VRIJE
UNIVERSITEIT
BRUSSEL

Proef ingediend met het oog op het behalen van de graad van Master in de fysica en sterrenkunde.

# PROBING THE CHARM QUARK YUKAWA COUPLING USING MACHINE LEARNING ALGORITHMS

Nordin Breugelmans

Mei 2021

Under the guidance of Prof. Dr. Jorgen D'Hondt and Dr. Seth Moortgat

**Faculty of Sciences and Bio-Engineering Sciences**

*This master's thesis came about during the period in which higher education was subjected to a lockdown and protective measures to prevent the spread of the COVID-19 virus. The process of formatting, data collection, the research method and/or other scientific work the thesis involved could therefore not always be carried out in the usual manner. The reader should bear this context in mind when reading this master's thesis, and also in the event that some conclusions are taken on board.*

## Abstract

In this thesis, a novel approach is introduced to constrain the value of the charm quark Yukawa coupling $y_c$ using the CMS detector at the CERN LHC. By targeting specifically the production of a Higgs boson in association with a jet from a charm quark (H+c), the jet flavour information extracted from state-of-the-art charm tagging algorithms on top of the event kinematical properties can be exploited to gain sensitivity to $y_c$. After creating dedicated simulations of the H+c process up to the detector level, both kinematical and jet flavour information are combined as input into an artificial neural network (ANN), which is trained on one hand to reduce the background processes, and on the other hand to provide an observable with an increased sensitivity to $y_c$. By means of a binned maximum likelihood fit, expected 95% confidence level intervals are derived on $\kappa_c = y_c/y_c^{\mathrm{SM}}$, assuming scenarios where 35.9, 300, or 3000 fb$^{-}1$ of integrated luminosity are collected. The results presented in this thesis demonstrate that, for these scenarios, a relative improvement in the sensitivity of up to 8%, 14%, and 18% can be obtained respectively using this ANN discriminator, as compared to traditional methods that only rely on the transverse momentum of either the leading jet or the Higgs boson candidate.

# Contents

# Chapter 1

# Introduction

## 1.1 The Standard Model of particle physics

The field of particle physics studies a wide variety of physical phenomena, from studying the smallest, most fundamental building blocks of matter with particle colliders, to understanding observations on a cosmological scale, such as the yet unknown nature of dark matter. Although these research subjects are seemingly different, they share a common goal: to increase our understanding of the fundamental laws of nature. Some of these laws manifest themselves only at the smallest scales, in the interactions between elementary particles. The efforts made by physicists throughout history to study these interactions have culminated in one of the most successful physical theories to date: the Standard Model (SM) of particle physics.

### 1.1.1 The particles of the Standard Model

The idea that all matter is constructed from a finite number of building blocks has been present for a large part of human history. Even in ancient Greece, followers of Aristotle's natural philosophy believed that matter is ultimately composed of a ratio of four elements: water, earth, fire and air [1]. The same nomenclature is used in modern chemistry, where an element describes a pure substance made of atoms containing the same number of protons. All chemical substances, from gas mixtures to metal alloys, are formed by different combinations of atoms. Yet atoms can be broken down into further constituents: The nucleus, which contains protons and neutrons, and one or more electrons that are bound to the nucleus. Both the fundamental particles and forces of particle physics come into play when investigating the inner workings of the atom. For example, the electromagnetic force binds electrons to the nuclei of atoms, the strong nuclear force holds the nuclei together, and the weak nuclear force governs the $\beta^-$ decay of unstable atoms. Of the particles mentioned so far, only the electron is thought to be fundamental. Protons and neutrons are both composed of up and down quarks. Together with electrons and electron neutrinos, which are emitted in nuclear $\beta^-$ decay, they form the first generation of elementary matter particles.

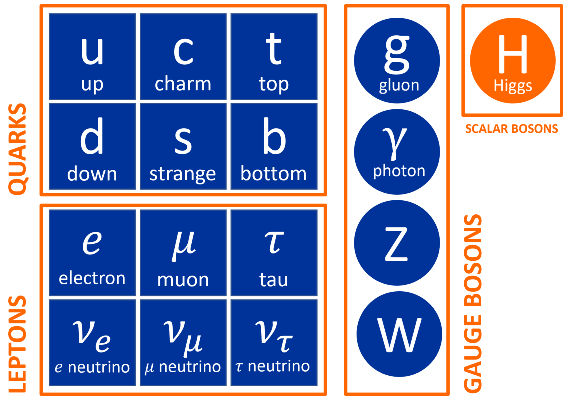



Figure 1.1: The elementary particles of the Standard Model.

One objective of the SM is to categorize the elementary particles. This allows for the SM to be summarized visually in a way similar to the periodic table, as seen in figure 1.1. A first subdivision of the particles of the SM is into fermions and bosons. The fundamental fermions of the SM are spin 1/2 particles that follow the Pauli exclusion principle, and hence obey Fermi-Dirac statistics. The fermions are made up of two families of particles: the quarks and the leptons. The first generation of fermions were mentioned earlier. The electron and electron neutrino form the first generation of the leptons, while the up and down quark form the first generation of the quarks. The SM includes two additional fermion generations. The second and third generation of the leptons consist of the muon and muon neutrino, and the tau lepton and tau neutrino. Likewise, the second and third quark generations are respectively composed of the charm and strange quarks, and the top and bottom quarks. Particles in higher generations have a higher mass than the corresponding particle in a lower generation, with the exception of the neutrinos, which are considered massless in the SM[1]. Each of the rows in the quark and lepton sector in figure 1.1 contain particles with the same electric charge. The up-type quarks have charge +2/3 while the down-type quarks have charge -1/3. Likewise, the neutrinos are neutral, while the remaining leptons having charge -1. The reason behind the distinction between leptons and quarks lies in the nature of their interactions. In the SM, these interactions are governed by three fundamental forces: the strong force, the weak force and electromagnetism. Due to their non-zero electric charge, the quarks and the charged leptons can interact electromagnetically. Quarks on the other hand carry an additional charge related to the strong force, sometimes referred to as the color charge, which allows them to have strong interactions as well. In the SM, each of these interactions is mediated by the exchange of bosons. The photon mediates the electromagnetic force, the W and Z bosons mediate the weak force and the gluons mediate the strong force. All of these particles are of integer spin, and follow Bose-Einstein statistics. Lastly,

[1]Contrary to the SM assumption, the phenomenon of neutrino oscillations, first discovered by the SNO [2] and Super-Kamiokande [3] Collaborations in 2001 and 1998, indicates that neutrinos are massive.

there is the Brout-Englert-Higgs boson, or the Higgs boson in short. The neutral scalar boson couples to all particles that have a non-zero mass. The particle was postulated by Robert Brout, Francois Englert [4] and Peter Higgs [5] in 1964, and is a key component of the Brout-Englert-Higgs mechanism, which is responsible for generating the mass terms of the fermions and gauge bosons of the SM. The Higgs boson represented the last missing piece of the Standard Model puzzle. In fact, its discovery was one of the main goals of the LHC. As a result, in July 2012, the ATLAS [6] and CMS [7] experiments at CERN's Large Hadron Collider announced the discovery of a new particle consistent with the Higgs boson predicted by the Standard Model.

### 1.1.2 The theoretical framework of the Standard Model

The mathematical description of the SM is constructed in the framework of Quantum Field Theory (QFT). In QFT, particles are described as fields extended over spacetime. The fermion fields are solutions of the Dirac equation of relativistic quantum mechanics, which describes the dynamics of spin 1/2 particles. Technically, these fields are expressed as Dirac spinors. The kinematics of these fields and their interactions are captured by the Lagrangian density $\mathcal{L}_{\text{SM}}$. From the Dirac Lagrangian density,

$$\mathcal{L}_{\text{D}} = \overline{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi, \tag{1.1}$$

one can derive the equations of motion of a free Dirac spinor field $\psi$. The adjoint spinor $\overline{\psi}$ is defined as $\psi^{\dagger}\gamma^{0}$, where $\gamma^{\mu}$ are the gamma-matrices and $\partial_{\mu}$ is the space-time derivative. The Dirac equation has both positive and negative energy solutions, resulting in particle and anti-particle spinors. Interaction terms can be added to this free-field Lagrangian by applying the gauge principle. Here, a global symmetry of the Lagrangian is transformed into a local symmetry by adding gauge fields to the theory. Remarkably, the theory of Quantum Electrodynamics (QED), and therefore Maxwell's equations, can be derived by applying this principle. Transforming the global U$(1)$ symmetry of equation (1.1) into a local U$(1)$ symmetry, requires the inclusion of one additional gauge field, namely the photon field. As a result from this inclusion, interaction terms are generated between the photon and the fermion fields. In other words, the generator of the U$(1)$ local gauge symmetry corresponds with the photon, the boson which mediates the QED interactions. Likewise, the theory of Quantum Chromodynamics (QCD) which describes the strong interactions, is derived by requiring (1.1) to be invariant under local SU$(3)$ gauge transformations. The eight generators of SU$(3)$ are associated with the eight massless gluons of QCD. The charge associated with SU$(3)$ is called the color charge, and is labelled as red, green or blue. A key principle in QCD is that of color confinement, which states that colored particles only appear in color-neutral composite states. An example of such a color-neutral state is the meson, composed of a quark and an antiquark of different flavours and opposite color charges (such as red and anti-red). QCD is an asymptotically free theory, which means that the strength of interactions between colored particles becomes weaker as the energy scale increases, or likewise as the length scale decreases. As a consequence, increasing amounts of energy are required to further separate a pair of quarks from one another. Ultimately, enough energy has become available to spontaneously produce another quark-antiquark pair from the vacuum. These quarks can then bind with the separated ones, and form two new quark pairs. Though an analytic description of this phenomenon has not yet been given successfully, the above example serves as a qualitative explanation for color confinement.

At high energies, electromagnetism and the weak interaction are unified into a single force, namely the electroweak force. The electroweak gauge fields are generated by the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge symmetry. The electroweak theory is a chiral gauge theory, where chiral right-handed components of Dirac fields transform as singlets under $\mathrm{SU}(2)$ while the left-handed components transform as doublets under $\mathrm{SU}(2)$. As a consequence, bare mass terms for fermions $\propto m\overline{\psi}\psi$ break the $\mathrm{SU}(2)$ gauge invariance of the theory. The full SM gauge group can be written down as $\mathrm{SU}(3)_{\mathrm{c}} \times \mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$.

### 1.1.3 Electroweak symmetry breaking

Consistency requires each of the terms in the SM Lagrangian density $\mathcal{L}_{\mathrm{SM}}$ to respect these symmetries. As a consequence, a bare mass term for the gauge fields is also not allowed. This is troublesome, since the W and Z bosons of the weak interaction are massive. The problem of adding mass terms in a gauge invariant way can be solved using the Brout-Englert-Higgs mechanism [4, 5], where a complex scalar field $\phi$ with a degenerate groundstate is added to the SM theory. The scalar field $\phi$ is a doublet under $\mathrm{SU}(2)$,

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix}, \tag{1.2}$$

where $\phi_i, i = 1, .., 4$ are real-valued scalar fields. The free field Lagrangian for the scalar doublet is given by

$$\mathcal{L}_{\mathrm{H}} = |\partial_\mu \phi|^2 - V(\phi), \tag{1.3}$$

where the scalar field potential $V(\phi)$ is defined by

$$V(\phi) = \mu^2 |\phi|^2 + \lambda |\phi|^4. \tag{1.4}$$

For $\mu^2 < 0$, the scalar field potential has a degenerate groundstate, with each of the minima satisfying $|\phi|^2 = v^2/2$, where $v$ is the vacuum expectation value (VEV) defined by $\sqrt{-\mu^2/\lambda}$. When the scalar field acquires this VEV, the Lagrangian in equation (1.3) is no longer invariant under the $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ gauge symmetry: The symmetry has been broken spontaneously. Because this happens in the electroweak sector, this process is also called electroweak symmetry breaking. To be precise, the electroweak gauge symmetry group $\mathrm{SU}_{\mathrm{L}}(2) \times \mathrm{U}_{\mathrm{Y}}(1)$ breaks to the $\mathrm{U}_{\mathrm{EM}}(1)$ gauge group of electromagnetism. Now the scalar field can be parametrized by

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ v + \eta + i\phi^3 \end{pmatrix}. \tag{1.5}$$

Goldstone's theorem states that for each continuous symmetry of the Lagrangian that has been broken, a Goldstone boson appears in the theory. These bosons are unphysical, and can be rotated away by means of an appropriate gauge transformation. After applying this transformation, the scalar doublet takes the following form:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \sigma \end{pmatrix}, \tag{1.6}$$

where $\sigma$ represents the scalar Higgs field. This particular gauge transformation is called the unitary gauge. The local gauge principle can again be applied to equation (1.3), such that the Lagrangian density of the scalar doublet is invariant under local $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ gauge transformations. This introduces interaction terms between the scalar doublet and the gauge bosons of the electroweak theory. From these interactions, mass terms arise for the massive gauge bosons in a gauge invariant way. These mass terms are related to the Higgs field VEV $v$ and the electroweak gauge couplings $g$ and $g'$ by the following relations:

$$m_W = \frac{vg}{2}, \quad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2}. \tag{1.7}$$

### 1.1.4 Yukawa interactions

In the previous section, mass terms were generated for the gauge bosons through electroweak symmetry breaking. What remains is to generate mass terms for the fermions in a gauge invariant way. This can be done by including interactions of the form $\propto \overline{\psi}\phi\psi$ in the SM theory. These are called the Yukawa interactions, named after the Japanese theoretical physicist Hideki Yukawa. In the SM, the Yukawa Lagrangian density takes the following form:

$$\mathcal{L}_\mathrm{Yukawa} = -y_\mathrm{d}^{ij}\overline{Q}_\mathrm{L}^i \phi d_\mathrm{R}^j - y_\mathrm{u}^{ij}\overline{Q}_\mathrm{L}^i \tilde{\phi} u_\mathrm{R}^j - y_\ell^{ij}\overline{\mathrm{L}}_\mathrm{L}^i \phi e_\mathrm{R}^j + (\mathrm{h.c.}), \tag{1.8}$$

where $\tilde{\phi} = i\sigma_2\phi$ and $\sigma_2$ is one of the Pauli matrices. The right-handed quark fields are represented by $u_R$ and $d_R$, with the former representing the up-type quarks and the latter the down-type quarks. Likewise, the charged right-handed lepton fields are represented by $e_R$. The index $i$ and $j$ run over the different generations of the quarks and leptons. The left-handed quark and lepton doublets are respectively given by $Q_\mathrm{L} = (u_\mathrm{L}, d_\mathrm{L})^\mathrm{T}$ and $\mathrm{L}_\mathrm{L} = (e_\mathrm{L}, \nu_\mathrm{L})^\mathrm{T}$. Most important to the scope of this thesis are the Yukawa couplings, denoted by $y$. Their values are a measure for how strongly the fermions couple to the scalar boson. In the unitary gauge, and after rotating into the mass eigenstates, the relation between the Yukawa couplings $(y_q)$ and the quark masses $(m_q)$ is given by

$$\mathcal{L}_\mathrm{Yukawa} \supset -\frac{y_q v}{\sqrt{2}}\overline{q}q = m_q\overline{q}q, \tag{1.9}$$

where $q$ is the quark field. In other words, the SM predicts a linear relationship between the Yukawa couplings $y$ and the fermion masses $m$,

$$y = \sqrt{2}\frac{m}{v}. \tag{1.10}$$

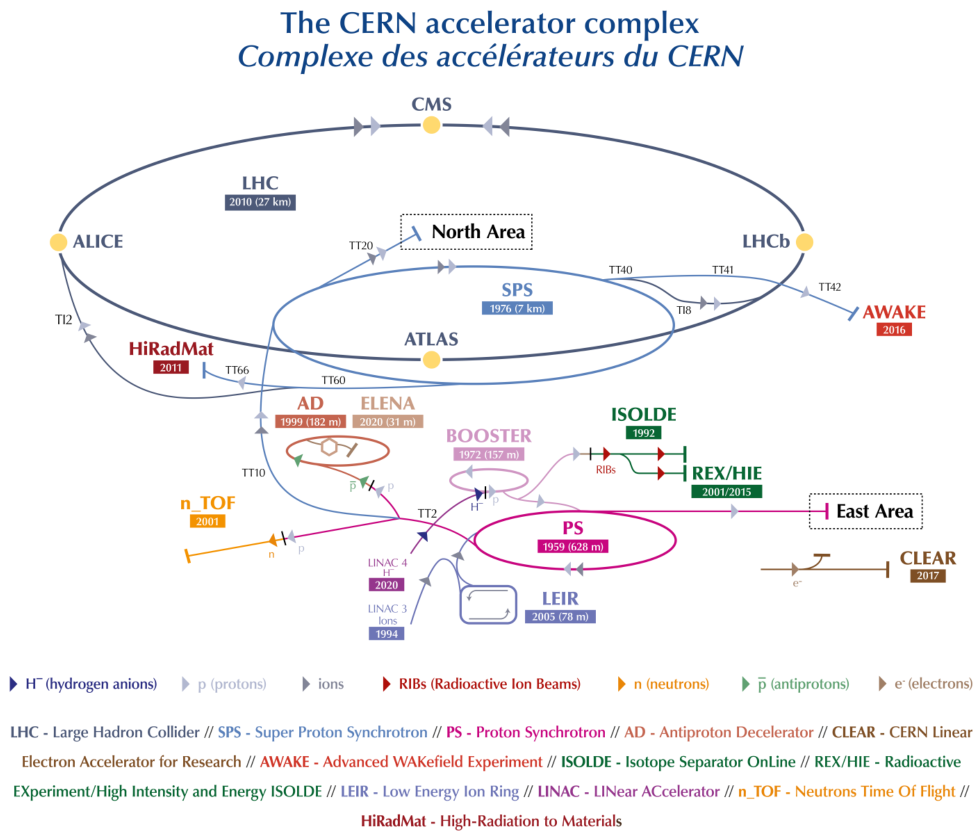



Figure 1.2: The CERN accelerator complex, taken from [8]. The yellow dots denote the locations on the LHC where the proton beams cross, and where the products of the collisions are observed by particle detectors.

