha(x) && (2.9)\\
&\neq \mathcal{L} && (2.10)
\end{aligned}$$

Local gauge invariance can not be reached with a Lagrangian describing the kinematics of a free spin-1/2 particle. By adding gauge fields, the invariance of the Lagrangian under local gauge transformations is retained, and in addition the interaction of a particle with the electromagnetic (photon) field can be described as following:

$$\mathcal{L} = -m\overline{\psi}\psi + \mathrm{i}\,\overline{\psi}\gamma^\mu\partial_\mu\psi + \underbrace{q\overline{\psi}\gamma^\mu\psi A_\mu}_{\text{Interaction}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{Photon field}}. \tag{2.11}$$

Local gauge invariance of a spin-1/2 particle interacting with the photon field in QED:

$$\mathcal{L} \to \mathcal{L}' \qquad \text{with} \qquad \psi \to \psi' = e^{i\alpha(x)}\psi \qquad \text{and} \qquad A \to A' = A - \frac{1}{q}\partial_\mu \alpha(x) \tag{2.12}$$

$$\mathcal{L}' = -m\overline{\psi'}\psi' + i\,\overline{\psi'}\gamma^\mu\partial_\mu\psi' + q\overline{\psi'}\gamma^\mu\psi' A'_\mu - \frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} \tag{2.13}$$

$$= -m\overline{\psi}\psi + i\,\overline{\psi}\gamma^\mu\partial_\mu\psi + q\overline{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}\gamma^\mu\psi\partial_\mu\alpha(x) - \overline{\psi}\gamma^\mu\psi\partial_\mu\alpha(x) \tag{2.14}$$

$$= \mathcal{L} \tag{2.15}$$

Using this principle three fundamental interactions can be introduced into the theory: the electromagnetic, the strong and the weak interaction. Introducing gauge fields to the QCD Lagrangian will generate eight gluons, which differ only in their colour charge composition. The electromagnetic and weak force can be unified. Local gauge invariance of the $SU(2)_L \times U(1)_Y$ group leads to the four gauge bosons of the electroweak unification.

2.1.2. Higgs mechanism in the standard model

2.1.2.1. Englert-Brout-Higgs mechanism

One of the fundamental problems of the SM was the absence of mass terms in the Lagrangian. In order to keep the SM Lagrangian renormalizable, gauge invariance has to be demanded. Artificially introducing mass terms to the SM Lagrangian violates this fundamental principle. Therefore a new mechanism was introduced by which the weak gauge bosons gain their mass [4–6]. A new $SU(2)$ scalar field ϕ is introduced. This scalar doublet reads:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\,\phi_2 \\ \phi_3 + i\,\phi_4 \end{pmatrix}. \tag{2.16}$$

A Lagrangian based on this field can be constructed, containing a term for the kinetic energy and one for the potential energy:

$$\mathcal{L}_{\text{Higgs}} = \underbrace{(D^\mu\phi)^\dagger(D_\mu\phi)}_{\text{Kinetic}} - \underbrace{\mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2}_{\text{Potential}}, \tag{2.17}$$

where D_μ denotes the covariant derivative

$$D_\mu = \partial_\mu - ig_1 W^a_\mu T_a - ig_2\frac{Y}{2}B_\mu. \tag{2.18}$$

The potential contains two parameters: λ and μ^2. The first parameter is real and positive and describes a self coupling term. The other term, containing μ^2, looks already like a mass-type term. If μ^2 is chosen to be positive, there is only one ground state possible: $\langle\phi\rangle_{min} = 0$. Unfortunately this will not solve the aforementioned mass problem. Choosing μ^2 to be negative leads to the following minimum of the potential:

$$\frac{\partial V}{\partial \phi^\dagger \phi} = \mu^2 + 2\lambda \phi^\dagger \phi \tag{2.19}$$

$$\Rightarrow |\phi_{min}| = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}, \tag{2.20}$$

where v denotes the vacuum expectation value (VEV). Without the loss of generality at the potential minimum the fields $\phi_i, i = 1..4$ can be chosen in the following way:

$$\phi_1 = \phi_2 = \phi_4 = 0 \quad \text{and} \quad \phi_3 = \phi_{min}. \tag{2.21}$$

Now the ground state of the potential is no longer symmetric under SU(2) rotations, but still keeping the renormalizability [7]. This choice leads to a ground state which can be expanded by using the Higgs field $\mathrm{h}(x)$:

$$\langle \phi \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \Rightarrow \quad \phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \mathrm{h}(x) \end{pmatrix}. \tag{2.22}$$

With $\phi(x)$ the Lagrangian $\mathcal{L}_{\text{Higgs}}$ in equation 2.17 can be expanded. The kinetic term $\left|\mathrm{D}_\mu \phi\right|^2$ reads:

$$\left|\mathrm{D}_\mu \phi\right|^2 = \frac{1}{8} v^2 g_2^2 \left|\mathrm{W}_\mu^1 + \mathrm{i}\, \mathrm{W}_\mu^2\right|^2 + \frac{1}{8} v^2 \left|g_2 \mathrm{W}_\mu^3 - g_1 \mathrm{B}_\mu\right|^2 + \ldots \tag{2.23}$$

$$= \frac{1}{2} m_\mathrm{W}^2 \mathrm{W}_\mu^+ \mathrm{W}^{-\mu} + \frac{1}{2} m_\mathrm{Z}^2 \mathrm{Z}_\mu \mathrm{Z}^\mu + 0 \cdot A_\mu A^\mu. \tag{2.24}$$

Here v denotes the VEV, g_1 and g_2 the coupling constants of the W and B fields, and $\mathrm{W}_\mu^{1,2,3}$ and B_μ the gauge fields of the electroweak theory. $\mathrm{W}_\mu^\pm, \mathrm{Z}_\mu$ and A_μ are the physical representations of these gauge fields. Finally it shows that the kinetic term directly rises mass terms for the weak gauge bosons, but not for the photon. The same expansion can be done for the potential of the Higgs Lagrangian:

$$V = \frac{1}{2} \mu^2 (v + \mathrm{h})^2 + \frac{1}{4} \lambda (v + \mathrm{h})^4 \quad \Rightarrow \quad m_\mathrm{h} = -\sqrt{2}\, \mu = v \sqrt{2\lambda}. \tag{2.25}$$

A very interesting property of the Higgs field appears: It comes along with a new particle with the mass $m_\mathrm{h} = v\sqrt{2\lambda}$, the SM Higgs boson. Additionally trilinear and quartic self-coupling terms in the expansion of the potential can be seen.

A mechanism is introduced, which spontaneously breaks the SU(2) symmetry, but not the U(1) or SU(3) symmetry. This leads to mass terms for the weak gauge bosons, with exception of the photon and the gluon, while keeping the SM Lagrangian renormalizable.

2.1.2.2. Yukawa interaction

In order to generate mass terms of the fermions, Yukawa interaction terms are added to the Lagrangian. Yukawa interaction describes the interaction between a scalar field (Higgs field) and a Dirac field (fermion field). Additionally the chirality of the fermions has to be taken into account. The Lagrangian for the fermion mass then reads:

$$\mathcal{L}_{m_f} = -\lambda_f \left(\psi_{\mathrm{L}}^{\dagger}\, \phi\, \psi_{\mathrm{R}} + \psi_{\mathrm{R}}^{\dagger}\, \phi^{\dagger}\, \psi_{\mathrm{L}}\right) \quad \text{with} \quad \phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \mathrm{h}(x) \end{pmatrix}. \tag{2.26}$$

Here ψ_{L} describes the left-handed doublet and ψ_{R} the right-handed singlet. After spontaneous symmetry breaking the fermion mass m_f and the coupling to the Higgs field $g_{\mathrm{h}ff}$ appear:

$$g_{\mathrm{h}ff} = \frac{m_f}{v}. \tag{2.27}$$

This principle now also generates mass for the fermions. Finally it can be concluded, that the coupling of fermions to the Higgs field is proportional to the fermion mass.

2.1.3. Limitations of the standard model

The SM is able to describe particle properties and their interactions very successfully. Unfortunately it still leaves a lot of open questions and thus is known to be incomplete, the main open questions can be found in Refs. [8–14]. A major discrepancy is found in the renormalized and fundamental value of the Higgs boson mass, known as the hierarchy problem. Additionally there is still no explanation as to why there are exactly three generations of quarks and leptons, and CP violation is absent in strong interactions. It also does not cover gravity at all. Astrophysical measurements show that there is a phenomenon called dark matter, which also can not be described by the SM. Another very fundamental observation is that the running coupling constants can not be unified for very high energies.
The SM has many more open questions and problems than the aforementioned ones. Therefore it is necessary to search for physics beyond the SM (BSM). A very promising and famous superset of BSM theories is supersymmetry (SUSY) [15]. A minimal extension to the known SM in terms of SUSY will be discussed in the following section 2.2.

2.2. The minimal supersymmetric standard model

The supersymmetric extension of the SM fixes most of the known problems of the SM, especially the aforementioned hierarchy problem. The main idea is to introduce a new supersymmetric partner to all known particles of the SM. These new particles follow an underlying symmetry transformation, which maps all quantum numbers except of the spin. By shifting the spin of all supersymmetric particles by $\frac{1}{2}$ all fermions become bosons and vice versa. In order to give unique names all SUSY fermion partners have the SM name with an additional "s" prefix: e.g. selectron, stau are sfermions. Similarly the SUSY boson partners are called gauginos and usually carry the SM name with an "ino" extension: e.g. photino or Wino.
Supersymmetry is not a single theory, but rather a combination of multiple theory approaches. This leads to a lot of specific SUSY theories with special assumptions. The most minimal extension to the SM is the minimal supersymmetric standard model (MSSM). It is a two Higgs doublet theory (2HDM type 2) with different non-vanishing vacuum expectation values (VEV) v_2 and v_1, which correspond to up-type quarks, and down-type quarks and all leptons respectively. These two VEVs can be parametrised with the SM Higgs VEV to

$v^2 = v_1^2 + v_2^2$. The two Higgs doublets create eight degrees of freedom, of which three are eaten up by Goldstone bosons to give masses to the Z and $W^{\pm}$ bosons in the same way as the SM Higgs mechanism. The other five degrees of freedoms result in five Higgs bosons:

$$h, H, A, H^+, H^-, \tag{2.28}$$

where h is a light SM-like $\mathcal{CP}$-even Higgs boson, H a heavier $\mathcal{CP}$-even one, A a heavy $\mathcal{CP}$-odd Higgs boson and $H^{\pm}$ as two charged Higgs bosons. The particles and their supersymmetric partners of the minimal supersymmetric standard model can be found in Fig. 2.1.

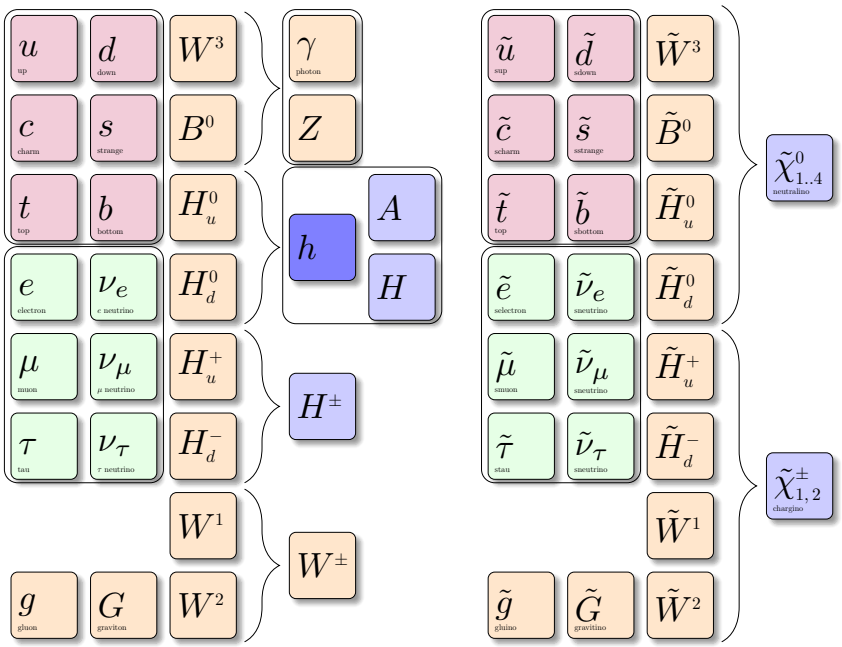


Figure 2.1.: Overview of particles (left) and their supersymmetric partners (right) in the minimal supersymmetric standard model [16].

As shown in Fig. 2.1 the extended Higgs sector of the MSSM has supersymmetric partners too, namely higgsinos. They mix together with the other gauginos, similar to the mixing in the electroweak sector, and form four neutral neutralinos $\widetilde{\chi}_1^0$, $\widetilde{\chi}_2^0$, $\widetilde{\chi}_3^0$ and $\widetilde{\chi}_4^0$, and four charginos $\widetilde{\chi}_1^+$, $\widetilde{\chi}_1^-$, $\widetilde{\chi}_2^+$ and $\widetilde{\chi}_2^-$. The MSSM can be extended to the full SUSY sector by including the graviton G and its supersymmetric partner gravitino $\widetilde{G}$.

2.2.1. Supersymmetric extension of the Higgs sector

The MSSM introduces two scalar Higgs doublets ϕ_1 and ϕ_2, which can be written as

$$\phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix}, \quad \phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix}, \tag{2.29}$$

where ϕ_1^+ and ϕ_2^+ denote the charged complex fields and ϕ_1^0 and ϕ_2^0 the neutral ones. The most general Lagrangian which is invariant under $SU(2)_L \times U(1)_Y$ transformation reads

$$\mathcal{L}_{2HDM} = \mathcal{L}_{SM} + \mathcal{L}_{Yukawa} + \mathcal{L}_{\phi}. \tag{2.30}$$

$\mathcal{L}_{\text{SM}}$ describes the interaction of the gauge bosons with the fermions, $\mathcal{L}_{\text{Yukawa}}$ the interaction between the Higgs bosons and fermions and $\mathcal{L}_\phi$ the scalar field with

$$\mathcal{L}_\phi = \sum_{i=1,2} \underbrace{\left(D_\mu \phi_i\right)^\dagger \left(D^\mu \phi_i\right)}_{\text{Kinetic}} - \underbrace{V_{\text{2HDM}}(\phi_1, \phi_2)}_{\text{Potential}}, \tag{2.31}$$

where D_μ denotes the extended covariant derivative of the SM 2.18. The most general potential without $\mathcal{CP}$-violation can be written as

$$V_{\text{2HDM}} = m_{11}^2 \left(\phi_1^\dagger \phi_1\right) + m_{22}^2 \left(\phi_2^\dagger \phi_2\right) + m_{12}^2 \left[\left(\phi_1^\dagger \phi_2\right) + \left(\phi_2^\dagger \phi_1\right)\right] \tag{2.32}$$

$$+ \frac{\lambda_1}{2} \left(\phi_1^\dagger \phi_1\right)^2 + \frac{\lambda_2}{2} \left(\phi_2^\dagger \phi_2\right)^2 + \lambda_3 \left(\phi_1^\dagger \phi_1\right) \left(\phi_2^\dagger \phi_2\right) \tag{2.33}$$

$$+ \lambda_4 \left(\phi_1^\dagger \phi_2\right) \left(\phi_2^\dagger \phi_1\right) + \frac{\lambda_5}{2} \left[\left(\phi_1^\dagger \phi_2\right)^2 + \left(\phi_2^\dagger \phi_1\right)^2\right], \tag{2.34}$$

where m_{ij} denotes mass like parameters and λ_i quartic couplings of the Higgs scalars.
The electroweak symmetry breaking works in the same way now as the one in the SM. Without the loss of generality the ground states of both doublets can be written as

$$\langle \phi_1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad \langle \phi_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix}, \tag{2.35}$$

where v_1 and v_2 are the two real VEV of the doublets, which fulfil $v_{\text{SM}} = \sqrt{v_1^2 + v_2^2} = 246\,\text{GeV}$. The neutral component of the two scalar fields can now be expanded around their minimum:

$$\phi_1 = \begin{pmatrix} \phi_1^+ \\ \frac{v_1 + \rho_1 + i\eta_1}{\sqrt{2}} \end{pmatrix}, \quad \phi_2 = \begin{pmatrix} \phi_2^+ \\ \frac{v_2 + \rho_2 + i\eta_2}{\sqrt{2}} \end{pmatrix}, \tag{2.36}$$

where ρ_1 and ρ_2 are two real scalar fields for the $\mathcal{CP}$-even eigenstates. The other two real scalar fields η_1 and η_2 describe the $\mathcal{CP}$-odd eigenstates of the neutral Higgs sector and introduce the G^0 Goldstone boson. Inserting the two expanded scalar fields again in the aforementioned potential V_{2HDM} leads to

$$V_{\text{2HDM}} = (\phi_1, \phi_2)\, \mathcal{M}_{\phi^\pm} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} + \frac{1}{2} (\rho_1, \rho_2)\, \mathcal{M}_\rho \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} + \frac{1}{2} (\eta_1, \eta_2)\, \mathcal{M}_\eta \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \tag{2.37}$$

with the mass-squared matrices:

$$\mathcal{M}_{\phi^\pm} = \left[m_{12}^2 - (\lambda_4 + \lambda_5) \frac{v_1 v_2}{2}\right] \begin{pmatrix} \frac{v_2}{v_1} & -1 \\ -1 & \frac{v_1}{v_2} \end{pmatrix}, \tag{2.38}$$

$$\mathcal{M}_\rho = \begin{pmatrix} m_{12}^2 \frac{v_2}{v_1} + \lambda_1 v_1^2 & -m_{12}^2 + (\lambda_3 + \lambda_4 + \lambda_5)\, v_1 v_2 \\ -m_{12}^2 + (\lambda_3 + \lambda_4 + \lambda_5)\, v_1 v_2 & m_{12}^2 \frac{v_1}{v_2} + \lambda_2 v_2^2 \end{pmatrix}, \tag{2.39}$$

$$\mathcal{M}_\eta = \frac{m_A^2}{v_1^2 + v_2^2} \begin{pmatrix} v_2^2 & -v_1 v_2 \\ -v_1 v_2 & v_1^2 \end{pmatrix}. \tag{2.40}$$

The three matrices $\mathcal{M}_{\phi^\pm}$, $\mathcal{M}_\rho$ and $\mathcal{M}_\eta$ can be diagonalised in order to obtain the physical scalars of the 2HDM. With a rotation of $\tan\beta = \frac{v_2}{v_1}$ the matrix $\mathcal{M}_{\phi^\pm}$ can be diagonalised, which leads to two Goldstone bosons $G^\pm$, which are eaten up by the $W^\pm$ gauge bosons, and the two charged Higgs bosons H^+ and H^-:

$$G^\pm = \cos(\beta)\phi_1^\pm + \sin(\beta)\phi_2^\pm, \tag{2.41}$$
$$H^\pm = -\sin(\beta)\phi_1^\pm + \cos(\beta)\phi_2^\pm. \tag{2.42}$$

This leads to the following mass relation of the two charged Higgs bosons:

$$m_{H^\pm}^2 = \left[\frac{m_{12}^2}{v_1 v_2} - \frac{\lambda_4 + \lambda_5}{2}\right] \underbrace{(v_1^2 + v_2^2)}_{=246\,\text{GeV}}. \tag{2.43}$$

The matrix $\mathcal{M}_\rho$ can be diagonalised by a rotation of the Higgs mixing angle α. The eigenstates of the diagonalised matrix generate the two physical $\mathcal{CP}$-even Higgs bosons h and H:

$$h = -\sin(\alpha)\rho_1 + \cos(\alpha)\rho_2, \tag{2.44}$$
$$H = \cos(\alpha)\rho_1 + \sin(\alpha)\rho_2, \tag{2.45}$$

with the corresponding masses:

$$m_h^2 = \frac{1}{2}\left(\mathcal{M}_{\rho,11} + \mathcal{M}_{\rho,22} - \sqrt{(\mathcal{M}_{\rho,11} - \mathcal{M}_{\rho,22})^2 + 4\mathcal{M}_{\rho,12}^2}\right), \tag{2.46}$$

$$m_H^2 = \frac{1}{2}\left(\mathcal{M}_{\rho,11} + \mathcal{M}_{\rho,22} + \sqrt{(\mathcal{M}_{\rho,11} - \mathcal{M}_{\rho,22})^2 + 4\mathcal{M}_{\rho,12}^2}\right), \tag{2.47}$$

where h corresponds to the SM Higgs boson in the case of $\cos(\beta - \alpha) = 0$ or $\sin(\beta - \alpha) = 1$ and H to a new heavier scalar boson.
The last matrix can be diagonalised again by using the angle β. The eigenstates generate a massless Goldstone boson G^0, which is absorbed by the Z boson, and a neutral $\mathcal{CP}$-odd boson:

$$G^0 = \cos(\beta)\eta_1 + \sin(\beta)\eta_2, \tag{2.48}$$
$$A = -\sin(\beta)\eta_1 + \cos(\beta)\eta_2, \tag{2.49}$$

with the corresponding mass:

$$m_A^2 = \left[\frac{m_{12}^2}{v_1 v_2} - \lambda_5\right](v_1^2 + v_2^2). \tag{2.50}$$

In total the eigenstates of the diagonalised $\mathcal{M}_{\phi^\pm}$, $\mathcal{M}_\rho$ and $\mathcal{M}_\eta$ matrices generate eight degrees of freedom. Three are Goldstone bosons, which are eaten up by the weak gauge bosons. The remaining five bosons of the 2HDM are two $\mathcal{CP}$-even Higgs bosons h and H, two charged Higgs bosons $H^\pm$ and one $\mathcal{CP}$-odd Higgs boson A.

2.2.2. Benchmark scenarios in MSSM

The full parameter space of the MSSM Higgs sector comprises 124 free parameters. Most of these parameters have impact on the Higgs sector only via higher order processes. Thus benchmark scenarios are used to investigate the MSSM. In these benchmark scenarios most parameters, as masses, coupling constants and mixing parameters of SUSY, are set to certain reasonable values. This allows to reduce the full parameter space to two parameters in leading order: $\tan\beta$ and m_A, which parametrises the masses of the remaining four Higgs bosons:

$$m^2_{h,H} = \frac{1}{2}\left(m^2_A + m^2_Z \mp \sqrt{\left(m^2_A + m^2_Z\right)^2 - 4m^2_A m^2_Z \cos^2 2\beta}\right), \tag{2.51}$$

$$m_{H^\pm} = \sqrt{m^2_A + m^2_W}. \tag{2.52}$$

This thesis focusses on the signal interpretation in the $m_h^{\text{mod}+}$ and hMSSM scenarios. A more detailed overview of MSSM scenarios can be found in Refs. [17–19].

2.2.2.1. $m_h^{\text{mod}+}$ scenario

The old m_h^{max} scenario of MSSM searches at the LEP collider is modified by changing the ratio of the on-shell stop mixing parameter to the third generation SUSY squarks from 2.0 to 1.5. This modification was introduced after the discovery of the SM Higgs boson in 2012 at LHC. Table 2.3 shows the assumed values for the remaining parameters in the $m_h^{\text{mod}+}$ scenario.

Table 2.3.: Parameter set for the $m_h^{\text{mod}+}$ scenario.

Parameter	Value	Brief explanation
M_{SUSY}	1000 GeV	Mass of the third generation squarks
$M_{\tilde{q}_{1,2}}$	1500 GeV	Mass of the first and second generation squarks
$M_{\tilde{l}_3}$	1000 GeV	Mass of the third generation sleptons
$M_{\tilde{l}_{1,2}}$	500 GeV	Mass of the first and second generation sleptons
A_b	$A_\tau = A_t = X_t + \mu\cot\beta$	Coupling constant to third generation SUSY particles
$A_{f_{1,2}}$	0	Coupling constant to first and second generation SUSY particles
μ	200 GeV	Higgsino mass parameter
M_2	1000 GeV	Gaugino mass parameters for U(1)
M_1	$\approx$ 478 GeV	Gaugino mass parameters for SU(2)
X_t^{OS}	$1.5 M_{\text{SUSY}}$	On-shell stop mixing parameter
$m_{\tilde{g}}$	1500 GeV	Gluino mass

2.2.2.2. hMSSM scenario

The focus of the hMSSM scenario is on the discovered Higgs boson in 2012. It is assumed that the low mass $\mathcal{CP}$-even Higgs boson of the MSSM corresponds directly to the discovered Higgs boson with its mass fixed to 125 GeV (h(125)) and thus is independent of $\tan\beta$ and m_A. Furthermore it is assumed that the SUSY mass scale for all other SUSY particles is so high that they will not be discovered at LHC, which leads to a large soft-SUSY scale breaking. In hMSSM the Higgs sector is only described by $\tan\beta$ and m_A and the assumption that M_{SUSY}

is larger than all other parameters. Thus this scenario is less "model-dependent" than other MSSM scenarios. In the hMSSM scenario the masses of the two $\mathcal{CP}$-even Higgs boson are described by

$$\mathcal{M}^2_{\mathrm{h,H}} = \mathcal{M}^2_{\mathrm{tree-level}} + \Delta\mathcal{M}^2. \tag{2.53}$$

Only the $\Delta\mathcal{M}_{22}$ component is considered, since all other corrections become irrelevant for $M_{\mathrm{SUSY}} > \mu$.

Chapter 3

The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) [20] is a particle collider located and operated by the Conseil Européen pour la Recherche Nucléaire (CERN). It was designed to perform proton-proton, proton-ion and ion-ion collisions at four different interaction points. In total there are seven experiments whereof ATLAS, CMS, LHCb and ALICE mark the biggest detectors, each of them located at one of the aforementioned interaction points. The design center-of-mass energy for proton-proton collisions amounts to 14 TeV at an instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ making it the most powerful accelerator on earth. The accelerator with its pre-accelerators and the experiments is depicted in Fig. 3.1.