As a consequence of equation (1.10), the heavier particles of the SM, such as the third-generation fermions, couple more strongly to the Higgs boson than the lighter particles. The values of the Yukawa couplings $y$ are free parameters of the SM theory, and therefore need to be probed experimentally. This can be done using the Large Hadron Collider at CERN, where beams of protons are collided in order to test the predictions of the SM at high energies.

## 1.2 Experimental setup

### 1.2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular particle accelerator located beneath the France-Switzerland border near Geneva. The collider was installed in an existing tun-


Figure 1.3: High-luminosity LHC project schedule, taken from [11].

nel that was previously in use by the Large Electron-Positron Collider (LEP). The LEP tunnel measures 26.7 km in circumference, and lies between 45 m and 170 m beneath the surface [9]. The LHC project was first approved by the CERN Council in 1994 with construction commencing in 1998. The aim of the LHC is to accelerate and collide two beams of protons at a centre-of-mass energy of up to 14 TeV. The beams cross in four points located around the collider, denoted by the yellow dots in figure 1.2. These form the locations of underground cavities wherein particle detectors are housed, like the Compact Muon Solenoid detector (CMS) or the ATLAS (A Toroidal LHC ApparatuS) detector. The beams consist of bunches of protons which are collided every 25 ns, with each bunch containing $1.15 \times 10^{11}$ protons with up to 2808 bunches per beam. During the first run of the LHC, also called Run-1, between 2011 and 2013, the LHC provided roughly 30 fb$^{-1}$ of integrated luminosity at 7 and 8 TeV. During Run-2, the LHC reached a peak luminosity of $2 \times 10^{34}$ cm$^{-2}$ s$^{-1}$ and a centre-of-mass energy of 13 TeV [10]. Run-2 provided over 130 fb$^{-1}$ of integrated luminosity. Run-3 is expected to start in 2022, and should together with Run-2 provide a combined integrated luminosity of over 300 fb$^{-1}$. Future prospects include the High Luminosity LHC (HL-LHC), where the LHC will be upgraded to achieve a luminosity of up to $5 \times 10^{34}$ cm$^{-2}$ s$^{-1}$ at the design centre-of-mass energy of 14 TeV [11]. In the HL-LHC, the combined integrated luminosity is expected to reach over 3000 fb$^{-1}$. The project schedule for the next runs of the LHC is given in figure 1.3.

### Proton acceleration

Protons are injected into the LHC through a series of smaller accelerators, as shown in figure 1.2. The first accelerator is Linac 2, a linear accelerator that drives the protons up to an energy of 50 MeV using radiofrequency (RF) cavities. These metallic chambers contain an electromagnetic field which oscillates at 400 MHz. If timed correctly, a proton arriving in the RF chamber will only feel a force in the forwards direction, and will hence be accelerated. After Linac 2, the protons reach the Proton Synchrotron Booster, a circular accelerator or synchrotron which accelerates the protons to an energy of 1.4 GeV. Next, the protons are accelerated by the much larger Proton Synchrotron to 25 GeV. The last step in the accelerator chain is the Super Proton Synchrotron, CERN's second-largest accelerator measuring nearly 7 kilometers in circumference, where the protons are accelerated to their LHC injection energy of 450 GeV. Finally, in the LHC itself, protons are further accelerated through 16 RF cavities in each round trip, before eventually reaching an energy of 6.5 TeV.

### Bending and focusing magnets

Essential to the working of the LHC are the dipole and quadrupole magnets, of which thousands are installed across the 27 km tunnel. The dipole magnets force the protons in a circular path, requiring a magnetic field strength of up to 8.3 T to do so, depending on the particle energy. This intense magnetic field is required to prevent the high-energy protons from colliding with the beam walls. Since the field intensity is directly proportional to the current travelling through the coil, the LHC's electromagnets demand up to 11.000 amperes of current. The internal resistance of the coils would greatly hinder the large current and cause massive energy losses. This is solved by using superconducting materials such that, by cooling the coils to a temperature of 1.9 K, the phenomenon of superconductivity can take over. At these low temperatures, current flowing through the coils would not encounter any resistance. Additionally, quadrupole magnets are utilised to act analogous to lenses, focusing the beams for example near collision points to increase the rate of interaction.

### Collider design

The LHC is a circular hadron collider, contrary to leptons used in the LEP collider. The LEP collider generated electron-positron collisions at a peak centre-of-mass energy of around 200 GeV. Despite having used the same tunnels in which now the LHC operates, it could not generate collision energies close to what the LHC is capable of. This is mainly because of the physical limits imposed by synchrotron radiation: the radiation emitted by charged particles when accelerated perpendicular to their velocity. The power $P$ emitted by a particle of mass $m$ undergoing synchrotron radiation scales as $P \sim m^{-4}$. Since the electron is around 2000 times lighter than the proton, the former will emit a much larger proportion of its energy as synchrotron radiation. Therefore, in order to achieve a higher collision energy, the LEP would have to counteract this effect

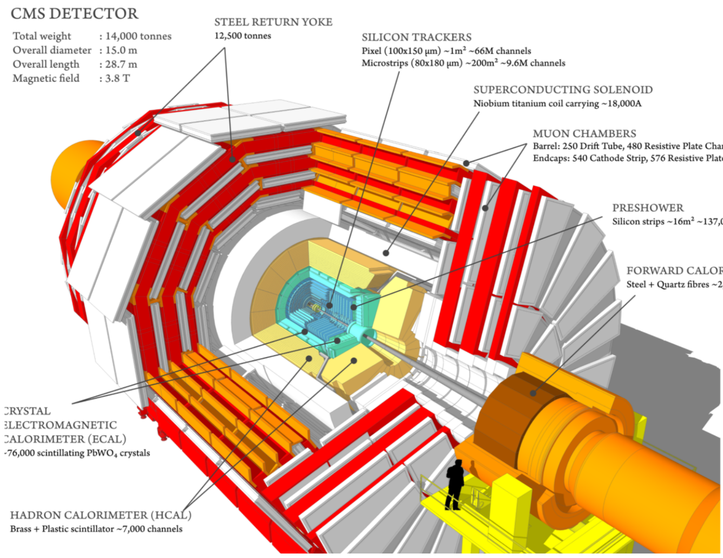



Figure 1.4: Schematic overview of the Compact Muon Solenoid detector, taken from [12].

either by putting more energy into the beam, or by increasing the radius of the collider. On the other hand, the electron-positron collider has the advantage that it provides clean event signatures and allows for precision measurements. This is less so in proton-proton colliders, since protons are composite particles. As a result, proton-proton collisions provide a larger variety of interactions at different center-of-mass energies, at the cost of having more complex events that are difficult to analyse. Crucial for the analysis of these events are the measurement of the energy and trajectory of stable particles emerging from the proton-proton collisions. This is handled by detectors like ATLAS and CMS, who are thought of as general-purpose detectors, capable of investigating a wide range of physics. Other detectors at the LHC serve a more specific goal, for example the LHCb detector which focuses on the interactions with b-hadrons, or the ALICE detector, which is built to investigate collisions of heavy-ions rather than protons.

### 1.2.2 The Compact Muon Solenoid Detector

The CMS detector has a cylindrical shape with a diameter of 15 metres and a length of 21 meters [13]. It is composed of several concentric layers of detector material suited for making energy measurements and recording particle trajectories, as seen in figure 1.4. The central component of the CMS detector is a superconducting solenoid with a diameter of 6 metres, capable of generating a 3.8 T magnetic field. The purpose of this field is to bend the trajectories of charged particles. From the bending direction and curvature, the charge and momentum of the charged particles can be inferred. The solenoid contains three other layers of detectors inside its coil.

### Inner tracker

The innermost layer is the silicon tracker, which surrounds the interaction point where the beams cross. It contains concentric layers of silicon pixels or strips, each producing a hit when a charged particle travels through them. The collection of hits can be connected to reveal the charged particle's trajectory. The tracker acceptance runs up to a pseudorapidity of $|\eta| < 2.5$. The pseudorapidity $\eta$ is a spatial coordinate derived from the polar angle $\theta$ of a particle's momentum with respect to the beam axis, and it is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{1.11}$$

### Electromagnetic calorimeter

The layer surrounding the tracker is the Electromagnetic Calorimeter (ECAL). This component measures the energy of electrons and photons, stopping them completely in the process[2]. The ECAL uses lead tungstate crystals, which have the property of producing light proportional to the particle's energy in the event of an electron or a photon passing through them. This light can then be detected by sensors on the back of the crystals.

### Hadronic calorimeter

Other particles, like hadrons, generally travel straight through the ECAL. The energy of hadrons is then measured by the Hadronic Calorimeter (HCAL), the detector surrounding the ECAL. The HCAL consists of alternating layers of absorbing and scintillating material, which makes it possible to measure the energies of the showers of particles produced when a hadron hits the absorbing material.

### Muon system

Finally, positioned at the exterior of the CMS solenoid, the muon chambers handle the detection of muons. Muon detection plays an important role in the CMS experiment, as implied by its name. Many physics analyses rely on the reconstruction of muons, including the one presented in this thesis. For example, one of the *golden channels* of the Higgs boson is the decay into a pair of Z bosons, which can themselves decay into 4 leptons. If all of the leptons are muons, then these will be relatively easy to detect. As a result of their higher mass, muons suffer less energy losses due to radiation or scattering processes. They are therefore capable of penetrating several metres of dense material, hence they can easily pass through the calorimeters. Since electrons do not travel past the ECAL, and tau leptons have a very short lifetime, any charged leptons detected in the muon chambers are very likely to be muons. Another advantage of muon final states is the higher resolution on the 4-lepton invariant mass, since muons suffer less energy loss from interactions with the tracker. This is especially relevant to the scope of this

[2]Occasionally, one of these particles can get through the ECAL. This incident is called *punch-through*.

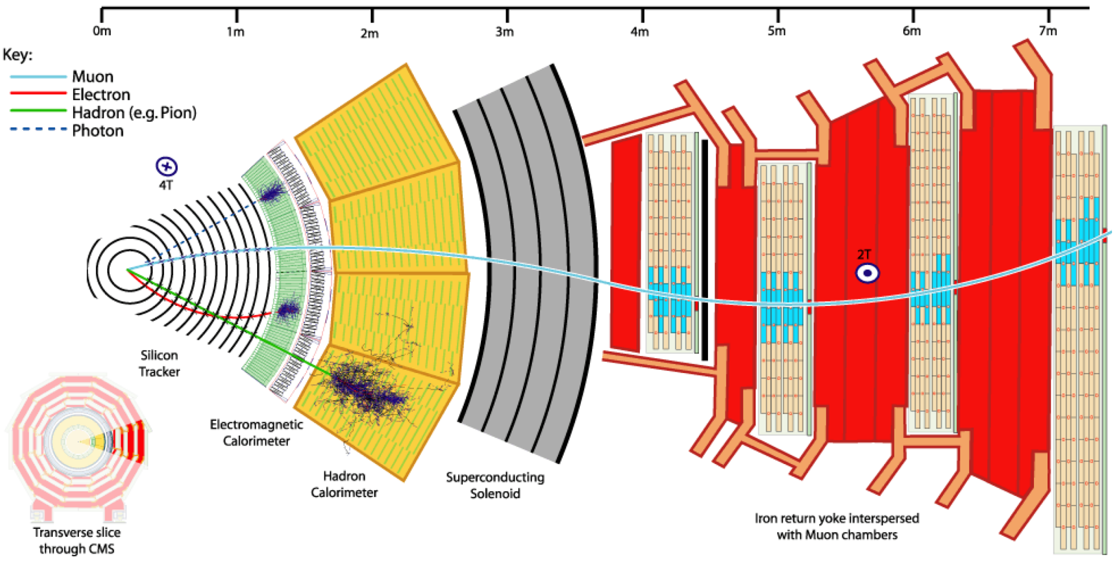



Figure 1.5: Schematic representation of a transverse slice of the CMS detector, and how particles interact with the different components of the detector. Image taken from [14].

thesis, where the four-muon final states will be used to reconstruct the Higgs boson. The CMS muon system consists of three different devices: drift tubes (DT), cathode strip chambers (CSCs) and resistive plate chambers (RPCs). A DT detector is comprised of drift cells. These cells are filled with a mixture of Ar and $CO_2$ gas, and contain a metal anode wire at the center. When a muon travels through this gas it leaves behind a track of electrons produced from the ionisation of the gas. These electrons then *drift* along the electric field lines towards the central wire. The information of where along the wire the electrons hit and the drift time of the electrons can be combined for several of these DTs to obtain a muon track. The cathode strip chambers are composed of seven trapezoidal cathode strip panels which form six gas gaps. Inside each of these gaps is positioned a plane of anode wires, perpendicular to the cathode strips. Its operation is similar to that of a DT. Finally, the resistive plate chamber is a detector consisting of two high-resistivity parallel plates, separated by a thin volume of gas. Electrodes on either side of the plates generate a large voltage in between. External detecting strips pick up the electrons produced by the muon track. The small gap between the plates allows for a good time resolution, contrary to the DTs and CSCs where the time resolution is constrained by the drift time of the electrons. Taken together, the three types of muon detectors provide the CMS detector with an excellent spacial and time resolution for the muon tracks, and allows for the high-quality reconstruction of muons.

**Particle identification**

The interactions of particles with a transverse slice of the CMS detector is visualised in figure 1.5. Particles can be identified by combining the information from the different components of the CMS detector. Charged particles have their trajectories mapped by the inner tracker, but this is not the case for neutral particles. Neutrinos are the only neutral leptons in the SM, but they rarely interact with matter and generally pass through the entire detector unaffected. Neutral hadrons however can also be produced, and will be stopped by the hadronic calorimeter. Thus, an energy deposit in the HCAL that does not match any of the trajectories in the tracker can be identified as belonging to a neutral hadron. Photons can be found in the same way, but this time by looking for energy deposits in the ECAL. For the identification of charged particles, the information from the inner tracker is crucial, since the direction in which a charged particle curves under the influence of the CMS magnetic field yields the charge sign of the particle. Electrons can then be identified for example by matching an energy deposit in the ECAL with a charged particle trajectory in the silicon tracker. Similarly, muons are identified by matching trajectories in the tracker with hits or tracks in the exterior muon chambers.

#### Trigger system

The LHC provides the CMS detector with an event rate of 40 MHz. It would be impossible to store and process the data supplied by the tracker, calorimeters and muon chambers for each event. As a consequence, the event rate needs to be reduced significantly without throwing out *interesting events*. This challenge is handled by the CMS trigger system. The 40 MHz detector output is reduced to a bandwidth of 100 kHz by the first level (L1) trigger. Pipelines provide a delay of around 3 $\mu$s in which the L1 trigger has to decide whether the event is interesting or not. This decision is made using only information from the calorimeters and the muon system. The silicon tracker supplies a large amount of data and requires complex algorithms in order to infer meaningful information from the data, which would take up too much time of the small (3 $\mu$s) window in which the L1 trigger has to operate. At this stage, only basic calculations can be made by the hardware components which are partly housed on the detector itself. When an event is deemed interesting by the L1 trigger (L1 accept), the readout process initiates, and the high-resolution data is transferred from the electronic components of the CMS detector to the High-Level-Trigger (HLT) processor farm. The HLT relies on software to execute complex algorithms using the complete read-out data. Here the event rate is decreased further to around 0.5 kHz before data storage.