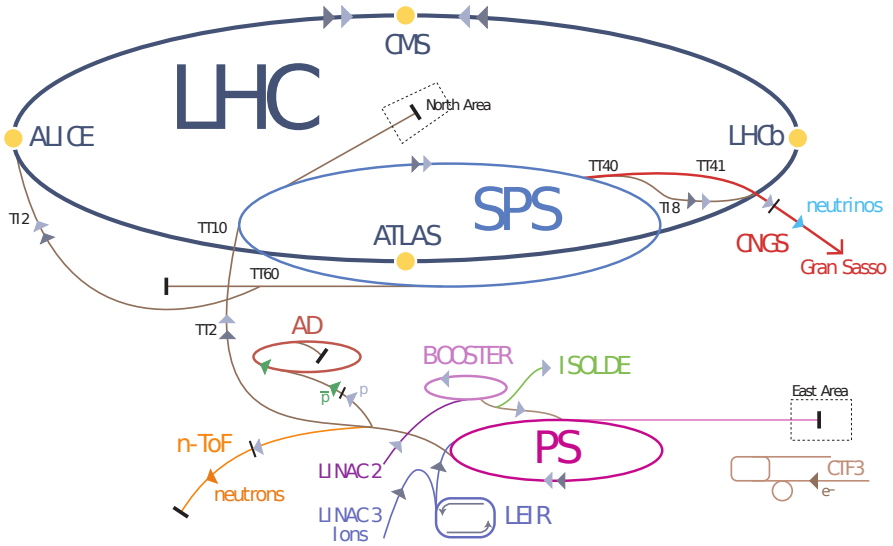


Figure 3.1.: Sketch of the LHC experiments including all pre-accelerators and all seven detector experiments [21].

The protons and ions need to pass a chain of pre-accelerators before reaching the LHC. Protons are obtained from a H_2 source (yearly sponsored by the general director of CERN) by injecting the gas into a very strong electric field, which separates successfully the atoms into protons and electrons. Afterwards these protons are accelerated at LINAC2 and smaller circular accelerator like Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). At this point the proton bunches have reached an energy of 450 GeV. It takes a few fills of the SPS into the LHC to reach the maximum capacity of 2808 bunches. Finally the LHC accelerates these bunches up to 6.5 TeV with radio frequency cavities, while keeping the bunches on their trajectory with 1232 superconducting dipole magnets. Each of those magnets builds up

a magnetic field of around 8.3 T. Orbit corrections and final focusing of the beams inside the detectors are performed using higher order multipole magnets.

While ATLAS and CMS were designed to perform searches for new particles and phenomena, the detectors LHCb and ALICE investigate specific phenomena. The main purpose of LHCb is to explain CP violation in B mesons, whereas ALICE focusses on quark-gluon plasma studies in heavy ion collisions. The two main reasons to build ATLAS and CMS were to find the Higgs boson and BSM phenomena like supersymmetry. A detailed description of the CMS detector can be found in chapter 4.

Chapter 4

The Compact Muon Solenoid (CMS) experiment

The multi-purpose Compact Muon Solenoid (CMS) detector was designed and built mainly to discover the Higgs boson and to search for evidence of physics beyond the standard model. It is able to detect particles coming from the proton-proton, proton-ions and ion-ion collisions at several TeV at the Large Hadron Collider (LHC) at CERN. The name of the detector suggests three main features. Firstly it is built more compact (smaller) than its bigger brother ATLAS. Secondly it has a special muon subdetector, which allows for a very efficient muon reconstruction. Last but not least a high magnetic field of 3.8 T is generated by a superconducting solenoid, which is located between the calorimeter subdetectors and the outermost muon chambers. The detector has a cylindrical shape with an onion-like structure of subdetectors in the barrel region. The detector is encapsulated at both ends with two endcaps. Fig. 4.1 shows the total detector complex. A more detailed description of the CMS detector can be found in the technical design reports [22–24].

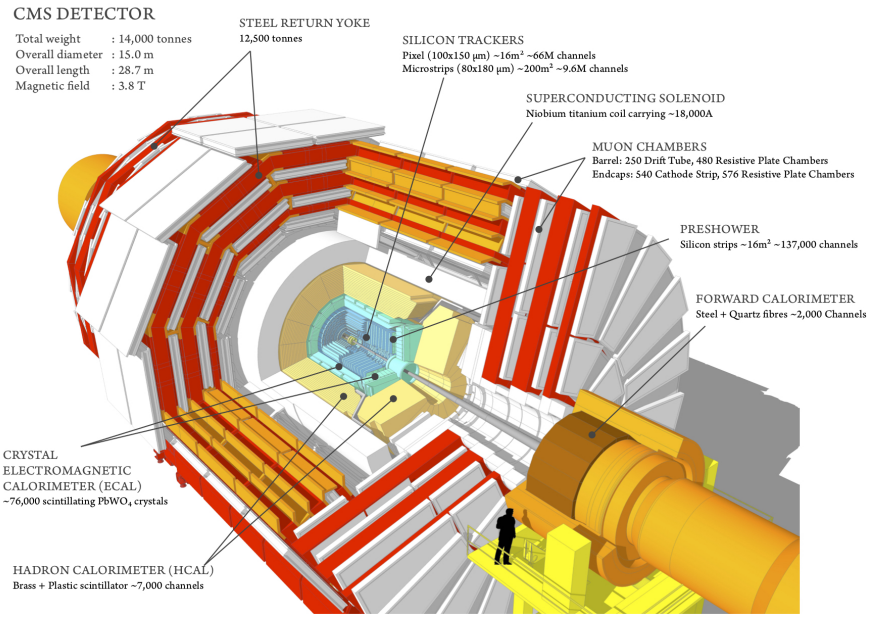


Figure 4.1.: Cutaway view of the CMS experiment containing all subdetector components [25].

4.1. Detector setup and geometry

Due to the cylindrical shape of the detector the coordinate system of CMS is described in cylindrical coordinates with the center of this system at the collision point. The x-axis points to the center of LHC, the y-axis to the surface and the z-axis into the beam direction. This allows the construction the cylindrical coordinates ϕ and θ. The azimuthal angle ϕ is measured in the x-y plane, whereas the polar angle θ is measured in the x-z plane. The polar angle is substituted with a new coordinate η, the pseudorapidity. Differences in η are invariant under Lorentz transformations. It is calculated in the following way:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{4.1}$$

With the definition of the azimuthal angle and the pseudorapidity an angular separation between two particles can be described as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{4.2}$$

It is common to reconstruct some quantities only in the transverse plane, e.g. the transverse momentum:

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{4.3}$$

4.1.1. Tracker

The innermost subdetector of CMS is the tracking system, which consists of the silicon pixel and the silicon strip detector. The total tracking system of CMS covers a region of pseudorapidities up to $|\eta| = 2.5$. It can reconstruct the trajectories of charged particles bended by the magnetic field by measuring small energy desposits of the traversing particles in the fine-grained silicon sensors. The energy desposition in the silicon is driven by ionization, which leads to an electric signal. The full three-momentum of a particle is reconstructed by measuring the coordinates of these signals and the track.
In the shutdown between 2016 and 2017 the pixel tracker subdetector received an upgrade. An additional layer in the barrel region and an additional disk in each endcap was installed. This upgrade was started after the data taking and detector simulation for Monte Carlo samples, which are used in this analysis. Therefore the ”old” pixel detector is described in the following section.

4.1.1.1. Pixel detector

The pixel tracker is the innermost tracking system and consists of three cylindrical layers in the barrel region and two disks at each of the two endcaps. The layout of the barrel and endcap pixel modules can be seen in Fig. 4.2.

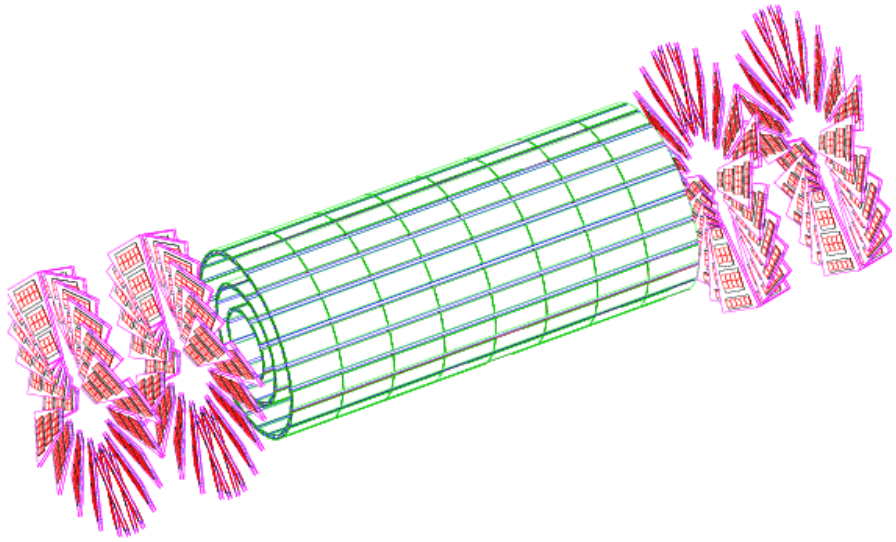


Figure 4.2.: Layout of the silicon pixel layers at CMS [22].

The barrel layers are located at radii of 4.4 cm, 7.3 cm, and 10.2 cm, of which each cylinder is 53 cm long. The pixel disks at the endcaps extend from 6 cm to 15 cm, located at $z = \pm 34.5$ cm and ± 46.5 cm. In total there are 768 pixel modules in the barrel region and 672 pixel modules in the endcap region. In order to obtain the best vertex resolution, the pixel size is designed to be $100 \times 150\,\mu\text{m}^2$. In the end the silicon pixel detector provides a spatial resolution of $10\,\mu$m in the $r - \phi$ plane and $20\,\mu$m along the z-axis.

4.1.1.2. Strip detector

The silicon strip detector is divided into four subregions and completely surrounds the silicon pixel tracker as shown in Fig. 4.3. The barrel region is divided into two subparts: the TIB (Tracker Inner Barrel) and the TOB (Tracker Outer Barrel). In the endcap region the silicon strip detector is divided into the TID (Tracker Inner Disks) and the TEC (Tracker End Cap). The TIB consists of four module layers. It provides a spatial resolution of 23 to 34 μm in the $r - \phi$ direction and 23 μm in z direction, while covering up to $|z| < 65$ cm. The TOB covers up to $|z| < 110$ cm with six module layers and provides a single-point resolution of 35 to 52 μm in the $r - \phi$ direction and 52 μm in z direction. Above $|\eta| \approx 1.4$ the TEC (9 module layers) and the TID (3 module layers) are located.
Since the strips can only measure one coordinate, some modules are double-sided and tilted by an angle of 100 mrad. The strip detector consists of 15400 modules in total. The TIB and TID systems provide single point resolutions of 23 to 35 μm, while the TEC system provides

a single point resolution of 35 to 53 μm. In the end the strip detector provides a final single-point resolution of 23 to 53 μm in the $r - \phi$ direction and 230 to 530 μm in the z-direction.

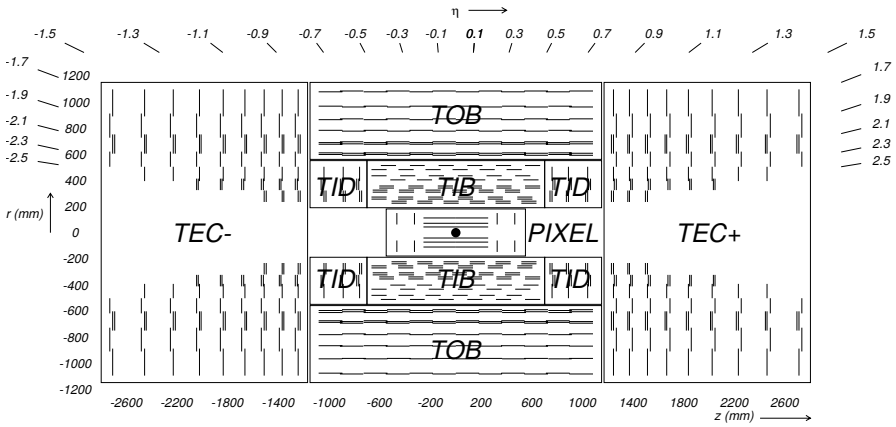


Figure 4.3.: Layout of the four silicon strip subdetector parts in the CMS tracker [26].

4.1.2. Calorimeters

The calorimeters of CMS are located outside of the tracking system and are designed to measure the energy of particles. Most particles will be stopped completely in the calorimeters. The calorimeters are splitted in two subdetectors, the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). While the ECAL measures mainly the energy of photons and electrons, the HCAL measures the energy of hadrons. Most of the time hadrons appear multiple times with other particles within a narrow cone. These structures are called jets. They are produced by the hadronization of quarks and gluons, since they can not exist freely due to the QCD confinement.

4.1.2.1. Electromagnetic calorimeter

The ECAL stops electrons and photons. Muons and most hadrons are not stopped entirely. They will leave energy deposition due to ionisation in the calorimeter. The full setup of the ECAL can be seen in Fig. 4.4.

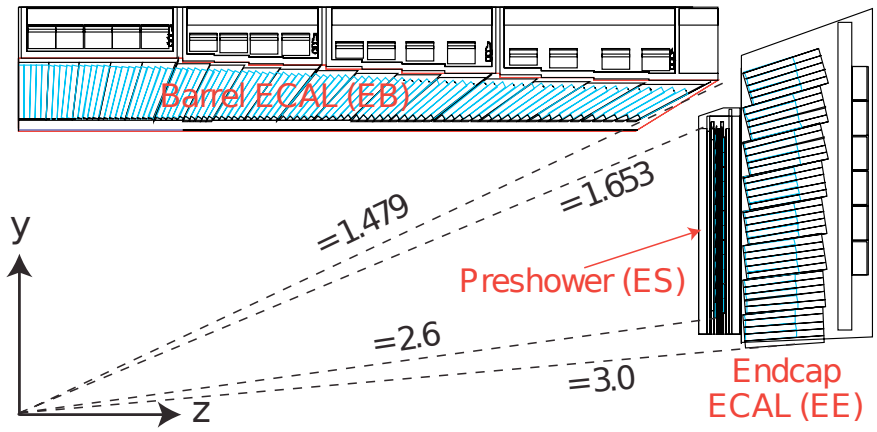


Figure 4.4.: Layout of the electromagnetic calorimeter and its three submodules in the barrel and endcap region [22].

The barrel region contains in total 61200 lead tungstate ($PbWO_4$) crystals and is encapsulated by 7324 crystals in each of the 2 endcaps. The crystals in the barrel region have front face cross-section of $22 \times 22\,\mathrm{mm}^2$ and a length of 230 mm corresponding to a radiation length of $25.8X_0 (X_0 = 0.89\,\mathrm{cm})$. Crystals in the endcap region ($1.479 < |\eta| < 3.0$) have a size of $28.6 \times 28.6\,\mathrm{mm}^2$ and a length of 220 mm corresponding to a radiation length of $24.7X_0$. In front of the endcap ECAL there is the so-called preshower. Its main purpose is to detect single photons coming from neutral pion decays.

The energy resolution contains three different terms. The first term is a stochastic term, which arises due to the counting of photons in the scintillators, the second term is a noise term and the last one a constant offset:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\,\%}{\sqrt{E/\,\mathrm{GeV}}}\right)^2 + \left(\frac{12\,\%}{E/\,\mathrm{GeV}}\right)^2 + (0.3\,\%)^2 . \tag{4.4}$$

4.1.2.2. Hadronic calorimeter

The HCAL is a sampling calorimeter, where brass absorber plates alternate with plastic scintillators. The fundamental principle of this subdetector is the same as the ECAL. Hadrons are stopped in this subdetector by depositing their energy through hadronic showers. The following Fig. 4.5 shows the full HCAL in the CMS coordinate system.

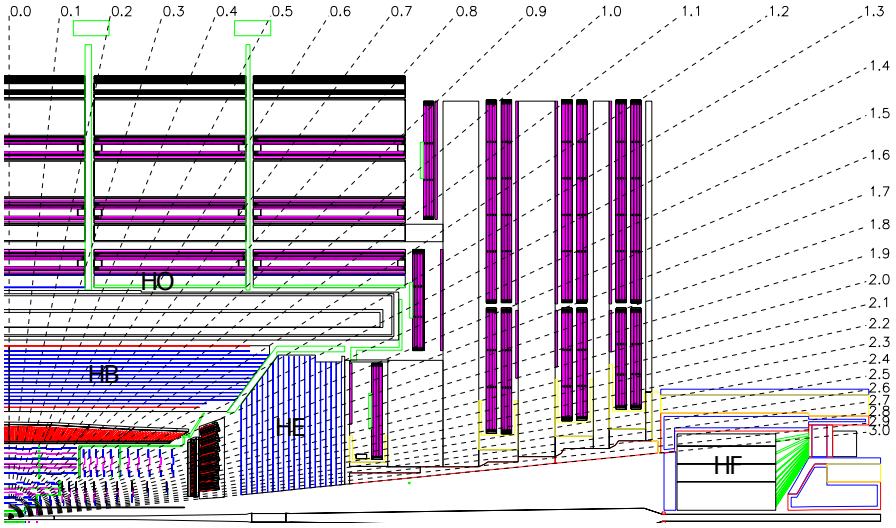


Figure 4.5.: Layout of the hadronic calorimeter and its three submodules in the barrel, endcap and forward region [26].

The calorimeter is splitted into three parts: the hadron outer ($|\eta| < 1.26$), the hadron endcap ($1.3 < |\eta| < 3.0$) and the hadron forward ($3.0 < |\eta| < 5.0$).

The energy resolution of the HCAL can be parametrized with:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{84.7\,\%}{\sqrt{E/\,\mathrm{GeV}}}\right)^2 + (7.4\,\%)^2 . \tag{4.5}$$

4.1.3. Solenoid

The superconducting solenoid magnet encapsulates the tracking system and the ECAL and HCAL. It provides a magnetic field of 3.8 T, which bends charged particles into the trans-

verse plane of the detector and allows for a measurement of the transverse momentum. The superconducting magnet is cooled down to 4.6 K in order to make the coil superconducting. Furthermore it stops remaining jets from the HCAL.

4.1.4. Muon chambers

The muon system of CMS is the outermost subdetector. All particles are stopped already in the calorimeters, except of neutrinos and muons. While the neutrinos leave the detector without any trace, the muons are the only particles reaching the muon chambers. Therefore this subdetector is specialised on muon detection with three different gaseous particle detectors. The fundamental detector principle is similar to the inner tracking system. Charged muons traverse the gas, ionise gas atoms and the charge carriers drift then in an electric field generating signals at anode wires and cathode strips. This induces a measurable current in the wire and provides spatial information of hits and tracks.
The full muon chamber system covers up to $|\eta| < 2.4$. In the barrel region drift tubes (DT) are installed up to a pseudorapidity of $|\eta| < 1.2$. The cathode strip chambers (CSC) are installed in the endcap region. Additional resistive plate chambers (RPC) are added in the DTs and CSCs. They have a very good time resolution, while the DT and CSC provide a good spatial resolution. A full sketch of the CMS muon chambers can be seen in Fig. 4.6.

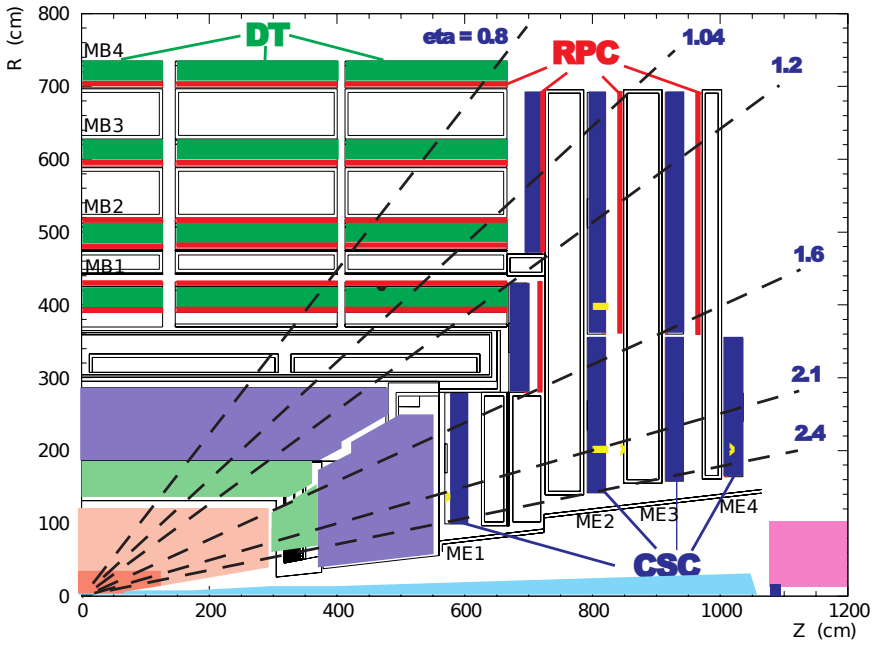


Figure 4.6.: Sketch of the muon subdetector system [22] in the CMS detector. The three different gaseous muon system parts are shown: The barrel drift tubes, the endcap cathode strip chambers and the resistive plate chambers.

4.2. Data acquisition

The LHC provides a bunch crossing rate of 40 MHz with multiple collisions within one crossing, which corresponds to a data rate of $\approx$ 40 TB/s. All detector signals can obviously not be stored on tapes with this rate. Therefore a multistage trigger system is installed in CMS.

First of all the rate is reduced from 40 MHz to 100 kHz by the Level-1 trigger. It is a hardware-based system using field-programmable gate arrays (FPGA). The surviving events are sent to another computing farm, which further reduces the rate to 1 kHz. This trigger system is called high-level trigger (HLT). In this step a first reconstruction step is performed. The HLT surviving events are finally written to tape and are available for a dedicated offline analysis. At the same time the trigger logic is used to select interesting events for the corresponding physics analysis.

Chapter 5

Analysis of the high mass $\mathrm{H} \rightarrow \mathrm{WW}$ search

With the discovery of the Higgs boson in 2012 at CMS and ATLAS modern experimental particle physics achieved a major breakthrough. Up to now the measured quantities like couplings to gauge bosons and fermions are compatible with the SM Higgs boson prediction. Moreover it is important to measure all properties, e.g. the $\mathcal{CP}$ quantum number, with high precision in order to check for physics beyond the SM in the Higgs sector. A more direct search for an extended Higgs sector is a search for additional Higgs bosons. Many theories predict multiple Higgs bosons. This thesis focusses on a model-independent search for an additional neutral scalar to the $\mathrm{h}(125)$ and to the interpretation of the results in two MSSM scenarios, which predict in total five Higgs bosons as described in section 2.2.
The analysis is performed in the $\mathrm{H} \rightarrow \mathrm{WW}$ channel, where both W bosons decay leptonically, but with different flavour. The final state, which is measured in the CMS detector is then an electron, a muon and two neutrinos, which escape the detector. Choosing the opposite flavour dileptonic final state reduces the $\mathrm{Z} \rightarrow ll$ (Drell-Yan) and multijet (QCD) background already heavily by nature. Furthermore a new analysis strategy using deep learning techniques is applied to the categorisation, which enables more control over the SM background processes. Fig. 5.1 shows the branching fraction of a single W boson decay.

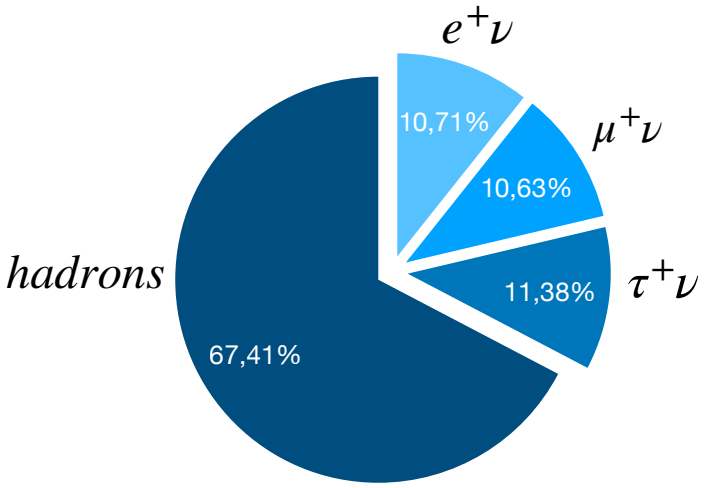
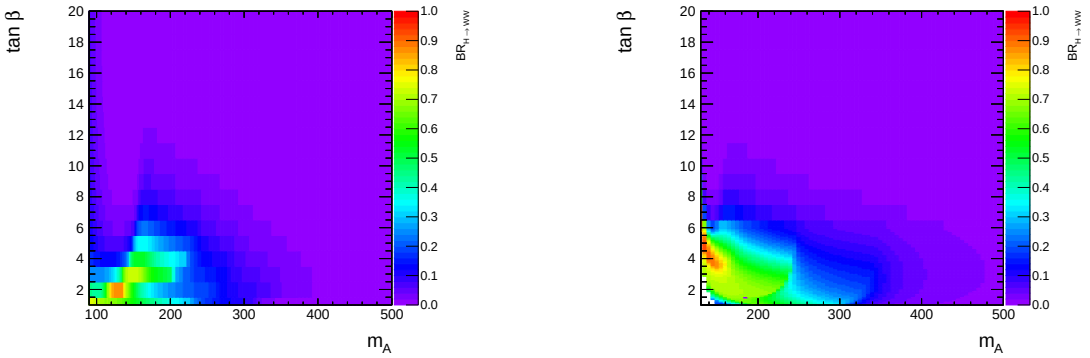


Figure 5.1.: Pie chart of the branching fractions of a single W boson decay in percent [27].

Assuming a branching fraction $\mathcal{BR}(H \to WW)$ of $\approx 21.4\%$, which corresponds to the SM Higgs boson, and using a branching fraction $\mathcal{BR}(W \to e(\mu)\nu_e(\nu_\mu))$ of $\approx 10.8\%(10.6\%)$ [27] the branching fraction of the chosen final state is $\mathcal{BR}(H \to WW \to e\mu\nu_e\nu_\mu) \approx 0.52\%$. The branching fraction of $H \to WW$ in MSSM scenarios is dependent on $\tan\beta$ and m_A. Fig. 5.2a and Fig. 5.2b show the branching fractions (colour-coded) for the m_h^{mod+} and hMSSM scenario.



(a) m_h^{mod+} (b) hMSSM

Figure 5.2.: Branching fractions of $H \to WW$ in the m_h^{mod+} and hMSSM scenario as a function of $\tan\beta$ and m_A.

The following sections describe the event selection of the $H \to WW$ process, the new deep learning strategy, the systematic uncertainties and the final statistical analysis.

5.1. Event selection

As aforementioned the W bosons decay into an electron and neutrino and a muon and neutrino as sketched in the Feynman diagram in Fig. 5.3. The three signatures in the detector of this final state are reconstructed in CMS by the particle flow algorithm [28]. Additional requirements are set in order to select muons and electrons originating from W decays. Those will be summarised in the following sections. More detailed information about the event selection and trigger efficiencies can be found in the CMS analysis note "CMS-AN-17-082" [29].

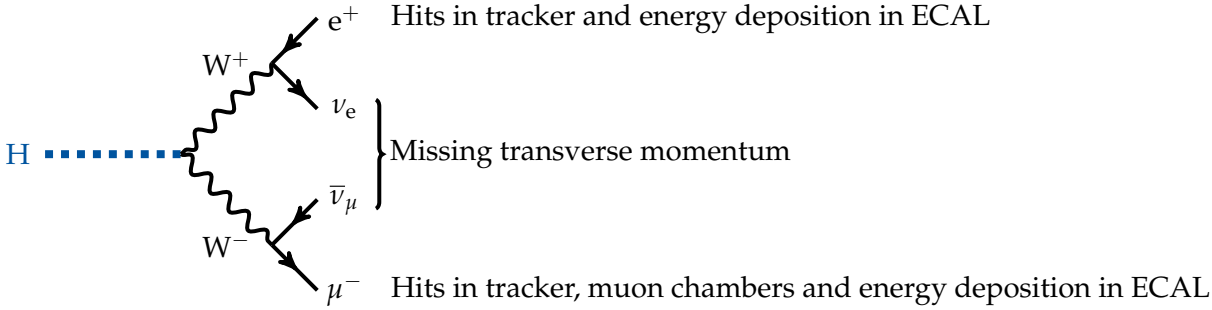


Figure 5.3.: Feynman diagram of the full $H \to W^+W^- \to e^+\mu^-\nu_e\bar{\nu}_\mu$ process and the signatures in the detector of each final state particle after reconstruction.

5.1.1. Trigger

In the $H \to WW \to e\mu\nu_e\nu_\mu$ analysis a combination of single and double triggers is used. The triggers for the signal selection using data measured in 2016 corresponding to an integrated luminosity of $\mathcal{L} = 35.9\,\text{fb}^{-1}$ can be found in table 5.1.

Dataset	Trigger path	Run range
SingleMuon	HLT_IsoMu22	Run2016B - Run2016H
	HLT_IsoTkMu22	
SingleElectron	HLT_Ele27_eta2p1_WPLoose_Gsf	
	HLT_Ele45_WPLoose_Gsf	
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	
DoubleEG	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	
MuonEG	HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL	
	HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	

Table 5.1.: HLT trigger paths for the different datasets recorded in 2016.

5.1.2. Muon selection

Muons are the only detectable particles, which reach the muon chambers. Muons leave a track in the tracker, energy deposition in the ECAL and HCAL, although they do not loose their entire energy, and hits in the muon chambers. Thus one can define two different kinds of muons in the reconstruction step: "Tracker Muons", which are reconstructed mainly by the tracking detector, and "Global Muons", which are reconstructed using the information of all subdetectors. A tight selection is required with the following cuts for muons.

- Reconstructed by particle flow as a "Global Muon"
- χ^2/degrees of freedom of the track fit < 10
- At least one hit in the muon chambers of the global track fit
- At least one hit in the pixel tracking subdetector
- At least five tracker layer hits (for a good p_T measurement)
- $p_T > 10\,\text{GeV}$
- $|\eta| < 2.4$
- $d_z < 0.1\,\text{cm}$ with respect to the primary vertex (d_z denotes track impact parameter in z-direction)
- $d_{xy} < 0.01\,\text{cm}$ for $p_T < 20\,\text{GeV}$ and $d_{xy} < 0.02\,\text{cm}$ for $p_T > 20\,\text{GeV}$ with respect to the primary vertex (d_{xy} denotes the track impact parameter in the transverse plane)

Additionally a tight particle flow based isolation with $\Delta\beta$ correction (for pile-up mitigation) is applied in a cone size of $\Delta R < 0.4$ in order to ensure that these muons do not originate from high mass quark decays. The cut on the isolation is: $ISO_{tight} < 0.15$.
The muon selection results in an efficiency of $95 - 99\%$ in almost all bins of p_T and η.

5.1.3. Electron selection

Electrons are completely stopped in the ECAL and thus will not reach the HCAL and the muon chambers. Their signature in the detector consists of a track in the pixel and strip detector and their full energy deposition in the ECAL. The electromagnetic showers in the ECAL come from Bremsstrahlung, pair-production and Compton scattering. Photon conversion, jets faking electrons and electrons from semileptonic b or c quark decays are the main background sources.
Electromagnetic variables serve for this identification with high discrimination power. The discriminating variables are constructed using the ratio of the energy deposition in the ECAL and HCAL, the information of the reconstructed supercluster by particle flow, isolation with a cone size of $\Delta R = 0.3$, tracking quality and the photon conversion veto. In the end a cut-based tight working point with an efficiency of $\approx 70\%$ is used for the analysis.

5.1.4. Missing transverse momentum

Two of the four final state particles are neutrinos. Neutrinos escape the detector completely without leaving a signature. They carry no electric charge, which means that they do not leave a track in the pixel and strip subdetectors. Furthermore neutrinos are stable in the scope of the SM, thus they do not produce secondary (detectable) particles by a decay. Last but not least the interaction probability of neutrinos with matter is extremely small. So far they are the only known particles, which escape the detector. Therefore the negative vectorial sum of transverse momenta of all reconstructed particle flow candidates gives the full transverse momentum of the missing neutrinos.
Unfortunately there are a lot of uninteresting sources for missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) like detector noise, cosmics and beam-halo scraping. Thus several algorithms are used to identify false $E_{\mathrm{T}}^{\mathrm{miss}}$.
In presence of pile-up the resolution of $E_{\mathrm{T}}^{\mathrm{miss}}$ gets worse very fast due to the additional energy from pile-up events. Therefore a dedicated algorithm, the so-called Pileup Per Particle Identification (PUPPI) [30] algorithm, is used to calculate $E_{\mathrm{T}}^{\mathrm{miss}}$.

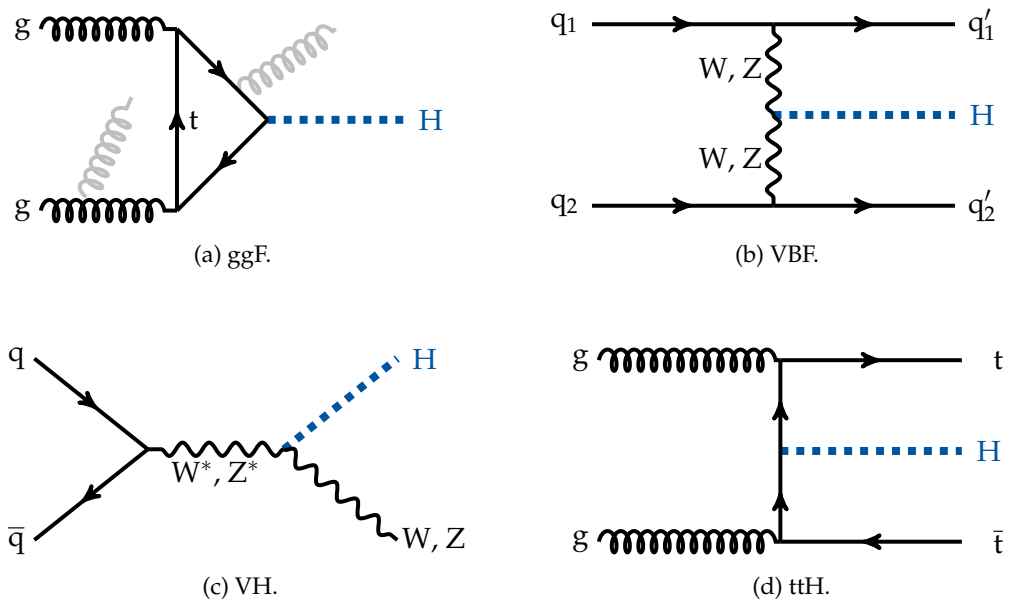
5.1.5. Jets and b quarks

Jets are used in this analysis for the event categorisation. They are reconstructed with the anti-kt algorithm [31] with a distance parameter of 0.4. Additionally a jet is required to pass the loose cut-based working point of the particle flow jet identification and a baseline selection of $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 5.0$.
Jets originating from b quarks have a secondary vertex distinguishable from the primary vertex. This effect is used to discriminate jets originating from b quarks from soft jets (from u, d or s quarks). Furthermore the kinematics of the decay products of B mesons is different, because of the high b quark mass. These informations are combined in a dedicated multivariate algorithm, called "cMVAv2" [32], which serves as a b-tagging algorithm. A loose working point is applied and a cut on the transverse momentum of the jet of 20 GeV.

5.2. Signal processes

A Higgs boson can be produced by four different production modes, which can be seen in Fig. 5.4. This analysis focusses only on two production modes: gluon-gluon fusion (ggF) and vector boson fusion (VBF).
The ggF process, sketched in Fig. 5.4a, is the most dominant production process at LHC. Two

gluons produce a top loop, which couples to the Higgs boson. The loop can be produced by any quark, but it was shown before, that the fermion mass is proportional to the coupling to the Higgs boson. This makes the heaviest fermion the most probable to participate in this loop: the top quark. As shown in the Feynman diagram additional jets can be radiated from the gluons or the loop causing a phenomenon called initial state radiation (ISR). The radiation of jets boosts the Higgs boson, which can be exploited in Higgs searches, since it enhances the transverse momentum of the Higgs boson and thus also the transverse momentum of the final state particles. This allows the electron and muon to pass the trigger thresholds. Unfortunately the cross section of the Higgs boson production including gluon radiation is reduced by α_s (strong coupling constant) for each additional vertex.
The second most prominent process is the VBF production, sketched in Fig. 5.4b. In comparison to the ggF production the cross-section is approximately a factor of 10 smaller for the SM Higgs boson. Two quarks from the protons radiate weak gauge bosons, which annihilate and produce a Higgs boson. The two quarks are then scattered into opposite directions, leading to a very clear signature of two forward jets and a very high invariant di-jet mass. For higher Higgs boson masses the VBF cross section rises in comparison to the ggF production mode and therefore plays a more important role in the high mass search.
The other two production modes are very small in comparison to the ggF and VBF mode. Higgsstrahlung (VH), shown in Fig. 5.4c, requires a quark and an anti-quark to annihilate. Since the LHC is a proton-proton collider, the anti-quark has to be a sea quark, which reduces the cross section of this mode. Higgsstrahlung would be the most prominent production mode at a proton anti-proton collider like Tevatron. Last but not least there is the quark associated Higgs production (ttH), shown in Fig. 5.4d. This process is suppressed, since it requires high energies to produce two top quarks in combination with a Higgs boson in the final state. Worth to mention here is the recent observation of this production mode at CMS and ATLAS [33].



(a) ggF. (b) VBF. (c) VH. (d) ttH.

Figure 5.4.: Feynman diagrams of the four main Higgs production modes at LHC.

5.3. Background processes

5.3.1. Non-resonant WW background

The most iconic background of the high mass H → WW search is the non-resonant WW background. Feynman diagrams of this background can be found in Fig. 5.5. The only difference to the signal are the kinematics of the two weak gauge bosons. Two processes are merged in this background: The WW → $2l2\nu$ at next-to-leading order (NLO) and the WW plus 2 jets at leading order (LO) process. The latter process contains two jets from either a two quark system or a quark and gluon system. The overlap of the merge is cleaned by using the former sample only for $m_{jj} < 100\,\text{GeV}$ (on generator level), and the latter one for $m_{jj} > 100\,\text{GeV}$ (on generator level).
The Monte Carlo samples with their corresponding branching fraction times cross section for this analysis can be found in table 5.2.

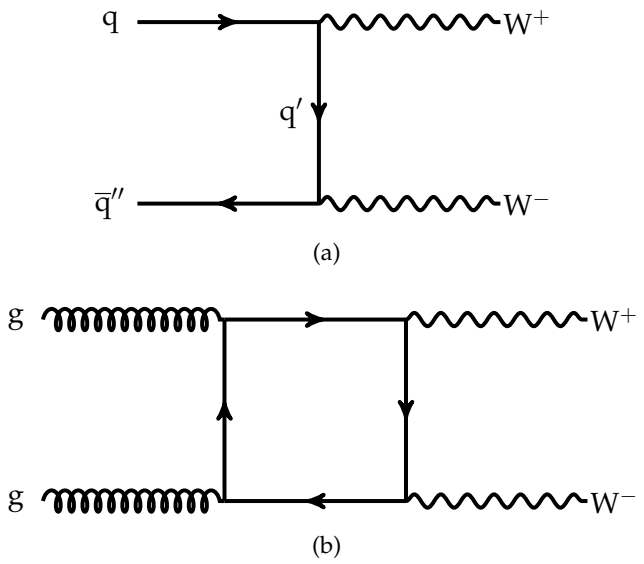


Figure 5.5.: Feynman diagram of the non-resonant WW background process at proton proton collisions.

Sample name	$\mathcal{BR} \times \sigma$ [pb]
WWTo2L2Nu_13TeV-powheg	12.178
WpWmJJ-QCD-noTop_13TeV-powheg	2.423
GluGluWWTo2L2Nu_MCFM_13TeV	0.5905

Table 5.2.: Used non-resonant WW Monte Carlo samples with the corresponding $\mathcal{BR} \times \sigma$ [pb] [34].

The shape of the kinematic distributions is fully estimated from MC simulation. The normalisation is left unconstrained in the final fit.

5.3.2. Top

The largest background arises from $t\bar{t}$ and single top production. The main $t\bar{t}$ processes can be found in Fig. 5.6 and the single top processes in Fig. 5.7. Top quarks decay mostly to a

bottom quark a W boson. The b quarks in the final state create a very unique jet topology, which is described in more detail in section 5.1.5. A dedicated b-tagging algorithm helps to suppress these backgrounds by applying a cut on the number b-tagged jets.
In this analysis a dedicated control region is constructed in order to obtain the normalisation of this background, which allows to fit an enriched and signal free control region to data. The shape of the kinematic distributions is taken from Monte Carlo simulation.

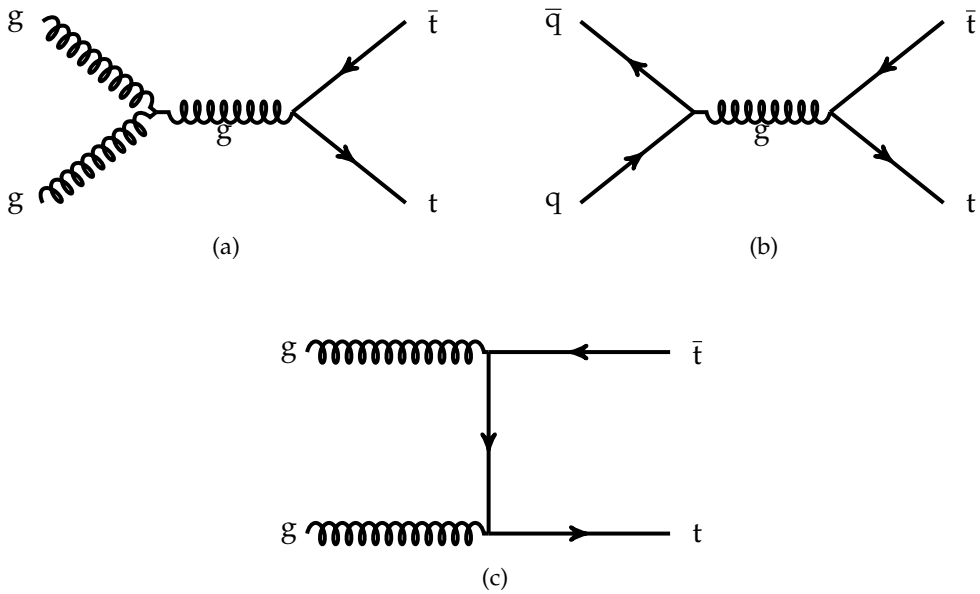


Figure 5.6.: Feynman diagrams of the three main $t\bar{t}$ production modes at LHC.

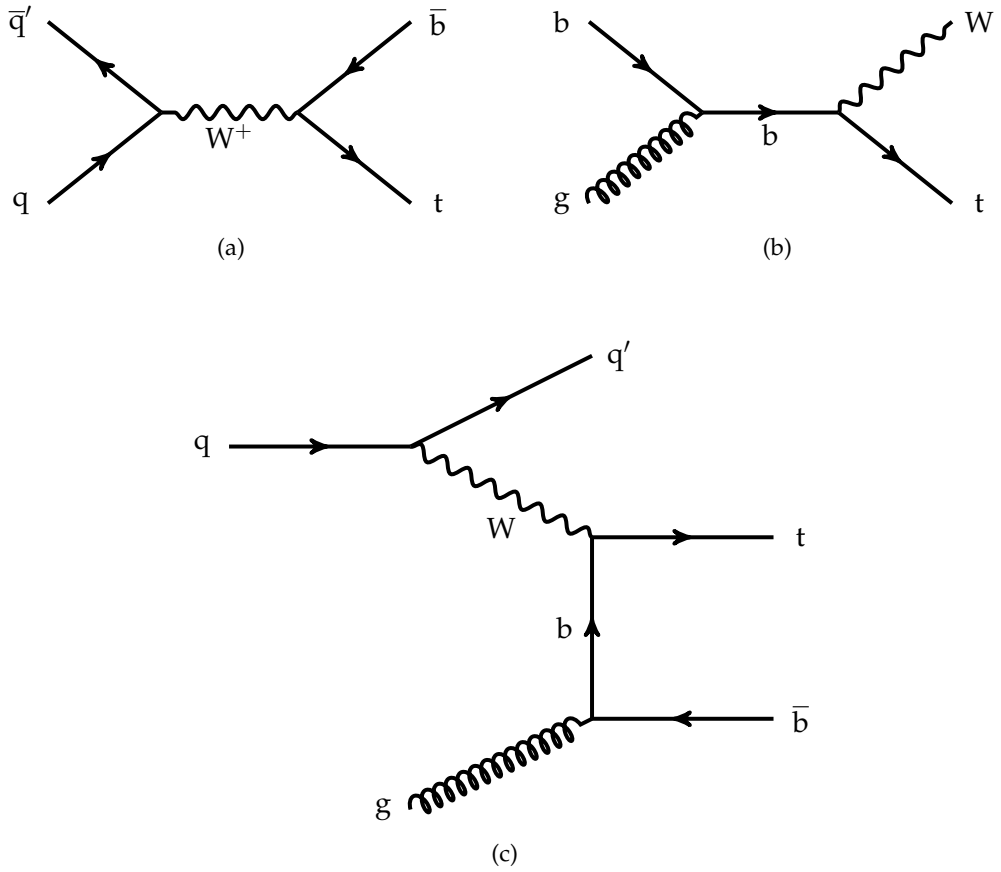


Figure 5.7.: Feynman diagrams of the three main single top production modes at LHC.

The Monte Carlo samples with their corresponding branching fraction times cross section can be found in table 5.3.

Sample name	$\mathcal{BR} \times \sigma$ [pb]
TTTo2L2Nu_13TeV-powheg	87.31
ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8 TuneCUETP8M1	35.85
ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8 TuneCUETP8M1	35.85

Table 5.3.: Used $t\bar{t}$ and single top Monte Carlo samples with the corresponding $\mathcal{BR} \times \sigma$ [pb] [34].

5.3.3. Drell-Yan

Another large background for this analysis is the $Z/\gamma^* \to \tau^+\tau^-$ Drell-Yan process, where tau leptons decay leptonically to an electron or a muon and two additional neutrinos each. The main difference to the signal process are the four additional neutrinos. The same flavour Drell-Yan processes $Z/\gamma^* \to \mu^+\mu^-$ and $Z/\gamma^* \to e^+e^-$ are heavily suppressed by choosing the opposite flavour final state, since the misidentification probability of muons and electrons is very small.
At LHC Z bosons are produced by quark anti-quark annihilation with very small transverse momentum at tree-level. The tree-level Feynman diagram of the $Z/\gamma^* \to \tau^+\tau^-$ background

can be seen in Fig. 5.8.

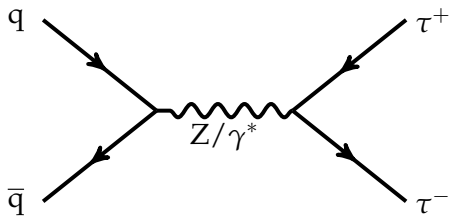


Figure 5.8.: Feynman diagram of the Drell-Yan background process at proton proton collisions.

In this analysis the normalisation of the Drell-Yan process is measured by a fit to data in a signal free and $Z \rightarrow \tau^+\tau^-$ enriched control region. The shape of this background is taken from Monte Carlo simulation. Since the transverse momentum is mismodeled by the MC@NLO [35, 36] Monte Carlo generator, events are reweighted accordingly. The samples used for the Drell-Yan background can be found in table 5.4.

Sample name	$\mathcal{BR} \times$ [pb]
DYJetsToTauTau_ForcedMuEleDecay_M-50_TuneCUETP8M1_13TeV	1867
DYJetsToLL_M-50 TuneCUETP8M1_13TeV	6025.26

Table 5.4.: Used Drell-Yan Monte Carlo samples with the corresponding $\mathcal{BR} \times \sigma$ [pb] [34] generated by MC@NLO.

5.3.4. W+jets

Another background arises from W+jets events when a jet is misidentified as a lepton. Two Feynman diagrams of the W+jets process are sketched in Fig. 5.9.

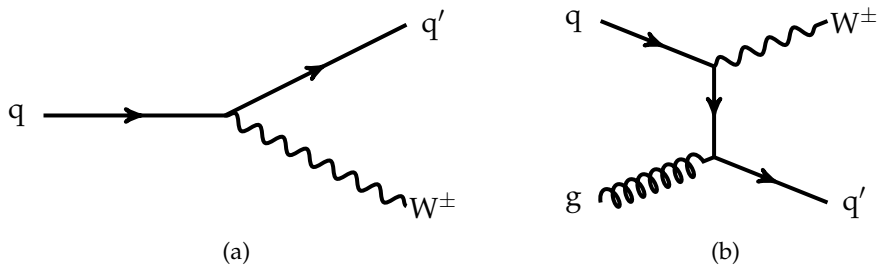


Figure 5.9.: Feynman diagrams of the W+jets background process at proton proton collisions.

The jet is misidentified as a lepton and the W decays to a real lepton and a neutrino, thus mimicking the desired final state. This background is critical for low Higgs masses. It is estimated using a data-driven method, which allows to measure the yield and the kinematic distributions of this background. The basic idea of this estimation is rather simple: a region (orthogonal to the signal region) in data is selected, where jets are faking leptons. This process enriched region is used to propagate the expected yield from the background in the

signal region.
Systematic uncertainties are introduced by using this method and discussed in section 5.8.

5.3.5. Other backgrounds

Apart from the non-resonant WW background other di-boson processes can be misidentified as the desired final state. The di-boson processes can be divided into two categories: a vector boson in combination with a photon and a vector boson in combination with a Z boson. The contribution of these processes is rather small, but non-negligible. The shape and normalisation of the di-boson backgrounds are taken fully from Monte Carlo simulation.
Another background process is the SM Higgs boson decaying into two W bosons or τ leptons and further into the opposite flavour final state of this analysis. Gluonfusion, vector boson fusion and Higgsstrahlung are considered as production modes for the SM Higgs boson. The shape and normalisation of the SM Higgs boson in all three production modes are fully estimated from Monte Carlo simulation for $H(125) \rightarrow WW$ and $H(125) \rightarrow \tau\tau$ decays.
The triple boson background is the smallest among all backgrounds. All combinations of three weak gauge bosons are taken into account, neglecting any tri-boson processes with photons. The shape and normalisation of the triple boson background is also fully taken from Monte Carlo simulation.

5.4. Analysis strategy using deep learning

5.4.1. Deep learning

In the last years deep learning techniques became more and more popular in computer science due to their huge and rapid success in speech recognition and image processing. Recently these methods are gaining more and more attention in modern particle physics with the most recent participation in the observation of the top quark associated Higgs boson production [33].
A very similar strategy with deep learning techniques is used in this analysis: an event-by-event process classification.
The idea of deep learning is to add more than one hidden layer to an artificial neural network and thus learn more features with these hidden layers in the data. It is important to understand the principle of artificial neural networks for this analysis. An artificial neural network is inspired by the structure of a human brain. Its basic structure is made up by layers, which contain neurones and connections between the neurones of different layers. Neurones are nothing else than placeholders for numbers and the connections are operations to change these numbers from one layer to the next one. Fig. 5.10 shows an example neural network for a binary classification task (two output nodes).

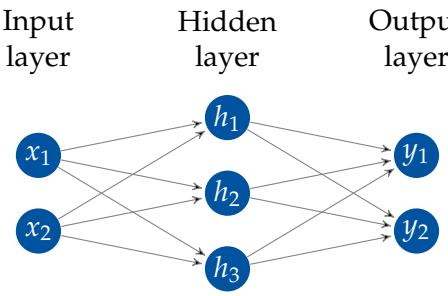


Figure 5.10.: Sketch of an artificial neural network with two input nodes, three nodes in the hidden layer and two output nodes.