## 1.3 Current status of Yukawa coupling measurements

### 1.3.1 The third-generation fermions

As described in section 1.1.4, the SM predicts a linear relationship between the mass $m$ of a fermion and the value of its Yukawa coupling $y$. In other words, fermions of higher mass will couple more strongly to the Higgs boson than lighter fermions. The SM however predicts neither the values of the fermion masses nor the values of the fermion Yukawa couplings. These parameters of the SM theory need to be established

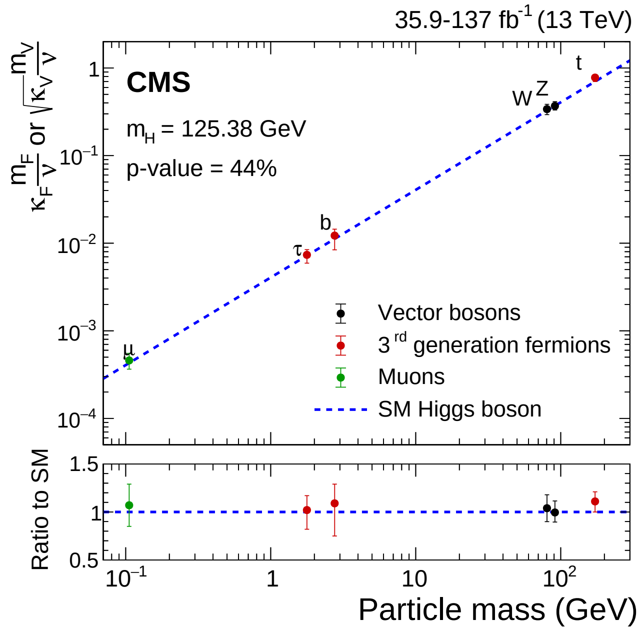



Figure 1.6: Yukawa couplings of the the third generation fermions and vector bosons compared to the SM prediction. The error bars represent 68% confidence level intervals. The lower panel shows the ratio of the measured couplings to the SM prediction. Image taken from [16]

through experimental means. At the time of writing, the Yukawa couplings of the third-generation of fermions have been found to be compatible with the SM prediction. The current estimates for the third generation fermion Yukawa couplings and the muon Yukawa coupling are shown in figure 1.6. This figure shows the values and 68% confidence level intervals of the coupling modifiers $\kappa_W, \kappa_Z, \kappa_t, \kappa_\tau, \kappa_b$ and $\kappa_\mu$. The $\kappa$ parameters are the measured coupling strength divided by the expected coupling strength in the SM, or $\kappa = y/y_{\text{SM}}$. The results in figure 1.6 were obtained through a $(M, \varepsilon)$ fit on a combined analysis of Higgs boson decay modes [15], where the previously mentioned parameters are related to the fermion and vector boson masses by

$$\kappa_{\text{f}} = v\frac{m_f}{M^{1+\varepsilon}} \quad \text{and} \quad \kappa_{\text{V}} = v\frac{m_V^{2\varepsilon}}{M^{1+2\varepsilon}}. \tag{1.12}$$

Here, $v = 246.22$ GeV is the vacuum expectation value of the Higgs field. $\kappa_f$ refers to fermions and $\kappa_V$ refers to the vector (W and Z) bosons. Note that, when $(M, \varepsilon) = (v, 0)$, the SM expectation $\kappa_i = 1$ is obtained. The Higgs decay to two fermions, $H \to f\bar{f}$ is directly sensitive to the fermion Yukawa coupling $y_f$. For a SM Higgs boson with a mass $m_H = 125.38 \pm 0.14$ GeV, which is currently the most precise measurement of $m_H$ [17], the highest expected decay branching fraction is expected for the Higgs boson decay into a bottom quark-antiquark pair, $H \to b\bar{b}$, with a branching fraction of $\mathcal{B}(H \to b\bar{b}) \approx 57.8\%$ [18]. The coupling modifier $\kappa_b$ in figure 1.6 was studied by observing the decay $H \to b\bar{b}$

of Higgs bosons produced in association with a vector boson [19]. Through the use of b-tagging algorithms, the large background of QCD can be rejected in order to retain a high enough signal sensitivity. Additionally, Higgs boson production in association with a vector boson gives an increased signal purity compared to the inclusive Higgs production case.
For the coupling modifier $\kappa_t$ of the top quark, a different channel was used. Since $m_H < 2m_t$, the Higgs boson cannot decay directly into a pair of top quarks. Therefore, the production of a Higgs boson in association with two top quarks [20] is used in the combined analysis. This production rate is directly sensitive to the value of $y_t$.
The coupling modifier $\kappa_Z$ can be studied through the $H \to ZZ^* \to 4\ell$ decay channel [15]. Despite its low branching fraction $\mathcal{B}(H \to ZZ^* \to 4\ell) \sim 0.03\%$, the four lepton decay channel provides a clean signal to be distinguished from other background processes. Only backgrounds which produce purely leptonic final states need to be taken into account, therefore a low background contamination is expected for this decay channel. Likewise, the $H \to WW^*$ decay channel [15] also has a low background final state with two leptons and two neutrinos. Additionally, this decay channel benefits from having the second highest expected branching fraction, even though $m_H < 2m_W$. One of the W bosons is produced off-shell, but this effect is countered by the high coupling strength of the W boson to the Higgs boson, which elevates the branching fraction to 21.6%. Another coupling modifier included in the fit is $\kappa_\tau$. Similar to the $H \to WW^*$ channel, the analysis of the $H \to \tau^-\tau^+$ decay channel also profits from having a relatively high expected branching fraction (6.4%), since the tau lepton is the most massive lepton. The measured value of its mass is $m_\tau = 1776.86 \pm 0.12$ MeV [18], which means that the tau lepton is the only lepton capable of decaying hadronically. Approximately 65% of tau lepton decays are into charged or neutral pions, which are observed as narrow, collimated jets containing few particles. In the combined analysis [15], four final states of the $H \to \tau^-\tau^+$ decay channel are used, including $e\mu, e\tau_h, \mu\tau_h$ and $\tau_h\tau_h$, where $\tau_h$ represents a tau lepton that underwent a hadronic decay.
Lastly, the $\kappa_\mu$ coupling modifier is studied via the $H \to \mu^+\mu^-$ decay channel. As was the case in measuring $y_Z$, the muon decay channel suffers from a low expected branching fraction of $\mathcal{B}(H \to \mu^+\mu^-) \approx 0.02\%$. The analysis includes the final state with two isolated and oppositely charged muons, and searches for a peak in the dimuon invariant mass spectrum around the mass of the Higgs boson. The results in figure 1.6 include the recent evidence for the observation of a Higgs boson decay to a pair of muons [16]. This concludes the discussion of the state-of-the-art coupling measurements of the third generation fermions, vector bosons and the muon. In the next section, the current constraints on the coupling of the charm quark will be discussed.

### 1.3.2 Measurements of the charm quark Yukawa coupling

From the direct search for Higgs boson decay to a bottom quark-antiquark pair [19, 21], and the production of a Higgs boson in association with a pair of top quarks [20, 22], the CMS and ATLAS Collaborations have found the Yukawa couplings of the third-generation quarks to be compatible with the SM prediction. Recently, in the direct search

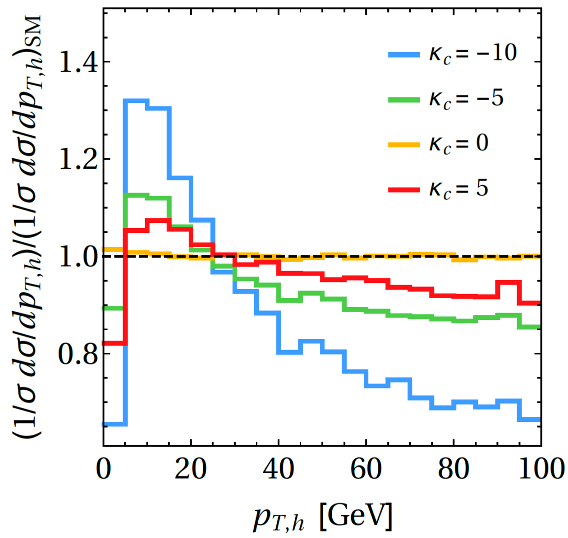



Figure 1.7: Normalised Higgs boson transverse momentum $(p_{T,h})$ distribution divided by the SM prediction for different values of $\kappa_c$. Image taken from [23].

for Higgs boson decay into a pair of muons, the CMS Collaboration found the coupling to the muon to also be consistent with the SM [16]. Apart from increasing the sensitivity on the above measurements, the next objective is to measure the Yukawa coupling of the charm quark. This will however be significantly more challenging compared to the measurement of the top and bottom quark couplings. The third-generation quarks are considerably heavier than the charm quark, and are therefore expected to have more interactions with the Higgs boson. The SM branching fraction of the Higgs boson decaying to a charm quark-antiquark pair is $\mathcal{B}(H \to c\bar{c}) \approx 3\%$, which is relatively small, especially when taking into account the large expected background from QCD multijet events. Additionally, this decay channel requires the reconstruction and identification of two jets formed by charm quarks, also known as charm jets. Distinguishing charm jets from jets that were formed by other partons, such as bottom quarks, is a challenging problem for which algorithms called c-taggers (or more general heavy-flavour taggers) are constructed. These algorithms are discussed further in section 3.4.

The first direct bound on the coupling of the charm quark to the Higgs boson was found by recasting the ATLAS and CMS findings on the $H \to b\bar{b}$ channel at 7 and 8 TeV, corresponding to an integrated luminosity of 5 and 20 fb$^{-1}$ [24] respectively. The approach here was to exploit the fact that jets originating from charm quarks may have been mistagged as bottom-jets. Hence, the existing analysis on the $H \to b\bar{b}$ could be recast to obtain similar information on the $H \to c\bar{c}$ rate. This resulted in a model-independent upper limit of $\kappa_c \lesssim 234$ at 95% confidence level (CL). In another approach, the bottom and charm quark couplings were probed by exploiting the transverse momentum distributions in Higgs production [23]. This method targeted Higgs plus jet events $(H+j)$, which receives contributions from several channels such as gluon fusion $(gg \to H+j)$ and production in association with a parton $(qg \to Hq$ and $q\bar{q} \to Hg)$. The sensitivity of the Higgs boson transverse momentum distribution to the modification of $\kappa_c$ is shown in figure 1.7. Using Run-2 data (2016) corresponding to an integrated luminosity of 35.9 fb$^{-1}$ at 13 TeV, the CMS Collaboration obtained a 95% confidence level (CL) interval $\kappa_c \in [-33, 38]$ using differential cross section measurements [25]. A direct search for the $H \to c\bar{c}$ channel was carried out by the CMS Collaboration, using proton-proton collision data corresponding to an integrated luminosity of 35.9 fb$^{-1}$, collected at a centre-of-mass energy of 13 TeV at the LHC [26]. In this analysis, a strict limit is imposed on the signal strength $\mu$, defined as

$$\mu = \frac{\sigma(\mathrm{VH}) \times \mathcal{B}(H \to c\bar{c})}{(\sigma(\mathrm{VH}) \times \mathcal{B}(H \to c\bar{c}))_{\mathrm{SM}}}. \tag{1.13}$$

The analysis targets Higgs bosons produced in association with a vector boson, the presence of which suppresses QCD multijet backgrounds. An observed (expected) upper limit on $\mu_{\mathrm{VH}(H\to c\bar{c})}$ was found to be 70 $(37^{+16}_{-11})$ at 95% confidence level (CL). A less sensitive limit was found by the ATLAS Collaboration using 36.1 fb$^{-1}$ of data collected with the ATLAS detector at 13 TeV [27], corresponding with an observed (expected) upper limit on $\mu_{\mathrm{VH}(H\to c\bar{c})}$ of 110 $(150^{+80}_{-40})$ at 95% CL.

## 1.4 Proposal of a new method to probe $\mathbf{y_c}$

In the previous sections it was mentioned that the direct detection of the Higgs boson decaying to a charm quark-antiquark pair $H \to c\bar{c}$ is extremely challenging. The channel suffers from a relatively low branching probability and a high background from QCD multijet events, and relies on the reconstruction of two charm-tagged jets. A new method is therefore proposed which relies on the observation of a Higgs boson in association with a charm-tagged jet $(gc \to Hc)$. This method has two main advantages. The first is that it relies on the reconstruction of one charm-tagged jet as opposed to two in the direct search. As a result, a higher purity charm-tagging algorithm can be used to further reduce the background from light and bottom jets, compared to the direct search method. The second, and perhaps the biggest advantage, comes from the fact that the sensitivity to $y_c$ does not originate from the Higgs boson decay, but from its production. Therefore, the Higgs boson can be reconstructed from a much cleaner decay mode, such as $H \to ZZ^* \to 4\ell$. As a consequence, the non-Higgs background is significantly reduced.

In the SM, there are three leading-order diagrams which contribute to the production of a Higgs boson in association with a charm quark. These are shown in figure 1.8. Only the two diagrams on the left contain vertices where a direct $y_c$-dependence enters, highlighted by the solid orange dots. The diagram on the right does however contain an effective vertex, where the quark triangle loop has been integrated out. In practice, it is true that charm quarks could also run in this loop and introduce a $y_c$-dependence,

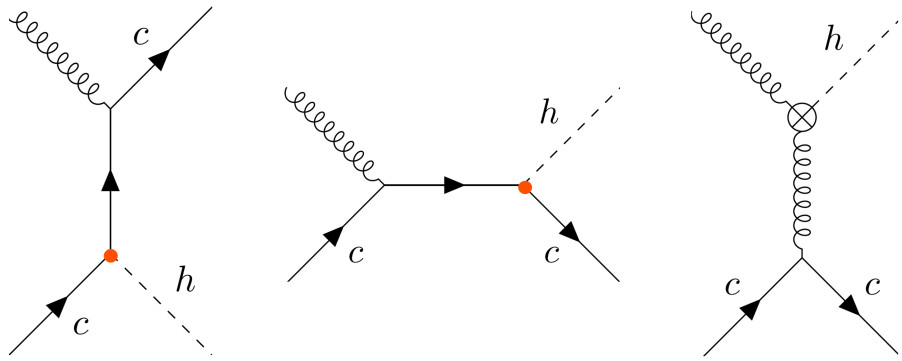



Figure 1.8: Leading-order diagrams contributing to the process $gc \to Hc$. The charm quark Yukawa coupling $y_c$ enters in the vertices highlighted by the solid orange dot.

but since the charm quark is much lighter compared to the top and bottom quarks, this contribution is expected to be relatively insignificant. Furthermore, because of the top quarks running in the loop, this diagram is expected to form the dominant contribution to the $H + c$ production cross section when $\kappa_c = 1$.

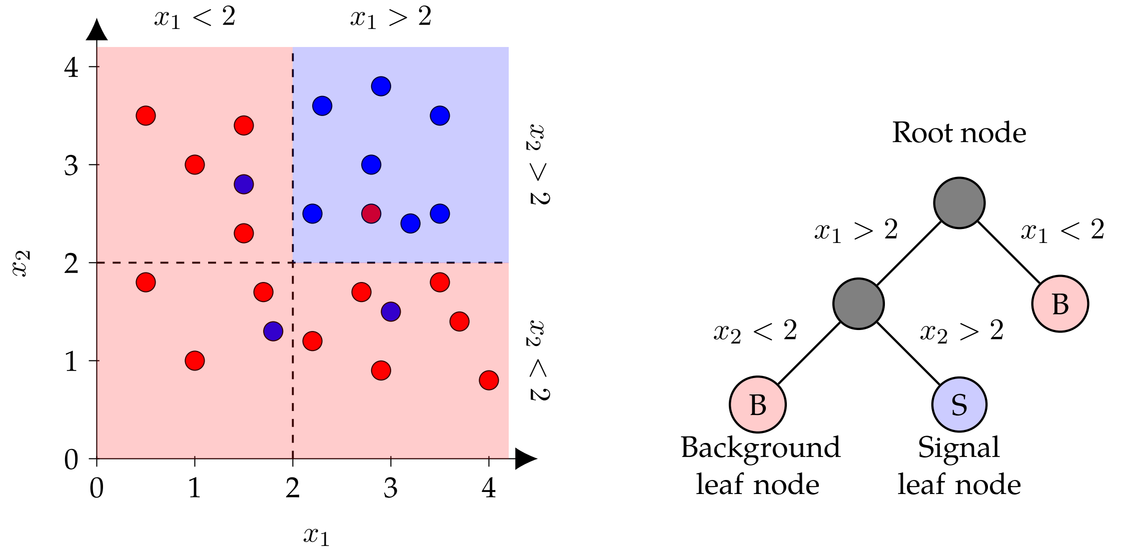
In this thesis, the above process in which a Higgs boson is produced together with a charm quark will be simulated up to the detector level. After applying an event selection, an artificial neural network (ANN) will be trained in order to further reduce the background from events containing a pair of Z bosons that were not produced in the decay of a Higgs boson. The outputs of this ANN will also be transformed into the `Hc` discriminator, which is a newly constructed observable with an optimised sensitivity to the charm quark Yukawa coupling $y_c$. A constraint will then be placed on $\kappa_c = y_c/y_c^{\mathrm{SM}}$ using a binned maximum likelihood fit to the distribution of this new observable. The result will be compared to the constraints that are obtained using only the reconstructed transverse momentum of the Higgs boson candidate or the transverse momentum of the most energetic jet, which are variables that were used in previous analyses [23, 28] to constrain $\kappa_c$. In the next chapter, a general overview of ML algorithms and techniques will be given, with a particular focus on artificial neural networks.