The mathematical description of this neural network is rather simple and can be easily explained step by step. The connections between two layers is performed mathematically by a multiplication of a weight w and the addition of a bias b. Consequently the connection of the input layer to the hidden layer can be written as:

$$\underbrace{\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}}_{\text{Hidden layer } h} = \underbrace{\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{pmatrix}}_{\text{Weights } W} \times \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\text{Input layer } x} + \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}}_{\text{Biases } b}. \tag{5.1}$$

In the same way the connection from the hidden layer to the output layer is described by the following equation:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}}_{\text{Output layer } y} = \underbrace{\begin{pmatrix} w'_{11} & w'_{12} & w'_{13} \\ w'_{21} & w'_{22} & w'_{23} \end{pmatrix}}_{\text{Weights } W'} \times \underbrace{\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}}_{\text{Hidden layer } h} + \underbrace{\begin{pmatrix} b'_1 \\ b'_2 \end{pmatrix}}_{\text{Biases } b'}. \tag{5.2}$$

Inserting equation 5.1 into equation 5.2 leads to a full description of the neural network from the input layer to the output layer. The neural network in its mathematical representation then reads:

$$y = W'(Wx + b) + b' \tag{5.3}$$
$$= W'Wx + W'b + b' \tag{5.4}$$
$$= W''x + b''. \tag{5.5}$$

Equation 5.5 shows that the full representation of the neural network is still a linear transformation of the input data. Unfortunately real world problems are not linear. Therefore a trick is applied by wrapping the output of each layer into a so-called activation function. These activation functions are non-linear and thus perform a non-linear mapping of each layer. Three simple activation functions are: ReLU, sigmoid and tanh:

$$\text{ReLU}: \quad \sigma(x) = \max(0, x), \tag{5.6}$$
$$\text{sigmoid}: \quad \sigma(x) = \frac{1}{1 + e^{-x}}, \tag{5.7}$$
$$\tanh: \quad \sigma(x) = \frac{e^{+2x} - 1}{e^{-2x} + 1}. \tag{5.8}$$

Now the only transition to a deep neural network (DNN) is to add more hidden layers. This is sketched in Fig. 5.11.

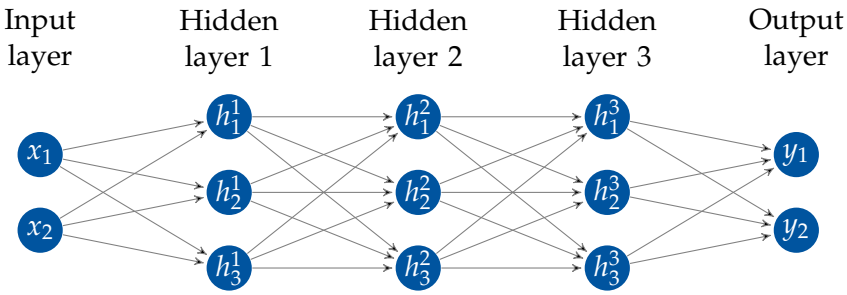


Figure 5.11.: Sketch of a DNN with two input nodes, three nodes in each hidden layer and two output nodes.

The additional hidden layers allow the DNN to learn and exploit more complex correlations and inner representations of the data.
During the so-called training step, all of these weights and biases are changed and tuned in a way that the real world problem is described in the best way. This is achieved in multiple optimisation iterations. First of all the weights and biases are initialised with certain values (e.g. randomised). Now the data will be propagated through the network once. At the end a loss (or objective) function (e.g. mean squared error) calculates the difference between the real output and the predicted output. Then the backpropagation algorithm [37] is used to calculate the gradient, which is applied by an optimisation algorithm in order to tune the weights and biases such that the objective function finds its global minimum. This procedure is performed in several iterations until the objective function converges and no further improvement in optimisation is achieved.

5.4.2. Deep learning in H → WW

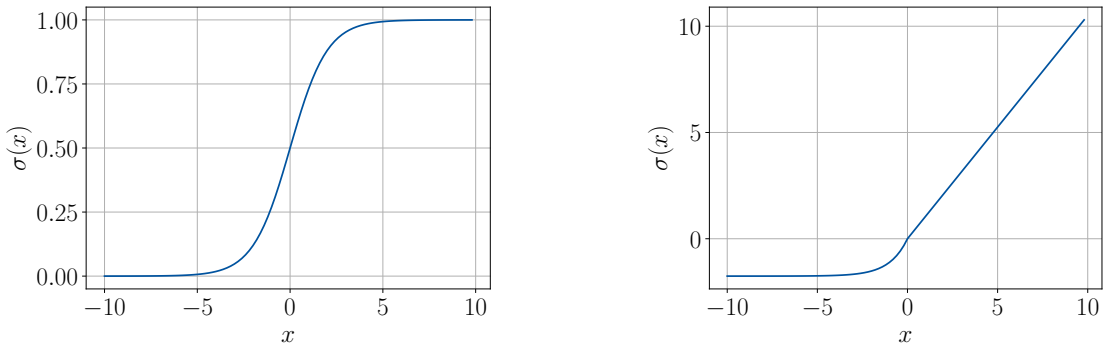
A deep neural network is built with the Keras library [38] using TensorFlow [39] backend. A fully connected network is chosen with 32 nodes in the input layer followed up by 10 hidden layer with 64 nodes each and finally ending in an output layer with 10 nodes. The number of input layer nodes corresponds to the training variables, whereas the number of nodes of the output layer corresponds to the number of classes (processes). Each layer has a weight initialisation based on the LeCun normal initialiser [40], which is necessary to set the first set of numbers of weights and biases and thus kick off the training procedure. This specific weight initialisation is a necessary feature for self-normalising neural networks [41]. After each layer an activation function is applied, which decides whether an artificial neurone fires: a scaled exponential linear unit activation (SELU) for the input and the hidden layer and a softmax activation for the output layer. The formulas for both activation functions can be seen in the following equation:

$$\text{softmax}: \quad \sigma(x_j) = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}}, \tag{5.9}$$

$$\text{SELU}: \quad \sigma(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases}, \tag{5.10}$$

where the index j corresponds to a certain output node and K denotes the total number of

output nodes. The SELU activation contains two empirical parameters, which are determined to be $\lambda \approx 1.0507$ and $\alpha \approx 1.6732$. Both activation functions are shown in Fig. 5.12.

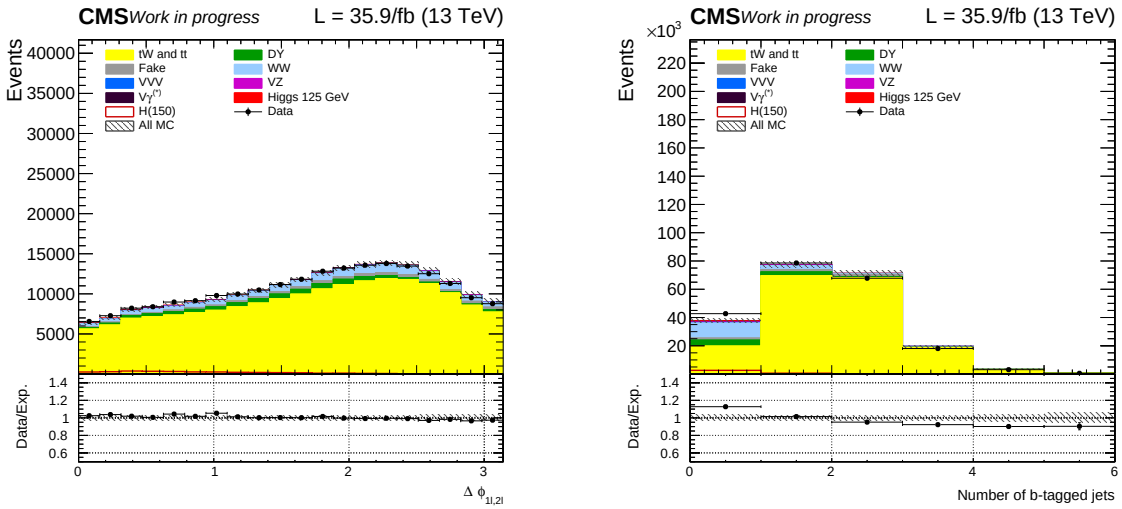


(a) Softmax activation function

(b) SELU activation function

Figure 5.12.: Sketches of the two non-linear activation functions used in the DNN approach of the H $\rightarrow$ WW analysis.

The softmax activation for the output layer is very important to ensure mutual exclusive classes. The output in each node of the output layer will be squashed into a range between 0 and 1. Furthermore the softmax activation has the property that the sum of all outputs of the output layer is equal to 1. Therefore the output can be interpreted as a probability, which is exploited for the event classification. The SELU activation function is very newly developed by Günter Klambauer, Thomas Unterthiner, Andreas Mayr and Sepp Hochreiter [41]. The idea is basically that the neurone activations of self-normalising neural networks automatically converge towards zero mean and unit variance without using algorithms with high computational costs (e.g. batch normalisation). The benefit of this self-normalisation after each layer is that the values of the training features remain in comparable ranges, it also avoids getting stuck in local minima during the training step and decreases the computation time of the whole training procedure.
The Adam optimiser [42] is chosen for the optimisation algorithm due to its performance robustness and task flexibility for optimising the cross entropy objective function, which is the most common objective function for a multi-class classification problem.
Additionally overtraining effects are reduced by the early stopping callback from Keras, which ensures that the training procedure stops as soon as the network overtrains.
Low-level variables are chosen as training variables to exploit their correlations and hence allow the network to learn from the very basic features. This ensures the largest degree of freedom during the training procedure. In addition to that high-level quantities are added based on their discrimination power. All training variables are and have to be well described in data and simulation. Before feeding these training variables into the neural network, all variables are transformed to a gaussian shape by removing the mean and scaling to unit variance. This effect is preserved by the SELU activation function as aforementioned, while propagating the data through the network. An example is the $\Delta\phi_{ll}$ distribution, which shows that signal is peaking near 0 and the background more to higher values, shown in Fig. 5.13a. Another example is the number of b-tagged jets, where none are expected for signal and at least one for the top background, shown in Fig. 5.13b.



(a) $\Delta\phi_{ll}$ distribution.

(b) Number of b-tagged jets distribution.

Figure 5.13.: Example of two training variables: $\Delta\phi_{ll}$ (left) and number of b-tagged jets (right) distribution.

The signal shape of the $\Delta\phi_{ll}$ distribution comes from the fact that the Higgs boson is a scalar, hence the azimuthal angle between the two leptons almost vanishes. This variable serves as a good discriminator for signal against other background processes. Fig. 5.13b clearly shows, that most top background is populated in the second and third bin. The other processes are expected to have nearly no b-tagged jets in the final state. Therefore this variable discriminates very good between top background events and other processes. All training variables can be found in an inclusive selection in the appendix A and a full list of (low- and high-level) variables is given below:

Angular and spatial information:

- $\Delta\eta_{jj}$
- $\Delta\phi_{l,j}$ for all combinations of leading/trailing lepton and leading/trailing jet
- $\Delta\phi_{ll}$
- $\Delta\phi_{l,E_{\mathrm{T}}^{\mathrm{miss}}}$ for leading and trailing lepton
- $\Delta\phi_{j,E_{\mathrm{T}}^{\mathrm{miss}}}$ for leading and trailing jet
- $\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$
- $\min(\Delta\phi_{l,E_{\mathrm{T}}^{\mathrm{miss}}})$
- ΔR_{ll}

Kinematic information:

- m_j for leading and trailing jet
- m_{jj}
- m_{ll}
- Collinear mass

- $m_{T,H}$
- $m_{T,i}$
- $m_{T,W}$ for leading and trailing W boson
- $p_{T,\mathrm{WW}}$
- $p_{T,ll}$
- $p_{T,j}$ for leading and trailing jet
- Missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed by PUPPI

Other:

- Number of b-tagged jets
- Number of jets
- cMVAv2 score for leading and trailing jet

The final output classes are corresponding to real world physics processes as aforementioned. The following list enumerates the chosen output nodes of the DNN architecture:

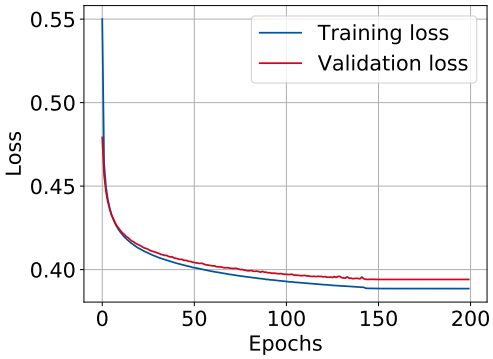
- 3 ggF signal classes: low, medium and high mass
- 3 VBF signal classes: low, medium and high mass
- Drell-Yan
- Top
- Non-resonant WW
- Misc

In total there are 6 signal classes, depending on the production mode and the mass regime. The low mass regime contains signal masses below 200 GeV, the medium mass regime between 200 and 500 GeV and the high mass regime 500 to 3000 GeV. Additionally the three main backgrounds, namely the Drell-Yan, Top and non-resonant WW process are assigned to an individual output node. Last but not least the remaining smaller backgrounds are merged together in a so-called "Misc" class, which allows the DNN during training to classify an event into this class instead of falsely classifying it and furthermore provides more degrees of freedom in the classification.
In order to ensure a statistically unbiased training and evaluation, all MC is split into a training subsample and an evaluation subsample, which in principle only allows then to use the evaluation subsample for further analysis. This is very inconvenient, because the full amount of events for the statistical analysis is unusable. Therefore all MC events are splitted by even (fold 0) and uneven (fold 1) event numbers. Then two DNN are trained on these subsamples independently and evaluated afterwards on the other one. In the end the two folds are merged again together. This trick ensures, that no events are lost and that the training and evaluation step is not correlated. The two DNNs are trained with the following set of hyper parameters:

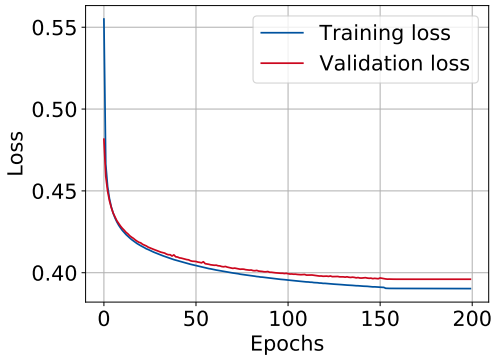
- Number of training epochs is 200,
- Learning rate is set to 0.000005,
- Batch size is chosen to be 1000.

The training performance of this multi-class classification problem is visualised by different

metrics. First of all the loss (objective) function is minimised during the training step. The minimisation as a function of the number of epochs can be found in Fig. 5.14a and Fig. 5.14b, which shows the expected behaviour for both folds.



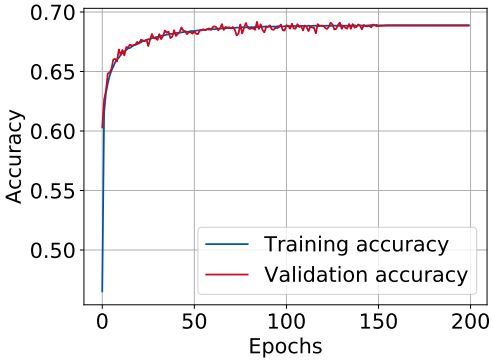
(a) Fold 0 loss.



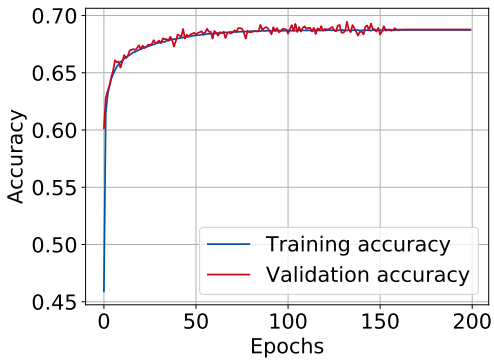
(b) Fold 1 loss.

Figure 5.14.: Loss as a function of the number of epochs. The trend of the loss function shows the expected minimisation behaviour.

Additionally the average accuracy to predict the classes correctly is evaluated with a small representative subsample, which is not used for training. Fig. 5.15 shows that the accuracy of correct predictions rises with the number of training iterations (epochs) for both folds. The rising accuracy in prediction success shows very clearly the learning process of the DNN.



(a) Fold 0 accuracy.

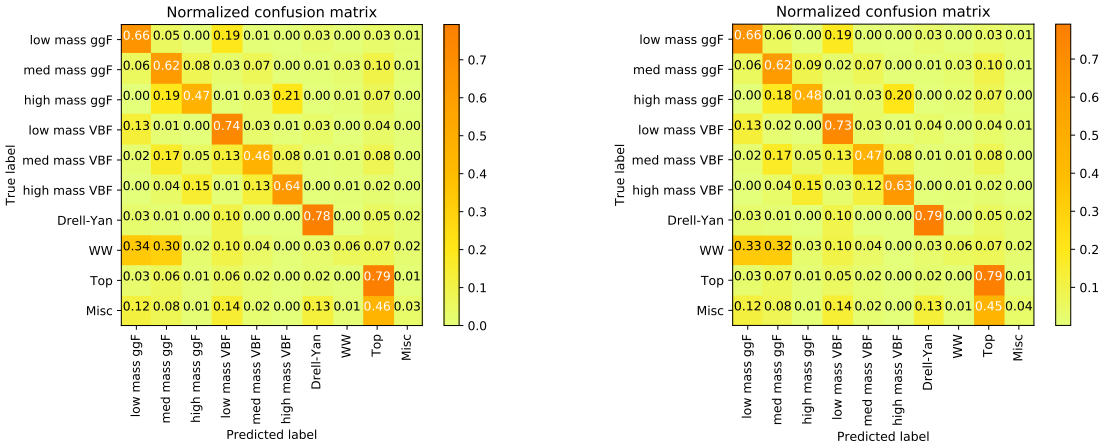


(b) Fold 1 accuracy.

Figure 5.15.: Average accuracy as a function of the number of epochs. The trend of the accuracy shows the expected learning behaviour.

Instead of the average accuracy the probabilities to predict one event correctly is shown in so-called confusion matrices. A confusion matrix shows on the y-axis the true class label, on the x-axis the predicted class label and on the z-axis the prediction probability. Fig. 5.16a and Fig. 5.16b show the prediction probabilities for both folds. The probabilities on the principal diagonal show the prediction probabilities for a correct classification, while off-diagonal entries are misclassification probabilities. The main background processes Drell-Yan and Top

have a very high prediction probability of around 80% to be correctly classified. The signal categories have high prediction probabilities between $\approx 45 - 75\%$. The most outstanding misclassification is the classification of a non-resonant WW event into the low and medium mass signal classes. This is expected, since the non-resonant WW process mimics the signal process as described in section 5.3 almost completely. Also the Misc class shows high off-diagonal entries, but this is also expected, since this class represents the merge of many different physics processes. Overall the confusion matrices look very promising, since it shows that the multi-class classification task was successfully solved by the DNN.



(a) Confusion matrix of fold 0.

(b) Confusion matrix of fold 1.

Figure 5.16.: Confusion matrix for both folds is shown. The perfect prediction of the DNN is an identity matrix. Overall the multi-class classification performance of one event propagating through the DNN looks very promising.

5.5. Discriminating variable

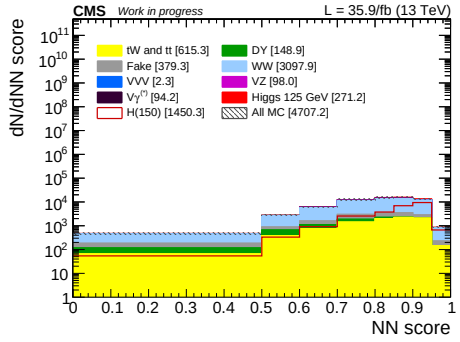
This analysis is not a counting experiment, but rather a shape analysis. Therefore a variable with high discrimination power between signal and background is needed. The most sensitive variable would be the reconstructed mass of the muon, electron and two neutrino system. Unfortunately this quantity is not possible to reconstruct due to the neutrinos. Hence a new variable is constructed, inspired by the classical definition of the transverse mass, denoted as improved transverse mass:

$$m_{\mathrm{T,i}} = \sqrt{\left(p_{ll} + E_{\mathrm{T}}^{\mathrm{miss}}\right)^2 + \left(p_{ll} + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\right)^2}. \tag{5.11}$$

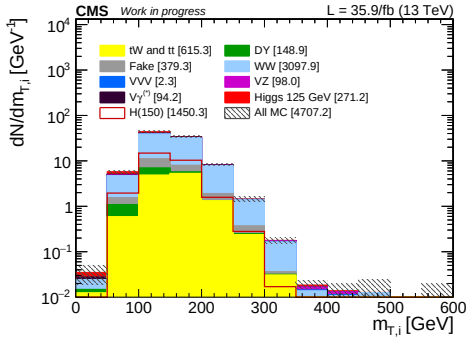
Additionally the DNN output prediction probability for each event in the three different signal mass regimes is used. The DNN probability for an event to be signal in a signal category is quite large already by definition, but it is larger for real signal than for background. This probability can be interpreted as a probability density function (PDF), which describes how signal-like an event in a signal category really is. Thus the constructed improved transverse mass is scaled by this PDF respectively for each mass regime. Three sensitive variables are built in the following way:

$$m_{\mathrm{T,i}}^{\mathrm{scaled}} = m_{\mathrm{T,i}} \cdot \text{DNN score}. \tag{5.12}$$

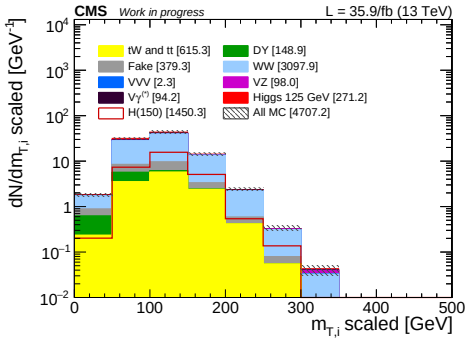
As an example the low mass 0 jet category is chosen to show the scaling procedure. Fig. 5.17a shows the DNN score, Fig. 5.17b the improved transverse mass distribution, and Fig. 5.17c finally the scaled transverse mass. The categorisation is described in detail in the section 5.6.



(a) Low mass 0 jet DNN score.



(b) Low mass 0 jet $m_{\mathrm{T,i}}$.



(c) Low mass 0 jet $m_{\mathrm{T,i}}$ scaled.

Figure 5.17.: Scaling of the improved transverse mass with the DNN score in the low mass 0 jet category.

5.6. Categorisation

Each event is propagated through the network, which is sketched in Fig. 5.18. Thus each event is assigned a certain probability to belong to a certain physical process. The highest probability determines then the process of the event. A single event can not land in two classes at the same time due to the choice of the last layer activation function, shown in Fig. 5.12a, which makes all classes mutual exclusive by definition. A baseline selection for different physics processes can be build with this strategy.
Furthermore basic cuts are required for all categories.

- Final state leptons are of opposite flavour and charge
- $m_{ll} > 12\,\text{GeV}$
- p_T of the leading lepton $> 25\,\text{GeV}$
- p_T of the trailing lepton $> 20\,\text{GeV}$
- p_T of a third lepton $< 10\,\text{GeV}$

The first requirement is directly given by the final state of this analysis. The remaining cuts on the transverse momentum assure that a third lepton of potential significance is rejected in an event and the other two remove low transverse momentum events, which survive the trigger logic, but do not carry new physics. The cut on the invariant dilepton mass rejects lower mass resonances.

The two different signal production modes are merged back together for each mass regime respectively after the classification of the DNN. In the end there are three signal classes and four background classes. These are further splitted into 0, 1, 2 jets and a VBF category, which results into twenty different categories in total. The VBF category is orthogonal to the 2 jets category by requiring an event to have $\Delta\eta_{jj} > 3.5$ and $m_{jj} > 500$, while the 2 jet category requires cuts of $\Delta\eta_{jj} \leq 3.5$ or $m_{jj} \leq 500$. This categorisation builds the baseline of the high mass $H \rightarrow WW$ search. The histograms of this categorisation can then be used for the statistical analysis in a simultaneous fit as described in section 5.7.

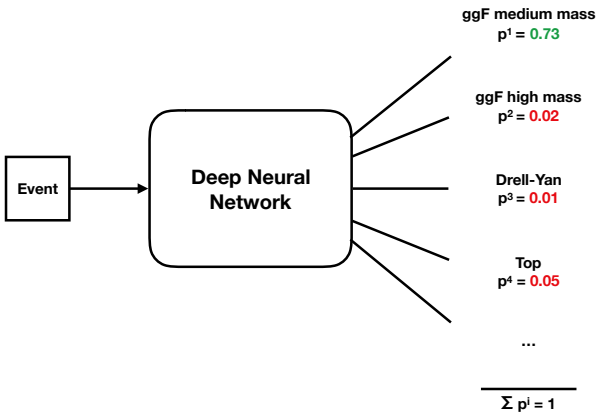
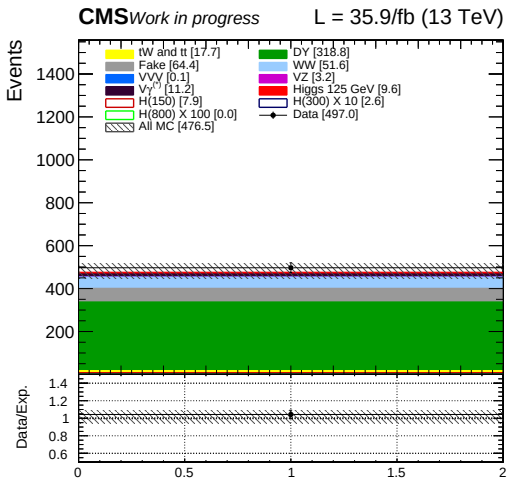
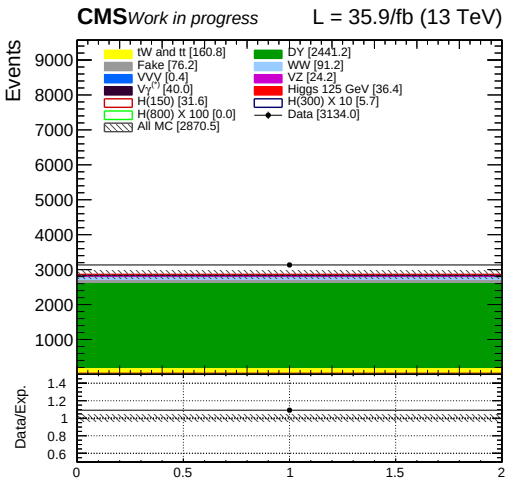


Figure 5.18.: Sketch of a single event propagating through the DNN. In this example the event will be classified into the ggF medium mass signal class.