# Chapter 2

# Machine learning and artificial neural networks

## 2.1 The concept of machine learning

The field of machine learning (ML) involves the study of computer algorithms that are capable of optimising themselves by learning from sample data. ML is seen as a subfield of artificial intelligence (AI), since the relevant computer algorithms mimic the way humans learn and therefore possess certain intelligence. To be more precise, a ML algorithm constructs a mathematical model from a training data sample in order to make predictions on a sample that it has not seen before. Different approaches can be used to train a ML algorithm, most notably supervised and unsupervised learning. In the former approach, the algorithm learns by comparing its output to the desired (true) output, whereas in the latter approach, this desired output is not available. Instead, in unsupervised learning, the ML algorithm attempts to recognize patterns or certain structures in the input data. One type of ML algorithms is the regression algorithm, which tries to predict the value of a continuous variable rather than a discrete categorization. A straightforward application of a regression model is in curve fitting, where the model attempts to find a continuous function that most adequately describes the distributed data. Particularly relevant to this thesis are classification algorithms or classifiers. The goal of a classifier is to predict in which category a certain event belongs, given the input data. An example of this is heavy-flavour tagging, where a classifier algorithm is constructed to label reconstructed jets according to the flavour of the quark from which the jet most likely originated. Generally, the output of a classifier can be interpreted as a probability for the event to belong to a certain category, for each of the possible output classes. If the algorithm is only allowed to decide between two complementary classes, then it is referred to as a binary classifier. Different ML approaches with varying levels of complexity can be used to tackle a classification problem. In the following section, models such as the decision tree and the artificial neural network will be described.




(a) (b)

Figure 2.1: (a) Qualitative example of a two-dimensional training sample used by a binary decision tree to distinguish between signal (blue) and background (red). (b) Visualisation of the sequence of binary logic that is imposed on the input vector to classify an event in a binary decision tree classifier.

## 2.2 The decision tree

A decision tree classifier is a ML algorithm that uses boolean logic to categorize data. To be more precise, a decision tree takes an input vector $\mathbf{x} = (x_1, x_2, \ldots, x_D)^{\mathrm{T}}$ of $D$ variables $x_i$, and assigns it to one of $N$ classes by passing the variables through a series of *decisions*. A binary decision tree, i.e. a decision tree for which $N = 2$, will be used to demonstrate the workings of this algorithm. Suppose the input vector only contains two variables, $x_1$ and $x_2$. The training sample can then be visualised as in figure 2.1a. The events that are labelled red represent the background, while the events that are labelled blue represent the signal. From this training sample, it is clear that most of the signal events are contained within the region defined by $x_1 \geq 2$ and $x_2 > 2$. The binary decision tree will attempt to find this region by imposing cuts on the input data of the training sample, such that an optimal separation between signal and background is achieved. After the decision tree has been constructed, it can be visualised as in figure 2.1b. At the root node, the entire training sample is split into two subsets by comparing the variable $x_1$ to the threshold value $T = 2$. The leaf nodes, which are the outermost nodes, are responsible for labelling the event as being either signal or background. In this example, the decision tree obtains a signal purity of 88% (in the training data) in the blue leaf node, which corresponds with the top-right region in the training sample graph. Any input data that has not been seen before by the binary decision tree classifier will undergo the same series of cuts that were imposed on the training sample. These new events will then be labelled depending on the leaf node they end up in. Both the chosen variable that is used to split the training sample and the corresponding threshold value at a given node are all trainable parameters which will be optimized during the training procedure.

The previous example offers a simple qualitative explanation for the workings of the decision tree. In most situations however, the training sample will be higher-dimensional and the structure of the decision tree will become more complex. In these situations, it becomes increasingly important to guide the construction of the decision tree by imposing certain restrictions on its growth. The tree can for example be required to not grow further than a set number of layers known as the tree depth. The final number of leaf nodes can also be constrained for this purpose. If the values of these parameters are not chosen optimally, the decision tree could become overtrained, a phenomenon that will be discussed further in section 2.5. The decision tree classifier can be optimised further through a method known as boosting. Instead of a single tree, an ensemble (called a forest) of multiple shallow decision trees is constructed. These trees are trained sequentially on weighted training samples, where higher weights are assigned to events that were incorrectly classified by the previous trees. The resulting tree is called a boosted decision tree (BDT).

## 2.3 Artificial neural networks

### 2.3.1 Single-layer perceptron

An artificial neural network (ANN) is a model composed of interconnected nodes or neurons, the function of which is inspired by neurons in biological brains. The concept of the artificial neuron was first developed in the 1950s and 1960s by Frank Rosenblatt [29], who developed the perceptron, a mathematical model that approximates the function of a biological neuron. The perceptron takes in a number $N$ of binary inputs, and returns a single binary output. To be precise, the input vector $\mathbf{x}$ is composed of $N$ variables $x_i \in \{0, 1\}, i \in \{1, \ldots, N\}$. Each of the input variables is associated with a weight $w_i$, which determines how important the respective input is to the output. These weights can be subject to change to allow for further optimisation of the perceptron. Essentially, the perceptron computes a weighted sum of the input variables,

$$S = \sum_{i=1}^{N} w_i x_i, \tag{2.1}$$

which is then compared to a threshold value $T$. The output of the perceptron is then generalised by

$$\text{output} = \begin{cases} 0 & \text{if } S \leq T \\ 1 & \text{if } S > T \end{cases} \tag{2.2}$$

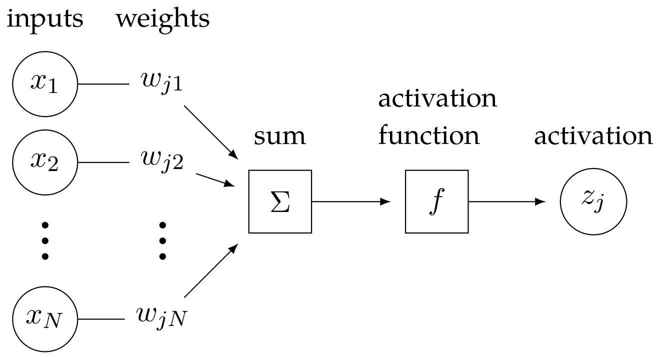



Figure 2.2: Diagram showing how the activation of a neuron in the first layer of an artificial neural network is calculated from the input variables and the associated weights.

In other words, the perceptron only outputs a '1' if the value of the weighted sum exceeds the threshold. This behaviour is analogous to a biological neuron outputting an electrical signal shaped as a voltage spike, in response to the receipt of similar signals from connections to neighbouring neurons. For this reason, another name for the perceptron is the artificial neuron. Since the only computation performed by the perceptron is the calculation of a weighted sum, it can only be applied successfully to linearly separable problems, i.e. problems where the signal and background can be separated by a single hyperplane. More complex problems can be handled by a network of artificial neurons, where the output of one layer of perceptrons is fed as an input to a subsequent layer of perceptrons. Non-linear behaviour can be introduced by a specific choice of the function that transforms the weighted sum into the output value of the artificial neuron. These functions are called the activation functions.

### 2.3.2 Construction of an artificial neural network

The perceptron in the previous section is one example of an artificial neuron. The output of an artificial neuron can be generalised by introducing the activation function $f$. This function is responsible for transforming the weighted sum into the output value of the neuron, the activation $z$. The conditional output of the single-layer perceptron is produced by the Heaviside step function $\theta(x)$, which is zero for negative arguments and one for positive arguments. In general, the activation $z$ of an artificial neuron is given by

$$z = f\left(\sum_{i=1}^{N} w_i x_i - T\right). \tag{2.3}$$

The threshold value $T$ is also referred to as the bias of the neuron. If chosen high enough, the neuron will mostly output negative values, which in the case of the Heaviside step function results in an output of zero. In the case of the single-layer perceptron, the bias term prevents the artificial neuron from *firing* unless the weighted sum constructed from the input variables has a high enough value to overcome this negative bias. Figure 2.2

shows the general form of the activation $z_j$ which is calculated from the input variables. A traditional feed-forward neural network is constructed out of layers of interconnected neurons. The first layer is known as the input layer, and features the variables $x_i$ of the input vector $\mathbf{x}$. Each of the neurons in the input layer simply outputs the value of the corresponding input variable to each of the neurons in the next layer. The layers in between the input and output layers of a neural network are known as the hidden layers. In the first hidden layer, the inputs are simply the weighted sums of the input variables. Each neuron in the hidden layer has a different set of weights associated with it. The activation $z_j$ of neuron $j$ in the first hidden layer can be calculated by

$$z_j = f\left(\sum_{i=1}^{N} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right), \tag{2.4}$$

with $N$ the dimension of the input vector $\mathbf{x}$. The weights of the first hidden layer can be represented by a weights matrix $\mathbf{W}^{(1)}$, where the entries in column $j$ represents the set of weights used by the $j$-th neuron in the first hidden layer. The threshold value $T$ has been replaced by an additional entry in the weights matrix for each hidden neuron. The notation can be simplified by introducing a dummy variable $x_0$ which is always equal to one. If we write $\mathbf{z} = (z_0, z_1, \ldots, z_M)^{\mathrm{T}}$, with $M$ the number of neurons in the first hidden layer, then equation (2.4) becomes

$$\mathbf{z} = f\left(\mathbf{W}^{(1)} \cdot \mathbf{x}\right), \tag{2.5}$$

In this notation, the activation function acts on each of the elements of the vector resulting from the dot product. The value of weight $w_{ji}$ can be interpreted as the strength of the connection between input neuron $i$ and hidden neuron $j$. In other words, a larger value of $w_{ji}$ indicates that the variable $x_i$ of the input vector $\mathbf{x}$ contributes more towards the activation of the hidden neuron $j$. The activations of the neurons in a second hidden layer can be calculated in the same manner. Instead of taking the vector $\mathbf{x}$ as input, the neurons in the second hidden layer use the activations of the neurons in the first hidden layer. Another weighted sum is computed, using another weights matrix $\mathbf{W}^{(2)}$ which represents the strength of the connections between the neurons of the first and second hidden layer. The output of the artificial neural network is calculated by the neurons in the output layer. The number of neurons in the output layer is equal to the number of categories that are required to tackle a certain classification problem. An example of a neural network with a single hidden layer is given in figure 2.3. The output values $y_k$ are often normalised by the final activation function, which can be different from the activation function that has been used throughout the network, such that $\sum_k y_k = 1$. Additionally, the output values are usually re-scaled to the interval [0,1]. This allows for a probabilistic interpretation of the network's output variables. For instance, the activation $y_i$ of the neuron $i$ in the output layer can be interpreted as the probability for the input event to belong to the $i$-th category. If $y_k$ is close to unity, then the ANN is very confident that the event belongs to the corresponding class. An example of an activation

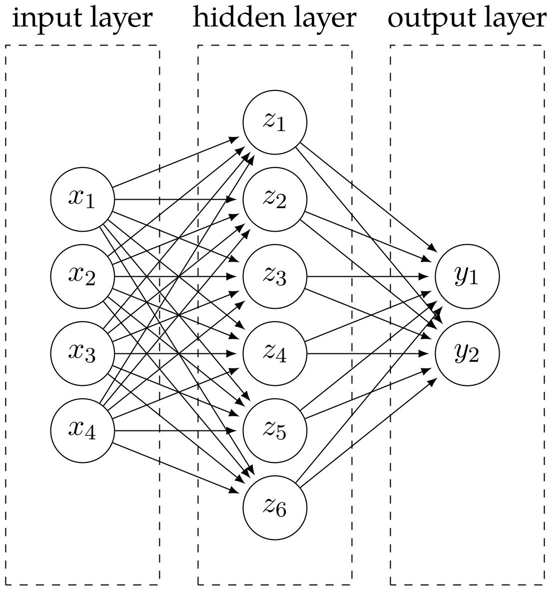



Figure 2.3: A diagram of an artificial neural network with one hidden layer containing six neurons and an output layer containing two neurons. This neural network takes four variables as an input.

function is the Rectifier Linear Unit or ReLu function, which is defined as

$$f_{\text{ReLu}}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \tag{2.6}$$

In other words, this function returns the argument if it is positive, and returns zero if this is not the case. Another activation function that is typically used in the output layer is the softmax function, which is defined as

$$f_{\text{SoftMax}}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{M} e^{z_j}}, \tag{2.7}$$

where $M$ is the number of neurons in the output layer. The term in the denominator ensures that the outputs of the network sum up to one.

### 2.3.3 Training a neural network

In supervised training, every event in the training sample has been labelled with the true category that the event belongs to. The performance of the ANN can be evaluated by comparing its predictions on a training set of input vectors $\{\mathbf{x}_n\}$ with the corresponding set of target vectors $\{\mathbf{t}_n\}$. Suppose the ANN is tasked with classifying an event in one of $N$ categories, with the event truly belonging to the $j$-th class. If the output of the ANN is the vector $\mathbf{y} = (y_1, y_2 \ldots, y_N)^{\text{T}}$, then the elements $t_i$ of the corresponding target vector

$\mathbf{t}$ are all equal to zero, except for $t_{i=j}$ which is set to one. Optimally, for this event, the neurons in the output layer should produce values that are somewhat in line with this target vector.

### The loss function

One way to quantify the performance of an ANN on a labelled sample is to introduce a loss function $L(\mathbf{w})$, where $\mathbf{w}$ represents the entire set of weights associated with the neural network. An intuitive version of such a loss function is the squared-error loss function, given by

$$L(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{M} \|\mathbf{y}_n - \mathbf{t}_n\|^2, \tag{2.8}$$

where $\mathbf{y}_n$ represents the output vector of the network in response to the input vector $\mathbf{x}_n$. If the difference between the target vector $\mathbf{t}_n$ and the network prediction $\mathbf{y}_n$ is small, then the norm of this difference will also be small. This corresponds with the network predicting a value close to one in the true category, labelled by the target vector, and values close to zero in all other categories. In equation (2.8), the squared differences between the target and output vector are calculated for each event in the labelled sample, and are then summed. Therefore, lower values of $L(\mathbf{w})$ are associated with a network that has a higher performance on the labelled sample. Another loss function is the categorical cross entropy function, which will be used to train the ANN in section 4.3.1. This loss function is defined by

$$L = -\sum_{n=1}^{N}\sum_{i=1}^{4}(\mathbf{t}_n)_i \log((\mathbf{y}_n)_i), \tag{2.9}$$

where $(\cdot)_i$ signifies the $i$-th component of a vector.

### Optimisation of the weights

For a fixed architecture, the performance of an artificial neural network is entirely dependent on the weights associated with each layer. As a consequence, the problem of training a neural network becomes a problem of finding a set of weights $\mathbf{w}$ that minimises the loss function. Because of the highly non-linear dependence on the weights, numerical approaches need to be used to find a solution to this optimisation problem. One approach is that of gradient descent, where the weights are updated in the direction of the negative gradient of the loss function,

$$\mathbf{w}' = \mathbf{w} - \eta\nabla L(\mathbf{w}), \tag{2.10}$$

where $\eta$ is the learning rate, which should be chosen small enough such that the weights are updated in small steps towards the minimum of the optimisation problem. Each cycle during which the ANN processes all the training data and updates its weights accordingly is known as an epoch. One can also split the training sample into smaller batches, enforcing an update of the weights after each batch has been processed. As a result, the weights are updated multiple times in each epoch, before reusing the same data again in the next epoch.

**Backpropagation**

Calculating the gradient in equation (2.10) is certainly a challenge. One approach which is widely used in training is backpropagation. The general idea behind backpropagation is to introduce a small change in the weights of the network layer-by-layer, observe the effect on the result of the loss function, and then update the weights accordingly. For a given neuron, the loss function is dependent on the weights corresponding to that neuron, and the weights of any neurons in the subsequent layers including the output layer. The problem simplifies when one starts out at the final hidden layer. Here, the loss function's dependence on the next layer's neurons is known, since the next layer is simply the output layer. By varying the weights of each neuron in the final hidden layer, the neural network makes slightly different predictions which are then compared to the targets by the loss function. The weights associated with these neurons are then updated such that lower values of the loss function are obtained. Now that the dependence of the loss function on these neurons is known, one can move back one layer and repeat the same process. This process, where one runs through the layers of the neural network in reverse and updates the weights layer-by-layer, is referred to as backpropagation. It is discussed in more detail in appendix A.

## 2.4 Receiver operator characteristic curve

The performance of a binary classification algorithm can be visualised using a receiver operating characteristic (ROC) curve. Here, the true positive rate (TPR) or signal selection efficiency is plotted against the false positive rate (FPR) or background acceptance for different discrimination thresholds. These quantities can be calculated after the classifier has made a series of predictions on a sample that contains events that were not used in training. The binary classifier outputs a probability $P$ for the event to belong to the signal class, with the complement $1 - P$ representing the probability for the event belonging to the background class. As mentioned before, these outputs are typically continuously distributed over the interval $[0, 1]$, contrary to the perceptron which outputs either one or zero. The output of the perceptron is straightforward to interpret: the event either belongs to a certain class or it does not. In general classifiers, this is not the case. Instead a threshold $T$ needs to defined, such that an event is predicted to belong to the signal class if $P > T$. The choice of $T$ is arbitrary, and different choices produce different performances of the classifier. When a signal event is classified as signal, then the prediction of the classifier is a true positive (TP). If that event is classified as background instead, then the prediction is a false negative (FN), which is also called a type II error. Likewise, a false positive (FP) prediction (or type I error) corresponds with a background event being classified as signal, while a true negative (TN) prediction means the background event was classified correctly. Using this terminology, the signal selection efficiency $\varepsilon$ can be defined as

$$\varepsilon = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}. \tag{2.11}$$

The signal selection efficiency $\varepsilon$ is also known as the probability of detection, since it is defined as the fraction between the number of correctly classified signal events and the total number of signal events. In the same way, the specificity or background rejection $r$ can be defined as the fraction between the number of correctly identified background events and the total number of background events. Relevant to the creation of the ROC curve is the background acceptance $1 - r$, defined by

$$1 - r = \frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{TN}}}. \tag{2.12}$$

The background acceptance is also known as the probability of false alarm, since it keeps track of the number of background events that were incorrectly classified as belonging to the signal class. For each value of $T \in [0, 1]$ the signal efficiency $\varepsilon$ and the background acceptance $1-r$ can be calculated. The ROC curve is then generated by plotting the points $(\varepsilon, 1 - r)$ for the different selection thresholds. The performance of a classifier can be quantified by introducing the area under the curve or AUC-value. A better performance is associated with classifiers whose ROC curves lie closer to the lower-right corner $(1, 0)$ of the $(\varepsilon, 1 - r)$ plot, since this point corresponds with a classifier that identifies 100% of the signal events correctly, with not a single background event being misidentified. Points on the diagonal $\varepsilon = 1 - r$ correspond with classifiers that are no different from a classifier which randomly distributes the events, i.e. a classifier that uses the result of a coin-toss to decide to which class an event belongs.

In the case of a multi-class or non-binary classifier, a ROC curve can still be used to visualise the performance of the model. In the general case of $N$ output classes, one can study the performance of the classifier for distinguishing between two classes $i$ and $j$ by constructing a binary discriminator

$$\text{i vs. j discriminator } = \frac{\mathrm{P}(i)}{\mathrm{P}(i) + \mathrm{P}(j)}. \tag{2.13}$$

This discriminator is naturally also distributed between 0 and 1, and can be used to create the ROC curve for distinguishing between these two classes. A multi-class classifier can also be converted into binary-classifier by constructing a ”1 vs. all” discriminator. Here, the performance of the classifier can be studied for distinguishing between a single output category $i$ and all other output categories. In this case the discriminator is simply defined as $\mathrm{P}(i)$.

## 2.5 Overtraining

An important concept in ML is that of overtraining or overfitting. In general, a model is overtrained when it has learned unwanted behaviour. An example of this is when a model learns the training data by heart instead of identifying any underlying patterns. This behaviour can show up in any supervised training problem. In a regression problem, such as one-dimensional curve fitting, the model attempts to find a function that most adequately describes the observed distribution of a continuous variable. These observations will have some inherent noise, which causes them to diverge from their *true* distribution. In extreme cases, an overfitted function will pass through each of the data points, and this way obtain a perfect match with the observations. Here, the model has decided to fit the noise of the data instead of finding the underlying distribution. The performance of this model on the training sample is perfect, but when another set of observations is pulled from the same distribution and is compared with the overfitted model, the performance drastically decreases. This is the defining signature of an overtrained model. For general ML algorithms, this can be quantified by introducing the loss and accuracy curves on a training and an independent validation sample. On the training sample, the value of the loss function should decrease after each subsequent training, while the accuracy should increase. After each training, the performance of the model can be checked in a separate validation sample. Here, the same trend should be observed. An overtrained model would have a very high accuracy and a low value of the loss function on the training sample, but a very low accuracy and a high value of the loss function on the validation sample.

A decision tree can also become overtrained. In extreme cases, the decision tree could grow indefinitely until splitting is no longer possible, and the leave nodes of the decision tree will contain a very small but pure subset of the training sample. This decision tree would have a perfect accuracy in classifying the events of the training sample, but would drastically underperform when confronted with a testing sample containing new events. This overtraining can be prevented by more carefully choosing the parameters that regulate the growth of a decision tree, such as limiting the maximum depth of the tree or requiring a minimal amount of events to be contained in the leaf nodes. Another procedure used to prevent overtraining is *pruning*, where some of the redundant nodes in a decision tree are *cut*, resulting in a less complex decision tree that is less sensitive to overtraining.

A popular way to avoid the overtraining of an artificial neural network is via *dropout*. This method is similar to *pruning* in decision trees. In each training epoch of the ANN, a certain percentage of neurons is randomly *frozen*. These neurons are effectively dropped out of the ANN during that epoch. As a result, the model is prevented from depending too much on a select few neurons, allowing it to optimise the other neurons as well.

# Chapter 3

# Simulation and reconstruction of proton-proton collisions

The Standard Model (SM) of particle physics has withstood decades of experimental verification. Despite the many successes of the SM, many its predictions still need to be tested and refined over an energy range spanning many orders of magnitude, with the highest energies being reached at particle colliders such as the Large Hadron Collider at CERN. At one of the four interaction points on the LHC, the CMS detector attempts to observe and reconstruct the outgoing particles resulting from the generated proton-proton (pp) collisions. In order to compare this experimental data with theory predictions, a detailed understanding of the pp collision process is needed. In the first part of this chapter, a description of this process will be given. The predictions of a theory model on this scattering process are then transformed into the signals that a detector would observe using Monte Carlo simulation techniques. This is discussed in the second part of this chapter. Finally, the methods used by the CMS detector to reconstruct physical objects from detector outputs are described in the final part of this chapter.

## 3.1 Proton-proton collisions

At the LHC, proton-proton collisions are used to test the limits of the SM. It is important that the pp collision process is well understood, such that precision measurements become possible. In the following paragraphs, the pp collision will be described in several steps, from the fundamental process described by theory models such as the SM Lagrangian, to the multitude of outgoing final-state particles that are observed in the detector.

### 3.1.1 The scattering process

Protons are composite particles, composed of two up quarks and one down quark. It is therefore more accurate to think of proton-proton collisions as collisions between their

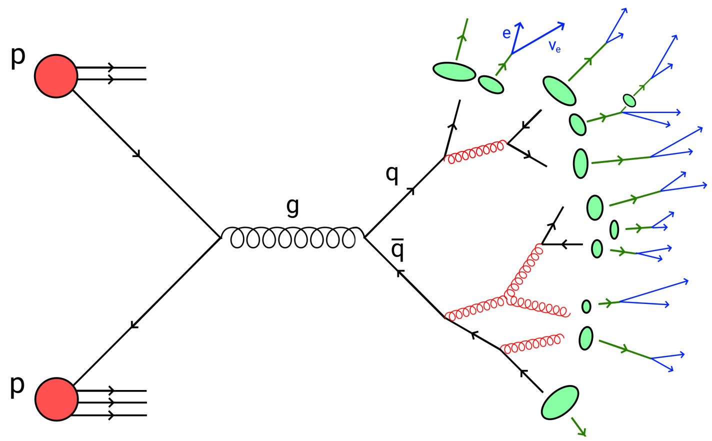



Figure 3.1: Schematic overview of the hard-scattering process in a proton-proton collision, followed by the final-state parton shower and hadronisation process.

constituent quarks. This alone increases the difficulty in describing the collision process, but there is an additional complication: protons are dynamic systems. On top of the up and down quarks, which are referred to as the valence quarks, there is a constant creation and annihilation of quark-antiquark pairs of any flavour within the proton system [30]. Additionally, gluons are exchanged constantly between the valence quarks and the additional sea quarks. It is therefore possible that, in a proton-proton collision, an interaction between two gluons takes place, each originating from one of the colliding protons. The initial state of the interaction process can be described by introducing the parton distribution functions of the proton, which give the probability of finding a parton (quark or gluon) within the proton with a certain fraction of the proton's energy.

#### Parton interactions

An interaction between partons can be broken down into several stages. Before entering the interaction, the partons can radiate gluons or photons. This is referred to as the initial-state radiation (ISR). Similarly, final-state radiation (FSR) can be emitted by the partons which leave the interaction. At sufficiently high energies, a gluon emitted by a quark could split into a quark-antiquark pair or into a pair of gluons, which can then themselves split into another pair of partons. As a result, the initial and final-state radiation can produce a cascade of particles. The shower of particles that results from the emissions from one initial parton is referred to as the parton shower.

#### Underlying event and pile-up

While proton bunches cross in the LHC interaction region, it is possible that more than one proton in the bunch interacts with protons in the crossing bunch. The interaction with the highest energy is the hard scattering event, while the larger number of additional lower-energy collisions are referred to as *pileup*. For the protons participating in the hard scattering, only one parton of each proton will be involved in the hard scattering event. The remnants of the protons will however also give rise to some activity in the detector, which is referred to as the *underlying event*. An example of a hard scattering event is visualized in figure 3.1.

### 3.1.2 Jets

The partons emerging from the hard scattering suffer energy losses as a consequence of their QCD and QED emissions. After a certain energy scale is reached, the particles leaving the parton showering process will start forming hadrons. This is a consequence of the principle of confinement, which states that colored particles cannot be isolated. Even though this process has not yet been described analytically from first principles, it can be described by using phenomenological models, called fragmentation functions, which give the fraction $x$ of the parton energy that is carried away by the hadron. Most of the hadrons formed by this process are unstable and will eventually decay into the particles that are either stable or live long enough to be detected. As a result, the presence of a parton in the final state of an interaction can be observed as a group of particles whose trajectories are roughly aligned. This stream of particles with similar trajectories is referred to as a jet. Any variables associated with the original parton have to be inferred from the properties of the jet. It is therefore important that jets are reconstructed optimally. For this purpose, different algorithms exist which will be explored further in later sections. The hadronisation of a parton is often accompanied by the production of an additional jet, which introduces additional challenges to reconstruct the kinematics of the original parton.

## 3.2 The simulation chain

The previous section highlighted the different steps in which a proton-proton collision is thought to take place. Starting from a theory model, such as the SM or an extension thereof, a differential cross section can be calculated for a process of interest. Most of the time, the particles that leave the fundamental interaction process can not be detected directly. Unstable particles will decay into other particles, gluons and quarks hadronize and form jets, and some particles like neutrinos traverse the detector without any interaction and hence without leaving a direct signature. Therefore, the predictions of a model cannot be compared directly to the results of the experiment. Only by simulating the different physical effects that the particles of interest undergo, from the interaction point all the way to the detector level, a comparison between theory and experiment can be made. These simulations rely on the use of Monte Carlo methods, which are techniques that use random number generators to obtain the stochasticity of physics phenomena, such as generating probability distributions predicted by theory. For the purpose of simulating a pp-collision, each of the steps described in the previous section is handled by a different simulation program. In this section, the methods that are used to simulate the collision process up to the detector level will be described.

### 3.2.1 Matrix element generator

At the heart of the proton-proton collision process lies an interaction between fundamental particles. The details of this interaction, such as the properties of the outgoing particles, are described by theory models such as the SM. A complete description of the interaction, i.e. a transition from an initial state to a final state, is given by the appropriate matrix elements. For simulation purposes, one usually starts out with the calculation of the matrix elements of the process of interest. This step involves the computations of tree-level, next-to-leading order (NLO) and sometimes even higher order (NNLO) cross-sections for this process. One example of a computer program capable of handling these computations is the `MadGraph5_aMC@NLO` matrix element generator [31], which is also used to generate the necessary simulations in this thesis. Essential to the working of the matrix element generator is the theory model that it needs to be supplied with. The theory model contains information on the Lagrangian of the theory and the parameters that are associated with it, such as the couplings and masses of the relevant particles. From this model, a set of Feynman rules is derived that will be used by the matrix element generator to reproduce the correct matrix elements. This procedure is automated by `FeynRules`, a Mathematica-based package [32]. The theory model, together with the process of interest, are supplied to the matrix element generator from which it computes the relevant Feynman diagrams. It then also calculates phase space integrals (based on the defined kinematic phase space) and allows to generate events whose kinematics follow the predicted probability distributions.

### 3.2.2 Parton shower simulation

As mentioned previously, the parton shower follows the hard scattering process in the pp collision. In some way, the parton shower is the first process which transforms the fundamental process of interest, which contains two or more outgoing particles, into the complex event that is observed in detectors. The main processes which contribute to the formation of a parton shower are gluon radiation from a quark $q \rightarrow qg$, a gluon splitting into two gluons $g \rightarrow gg$ and a gluon splitting into a quark-antiquark pair $g \rightarrow q\bar{q}$. With each of these processes is associated a probability of occurrence, which can be calculated from the DGLAP (Dokshitzer – Gribov – Lipatov – Altarelli – Parisi) evolution equations [33]. Any partons involved in the hard scattering will undergo these branchings until a certain energy scale $Q^2$ is reached. In the case of a final-state parton shower, this energy scale is the QCD scale, $Q^2 = \lambda^2_{\mathrm{QCD}}$, where the strong coupling constant becomes close to unity and therefore results from perturbation theory can no longer be applied. It is assumed that the hadronisation process kicks in when the partons of the shower have evolved downwards in energy to the QCD scale. The parton shower can be simulated using a program such as `Pythia8` [34], which is used for the hard scattering events generated by `MadGraph5_aMC@NLO`.

### 3.2.3 Hadronisation

The next important step in the pp collision is the hadronisation of the partons that leave the interaction, along with the colored particles that are produced by the parton shower. Following the principle of confinement, the partons cannot be isolated and as a consequence start forming hadrons by producing quark-antiquark pairs from the vacuum. This process cannot yet be described analytically from first principles. There do however exist two models that provide an accurate description: the string model and the cluster model. Both use an iterative approach to describe the hadronisation process. In the string model, when the distance between a quark-antiquark pair increases, energy is lost to the colour field, whose field lines connect the pair in a string-like configuration [35]. With this string is associated a uniform energy per unit length, which causes the string to break up into a quark-antiquark pair when sufficient energy has become available. These string breakups will keep occurring as long as the partons have enough energy. Eventually, a set of colorless hadrons is obtained. In the cluster model, the color connections of the particles in the parton shower are used to form color-singlet states. Any gluons that are present in the shower will split into a quark-antiquark pair and form color-singlet states with surrounding quarks. This way, the partons are clustered into colorless hadrons.

### 3.2.4 Detector simulation

The stable particles which emerge from the pp collision are observed by the CMS detector, which is comprised of several layers of detector material. Therefore, a detailed model of the CMS detector is needed to describe how these final-state particles interact with the detector material. Such a model is included in the `Geant4` simulation toolkit [36], which is generally used to simulate the passage of particles through matter. The model takes into account the effect of the intense magnetic field of the CMS detector on the trajectories of charged particles, and also handles the interactions of the particles with the dead zones of the detector, for example the support structures.

## 3.3 Reconstruction of physics objects

For each collision event that takes place, the CMS detector outputs a variety of signals from each of its components. From the energy clusters in the electromagnetic and hadronic calorimeters to the hits produced in the inner tracker, all of these signals still need to be somehow converted into physics objects before any analysis can be started. In this section, the reconstruction algorithms that are used to obtain these physics objects are described. At the CMS detector, the signals from each of the detector components are collected and used to obtain an improved event description, in which the properties of each detected final-state particle are reconstructed. The approach that is used in the CMS detector to achieve this reconstruction is called the particle-flow (PF) reconstruction algorithm [37].

### 3.3.1 Track reconstruction

An important feature of the inner tracker is that it allows for the measurement of the momentum of the charged particles that pass through it. From the geometry of the track, a set of helix parameters can be derived that directly relate to the particle momentum. It is therefore important that this trajectory is reconstructed accurately. In the PF algorithm, this is done using the Kalman Filtering (KF) method, where the tracks are reconstructed in three separate stages. The first stage is the *initial seed generation*, where a few hits from the inner tracker are used to obtain a first estimation of the charged particle's trajectory. Next is the *trajectory building* stage. Here, hits are gathered from all tracker layers along the particle trajectory that was seeded in the first stage. The final particle trajectory is then fitted to this collection of hits to obtain the origin, transverse momentum and direction of the charged particle.

### 3.3.2 Muon reconstruction

Essential to the analysis presented in this thesis is the reconstruction of muons. The momentum of the muon is obtained through the reconstructed trajectory of the muon, obtained from hits in the inner tracker and the exterior muon chambers. Dependent on the way the muon trajectory is reconstructed, the high-level muon physics object can be one of three types. The first is the *standalone muon*, where exclusively hits from the muon system are clustered to form track segments, which are ultimately fitted to obtain a *standalone-muon track*. Tracks in the inner tracker (inner tracks) are matched with each of these standalone-muon tracks if they are compatible. A fit to the combined collection of hits from the standalone-muon track and inner track results in a *global-muon track*, which corresponds with the second muon type, the global muon. The third type is the *tracker muon*. Here, inner tracks are matched with track segments in the muon system. If a match is obtained with just one of these segments, then the inner track becomes known as a *tracker muon track*. The benefit of this inside-out approach is that, for low muon momenta, the tracker muon reconstruction is more efficient than the global muon reconstruction, since the former requires only one muon segment in the exterior muon system. The global muon reconstruction on the other hand has a high efficiency for muons penetrating through multiple muon stations. Almost all muons produced within the acceptance of the muon system are reconstructed either as a global muon or as a tracker muon [37]. A large amount of muons are even reconstructed as both. Global and tracker muons that share the same inner track are merged into one muon object.

#### Selection quality

Different reconstruction qualities are assigned to the muon objects. A selection requirement that will be used in later sections of this thesis is the *tight muon* ID. A tight muon is a global muon with a higher track fit quality of $\chi^2/d.o.f. < 10$ and with at least one muon chamber hit included in the fit [38]. The muon is also required to be reconstructed inside-out as a tracker muon, using more than 10 inner-tracker hits.