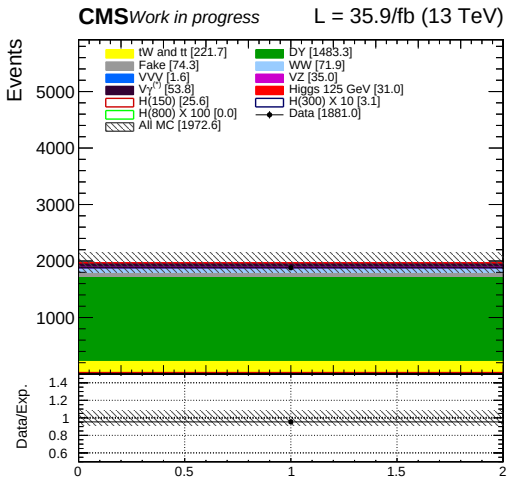
The Drell-Yan control regions (CR) are made up by the aforementioned DNN classification and jet multiplicity. Furthermore a b-jet veto is applied with a low working point. The resulting control regions can be found in Fig. 5.19 and Fig. 5.20 for each jet multiplicity respectively. They show Drell-Yan enriched categories with barely any contamination from other processes, hence validating the DNN categorisation strategy. The agreement between data and MC is very good, except of the VBF category, seen in Fig. 5.19d and Fig. 5.20d. Here the statistics is very low.



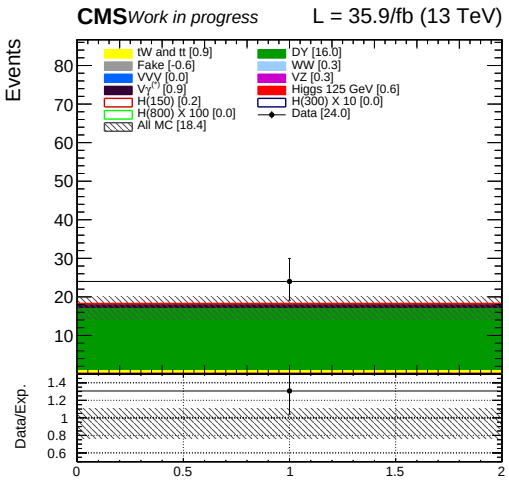
(a) Drell-Yan 0 jet.



(b) Drell-Yan 1 jet.

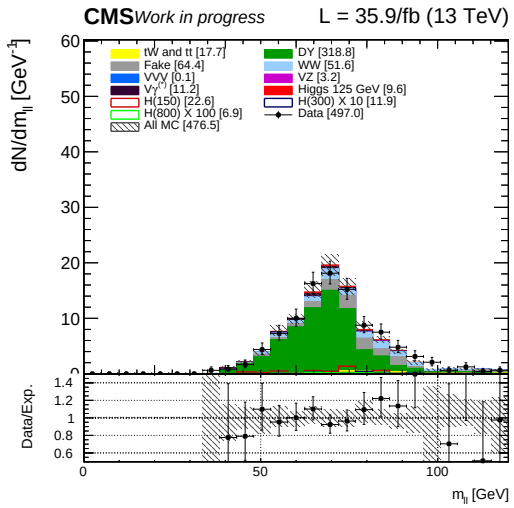


(c) Drell-Yan 2 jets.

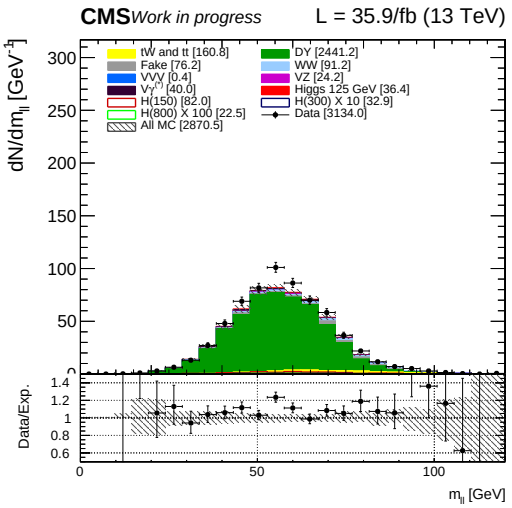


(d) Drell-Yan VBF.

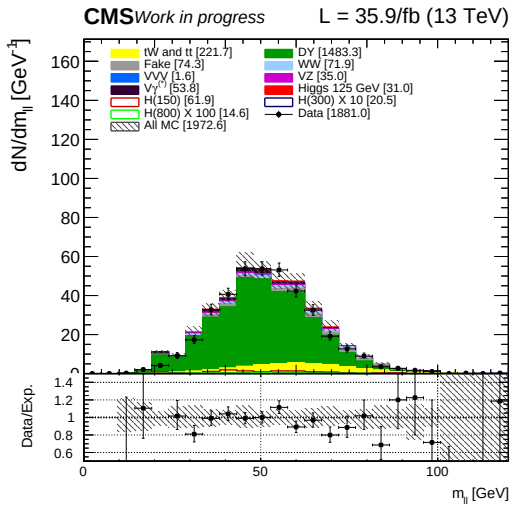
Figure 5.19.: Control regions of a one bin histogram for the four different Drell-Yan categories.



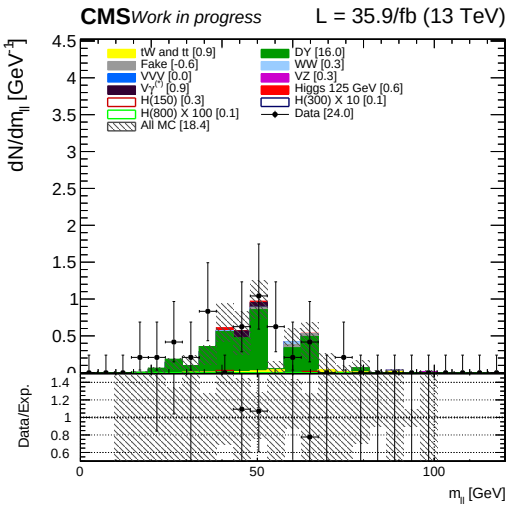
(a) Drell-Yan 0 jet.



(b) Drell-Yan 1 jet.



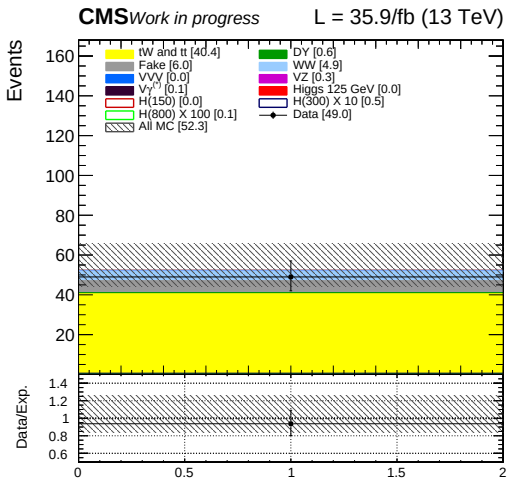
(c) Drell-Yan 2 jets.



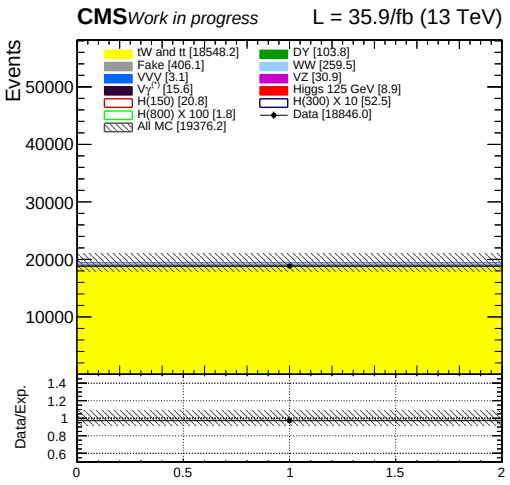
(d) Drell-Yan VBF.

Figure 5.20.: Control regions of the invariant dilepton mass for the four different Drell-Yan categories.

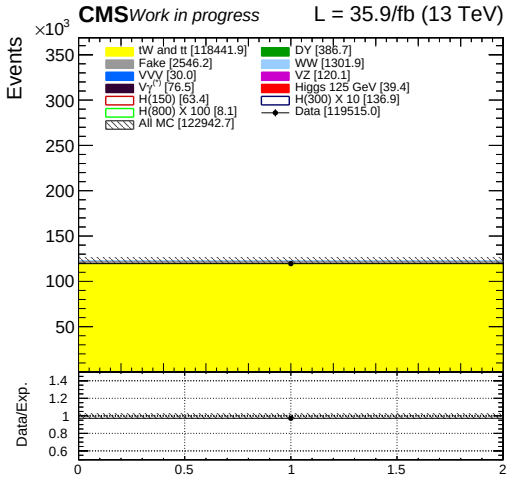
Similar to the Drell-Yan control region, a Top control region can be constructed by selecting the DNN Top output node. A very important difference to the Drell-Yan CR is to drop the b-jet veto. Fig. 5.21 and Fig. 5.22 show the Top CR for each jet multiplicity respectively. Also here top process enriched categories can be seen, thus also validating the DNN classification strategy. The agreement between data and MC is again very good. The purity of the 0 jet category is a bit worse than in the other categories, also lacking in statistics. This is expected, since the 0 jet top control region is very difficult to control.



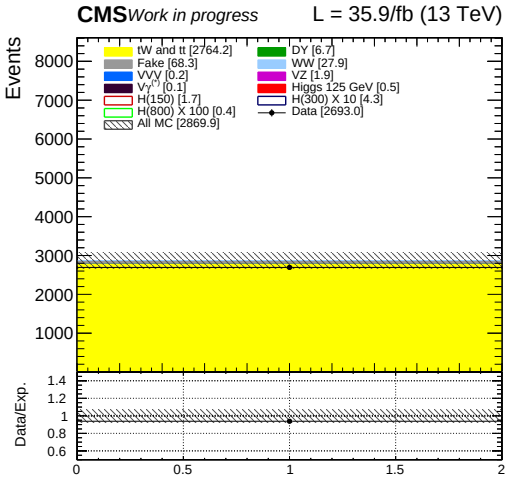
(a) Top 0 jet.



(b) Top 1 jet.

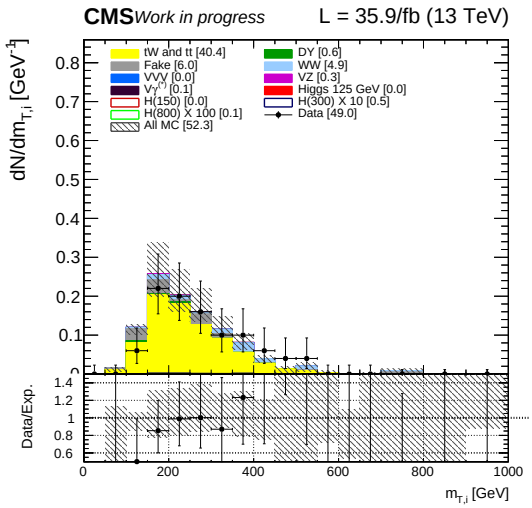


(c) Top 2 jets.

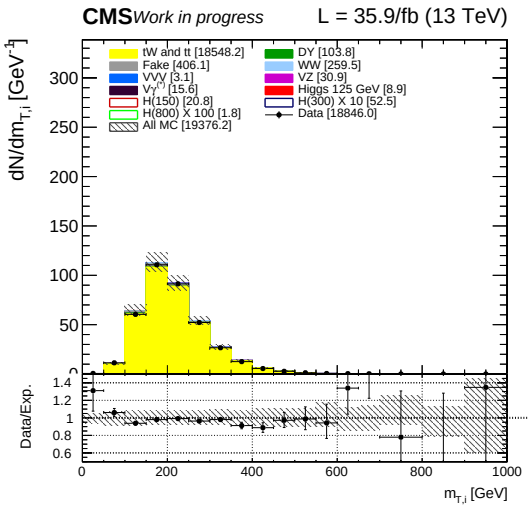


(d) Top VBF.

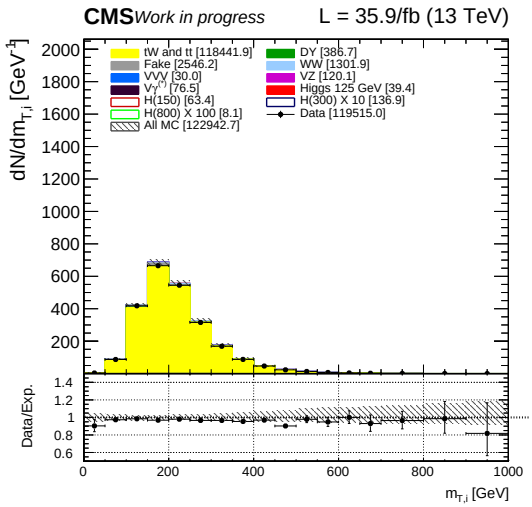
Figure 5.21.: Control regions of a one bin histogram for the four different Top categories.



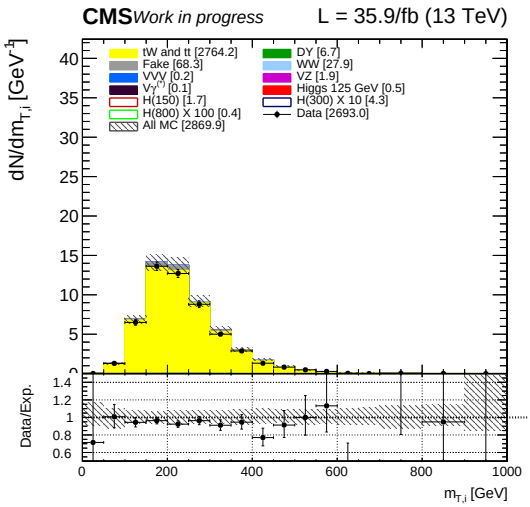
(a) Top 0 jet.



(b) Top 1 jet.



(c) Top 2 jets.



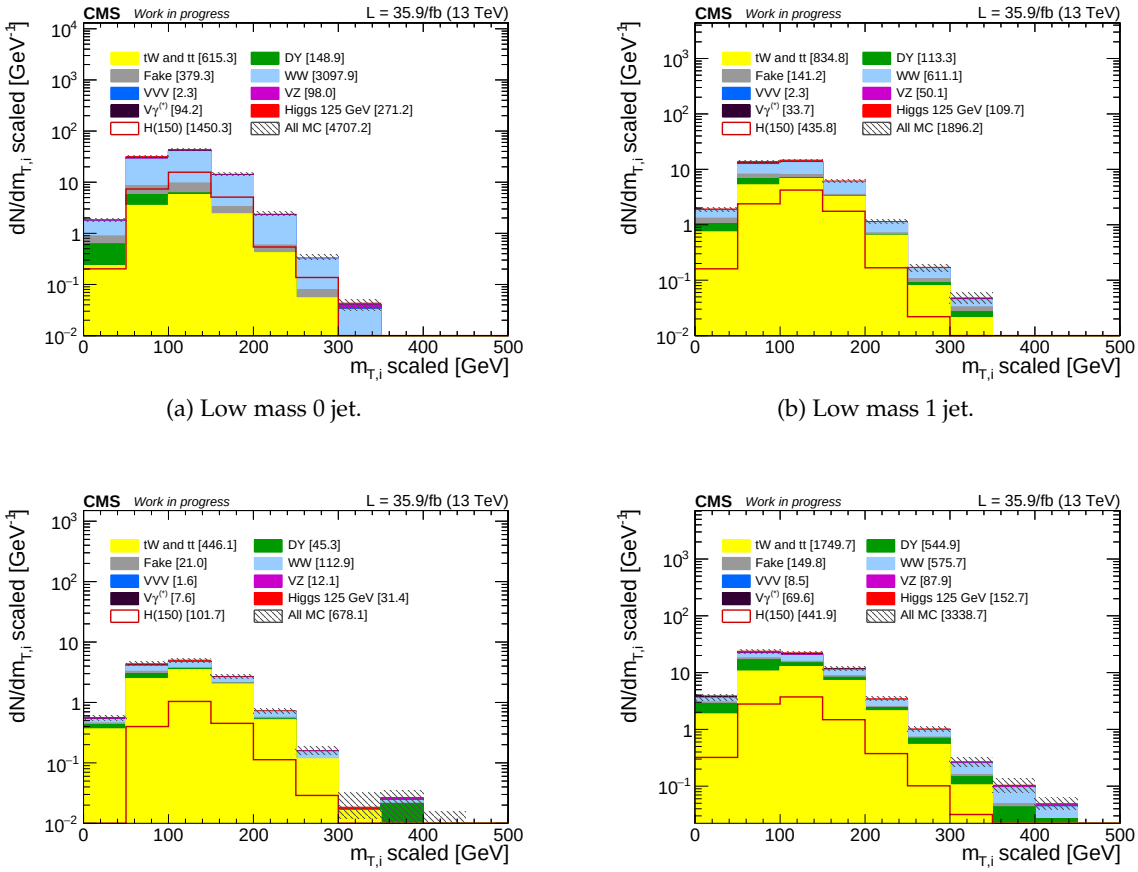
(d) Top VBF.

Figure 5.22.: Control regions of the improved transverse mass for the four different Top categories.

No other background control region is built, since the non-resonant WW background has too much signal contamination and the Misc class does not provide a CR for one certain process. The events of these two classes are further neglected in the analysis.

The remaining twelve signal classes are only different from each other in signal mass range and jet multiplicity. Again the corresponding output node of the DNN was chosen to build the following categories. A b-jet veto is applied on all signal regions.

Fig. 5.23 shows the scaled improved transverse mass distributions, which is described in more detail in section 5.5, in the low mass regime for the four different signal categories.



(a) Low mass 0 jet.

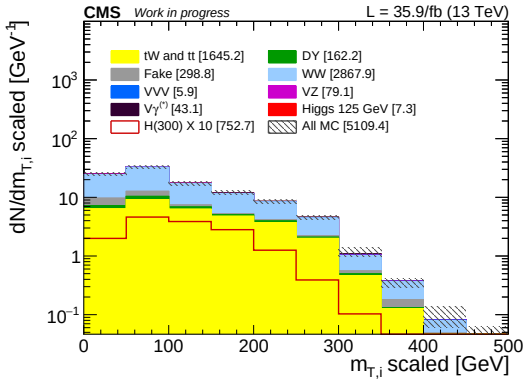
(b) Low mass 1 jet.

(c) Low mass 2 jets.

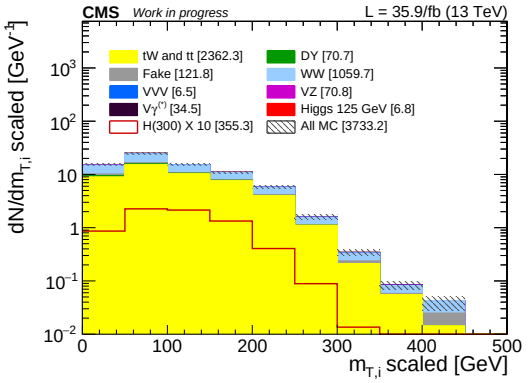
(d) Low mass VBF.

Figure 5.23.: Signal regions of the scaled improved transverse mass for the four different low mass signal categories.

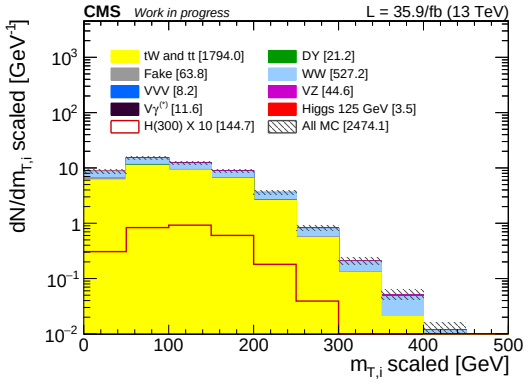
Fig. 5.24 shows the scaled improved transverse mass distributions in the medium mass regime for the four different signal categories.



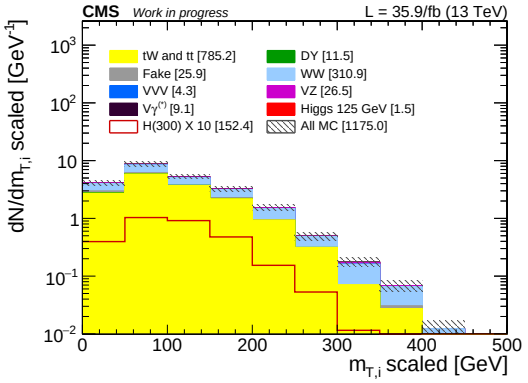
(a) Medium mass 0 jet.



(b) Medium mass 1 jet.



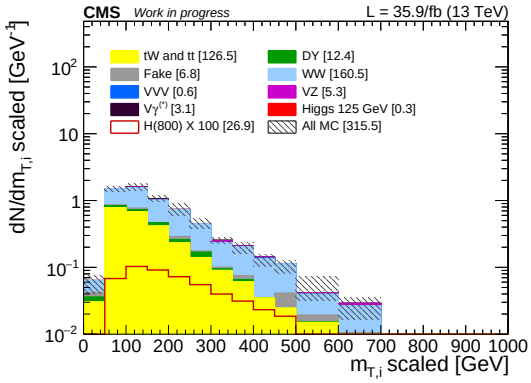
(c) Medium mass 2 jets.



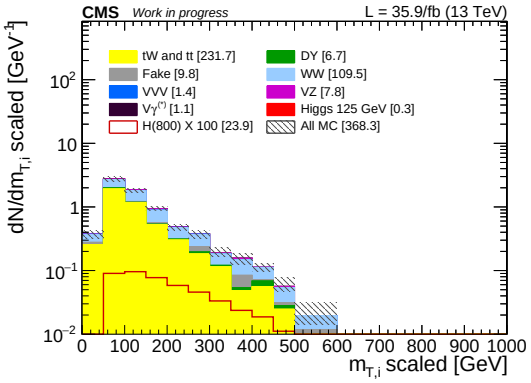
(d) Medium mass VBF.

Figure 5.24.: Signal regions of the scaled improved transverse mass for the four different medium mass signal categories.

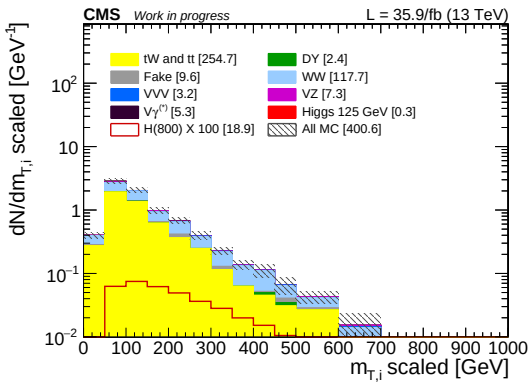
Fig. 5.25 shows the scaled improved transverse mass distributions in the high mass regime for the four different signal categories.



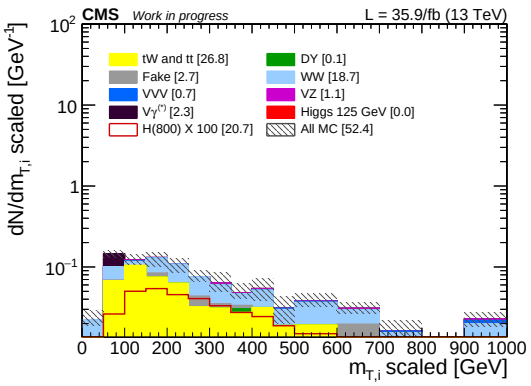
(a) High mass 0 jet.



(b) High mass 1 jet.



(c) High mass 2 jets.



(d) High mass VBF.

Figure 5.25.: Signal regions of the scaled improved transverse mass for the four different high mass signal categories.

5.7. Statistical inference

In order to quantify the performance of the analysis and compare it to previously published results, expected limits are computed based upon data corresponding to the published run range at $\sqrt{s} = 13\,\text{TeV}$ in the year 2016. During the time of the Higgs boson search a new method was developed by ATLAS and CMS, the so-called CL_s-method [43]. This method is implemented in the Combine software package, which is based on RooStats [44] and RooFit. A likelihood for the signal+background hypothesis and the background-only hypothesis is calculated. A signal can be excluded at a certain confidence level (usually chosen to be 95%), if the probability that an observed number of events fits to the signal+background hypothesis is sufficiently small.
Systematic uncertainties are introduced as nuisance parameters θ. The likelihood observing the data given the signal and background predictions then reads:

$$\mathcal{L}\left(n|\mu,\theta\right) = \prod_i \frac{\left(\mu s_i + b_i\right)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \cdot p\left(\tilde{\theta}|\theta\right), \tag{5.13}$$

where s_i and b_i correspond to the signal and background prediction of a certain bin respectively, n_i to the number of data events in this bin, μ to the signal strength multiplier, and the systematic uncertainties are introduced in the likelihood as probability density functions $p\left(\tilde{\theta}|\theta\right)$. The Poission distributions are multiplied for each bin i. With the likelihood the LHC test statistic $\tilde{q}_\mu$ can be constructed:

$$\tilde{q}_\mu = -2\ln\frac{\mathcal{L}\left(n|\mu,\tilde{\theta}_\mu\right)}{\mathcal{L}\left(n|\hat{\mu},\hat{\theta}\right)}, \tag{5.14}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the values for the maximised likelihood. The maximum likelihood value of the signal strength multiplier is further constrained to fulfil $0 \leq \hat{\mu} \leq \mu$. The nuisance parameter, which maximises the likelihood for a given signal strength multiplier, is given as $\tilde{\theta}_\mu$. Equation 5.14 transforms to the χ^2 distribution in the limit of large n. The probability to obtain a larger value of $\tilde{q}_\mu$ than the observed test statistics reads:

$$\text{CL}_{\text{s+b}}(\mu) = p\left(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{signal} + \text{background}\right) \tag{5.15}$$

$$\text{CL}_{\text{b}}(\mu) = p\left(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}}|\text{background}\right) \tag{5.16}$$

A signal strength can then safely be excluded with a certain confidence level (CL) α:

$$\text{CL}_{\text{s}}(\mu) = \frac{\text{CL}_{\text{s+b}}(\mu)}{\text{CL}_{\text{b}}(\mu)} \leq 1 - \alpha, \tag{5.17}$$

where α is commonly chosen to be 95% in high energy physics.
For the calculation of model-dependent exclusion limits the Tevatron (TEV) test statistic is used:

$$\tilde{q}_\mu = -2\ln\frac{\mathcal{L}\left(n|\mu,\tilde{\theta}_\mu\right)}{\mathcal{L}\left(n|0,\hat{\theta}\right)}. \tag{5.18}$$

The main difference between Eq. 5.18 to Eq. 5.14 is that the signal strength modifier is set to 0. It ensures, that the theory model is either excluded or not and nothing in between. In the end the same calculation for CL_s, see Eq. 5.17, is performed to calculate exclusions limits for the model-dependent scenarios.