### 3.3.3 Jet reconstruction

Partons produced in proton-proton collisions give rise to sprays of collimated particles known as jets. A jet can have different features depending on the properties of the particle from which it originated. It is therefore important that the outgoing particles are clustered together correctly, such that the jet can be optimally reconstructed. One clustering algorithm adopted by the CMS collaboration is the *anti-$k_T$* algorithm. It can be seen as belonging to a broader class of sequential recombination jet algorithms [39]. In these algorithms, two distance measures are defined as

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{\Delta_{ij}^2}{R^2} \quad \text{and} \quad d_{iB} = k_{ti}^{2p}, \tag{3.1}$$

where $\Delta_{ij}^2 = \Delta y_{ij}^2 + \Delta\phi_{ij}^2$ and $k_{ti}$, $y_i$ and $\phi_i$ are respectively the transverse momentum, rapidity and azimuth of particle $i$. The CMS collaboration uses a default value of 0.4 for the radius parameter $R$, which is a measure for the radius of the conical shape that defines the geometry of a jet. The general approach of this algorithm is to loop over each pair of reconstructed particles, and compare the values of $d_{ij}$ and $d_{iB}$. If the smallest distance is $d_{ij}$, the two particles are merged into a new entity, the *pseudo-jet*. The algorithm then proceeds in the same way by considering the pseudo-jet as another particle, such that it can cluster together with other particles or pseudo-jets. If instead $d_{iB}$ was the smallest distance, then the particle (or pseudo-jet) is considered a reconstructed jet and is removed from the list of entities.

## 3.4 Jet flavour-tagging algorithms

### 3.4.1 Heavy-flavour jets

Knowledge of the flavour of the quark that initiated a jet is vital in many situations. An algorithm that attempts to find the *jet flavour* using its observed properties is known as a jet flavour *tagger*. Important to the subject of jet flavour tagging is the terminology described below, which is used to denote the jet flavours in simulated events. In the simulation, the flavour of the jet can be defined by a procedure called *ghost-matching*. In this procedure, the reconstructed jets are reclustered, but this time together with the generator-level hadrons. These hadrons are also called *ghost hadrons*, because their momenta are set to a negligibly small number such that the reconstructed jet momentum is not affected. The flavour of the reconstructed jet in a simulation is then one of the following options:

**b jet:** The reconstructed jet contains at least one ghost hadron containing a bottom quark, or a b (ghost) hadron in short.

**c jet:** The reconstructed jet contains at least one c (ghost) hadron, and no b (ghost) hadrons. This is important, since a bottom quark is heavier than the charm quark and therefore a b hadron can decay into a c hadron. It is therefore possible that a b jet contains c hadrons, but not vice-versa.

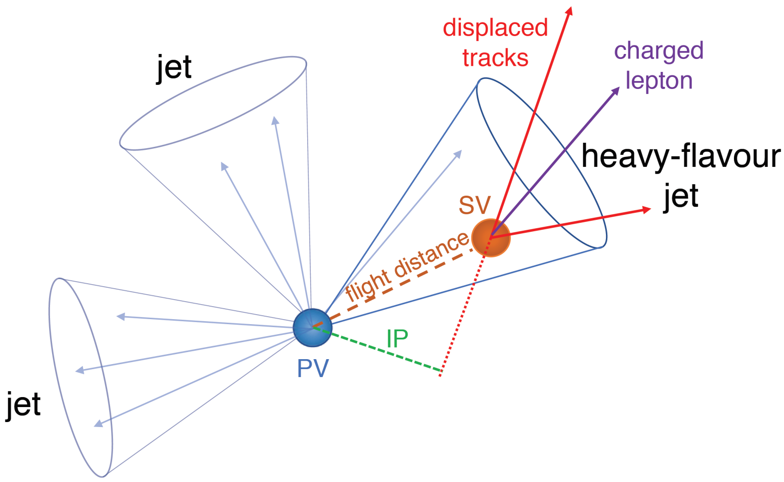



Figure 3.2: Depiction of a heavy-flavour jet produced from the decay of a b or c hadron, which displaces the charged-particle tracks from the primary interaction vertex (PV) to a secondary vertex (SV). The track displacement is characterised by the impact parameter (IP), defined as the distance between the PV and the point of closest approach of the displaced tracks. Image taken from [40].

**light (udsg) jet:** The reconstructed jet does not contain any c or b (ghost) hadrons. As a result, the jet was generated either from a gluon or from one of the lighter quarks (u,d,s), and is therefore referred to as a light jet.

The b and c jets are collectively known as the heavy flavour (HF) jets. An algorithm whose main objective is to identify c jets over a background of b jets and light jets is referred to as a c-tagger. Likewise, a b-tagger attempts to identify b jets over a background of c jets and light jets. The development and further optimisation of b-taggers and c-taggers is essential, since they play an important role in many physics analyses. An example application is in the study of the properties of the top quark. The lifetime of the top quark is shorter than the timescale at which hadronisation occurs, and therefore top quarks are the only quarks that do not hadronize. Instead, in nearly all cases, the top quark decays into a W boson and a bottom quark. Any analysis related to the top quark that makes use of this decay is therefore heavily dependent on b-taggers. Likewise, c-taggers are used in the analysis presented in this thesis, and are vital to identifying the charm quark that is produced along with a Higgs boson in the process $qg \to Hc$.

### 3.4.2 Properties of heavy-flavour jets

Quarks leaving the hard scattering will form a jet of hadrons. Hadrons containing bottom quarks are called b hadrons and have relatively long lifetimes of the order of 1.5 ps [18], and are therefore capable of travelling a small distance (up to about a cm) away from the collision point, which is also referred to as the primary interaction vertex (PV). The decay of the b hadron results in charged-particle tracks that are displaced from the

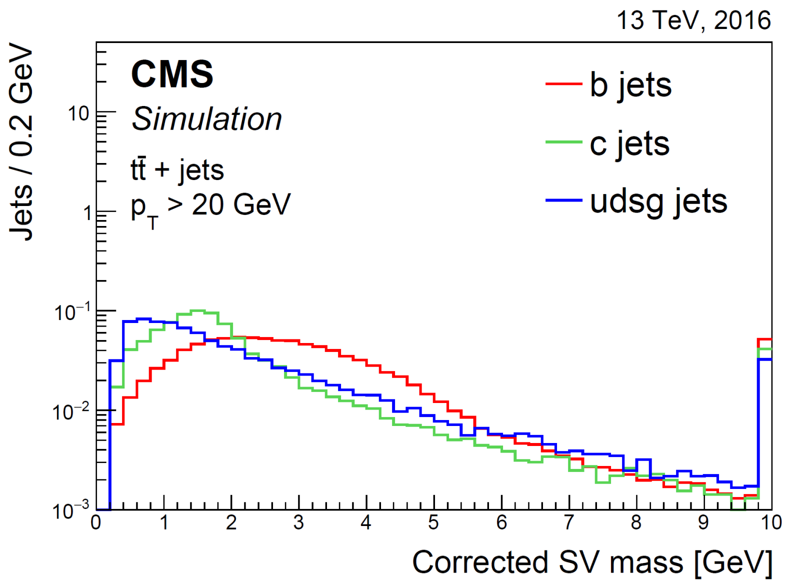



Figure 3.3: Secondary vertex (SV) mass $m_{\mathrm{SV}}$ distribution for bottom (red), charm (green) and light jets (blue) in $t\bar{t}$ events, taken from [40]. The last bin contains the overflow of jets for which the SV mass was greater than 10 GeV.

PV. Because of this, a secondary vertex (SV) can be reconstructed from the displaced inner tracks. Since the lifetime of b hadrons is typically an order of magnitude higher than the lifetime of c hadrons, a longer flight distance from the PV to the SV is a strong indicator towards b jets. The track displacement can be characterised further by introducing the impact parameter (IP), as in figure 3.2, which is defined as the distance between the PV and the point of closest approach of the displaced tracks. Another important variable in distinguishing b jets from c jets is the SV mass $M_{\mathrm{SV}}$, which is defined as $\sqrt{M_{\mathrm{SV}} + p^2 \sin\theta} + p\sin\theta$, where $M_{\mathrm{SV}}$ and $p$ are respectively the invariant mass and momentum of the tracks associated with the SV, and $\theta$ is the angle between the SV and the flight direction (the vector which points from the PV to the SV) [40]. The SV mass $M_{\mathrm{SV}}$ is directly related to the mass of the heavy hadron and is therefore an important discriminating variable. This can be seen in figure 3.3, which respectively shows the SV mass distributions for bottom, charm and light jets in red, green and blue. The shapes as well as the peaks of the SV mass distributions are different for each of the jet flavours. As a result, the SV mass is indeed a very useful variable for jet flavour tagging. Another property of heavy-flavour jets is that, because of the higher mass of the bottom and charm quarks, more energy is redistributed among the decay products of the heavy-flavour hadron, and therefore a larger transverse momentum relative to the jet axis is expected.

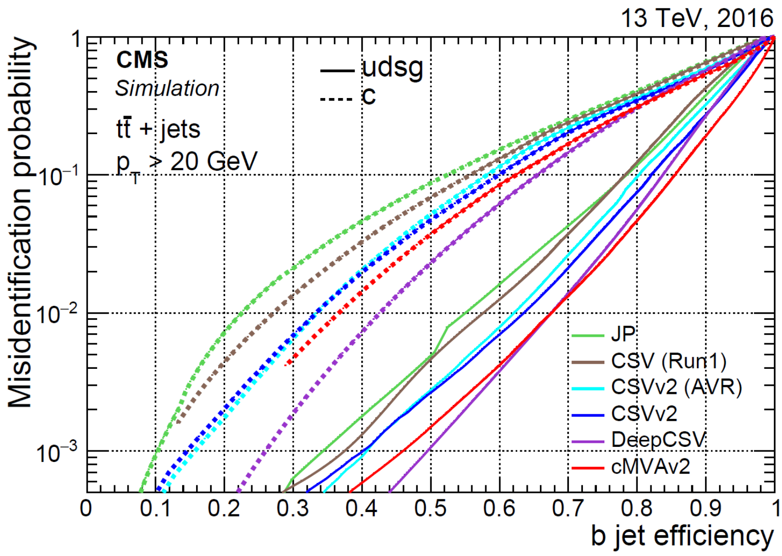



Figure 3.4: The ROC curve for several tagging algorithms including DeepCSV. Image taken from [40].

### 3.4.3 The DeepCSV tagger

The heavy-flavour tagger currently in use by the CMS collaboration is DeepCSV. The tagger is an artificial neural network (ANN) classifier which takes in a collection of track and SV related variables and has four output classes, corresponding with the possible flavours of the jet (b, c and udsg) and an additional $bb$ output class. This last category targets jets containing two b hadrons formed from the gluon splitting process $g \to b\bar{b}$. To be precise, the architecture of DeepCSV is a feed-forward artificial neural network [41] with four hidden layers containing one hundred interconnected neurons per layer, and therefore qualifies as a deep neural network. DeepCSV was trained on a mixture of simulated $t\bar{t}$ and QCD multijet events with jet transverse momentum between 20 GeV and 1 TeV [40]. Figure 3.4 shows the ROC curve of the DeepCSV tagger compared to other tagging algorithms which were previously used by CMS during the Run-1 of the LHC. The outputs of the heavy-flavour jet tagger can be interpreted as probabilities $P(b), P(c), P(udsg), P(bb)$, one for each of the jet flavour categories. From these outputs, discriminating variables can be derived whose aim is to distinguish between the output categories (or combinations thereof). These variables are also called *discriminators*. One example is the b-tagging discriminator used in DeepCSV, which is simply defined as $\mathrm{P}(b) + \mathrm{P}(bb)$. This discriminator distinguishes bottom jets from charm and light jets. For c-tagging purposes, DeepCSV uses two separate discriminators. One is the charm versus light (CvsL) discriminator, which is defined as

$$\text{CvsL discriminator} = \frac{\mathrm{P}(c)}{\mathrm{P}(c) + \mathrm{P}(udsg)}. \tag{3.2}$$

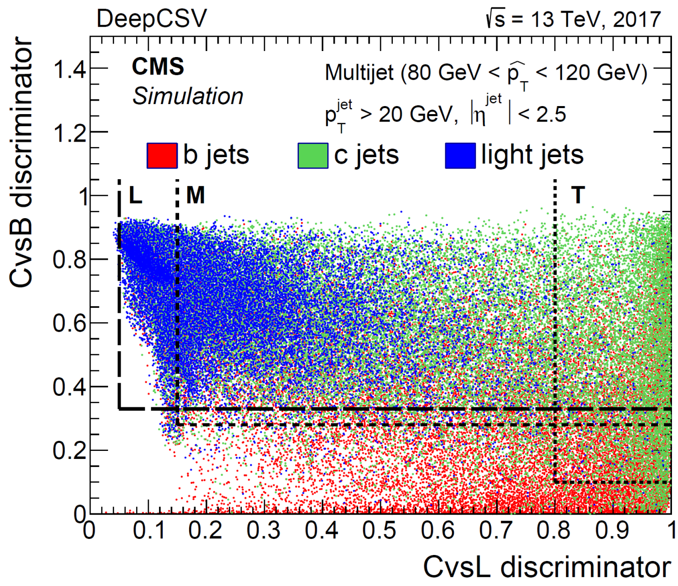



Figure 3.5: Two-dimensional distribution of the CvsL and CvsB discriminators from the DeepCSV model for heavy- and light-flavour jets taken from simulated multijet events. Image taken from [42].

The CvsL discriminator aims to distinguish charm jets from light jets. If DeepCSV is confident that the reconstructed jet is a charm jet, then it will output a value close to one for $\mathrm{P}(c)$ and a lower value for $\mathrm{P}(udsg)$. As a result, the CvsL discriminator will have a value close to unity for jets that DeepCSV predicts likely to be charm jets. Similarly, a charm versus bottom (CvsB) discriminator can be defined,

$$\text{CvsB discriminator } = \frac{\mathrm{P}(c)}{\mathrm{P}(c) + \mathrm{P}(b) + \mathrm{P}(bb)}. \tag{3.3}$$

The CvsB discriminator attempts to distinguish charm jets from bottom jets. Just like the CvsL discriminator, the CvsB discriminator is close to unity for jets that are likely to be charm jets as predicted by DeepCSV. Values close to zero then indicate that the jet is likely to be a bottom jet. In figure 3.5, the distribution of heavy- and light flavour jets in the two-dimensional phase space of the CvsL and CvsB discriminators is shown. The scattered red, green and blue points respectively correspond with bottom, charm and light jets. Charm jets are mostly concentrated near high values of the CvsL and CvsB discriminators, i.e. near the top-right corner of figure 3.5. The bottom jets are mostly distributed in the bottom-right corner, and the light jets in the top-left corner.

# Chapter 4

# Constraining the charm Yukawa in $\mathrm{H} + \mathrm{c}$ production

The remarkable discovery of the Higgs boson at the CERN LHC in 2012 [6, 7] signaled the start of a new series of precision measurements on the properties of this new particle. Most importantly are the interactions between the Higgs boson and the other fundamental particles of the SM, which are characterised by their coupling strength. The CMS and ATLAS Collaborations have found the couplings of the third generation fermions and the muon to be consistent with the SM prediction [16, 19–21, 27]. Naturally, the next objective is to probe the coupling of the Higgs boson to the charm quark, which together with the strange quark makes up the second generation of quarks. Previous attempts at constraining this coupling relied on the direct detection of the Higgs boson decay into a charm quark-antiquark pair, $H \to c\bar{c}$ [26]. This decay channel however suffers from a relatively low branching rate and a large QCD background, and additionally relies on the reconstruction of two charm-tagged jets. In this thesis, a new method is proposed where the coupling of the charm quark to the Higgs boson is measured in the process $gc \to Hc$, i.e. in the production of a Higgs boson in association with a charm quark. This was discussed in more detail in section 1.4. In this chapter, the sensitivity of the process $gc \to Hc$ to the charm quark Yukawa coupling $y_c$ will be explored. In the first part, a simulation of this process and the relevant backgrounds will be obtained up to the detector level. Afterwards, an artificial neural network will be trained using kinematical event properties as well as charm tagging variables to improve the event selection and to obtain a variable with increased sensitivity to $y_c$. Novel to this approach is the use of the charm tagging information, which will provide a significant contribution to the $y_c$ sensitivity. Finally, constraints on the coupling $y_c$ will be obtained using a binned likelihood fit, and will be compared to results using only either the leading jet $p_T$ or the Higgs boson $p_T$ in the fit.

## 4.1 Simulated datasets

### 4.1.1 Signal sample

As was outlined in the first chapter, the charm quark Yukawa coupling $y_c$ will be probed in the production of a Higgs boson in association with a charm quark, or $gc \to Hc$. The Higgs boson will be reconstructed through the $H \to ZZ^* \to 4\mu$ decay channel. The sensitivity of this process to changes in $y_c$ will be measured by varying $\kappa_c$ in simulation. In practice, this is done by supplying the matrix element generator `MadGraph5_aMC@NLO` with a theory model that allows for the modulation of $\kappa_c$. This model is the `Higgs Effective Couplings` model from `FeynRules` [43], which describes the effective couplings of the Higgs boson to gluons and photons. In this model, the top-quark loop in the third diagram in figure 1.8 is integrated out and replaced by an effective vertex. By providing the matrix element generator with this theory model, the desired Feynman diagrams can be generated. The resulting $gc \to Hc$ dataset consists of roughly 500,000 events, simulated with a minimum jet transverse momentum $p_T > 20$ GeV, a maximum jet pseudorapidity $|\eta| < 3.0$ and a minimum angular distance $\Delta R$ between the outgoing particles of 0.5. The angular distance $\Delta R$ is defined by

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{4.1}$$

where $\Delta\phi$ is the difference in azimuthal angle $\phi$ and $\Delta\eta$ the difference in pseudorapidity $\eta$. Instead of running the simulation multiple times for different values of the $\kappa_c$ parameter, different sets of weights are generated, each corresponding with a different $\kappa_c$ [44]. Using the `MadGraph5_aMC@NLO` matrix element generator and the `Higgs Effective Couplings` model, the interaction $gc \to Hc$ was simulated at leading order in QCD, including a simulation of the parton shower (up to the hadronisation stage) with `Pythia8`. Afterwards, the interactions of the collection of outgoing final-state particles with the CMS detector is simulated by `Geant4`. The result of this simulation is then passed through the entire CMS reconstruction chain, until finally a sample ready for analysis is obtained. Figure 4.1 shows the different steps that were required to obtain the desired $gc \to Hc$, $H \to ZZ^* \to 4\mu$ sample, simulated up to the detector level.

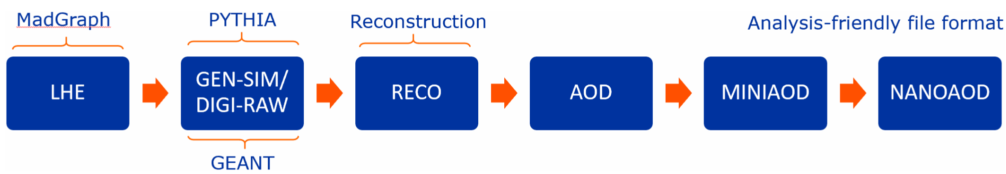



Figure 4.1: Diagram illustrating the different file formats that are generated by the respective simulation programs to obtain a simulation up to the detector level, ready for a physics analysis.

### 4.1.2 Background samples

The backgrounds included in this analysis are separated into two categories. The first category is the expected background from Z boson pairs from the decay of a Higgs boson that was not produced in association with a charm quark, or $H \to ZZ$. These background Higgs production mechanisms include: Gluon fusion (GF), vector boson fusion (VBF), Higgs bosons produced in association with a vector boson ($ZH$ or $W^{\pm}H$), and Higgs bosons produced in association with a top quark pair ($ttH$). The second category of background processes included in this analysis is the background from $Z$ boson pairs that were produced through other mechanisms. These include the $gg \to ZZ$ and $q\bar{q} \to ZZ$ processes. Lastly, the background from the Drell-Yan processes $qg \to Zq$ and $q\bar{q} \to Zg$ are also included. These are referred to as $Z+X$ or Z boson plus jet events, since both interactions produce a parton in the final state which is observed as a jet. The list of simulated processes and the corresponding category labels can be found in table 4.1.

Table 4.1: Background processes used in the analysis of the $gc \to Hc$ channel.

| Category | Channels | Decay channel | Cross section [pb] |
|---|---|---|---|
| **H-bkg** | $H \to ZZ \to 4\ell$ (VBF) | | 0.000986 |
| | $H \to ZZ \to 4\ell$ (GF) | | 0.0127 |
| | $ttH$ | $H \to ZZ \to 4\ell$ | 0.000393 |
| | $ZH$ | | 0.000668 |
| | $W^+H$ | | 0.000218 |
| | $W^-H$ | | 0.000138 |
| **Non-H bkg** | $Z+X$ | $Z \to 2\ell$ | 6435 |
| | $q\bar{q} \to ZZ \to 4\ell$ | | 0.0133 |
| | $gg \to ZZ \to 4\mu$ | | 0.00158 |
| **H+c** | $gc \to Hc$ | $H \to ZZ \to 4\mu$ | 0.00000587 |

## 4.2 Event selection

### 4.2.1 Signal topology

The event topology associated with the signal process $gc \to Hc, H \to ZZ^* \to 4\mu$ is relatively straightforward. At least one charm jet is expected from the charm quark that is produced along with the Higgs boson. The Higgs boson is reconstructed through the four-muon final state. Here, a large background is expected from $Z$ boson pairs that were not produced by a Higgs boson, since they can also produce a four-lepton final state. The inclusion of the $Z$ plus jet background is also relevant, even though this process cannot physically lead to a four-muon final state. It is however possible that two additional *fake muons* are reconstructed along side a pair of muons from the Z boson decay. These fake muons are particles that were reconstructed as muons, but were misidentified. Even though this misidentification is unlikely, the large cross section of the $Z$ plus jet process can result in a significant amount of events containing fake muons.

The background from Z boson pairs that were not produced from Higgs boson decay can be reduced by reconstructing the invariant mass of the Higgs boson in the four-muon channel. The four-muon invariant mass $m_{4\mu}$ distribution should peak around the Higgs boson mass $m_H \approx 125$ GeV. Therefore, by keeping only events where the reconstructed $m_{4\mu}$ lies around the measured value of the Higgs boson mass $m_H$, the background from Z boson pairs should be reduced significantly.

In order to construct high-level variables such as the four-muon invariant mass $m_{4\mu}$, the necessary physics objects need to be available. Only events that contain at least one reconstructed jet with $p_T > 20$ GeV and $|\eta| < 2.4$ are selected. From this reconstructed jet, essential charm-tagging information can be obtained from the DeepCSV tagger. The selection also requires at least four reconstructed muons with $|\eta| < 2.4$ and $p_T > 10$ GeV. Each of the reconstructed muons are required to be *tight* muons. This higher reconstruction quality ensures that the background from the $Z$ boson plus jet channel is reduced, since less fake muons should be selected.

### 4.2.2 Z boson reconstruction

The physics objects resulting from the CMS reconstruction chain can be analysed using the ROOT framework. From the reconstructed muon objects, $Z$ boson candidates can be constructed. The procedure used in this thesis to select $Z$ boson and $Z$ boson pair candidates was inspired by the selection that was adopted by the CMS collaboration in the measurement of the properties of the Higgs boson in the four-lepton final state [45]. The algorithm that handles the construction and reconstruction of $Z$ and $ZZ$ candidates is the following:

1. A $Z$ candidate is constructed by merging a pair of opposite-sign muons.
2. Each of the non-overlapping $Z$ candidates are merged into $ZZ$ candidates. Each of the $ZZ$ candidates must contain at least two muons with $p_T > 20$ GeV, and one $Z$ candidate with an invariant mass $m_{2\mu}$ between 70 and 110 GeV. If multiple $ZZ$ candidates are obtained, then the one that contains the $Z$ candidate with an invariant mass closes to the Z boson mass is selected.

From the constructed $ZZ$ candidate, a four-muon invariant mass $m_{4\mu}$ can be calculated. Only events where the invariant mass of the $ZZ$ candidate lies between 100 and 150 GeV are selected, to reduce the background from $Z$ boson pairs that were not produced from Higgs boson decay. This selection criterion was purposefully chosen to not be too tight, such that the ANN can still exploit this variable in order to produce network outputs which will then be used to reduce the non-H background further. The invariant mass $m_{4\mu}$ spectrum of both the H background and H+c samples peak near the Higgs boson mass, while this is not the case for samples from the non-H background category. This can be seen in figure 4.2, where the $m_{4\mu}$ spectrum of three samples (one of each category) is shown.

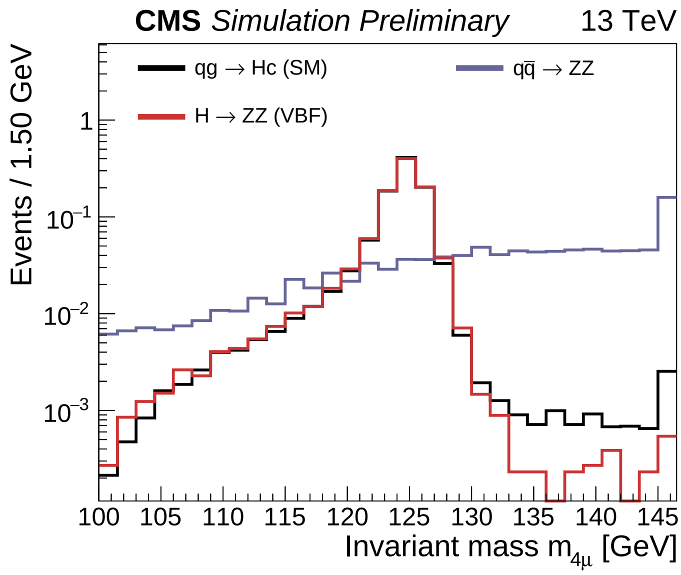



Figure 4.2: Normalised four-muon invariant mass $m_{4\mu}$ distribution for the $H \to ZZ$ (VBF) process, which belongs to the H background category, and the $q\bar{q} \to ZZ$ process, which belongs to the non-H background category. These are compared with the distribution from the signal H+c sample. The last bin contains the overflow of events.

### 4.2.3 First event selection

The simulated samples of the background processes contain a different number of simulated events. In order to obtain a fair comparison between the signal and background processes, the total number of simulated events needs to be converted into an expected amount of events,

$$N_{\text{exp}} = \sigma L \times \frac{N_{\text{MC}}^{\text{selected}}}{N_{\text{MC}}}, \tag{4.2}$$

where $N_{\text{exp}}$ is the expected number of events, $\sigma$ is the cross section of the process and $L$ is the integrated luminosity, which is defined as the integral of the beam luminosity over the data taking period. The product of the integrated luminosity and the cross section is then multiplied by the fraction of simulated events that were selected to obtain the expected number of events. In this analysis, an integrated luminosity $L$ of 300 $fb^{-1}$ is used, which corresponds with the expected LHC integrated luminosity at the end of Run-3 [46].

Table 4.2: The background and signal data samples that will be used in the analysis of the signal $qg \to Hc$ process. Each of these processes feature inclusive decays, except for the signal where $H \to ZZ^* \to 4\mu$.

| ***Simulated events*** | | | |
|---|---|---|---|
| Process | Sample size | Events left after initial selection | Events expected (300 fb$^{-1}$) |
| $H \to ZZ$ (GF) | 965,198 | 22,663 | 89.45 |
| $H \to ZZ$ (VBF) | 928,272 | 25,871 | 8.24 |
| $gg \to ZZ$ | 971,500 | 15,888 | 8.09 |
| $q\bar{q} \to ZZ$ | 16,075,000 | 6,016 | 1.49 |
| $Z + X$ | 25,757,729 | 38 | 2,850 |
| $ttH$ | 492,273 | 8,142 | 1.95 |
| $W^-H$ | 200,000 | 5,143 | 1.06 |
| $W^+H$ | 299,710 | 6,939 | 1.51 |
| $ZH$ | 494,864 | 6,162 | 2.50 |
| $qg \to Hc$ | 465,916 | 107,625 | 0.41 |

## 4.3 Improving event selection with an ANN

In the previous section, cuts were imposed on the four-muon invariant mass spectrum $m_{4\mu}$ to reduce the background from $Z$ boson pairs that were not produced by Higgs boson decay. The results of this cut are given in table 4.2. In this section, an artificial neural network (ANN) will be trained in order to further reduce this non-H background. At the same time, the ANN will attempt to identify events that are sensitive to the charm quark Yukawa coupling $y_c$. This corresponds with distinguishing between events produced from the effective diagram in figure 1.8, and events produced via the two diagrams on the left, where there is an explicit $y_c$ dependence. The ANN will be implemented in python using the Keras deep learning library [47], which is interfaced with TensorFlow [48] as a backend.

### 4.3.1 Training an ANN

An ANN will be trained on the simulated signal and background datasets. The outputs of the ANN will be transformed into a `ZZnoH` and a `Hc` discriminator. The former discriminator aims at obtaining an optimal separation between events that contain a pair of $Z$ bosons that were not produced from Higgs boson decay and the pairs that were produced by this decay. The information provided by the `ZZnoH` discriminator will then be included in the event selection as an additional selection requirement to further reduce the background from non-H events. The `Hc` discriminator is responsible for differentiating between events that are directly sensitive to $y_c$ and the events that are not. In other words, this discriminator will attempt to identify H+c events that are directly sensi-

Table 4.3: The default network properties of the ANN's trained in section 4.3.1.

| *Network properties* | *Default setting* |
|---|---|
| Number of epochs | 100 |
| Batch size | 256 |
| Loss function | Categorical cross-entropy |
| Learning rate | 0.0005 |
| Dropout | 0.1 |
| Hidden layer activation function | ReLu |
| Output layer activation function | Softmax |

tive to $y_c$ over a background from H and non-H events, and also a background of H+c events produced by the effective diagram in figure 1.8. By construction, the Hc discriminator should be more sensitive to changes in $\kappa_c$ than any of the variables included in the training of the ANN. As a matter of fact, the discriminator will be defined from the ANN outputs, which are themselves calculated using the information provided by the input variables. The Hc discriminator can hence be seen as an ideal combination of the input variables, resulting in a variable that is more sensitive to $\kappa_c$.

#### Output categories

The ANN will be trained to classify events into four output categories. The first two are dedicated to the events that originate from the $gc \to Hc$ process. One of these categories exclusively contains events that are directly sensitive to $y_c$, labelled `H+c sens.`, while the other category contains events from the insensitive effective diagram in figure 1.8, labelled `H+c insens.`. These categories can be distinguished from each other in the leading-order samples using the $\kappa_c$ weights that are provided with each event. After all, owing to the re-weighting [44] procedure that is used in the generation of the H+c sample, the sensitive diagrams have a weight proportional to $(\kappa_c)^2$. The other two output categories correspond with the two groups of background processes. The first of these two contains background events from $Z$ boson pairs are produced from Higgs boson decay, labelled `H bkg`, while the other category contains background events from $Z$ boson pairs that are produced through another process, labelled `non-H bkg`. Each of the ANN outputs can be interpreted as a probability P(H+c sens.), P(H+c insens.), P(H bkg) or P(non-H bkg) for an event to belong to a certain category.

#### Input variables

The ANN is composed of one hidden layer with 20 neurons and four output neurons corresponding with the four categories outlined above. Initially, the architecture of the ANN will remain unchanged such that the effect of adding different input variables on the model's performance can be studied. After having found a set of input variables that results in an optimal performance of the ANN, different architectures will be considered to further improve this performance. The model uses a Rectifier Linear Unit or ReLu as an activation function for the hidden layer, and a softmax function in the output layer. In order to prevent overtraining, the model adopts a dropout procedure where each epoch 10% of the neurons have their weights set to zero. Table 4.3 contains the network properties that will be used for each of the models that are described in the following paragraphs.

The number of neurons in the input layer corresponds with the number of input variables that the ANN has access to. The total collection of variables that can be made available to the ANN in training are the following:

**Jet variables:** For each event, information on three reconstructed jets is stored in two different rankings: the $p_T$ ranking and the CvsL discriminator ranking. For each of these rankings, information on the $p_T$, $\eta$ and CvsL value of the three leading jets of each ranking are available, for a total of eighteen variables. If an event contains three or fewer reconstructed jets that meet the selection requirements described in section 4.2.1, then these will be sorted both by their $p_T$ (to obtain a $p_T$ ranking) and by their CvsL discriminator value (to obtain a CvsL ranking). The jet with the highest value of $p_T$ (CvsL discriminator) is referred to as the leading-$p_T$ (CvsL) jet. The CvsL ranking is physically more relevant, since reconstructed jets with a higher value of the CvsL discriminator are more likely to be charm jets. The variables associated with this jet, for example the CvsL discriminator value itself, will contribute significantly towards the identification of the H+c process from the H and non-H background events.

**Z and ZZ candidate variables:** The invariant mass of the selected $ZZ$ candidate, and of the individual $Z$ candidates it is composed of, as well as the $\Delta R$ between the muons that make up the $Z$ candidates.

**Muon variables:** The $p_T$ and $\eta$ of the reconstructed muons that form the $ZZ$ candidate, as well as the $\Delta R$ between each of the muon objects.

**Other variables:** The $\Delta R$ between the leading jet (of each ranking) and each of the muons.

Not all of these variables will be included in the training of the ANN. Instead, different groups of variables will be added, and the change in the performance of the ANN will be followed closely. Through this procedure, the set of input variables that amounts to an optimal performance of the ANN will be found.