5.8. Systematic uncertainties

Systematic uncertainties are introduced as nuisance parameters in the likelihood function. They can affect the normalisation and shape of the fitted distributions. Correlations between systematic uncertainties in different categories are taken into account.

The normalisation and kinematic shapes of the W+jets background are derived from a control region in data using the data driven method, which is briefly described in section 5.3. Propagating the systematic uncertainties of this method into the signal region yields an uncertainty for electrons (muons) of about 15% (3%).

For the top background normalisation additional categories are defined, which are simultaneously fitted with the signal region. These categories need to be enriched with top events and have only very little signal contamination. The normalisation of this background is correlated to the signal regions in the different jet multiplicities and the three different mass regimes. During the fit the normalisations of the top background for each category are left unconstrained. The same procedure is done for the normalisation of the Drell-Yan background.

Similarly the normalisation of the non-resonant WW background is left unconstrained in the fit in all categories.

The normalisation uncertainty of the $W\gamma^{(*)}$ is fixed to a conservative uncertainty of around 25%. A three lepton control region is built for the WZ and $Z\gamma^*$ background normalisation and yields constrained uncertainties of 6% and 17% respectively. This three lepton control region is taken from the SM $H \rightarrow WW$ analysis [45]. The ZZ background is treated with a normalisation uncertainty of 10%.

Other experimental uncertainties are calculated by smearing and scaling of certain variables and then a recalculation of all correlated variables. The following uncertainties are all treated as shape and normalisation uncertainties, except of the uncertainty of the luminosity, which is only treated as a systematic effect on the normalisation.
The systematic uncertainty of the luminosity is 2.5%. The trigger introduced a systematic uncertainty of around 2%, by varying the tag selection and Z window in the tag and probe calculation. Furthermore an uncertainty on the electron (muon) identification efficiency of around $2-5\%$ ($1-2\%$) is applied. The electron energy scale and muon momentum scale get an additional uncertainty due to different detector effects. They amount to the order of 0.2% for muons and $0.6-1\%$ for electrons in the signal region. In order to cover small disagreements for electrons with low transverse momentum or large eta a systematic uncertainty for the electron modelling is added. In this analysis the mismodelling of $E_{\mathrm{T}}^{\mathrm{miss}}$ has a systematic effect, which rises due to additional contributions from pile-up events and mismeasurements of individual particles. The jet energy scale uncertainties are also taken into account and amount to around 10% in the signal region. The scale factors correcting the b-tagging efficiency and mistagging rate introduce an additional nuisance, and thus are varied within their uncertainties.

Furthermore theoretical systematic uncertainties are used in the analysis coming from mismodelling of event generators or the missing knowledge on higher-order corrections. The mismodelling of the underlying events (UE) and the parton shower (PS) are calculated by

comparing two different event generators. This results in an impact on the expected on the signal yields of 5% (5%) for UE tuning and about 7% (10%) for parton shower description in the ggF (VBF) production mode. There are some theory uncertainties, which only affect certain backgrounds. The ratio between the single top and top pair process cross section is varied by the uncertainty on the ratio between their cross sections. The single top and top pair cross sections are respectively 71.70 pb and 831.76 pb, with relative uncertainties of 6% and 5% respectively, which leads to an uncertainty on the ratio of 8%. In order to reweight the LO gg $\rightarrow$ WW cross section to higher orders a k-factor with an uncertainty of around 15% is taken into account. The uncertainty related to missing higher order corrections of the qqWW process is modelled by varying the factorisation and renormalisation scale by a factor of 2. For the signal at different masses, uncertainties related to QCD and parton distribution function scales are included. They have a direct impact on the production cross section and acceptance of the signal process. Those uncertainties are taken from the Yellow Report 3 [46].

Finally, so-called bin-by-bin uncertainties are included arising from limited simulation statistics for each bin of the discriminant distributions in each category independently.

5.9. Signal interpretation

The signal model relies on two parameters: a scale factor of the couplings of the high mass resonance with respect to the SM, C', and the branching fraction to non-SM decays modes, BR_{new}. The modified signal strength and width can then be parametrised with these two parameters in the following way:

$$\mu' = C'^2 \left(1 - BR_{new}\right), \tag{5.19}$$

$$\Gamma' = \Gamma_{\mathrm{SM}} \frac{C'^2}{1 - BR_{new}}. \tag{5.20}$$

For this analysis C' is chosen to be 1 and BR_{new} to be 0.
This results in a non-negligible width Γ', which leads to interference effects between the ggWW continuum and the H(125) off-shell tail in the ggF production case and to interference effects between the high mass VBF signal, the WW plus two quarks and the VBF H(125) process in the VBF production mode. These interference effects have been taken into account in this analysis and the corresponding signal samples have been reweighted using the MELA package [47]. The yield has been adapted to prevent possible negative probability distribution functions arising from the interference. The new yield introduced in the final fit reads:

$$Yield = \sqrt{\mu} \times (S + B + I) + (\mu - \sqrt{\mu}) \times S + (1 - \sqrt{\mu}) \times B, \tag{5.21}$$

where S denotes the signal yield, B the background yield and I the interference yield.

Chapter 6

Results

The final binned fit is performed with the discriminating variable described in section 5.5. A comparison between prefit and postfit distributions can be found in the appendix B, where a signal mass hypothesis of $m_{\mathrm{H}} = 300\,\mathrm{GeV}$ and interference effects with the background are shown. Expected exclusion limits with a 95% confidence level are calculated in a model-independent and different model-dependent scenarios. The model-independent exclusion limits are calculated for several mass points between 200 GeV and 3000 GeV, while the model-dependent limits also include lower signal mass hypotheses. The performance of this analysis is compared with a cut-based analysis [34]. Fig. 6.1 shows the exclusion limits for the two different analysis strategies and the theory prediction for an additional resonance, namely a heavy $\mathcal{CP}$-even Higgs boson.

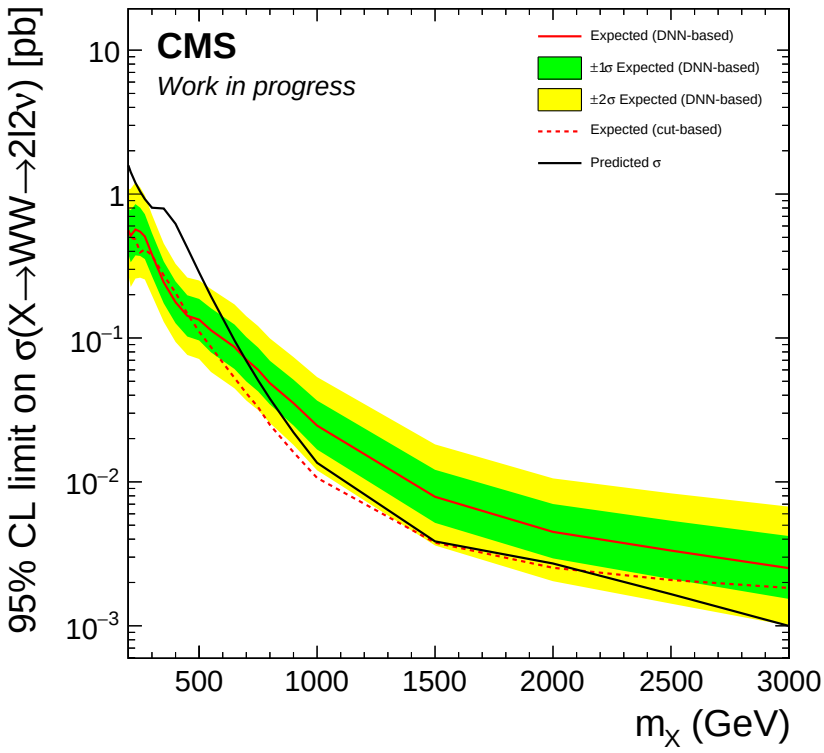


Figure 6.1.: Exclusion limit with 95% confidence level on the cross section of an additional heavy $\mathcal{CP}$-even Higgs boson for the DNN-based and cut-based analysis.

The DNN-based analysis, which is presented in this thesis, is sensitive to exclude a signal hypothesis up to a signal mass of around 700 GeV. The cut-based analysis is sensitive to signal masses up to 2000 GeV. It is important to note that the significance of a potential signal between $1000 - 2000$ GeV is extremely low for the cut-based, since the expected exclusion limit is almost identical with the theory prediction. In the lower mass regime both analysis approaches are very compatible, in some regions the DNN-based analysis outperforms the cut-based analysis. In the model-independent search the DNN-based analysis serves as a good cross-check analysis, delivering similar expected exclusion limits.

An easy explanation on why the cut-based analysis outperforms the DNN-based at higher masses is, that high mass signals can be easily isolated from background with basic cuts. Furthermore multiple DNN classes introduce misclassification probabilities, even for higher signal masses. The main idea of the DNN-based analysis is to gain more control over the SM backgrounds, which are also barely present at higher signal masses. Thus it is expected that the DNN-based analysis gains sensitivity especially for lower signal masses. This can already be seen in some regions in the model-independent expected exclusion limits below 500 GeV. The effect can even be quantified by looking at the total impact of the nuisances on the parameter of interest, which is the signal strength in this case. Fig. 6.2 shows the thirty most important nuisances ranked by their impact.

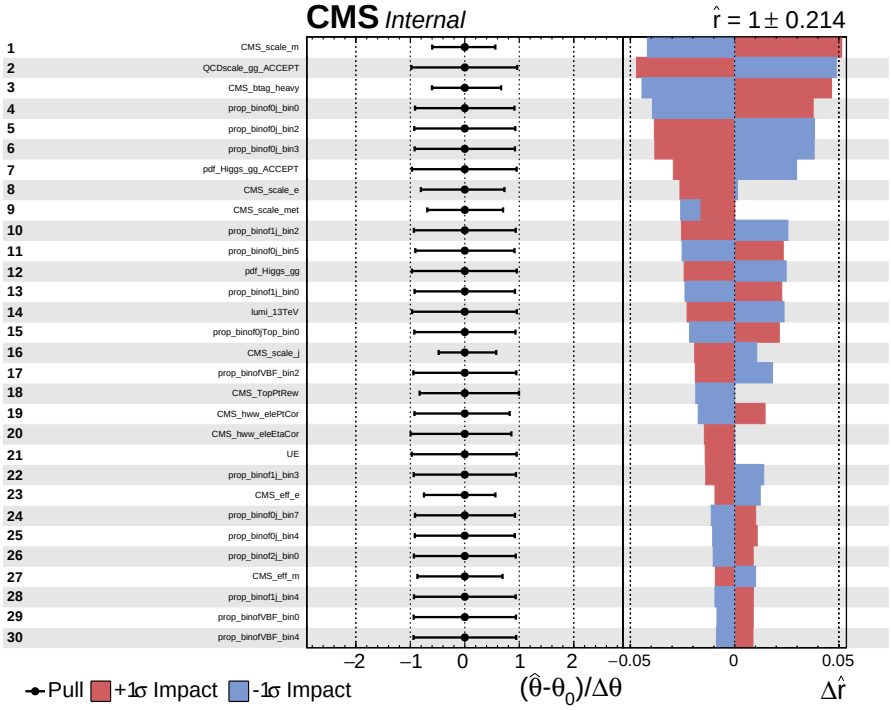


Figure 6.2.: Impacts and pulls of the main nuisances for a signal mass hypothesis of 300 GeV.

Additionally one can see in the uncertainty of the pulls of each nuisance, that all systematics have been well estimated and some are even constrained by the fit, since their assumed uncertainty was too conservatively estimated. The nuisance with the highest impact on the signal strength is ranked first, which is the muon momentum scale *CMS_scale_m*. Directly followed by higher order QCD corrections on the ggF signal *QCDscale_gg_ACCEPT* and b-

tagging efficiency *CMS_btag_heavy*. Their impact on the signal strength is a bit less than 5% respectively. Furthermore it can be seen that some nuisances have impacts in only one direction. This is well understood in this analysis, since the systematics introduced by underlying events *UE* and parton shower modelling *PS*, top p_T reweighting *CMS_TopPtRew* and large η electron corrections *CMS_hww_eleEtaCor* are only estimated in one direction. Sometimes this effect is also seen in nuisances, which have been constrained by the fit. This is visible for the $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainty *CMS_scale_met*. The signal strength sensitivity for a mass hypothesis of 300 GeV is extracted from the fit and measured to be $\hat{r} = 1 \pm 0.214$, whereas the cut-based analysis achieves a sensitivity of $\hat{r} = 1 \pm 0.237$ [34]. The uncertainty on this mass point is improved by around 11% by introducing the DNN-based analysis strategy for this certain mass hypothesis. Furthermore a lot of bin-by-bin uncertainties, marked by the *prop*-prefix, are found to have a huge impact on the final result. The pulls and impacts for a mass hypothesis of 500 GeV and 2000 GeV can be found in the appendix C.

Furthermore it was checked how much influence the systematic uncertainties have in the final fit. This is achieved by "freezing" all nuisances, which means that the fit has no handle anymore on these parameters and they are fixed. Fig. 6.3 shows the DNN-based expected exclusion limit with (dashed red line) and without (solid red line) freezing the systematic uncertainties.

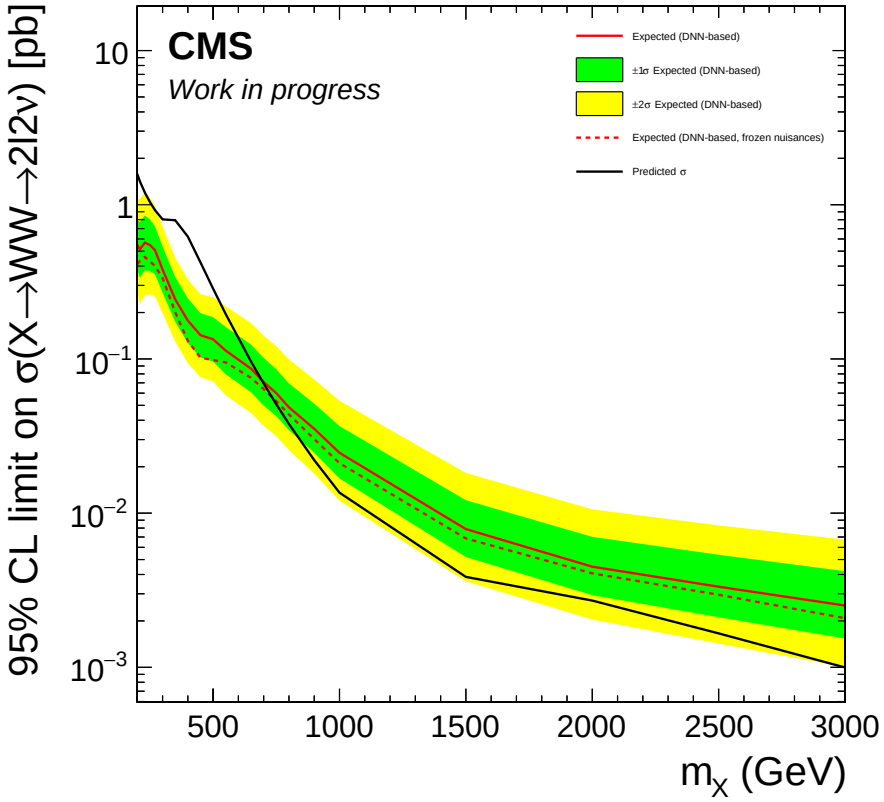
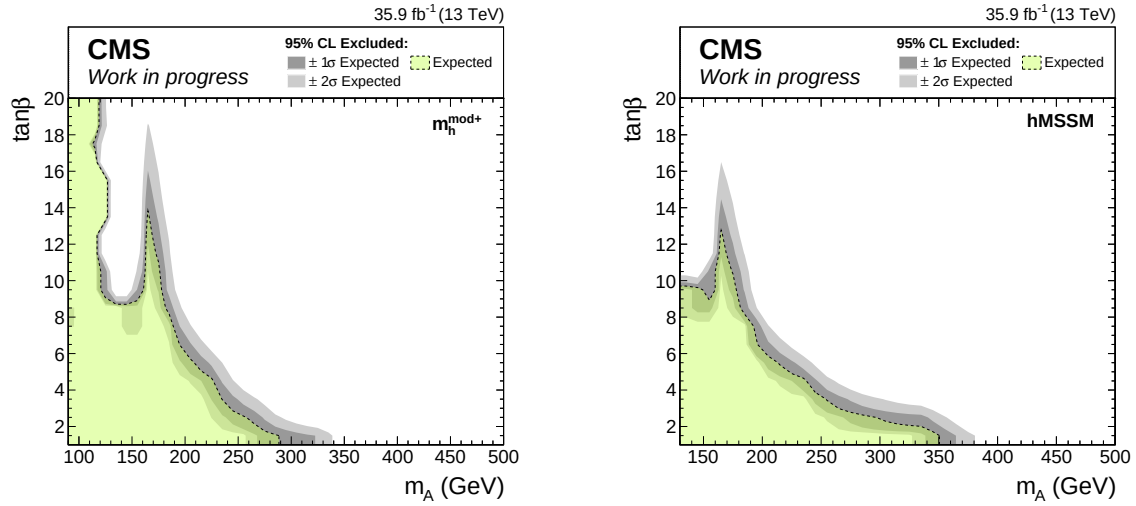


Figure 6.3.: Exclusion limit with 95% confidence level on the cross section of an additional heavy $\mathcal{CP}$-even Higgs boson with (dashed red line) and without (solid red line) frozen nuisances.

The difference in the expected exclusion limit with and without freezing all nuisances is on

average around 25%, while the maximum difference of 50% can be found at a mass hypothesis of $400 - 500\,\text{GeV}$.

Additionally expected exclusion limits in two MSSM scenarios are set. Here, expected exclusion limits are not determined as a function of the Higgs boson mass, but rather on the model itself. The parametrisation is described in section 2.2.2. The mass, cross-section and branching fraction of the heavy $\mathcal{CP}$-even Higgs boson is parametrised by $\tan\beta$ and m_A. Fig. 6.4a and Fig. 6.4b show the expected exclusion limits in the $m_\text{h}^{\text{mod}+}$ and hMSSM scenario at 95% confidence level.



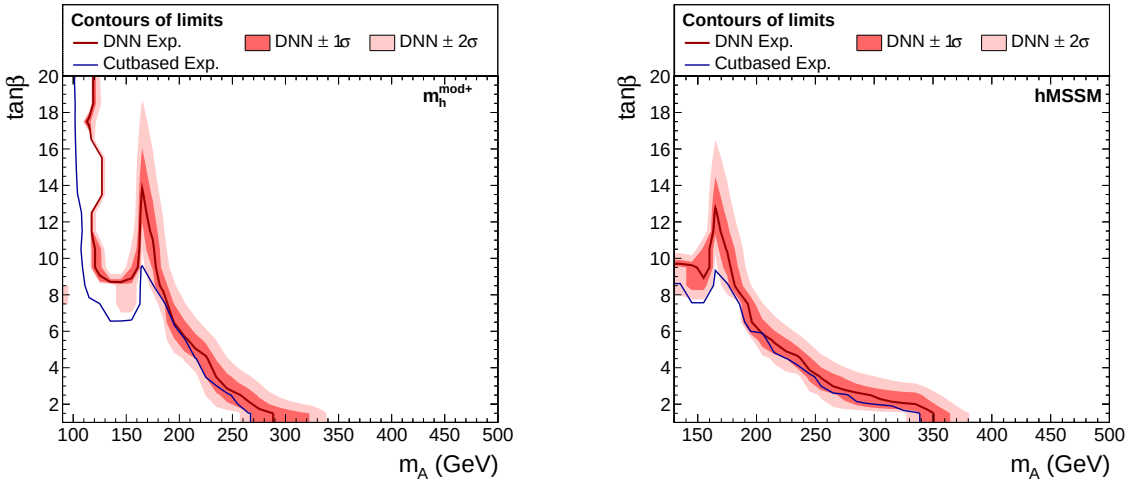
(a) $m_\text{h}^{\text{mod}+}$ scenario.

(b) hMSSM scenario.

Figure 6.4.: Expected exclusion limits at 95% confidence level in the $m_\text{h}^{\text{mod}+}$ (left) and hMSSM (right) scenario.

The expected excluded phase space is marked in green, while the 1σ and 2σ uncertainty bands are colored in grey. For each point in the two dimensional plane the corresponding mass hypothesis of the heavy $\mathcal{CP}$-even Higgs boson is used, see Eq. 2.51, to calculate the CL_s value of the benchmark scenario. Low values of $\tan\beta$ and m_A are expected to be excluded and thus complements the current $\text{h}, \text{H}, \text{A} \rightarrow \tau\tau$ MSSM search [48]. A peak signature at $m_A \approx 160\,\text{GeV}$ is noticeable. It comes from the decay to two on-shell W bosons at masses higher than two W bosons. This effect enhances the branching fraction significantly, which results in stronger expected exclusion limits.

The performance of this analysis in the two MSSM scenarios is again compared to the aforementioned cut-based analysis strategy. Fig 6.5a and Fig. 6.5b show the contours of both analysis strategies in each MSSM scenario.



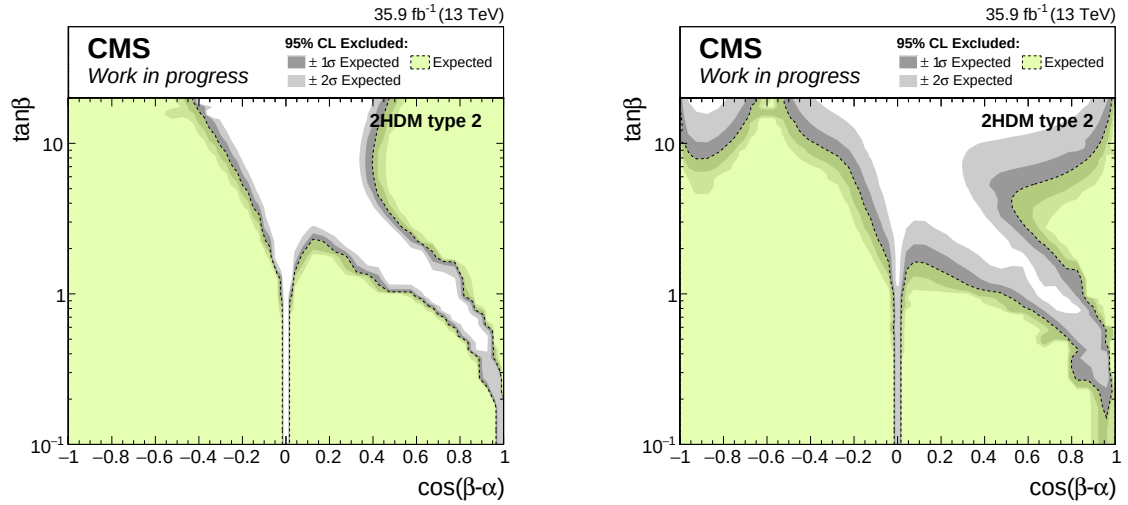
(a) m_h^{mod+} scenario.

(b) hMSSM scenario.

Figure 6.5.: Expected exclusion limits at 95% confidence level in the m_h^{mod+} (left) and hMSSM (right) scenario for the DNN-based (red) and the cut-based (blue) approach.

The red solid line shows the contour of the expected exclusion from Fig. 6.4a and Fig. 6.4b, while the blue solid line denotes the expected exclusion limit contour of the cut-based analysis. The DNN-based analysis shows an overall improvement in all regions of the two dimensional parameter space. The improvement is especially visible in the peak region, where the heavy Higgs boson decays to two on-shell W bosons. Here there is a relative improvement of $\approx$ 50% in the m_h^{mod+} scenario and $\approx$ 35% in the hMSSM scenario. Also the expected exclusion limit on m_A for very low values of $\tan\beta$ could be improved by using the DNN-based analysis strategy.

The MSSM is a special case of the more general 2HDM type 2 model. Therefore expected exclusion limits are calculated in the two dimensional parameter space spanned up by $\tan\beta$ and $\cos(\beta-\alpha)$, where α and β denote the mixing angles of the 2HDM, introduced in section 2.2.1. The remaining free parameters of the 2HDM are the two VEVs, the five Higgs boson masses and the mass parameter m_{12}. The light Higgs boson mass is fixed to 125 GeV, the VEVs are fixed by $\sqrt{v_1^2+v_2^2} = 246\,\text{GeV}$ and the mass parameter m_{12} is reparametrised to $m_{12} = \frac{m_A}{\sqrt{\sin\beta\cos\beta}}$. Fig. 6.6a and Fig. 6.6b show the expected exclusion limits for $m_A = m_H = m_{H^\pm} = 200\,\text{GeV}$ and $m_A = m_H = m_{H^\pm} = 300\,\text{GeV}$ respectively.



(a) 2HDM type 2 (200 GeV).

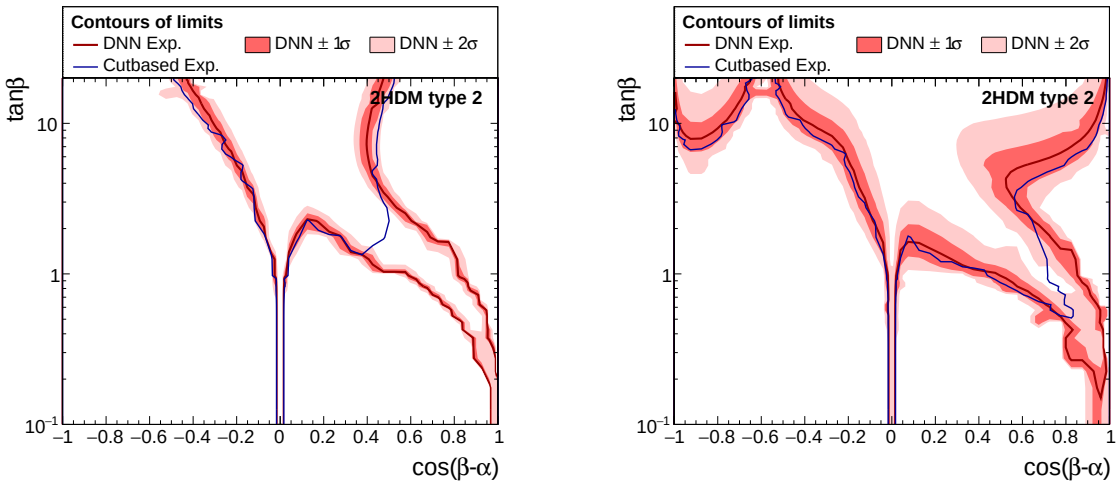
(b) 2HDM type 2 (300 GeV).