### 4.3.2 Selection of the sensitive input variables

In distinguishing between `non-H bkg` and `H bkg` events, the variable that is expected to make a large positive contribution to the ANN's performance is the four-muon invariant mass $m_{4\mu}$. Indeed, when the ANN was trained using only this variable as input, it was

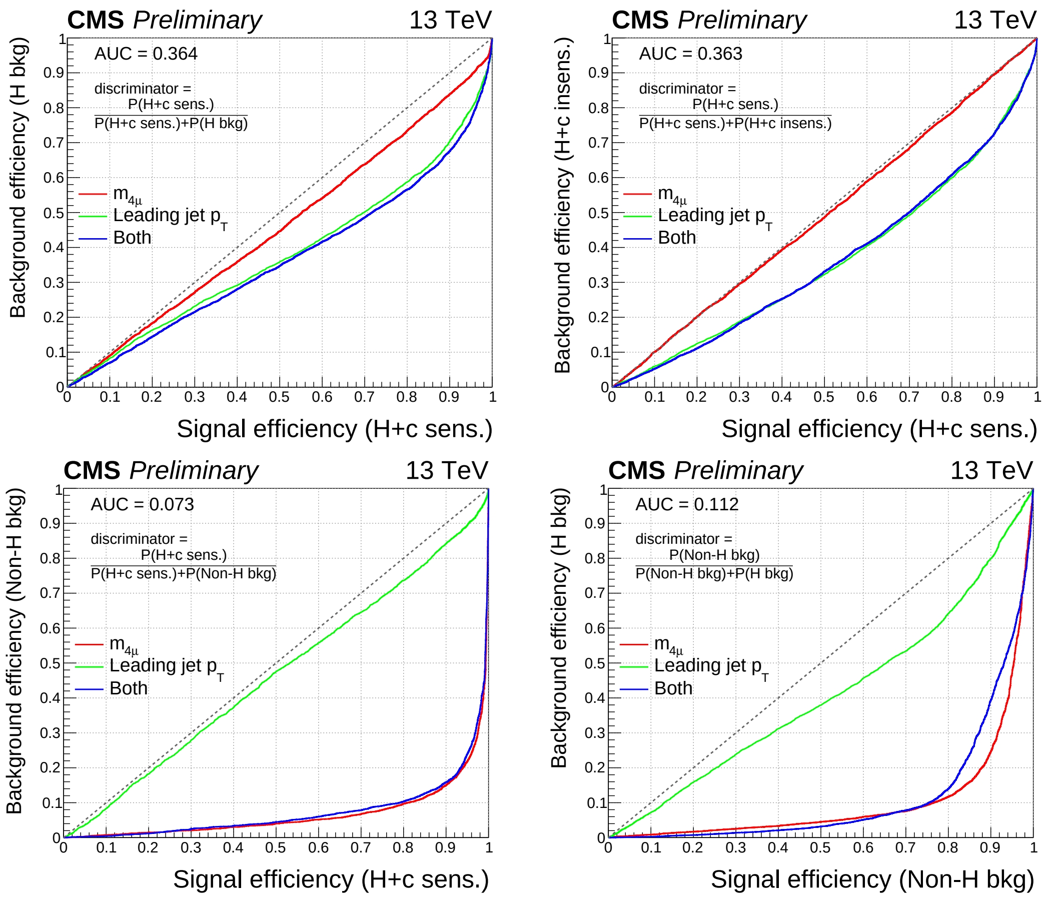



Figure 4.3: The ROC curves of an artificial neural network trained using either the four-muon invariant mass $m_{4\mu}$ (red), the leading jet transverse momentum $p_T$ (green) or both (blue). The AUC value shown in these figures belongs to the model that uses both variables.

able to improve upon the event selection that was implemented earlier. For comparison, another model was trained using only the transverse momentum $p_T$ of the leading jet in the $p_T$ ranking. The performance of both these models is shown respectively by the red and green curves in figure 4.3. Additionally, a model that uses both the invariant mass $m_{4\mu}$ and the leading jet $p_T$ was trained to compare to the previous two models. The performance of this last model is given by the blue curve in figure 4.3. In each of the following paragraphs, the layout for the displayed ROC curves is the following: The figure in the top-left corner shows the performance of the models for discriminating between `H+c sens.` events and `H bkg` events. In the figure in the top-right corner the discrimination between `H+c sens.` and `H+c insens.` events is shown. The figure in the bottom-left corner corresponds with the discrimination between `H+c sens.` and `non-H bkg` events. Lastly, in the figure in the bottom-right, the performance of the models is shown for the identification of `non-H bkg` events over a background of `H bkg` events. From the ROC curves in the bottom row of figure 4.3, it becomes clear that the four-muon invariant mass $m_{4\mu}$ is a powerful variable for discriminating `H+c` from `non-H` events, and for discriminating `non-H bkg` from `H bkg` events. In other words, the invariant mass $m_{4\mu}$ contributes heavily towards the model's ability to distinguish H events, i.e. events containing a Higgs boson, from non-H events. The same model however fails to distinguish between those H+c events that are sensitive to $y_c$, and those that are not, as seen in the top-right corner of figure 4.3. For this discrimination, the leading jet $p_T$ variable seems to provide more useful information to the ANN compared to the invariant mass $m_{4\mu}$. The model seems to recognise a different kinematic behaviour associated to the produced charm quark in the sensitive diagrams compared to the insensitive diagram, and exploits this to achieve better performance. The leading jet $p_T$ however does not seem to contribute much towards the model's performance in distinguishing `H+c sens.` from `non-H bkg` events, as seen in the bottom-left of figure 4.3. The blue curve in figure 4.3 shows the performance of an ANN that was trained using both of these input variables. As a result, this model uses the information of both variables optimally, and achieves an overall better performance than any of the models separately.

#### Further comparison after including more jet-related variables with different jet rankings

Following these results, another model was trained using the invariant mass $m_{4\mu}$ and the leading jet $p_T$, $\eta$ and CvsL discriminator value as input variables. For this model, as well as for the previous ones, the leading jet was selected from the $p_T$ ranking. In other words, the leading jet $p_T$ is defined here as the transverse momentum $p_T$ of the reconstructed jet that has the highest $p_T$ out of all selected reconstructed jets in an event. An ANN was trained that used the $p_T$, $\eta$ and CvsL variables associated with the leading-$p_T$ jet, along with the four-muon invariant mass $m_{4\mu}$. The additional charm-tagging information should aid the model in distinguishing between the $Hc$ signal events and the $H$ background events. Another model was trained using the same input variables, but with additional information from the leading-CvsL jet, i.e. the jet with the highest value of the CvsL discriminator. The performance of these models is shown respectively by the red and green curves in figure 4.4. As expected, both models clearly benefit from the additional charm-tagging information. The model that also uses the variables associated with the leading-CvsL jet outperforms the model that only uses the leading-$p_T$ jet variables. This effect is most pronounced in the top-left of figure 4.4, which shows the performance of the model for discriminating between `H+c sens.` and `H bkg` events. It therefore seems that the information from the additional reconstructed jet with a high CvsL discriminator value contributes significantly towards the model's performance. This is no surprise, since the variables associated with the leading-CvsL jet are more likely to belong to a charm-tagged jet, and are therefore more physically relevant for discriminating between $Hc$ signal and $H$ background events.

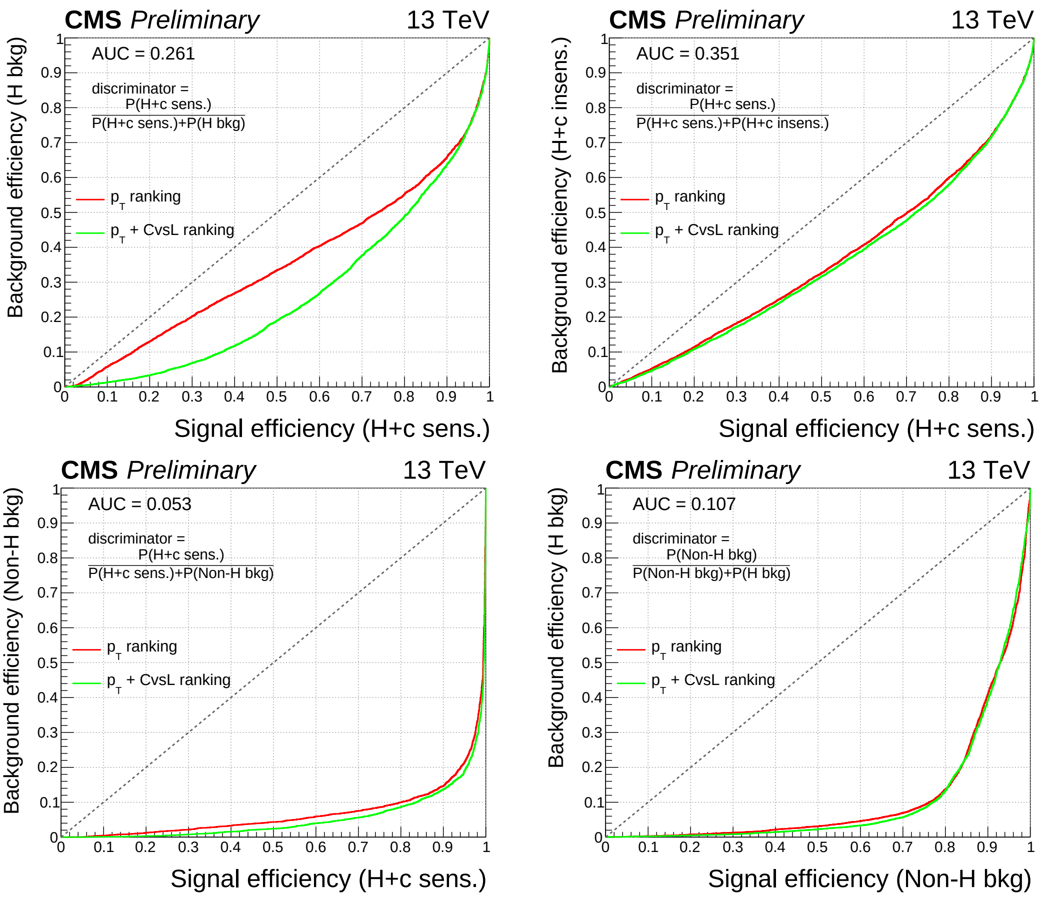



Figure 4.4: The ROC curve of the neural network trained using the variables associated with leading jet according to the transverse momentum ranking (red) compared to another model that also used the CvsL discriminator ranking (green). The AUC value shown in these figures belongs to latter model.

### Adding sub-leading jet information

Adding the variables associated with the sub-leading jets of both rankings resulted in a slightly better performance of the ANN. It was however found that the information provided by the sub-sub-leading jet did not have any significant effect on the performance. Likely, this is a result of the fact that more and more duplicate information is being added to the ANN. It is possible that in some events the leading-$p_T$ jet and leading-CvsL jet are one and the same. When adding the information associated with the sub-leading and sub-sub-leading jets, the probability of having duplicate information increases. In exception of the sub-sub-leading jet variables, all of the possible input variables discussed in the previous section were added as input to the ANN, and resulted in a noticeable increase in performance. In figure 4.5, a comparison is shown between the first two models, which used only the four-muon invariant mass $m_{4\mu}$ and

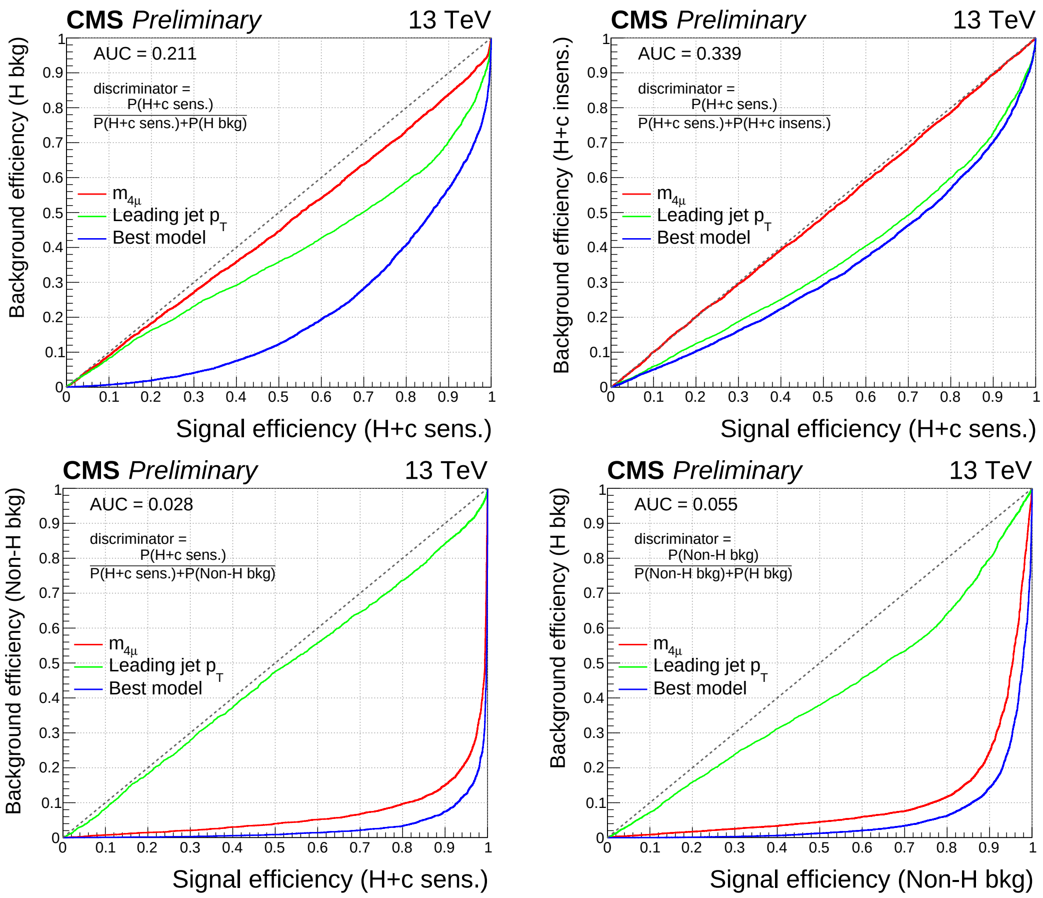



Figure 4.5: The performances of the models using either the four-muon invariant mass $m_{4\mu}$ or the leading jet transverse momentum $p_T$ are compared with the best performing model so far, which uses the variables described in section 4.3.2. The AUC value shown in the upper-left corner of each of these figures corresponds with the best performing model.

the leading-$p_T$ jet $p_T$, and the ANN that obtained the best performance so far.

### ANN architecture

Finally, after having acquired an optimal set of input variables, the architecture of the ANN could be tweaked to further improve its performance. During the previous training stages, the ANN was comprised of a single hidden layer containing 20 neurons. The change in performance in response to varying the number of neurons in the first hidden layer was studied, along with the effects of adding a second hidden layer. For most architectures, the performance of the ANN was unchanged with respect to the initial architecture. A significant improvement was observed when training the ANN with

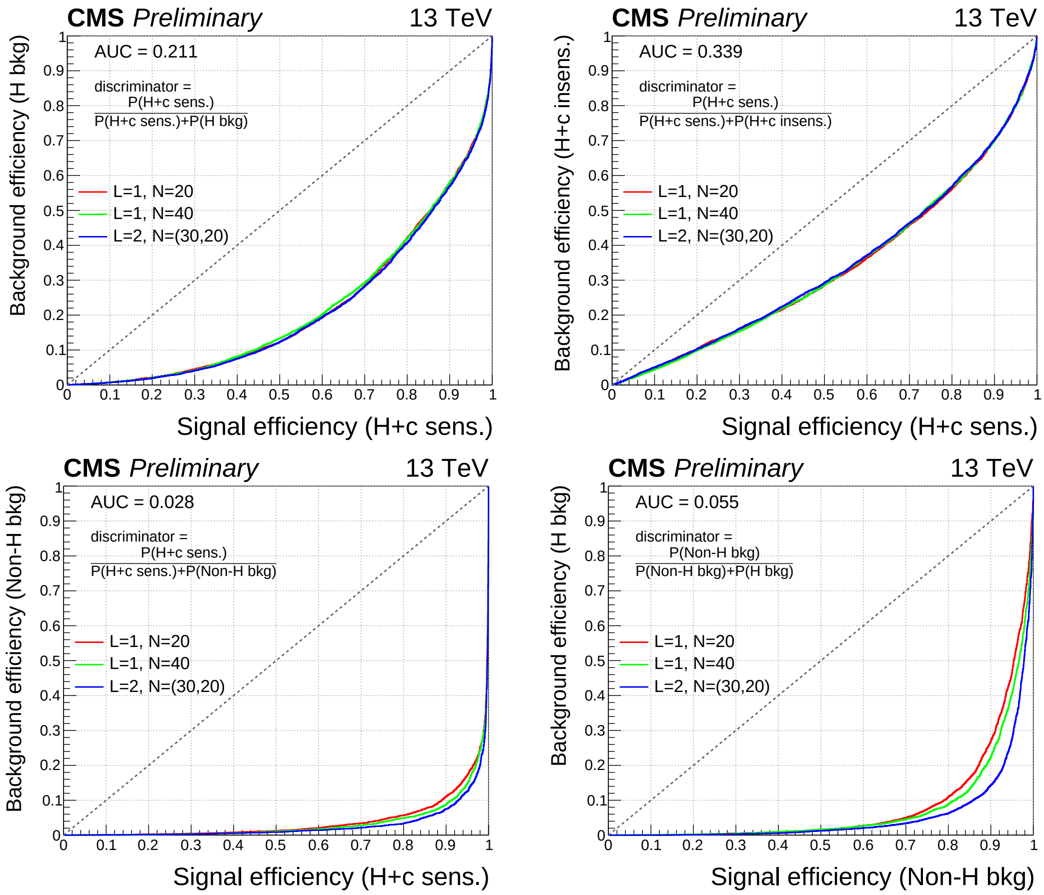



Figure 4.6: Comparison between the performances of a neural network using the same optimal set of input variables described in section 4.3.2, but using a different number of layers $L$ and a different number of neurons $N$ per layer. The AUC value shown in the upper-left corner of each of the figures corresponds with the $L = 2$, $N = (30, 20)