Figure 6.6.: Expected exclusion limits at 95% confidence level in the 2HDM type 2 model (left: free mass parameters fixed to 200 GeV; right: free mass parameters fixed to 300 GeV).

In both figures two remarkable regions can be found, which are not expected to be excluded. First of all there is the alignment limit of $\cos(\beta - \alpha) = 0$, where the mixing between the two Higgs doublets goes to 0 and the light Higgs boson becomes indistinguishable from the SM Higgs boson:

$$h_{SM} = h \sin(\beta - \alpha) + H \cos(\beta - \alpha). \tag{6.1}$$

The coupling of $H \to VV$ completely vanishes. The other region is the limit for $\alpha \to 0$. The coupling strength to up-type quarks is proportional to $\frac{\sin\alpha}{\sin\beta}$, hence vanishing for $\alpha \to 0$. Thus the only production modes in this region are the ggF production mode, driven by b quarks, and the VBF production mode, which reduces heavily the total production cross section of a heavy Higgs boson in this region. Furthermore a performance comparison to the cut-based analysis is done. Fig. 6.7a and Fig. 6.7b show the contours of the expected exclusion limits from Fig. 6.6 for the cut-based and DNN-based analysis.



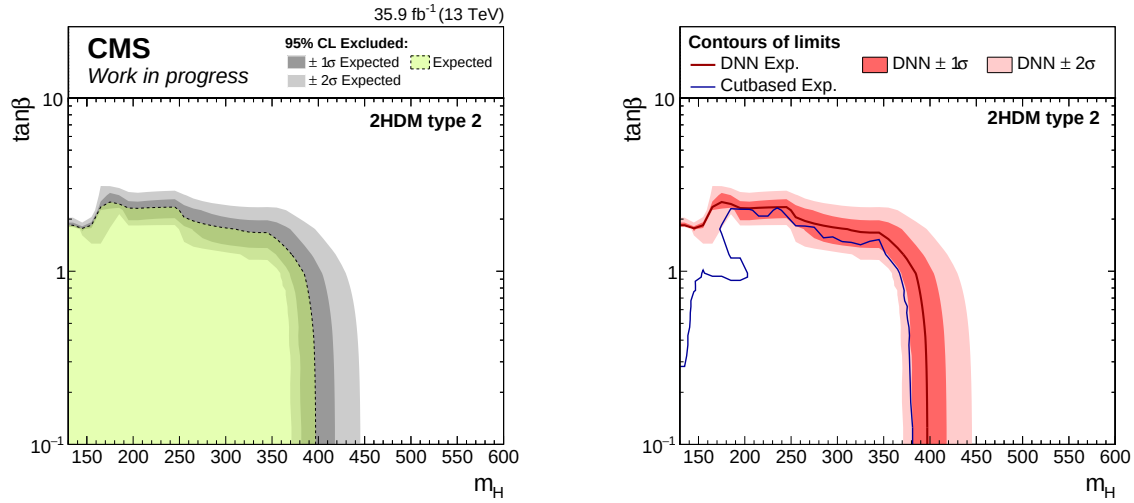
(a) 2HDM type 2 (200 GeV).

(b) 2HDM type 2 (300 GeV).

Figure 6.7.: Comparison of expected exclusion limits at 95% confidence level in the 2HDM type 2 model for the DNN-based (red) and cut-based (blue) analysis strategy (left: free mass parameters fixed to 200 GeV; right: free mass parameters fixed to 300 GeV).

The cut-based analysis shows a better expected exclusion in both figures in the decoupling limit, which shows, that the cut-based analysis is more sensitive to the VBF production mode. This gives a hint on why the DNN-based model-independent performs worse at higher mass hypothesis, since there the VBF production mode becomes the most relevant one. In all other regions of the 2HDM type 2 expected exclusion limits the DNN-based analysis performs better.

The 2HDM type 2 model can also be expressed by $\tan\beta$ and m_{H}. Here, $\cos(\beta-\alpha)$ is set to 0.1, the light Higgs boson mass and the VEVs are fixed in the same way as before and the remaining mass parameters are set to be equal to m_{H}. Fig. 6.8a and Fig. 6.8b show the expected exclusion limits of the reparametrised 2HDM type 2 model and the comparison to the cut-based analysis.



(a) Reparametrised 2HDM type 2.

(b) Comparison to cut-based in 2HDM type 2.

Figure 6.8.: Expected exclusion limits at 95% confidence level in the reparametrised 2HDM type 2 model (left) and the comparison between the DNN-based (red) and cut-based (blue) analysis strategy (right).

Fig. 6.8a shows that the DNN-based analysis strategy is sensitive to exclude the parameter space up to roughly $\tan\beta = 2$ and $m_{\mathrm{H}} = 400\,\mathrm{GeV}$. The DNN-based analysis is more sensitive over the full parameter space than the cut-based analysis, which is shown in Fig. 6.8b. Especially the gap at $m_{\mathrm{H}} \leq 200\,\mathrm{GeV}$ is closed by the DNN-based analysis.

Chapter 7

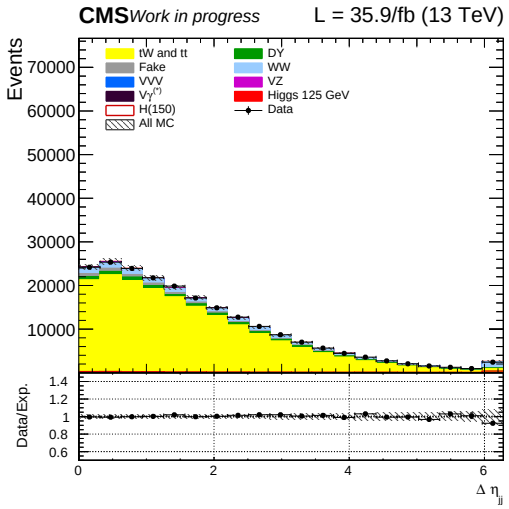
Conclusion and outlook

A search for a heavy $\mathcal{CP}$-even Higgs boson in $H \to W^{\pm}W^{\mp} \to e^{\pm}\mu^{\mp}\nu\nu$ decays using modern deep learning techniques is presented for recorded data of $35.9\,\text{fb}^{-1}$ at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ in the year 2016. The deep learning methods are used for a process enriched multi-class classification and are validated to work as expected. Additionally an improved discriminating variable is used for the final fit, which is constructed by scaling the improved transverse mass of the dilepton and $E_{\text{T}}^{\text{miss}}$ system, by the DNN response of the corresponding mass regime. The DNN response then acts as a probability density function, which describes how signal-like an event in the signal category really is.
In the model-independent search comparable expected limits on the cross-section between the presented DNN-based analysis and a cut-based analysis are achieved. Unfortunately the DNN-based analysis can not compete with the cut-based analysis for very high masses, but outperforms it between a mass hypothesis of $300 - 500\,\text{GeV}$. Furthermore an uncertainty breakdown is presented. It is found that systematic uncertainties introduced by the muon momentum scale, higher order QCD corrections and b-tagging efficiency act as the limiting systematics in this analysis. For a higher mass hypothesis bin-by-bin uncertainties become more and more relevant and limit the performance of the analysis, which can be seen in Fig. C.2. By freezing all systematic uncertainties the overall impact of nuisances are studied. With the freezing the expected exclusion limits improved on average by 25%, with a maximum of $\approx 50\%$ at a mass hypothesis between $400 - 500\,\text{GeV}$.
Additionally expected exclusion limits are calculated in model-dependent MSSM scenarios, namely the hMSSM and $m_{\text{h}}^{\text{mod}+}$ scenario, and in the more general 2HDM type 2 model. In both MSSM scenarios an overall improvement in the expected exclusion limits is achieved. Especially in the region, where the heavy Higgs decays to two on-shell W bosons, an improvement of $\approx 50\%$ in the $m_{\text{h}}^{\text{mod}+}$ scenario and $\approx 35\%$ in the hMSSM scenario is obtained. An improvement for the 2HDM type 2 expected exclusion limits is achieved with the exception of the region, where α goes to 0. This gives a hint on why the DNN-based analysis suffers for high mass hypotheses in the model-independent interpretation. The DNN-based analysis strategy shows an overall improvement in the reparametrised 2HDM type 2 interpretation.
The DNN-based analysis will be followed up for the upcoming data taking periods. Since deep learning algorithms profit a lot from high training statistics, this strategy is expected to provide even more improvement with more integrated luminosity from the data taken in 2017, 2018 and especially the full Run 2 combination. Furthermore systematic uncertainties related to background normalisation are also expected to be reduced, since they are derived from control regions. The relative statistical uncertainties will decrease by $\sqrt{N}$, where N denotes the number of events. In the end the new proposed analysis strategy using deep

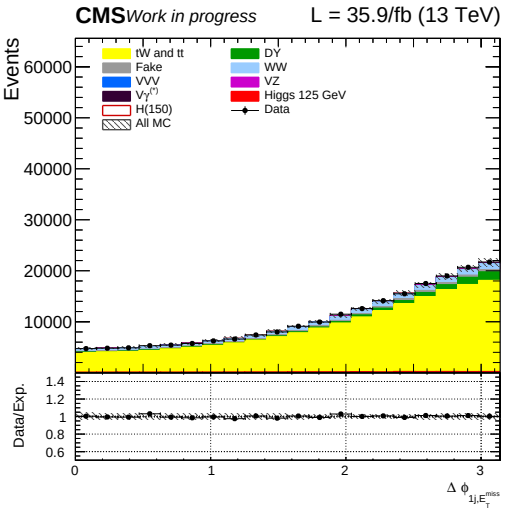
learning techniques anticipates very promising upcoming results for the future.

Appendix A

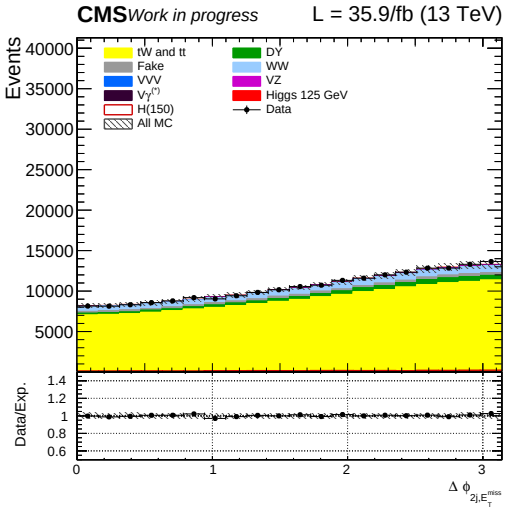
Deep neural network training variables



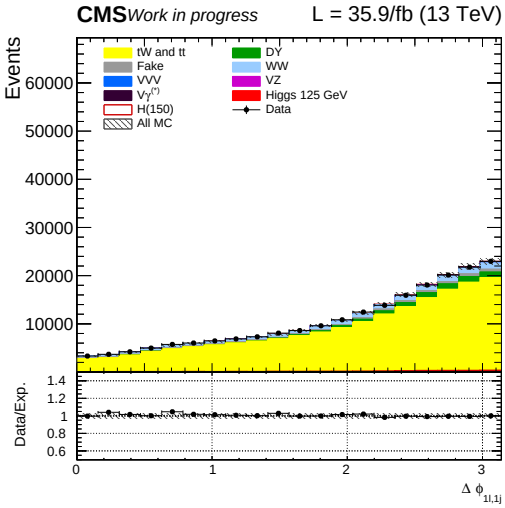
(a) $\Delta\eta_{jj}$.



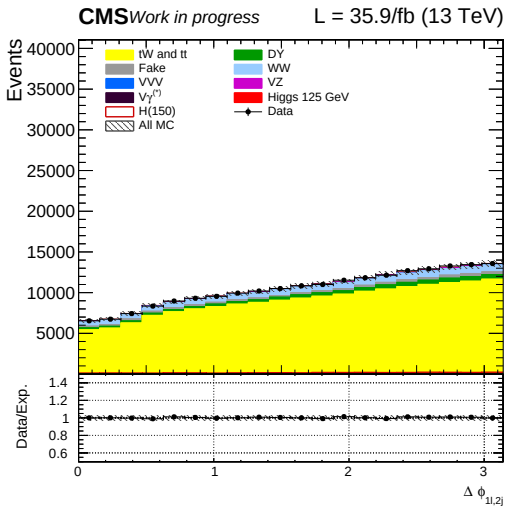
(b) $\Delta\phi_{1j,E_{\mathrm{T}}^{\mathrm{miss}}}$.



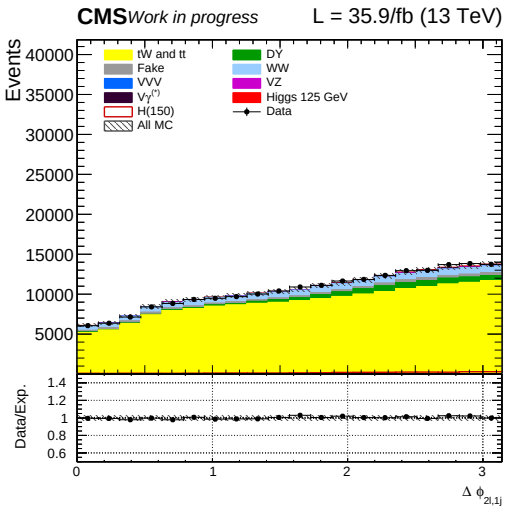
(c) $\Delta\phi_{2j,E_{\mathrm{T}}^{\mathrm{miss}}}$.



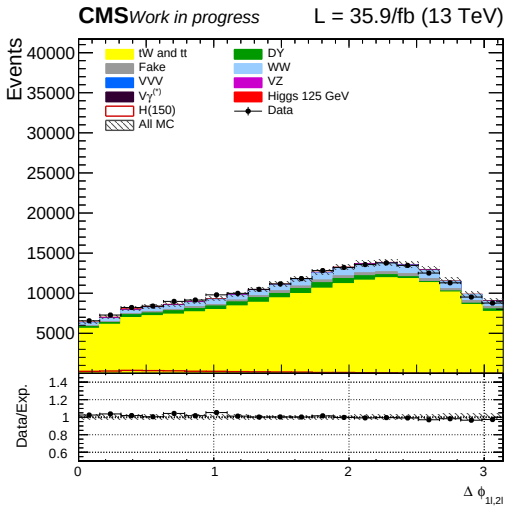
(d) $\Delta\phi_{1l,1j}$.



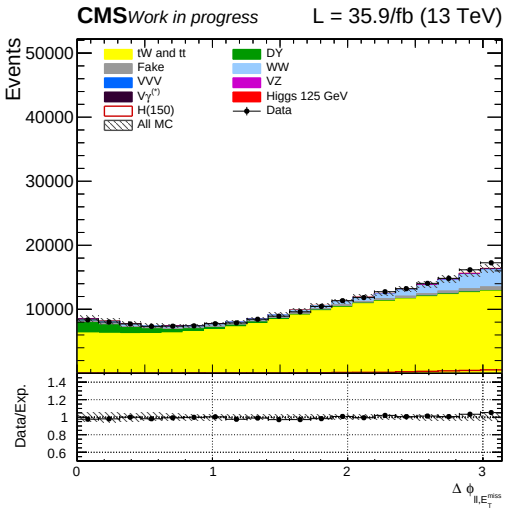
(a) $\Delta\phi_{1l,2j}$.



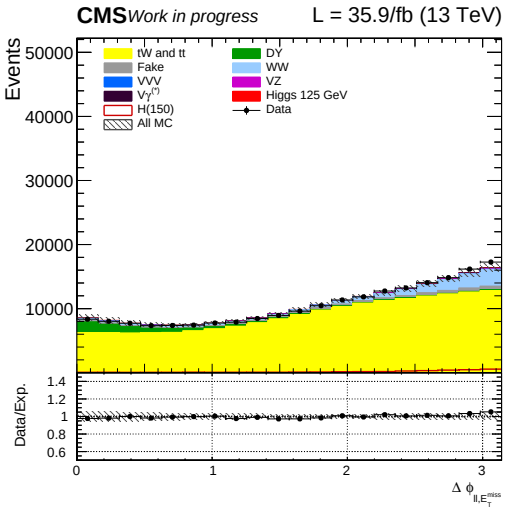
(b) $\Delta\phi_{2l,1j}$.



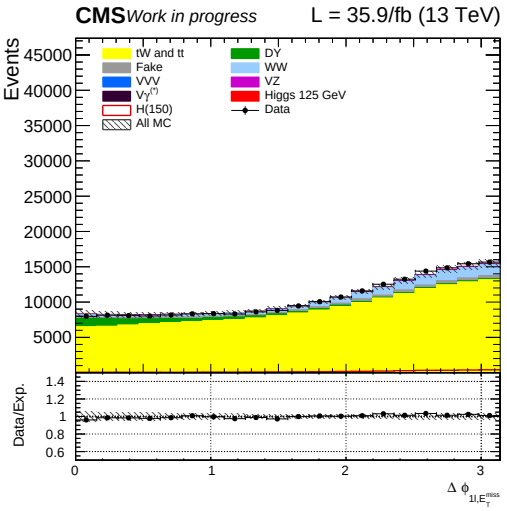
(c) $\Delta\phi_{ll}$.



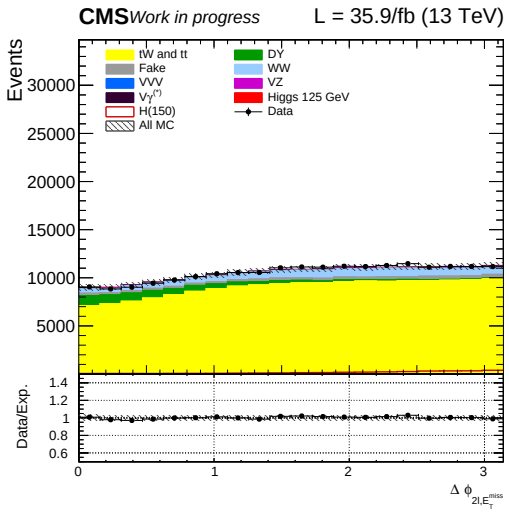
(d) $\Delta\phi_{ll,E_{\mathrm{T}}^{\mathrm{miss}}}$.



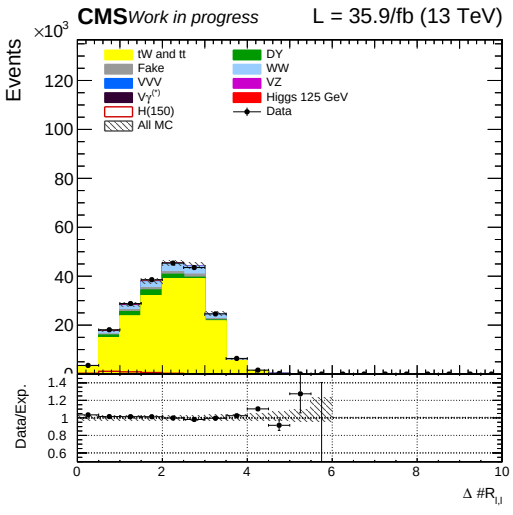
(e) $\min(\Delta\phi_{l,E_{\mathrm{T}}^{\mathrm{miss}}})$.



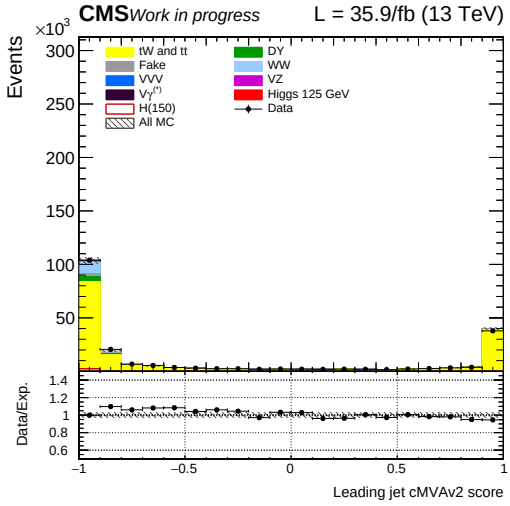
(f) $\Delta\phi_{1l,E_{\mathrm{T}}^{\mathrm{miss}}}$.



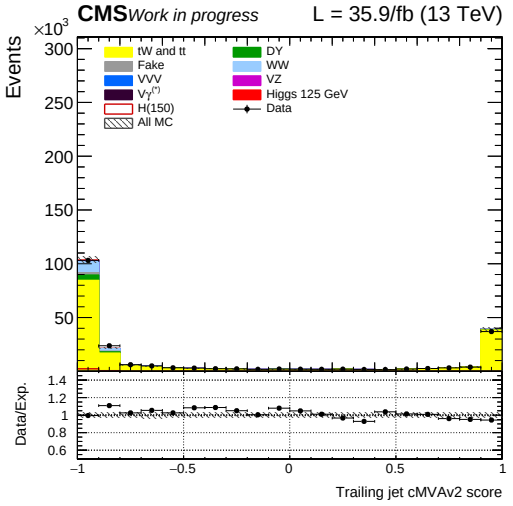
(a) $\Delta\phi_{2l,E_{\mathrm{T}}^{\mathrm{miss}}}$.



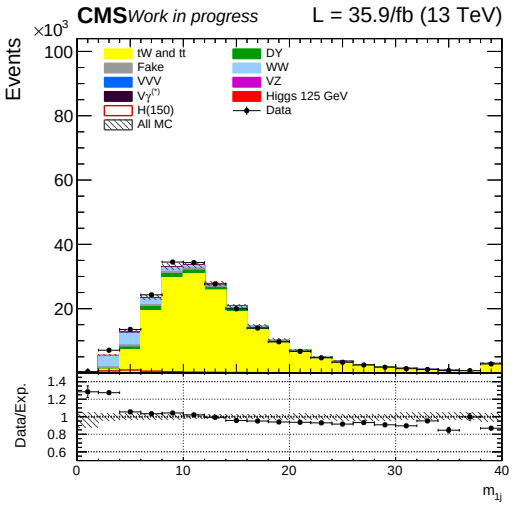
(b) ΔR_{ll}.



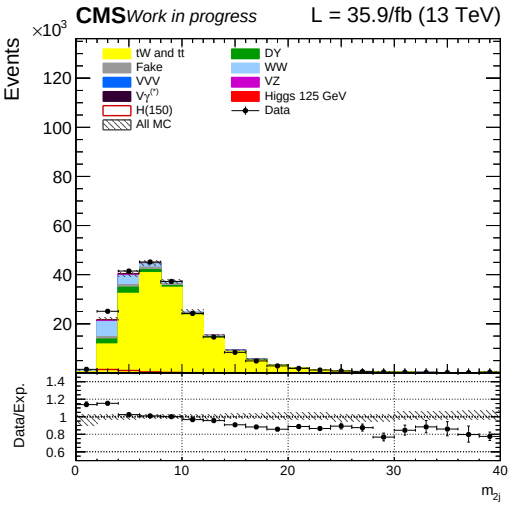
(c) Leading jet cMVAv2 score.



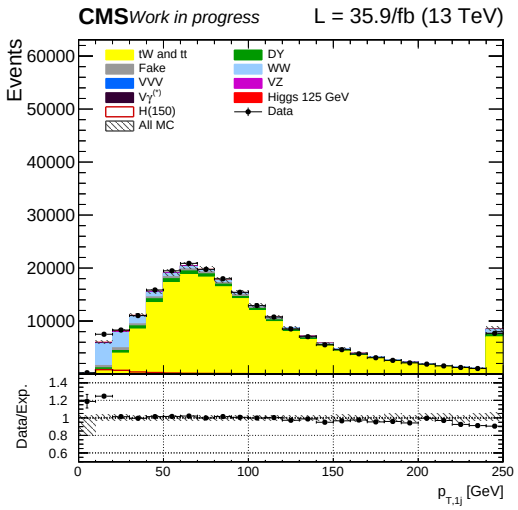
(d) Trailing jet cMVAv2 score.



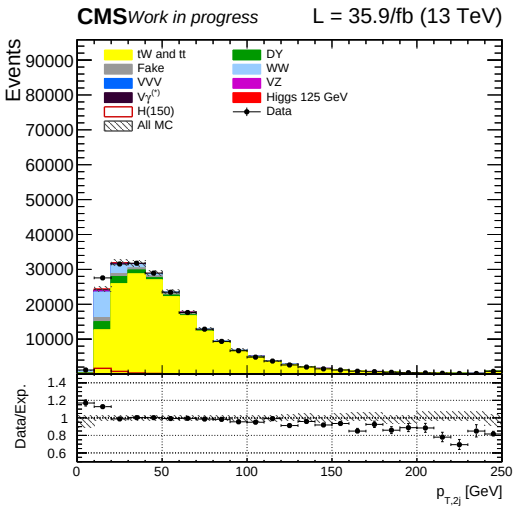
(e) m_{1j}.



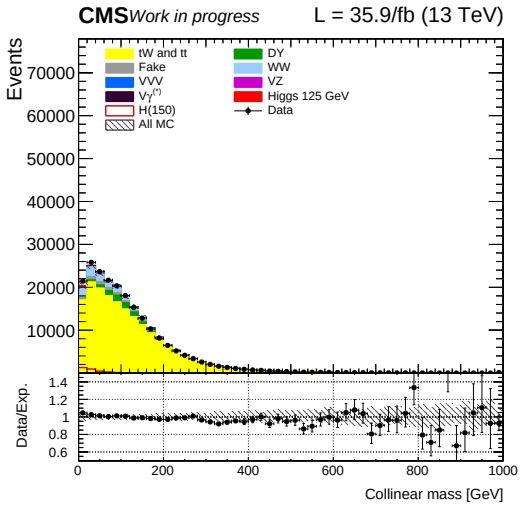
(f) m_{2j}.



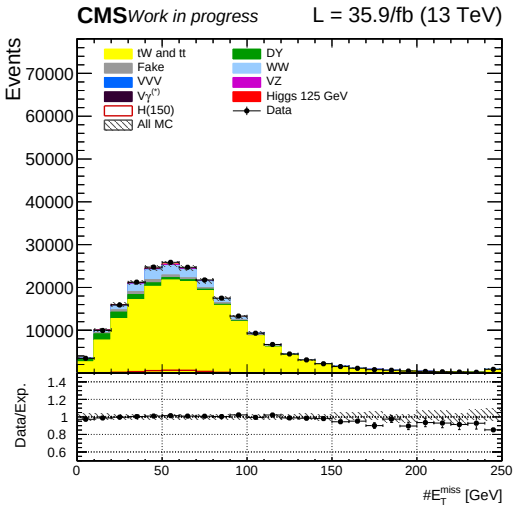
(a) $p_{T,1j}$.



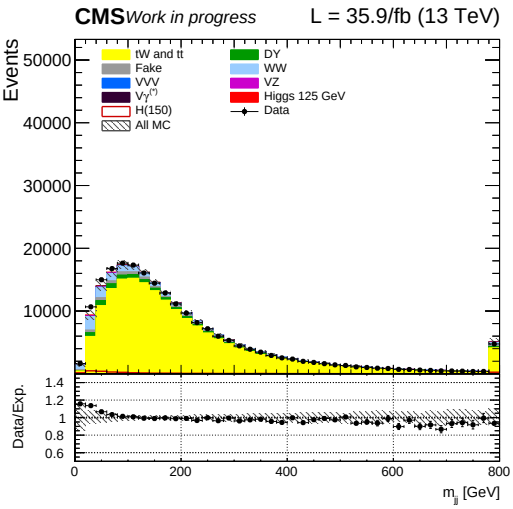
(b) $p_{T,2j}$.



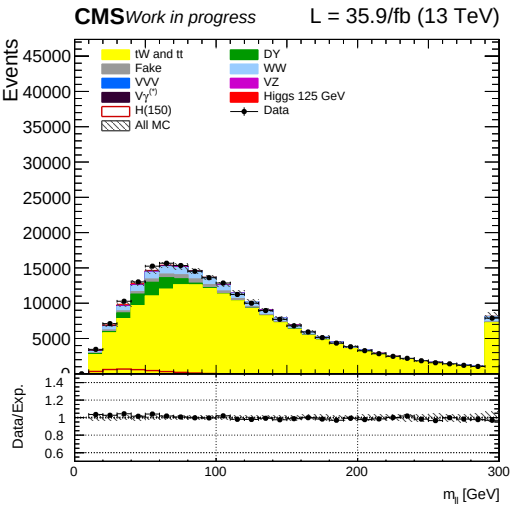
(c) Collinear mass.



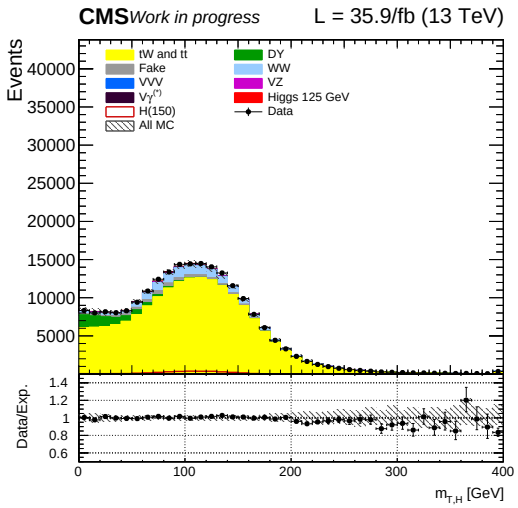
(d) $E_{\mathrm{T}}^{\mathrm{miss}}$.



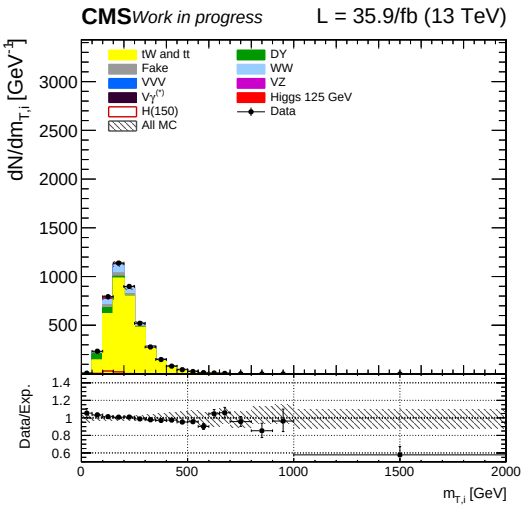
(e) m_{jj}.



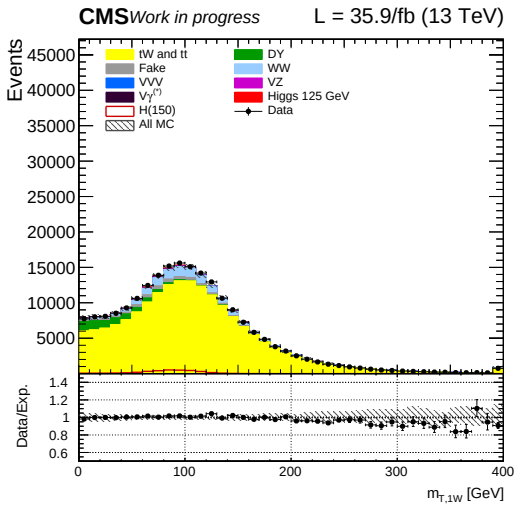
(f) m_{ll}.



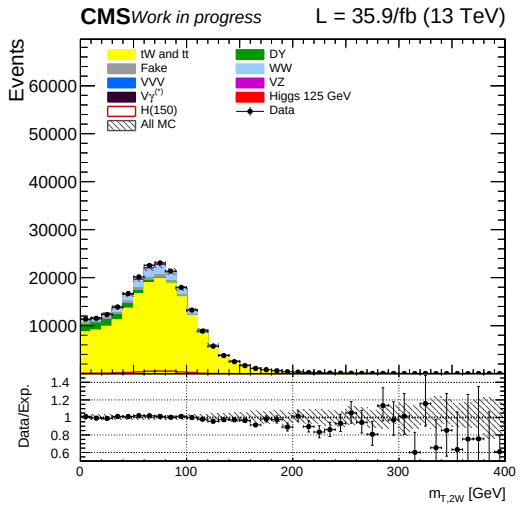
(a) $m_{T,H}$.



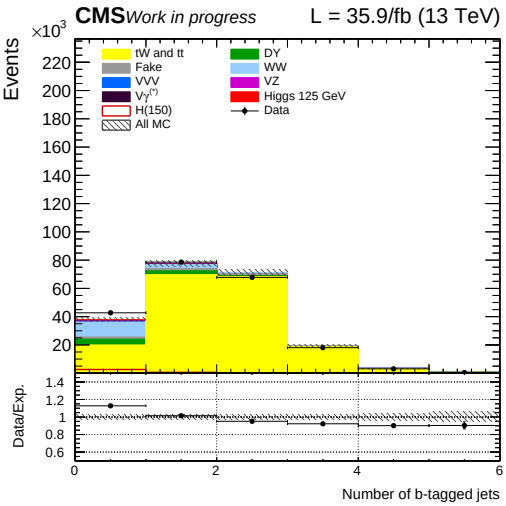
(b) $m_{T,i}$.



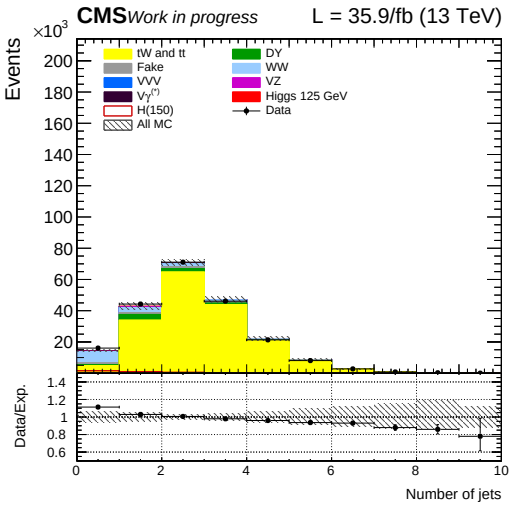
(c) $m_{T,1W}$.



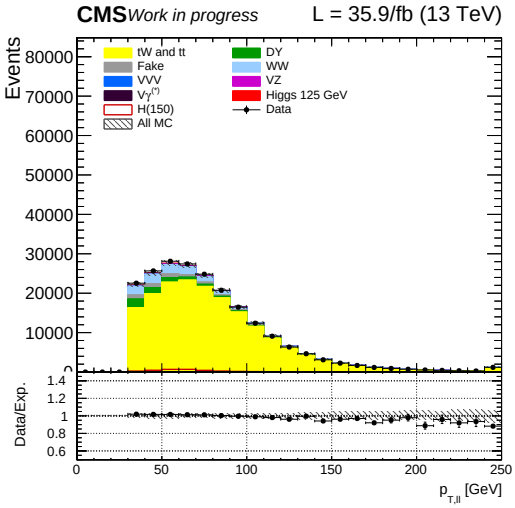
(d) $m_{T,2W}$.



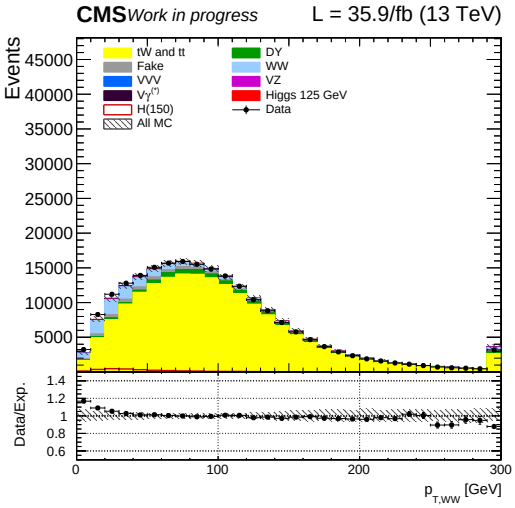
(e) Number of b-tagged jets.



(f) Number of jets.


CMS Work in progress
L = 35.9/fb (13 TeV)
Events
tW and tt
Fake
VVV
Vγ(*)
H(150)
All MC
DY
WW
VZ
Higgs 125 GeV
Data
Data/Exp.
p_T,ll [GeV]

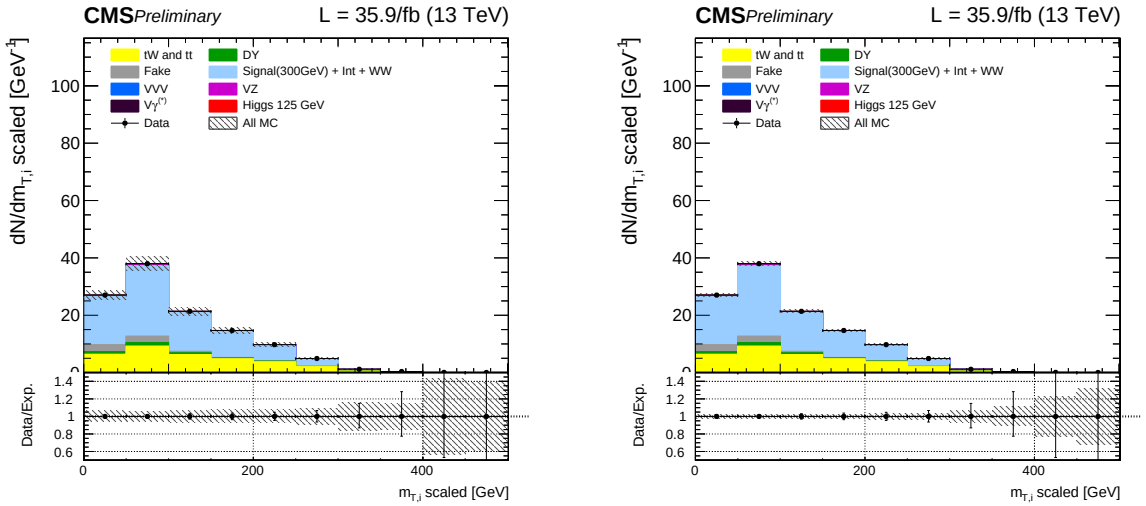
(a) $p_{T,ll}$.


CMS Work in progress
L = 35.9/fb (13 TeV)
Events
tW and tt
Fake
VVV
Vγ(*)
H(150)
All MC
DY
WW
VZ
Higgs 125 GeV
Data
Data/Exp.
p_T,WW [GeV]

(b) $p_{T,WW}$.

Appendix B

Prefit and postfit distribution for $m_{\mathrm{H}} = 300\,\mathrm{GeV}$



(a) Prefit scaled improved transverse mass distribution.

(b) Postfit scaled improved transverse mass distribution.

Figure B.1.: Prefit (left) and postfit (right) distribution of the scaled improved transverse mass.

Appendix C

Impacts and pulls for $m_{\mathrm{H}} = 500\,\mathrm{GeV}$ and $m_{\mathrm{H}} = 2000\,\mathrm{GeV}$

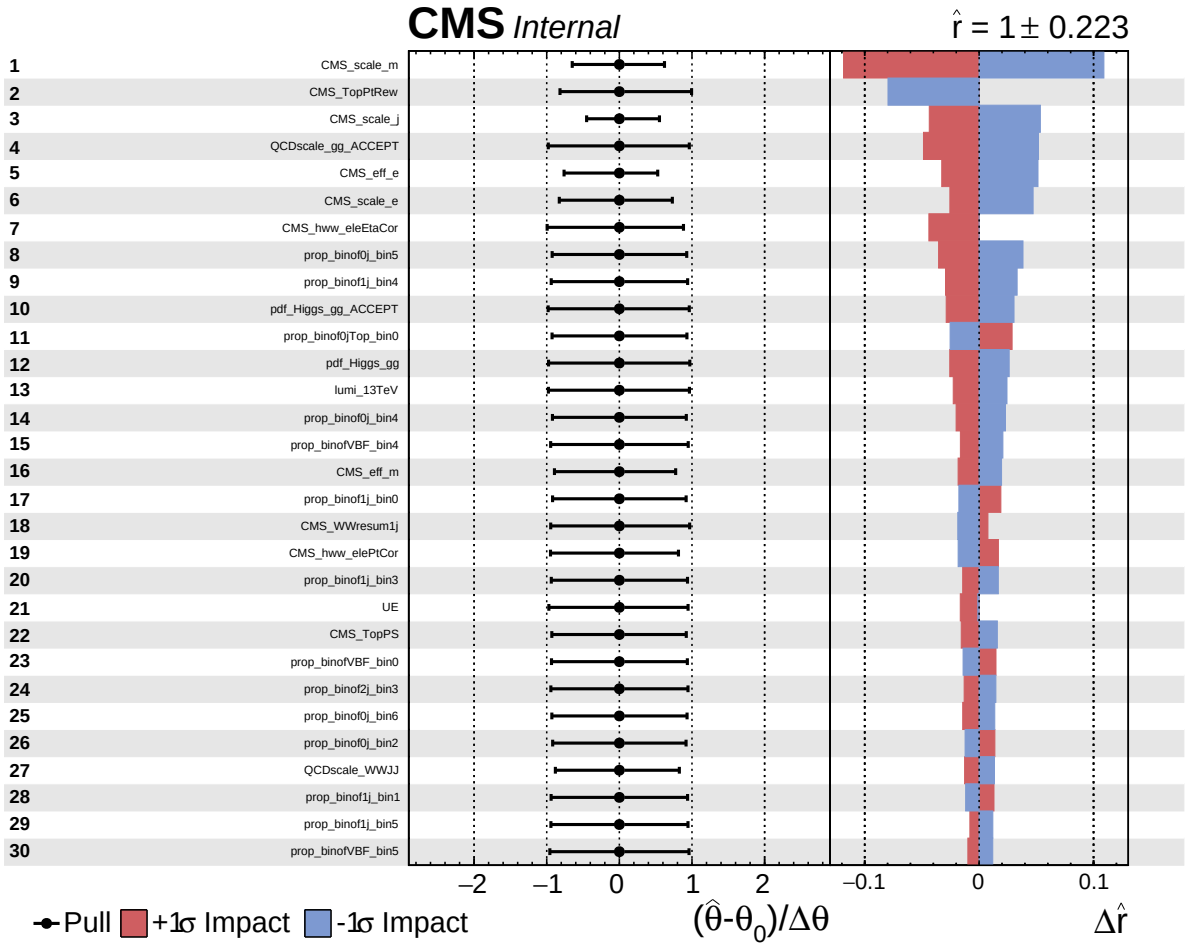


Figure C.1.: $m_H = 500\,\mathrm{GeV}$.

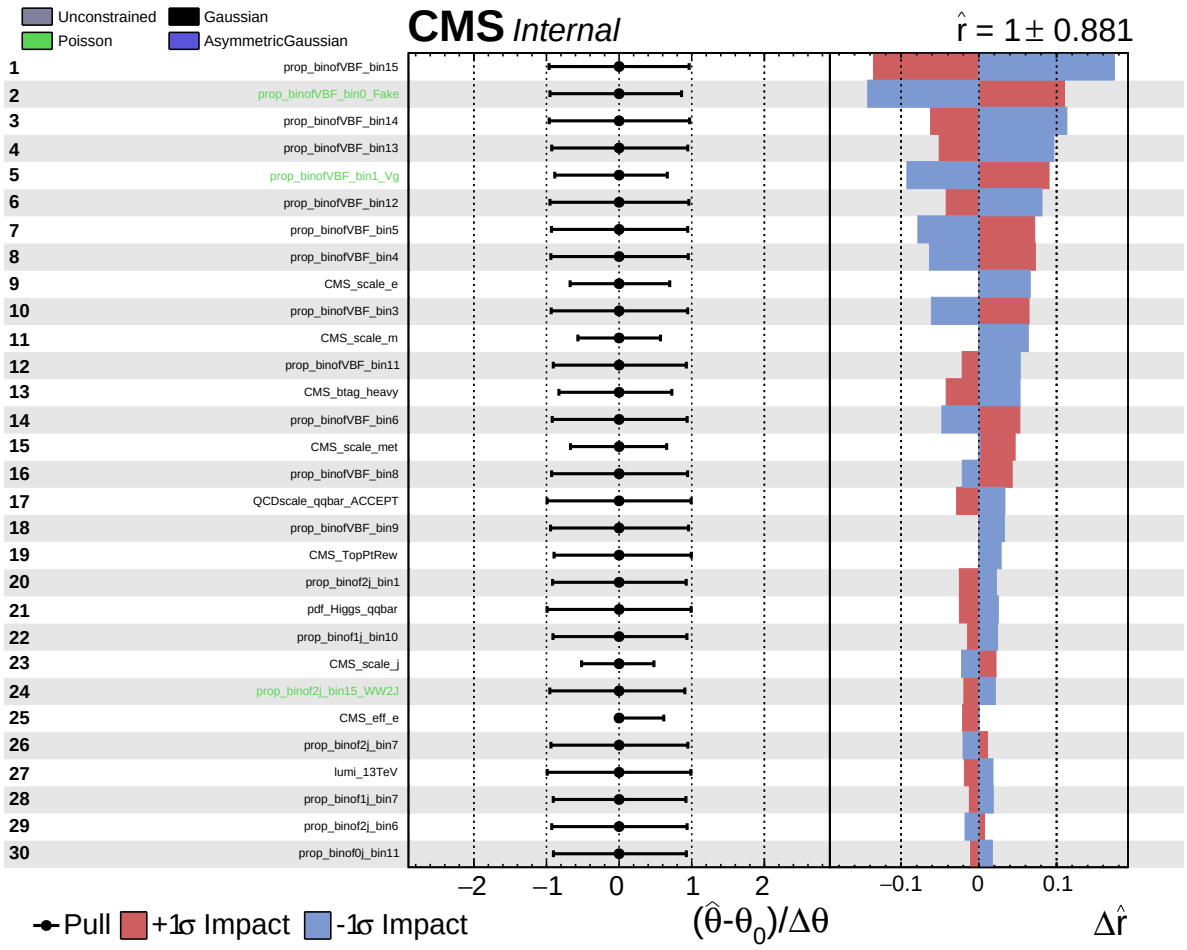


Figure C.2.: $m_H = 2000\,\mathrm{GeV}$.

References

[1] S. L. Glashow, "Partial-symmetries of weak interactions", *Nuclear Physics* **22** (1961), no. 4, 579 – 588, doi:10.1016/0029-5582(61)90469-2.

[2] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (11, 1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[3] A. Salam, "Weak and Electromagnetic Interactions", *Conf.Proc.* **C680519** (1968) 367–377.

[4] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[5] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (11, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[7] G. 't Hooft and M. Veltman, "Regularization and renormalization of gauge fields", *Nuclear Physics B* **44** (1972), no. 1, 189 – 213, doi:https://doi.org/10.1016/0550-3213(72)90279-9.

[8] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters", *Astron. Astrophys.* **594** (2016) A13, doi:10.1051/0004-6361/201525830, arXiv:1502.01589.

[9] R. Massey, T. Kitching, and J. Richard, "The dark matter of gravitational lensing", *Rept. Prog. Phys.* **73** (2010) 086901, doi:10.1088/0034-4885/73/8/086901, arXiv:1001.1739.

[10] M. Persic, P. Salucci, and F. Stel, "The Universal rotation curve of spiral galaxies: 1. The Dark matter connection", *Mon. Not. Roy. Astron. Soc.* **281** (1996) 27, doi:10.1093/mnras/281.1.27,10.1093/mnras/278.1.27, arXiv:astro-ph/9506004.

[11] A. D. Sakharov, "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe", *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35, doi:10.1070/PU1991v034n05ABEH002497. [Usp. Fiz. Nauk161,no.5,61(1991)].

[12] G. R. Farrar and M. E. Shaposhnikov, "Baryon asymmetry of the universe in the minimal Standard Model", *Phys. Rev. Lett.* **70** (1993) 2833–2836, doi:10.1103/PhysRevLett.71.210.2,10.1103/PhysRevLett.70.2833, arXiv:hep-ph/9305274. [Erratum: Phys. Rev. Lett.71,210(1993)].

[13] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and Antimatter in the Universe", *New J. Phys.* **14** (2012) 095012, doi:10.1088/1367-2630/14/9/095012, arXiv:1204.4186.

[14] M. C. Gonzalez-Garcia and Y. Nir, "Neutrino masses and mixing: Evidence and implications", *Rev. Mod. Phys.* **75** (2003) 345–402, doi:10.1103/RevModPhys.75.345, arXiv:hep-ph/0202058.

[15] S. P. Martin, "A Supersymmetry primer", doi:10.1142/9789812839657_0001,10.1142/9789814307505_0001, arXiv:hep-ph/9709356. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[16] "Particles and their supersymmetric partners in MSSM", September, 2018. Provided by Sebastian Wuchterl.

[17] E. Bagnaschi et al., "Benchmark scenarios for low tan β in the MSSM",.

[18] M. Carena et al., "MSSM Higgs Boson Searches at the LHC: Benchmark Scenarios after the Discovery of a Higgs-like Particle", *Eur. Phys. J.* **C73** (2013), no. 9, 2552, doi:10.1140/epjc/s10052-013-2552-1, arXiv:1302.7033.

[19] A. Djouadi et al., "The post-Higgs MSSM scenario: Habemus MSSM?", *Eur. Phys. J.* **C73** (2013) 2650, doi:10.1140/epjc/s10052-013-2650-0, arXiv:1307.5205.

[20] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (2008), no. 08, S08001.

[21] C. Lefevre, "LHC: the guide (English version). Guide du LHC (version anglaise)",.

[22] A. de Roeck et al., "CMS Physics Technical Design Report Volume I: Detector Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006.

[23] CMS CollaborationM. Grunwald, et al., "CMS physics Technical Design Report, Volume II: Physics Performance", volume 34. 2007.

[24] G. L. Bayatyan et al., "CMS computing: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 2005. Submitted on 31 May 2005.

[25] T. Sakuma, "SketchUpCMS Gallery". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMSGallery#Cutaway_View, Accessed 23 May 2018.

[26] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, doi:10.1088/1748-0221/3/08/S08004.

[27] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (Aug, 2018) 030001, doi:10.1103/PhysRevD.98.030001.

[28] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[29] CMS Collaboration, "Common analysis object definitions and trigger efficiencies for the H $\rightarrow$ WW analysis with 2016 full data". CMS AN-17-082 (unpublished, version 4).

[30] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[31] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[32] ATLAS, CMS Collaboration, L. Scodellaro, "b tagging in ATLAS and CMS", in *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. arXiv:1709.01290.

[33] CMS Collaboration, "Observation of $t\bar{t}H$ production", *Phys. Rev. Lett.* **120** (2018), no. 23, 231801, doi:10.1103/PhysRevLett.120.231801,10.1130/PhysRevLett.120.231801, arXiv:1804.02610.

[34] CMS Collaboration, "Search for high mass resonances using WW fully leptonic decays with 2016 data". CMS AN-17-243 (unpublished, version 5).

[35] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[36] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", *Nature* **323** (10, 1986) 533 EP –.

[38] F. Chollet et al., "Keras". https://keras.io, 2015.

[39] M. Abadiet al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", 2015. Software available from tensorflow.org.

[40] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient BackProp", in *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*, pp. 9–50. Springer-Verlag, London, UK, UK, 1998.

[41] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks", *CoRR* **abs/1706.02515** (2017) arXiv:1706.02515.

[42] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", *CoRR* **abs/1412.6980** (2014) arXiv:1412.6980.

[43] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug, 2011.

[44] L. Moneta et al., "The RooStats Project", arXiv:1009.1003v2.

[45] CMS Collaboration Collaboration, "Higgs to WW measurements with 15.2 fb^{-1} of 13 TeV proton-proton collisions", Technical Report CMS-PAS-HIG-16-021, CERN, Geneva, 2017.

[46] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", doi:10.5170/CERN-2013-004, arXiv:1307.1347.

[47] A. V. Gritsan, R. Röntsch, M. Schulze, and M. Xiao, "Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques", *Phys.*

Rev. **D94** (2016), no. 5, 055023, doi:10.1103/PhysRevD.94.055023, arXiv:1606.03107.

[48] 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* **09** (2018) 007, doi:10.1007/JHEP09(2018)007, arXiv:1803.06553.

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)
Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
Para. 156 StGB (German Criminal Code): False Statutory Declarations
Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence
(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.
(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature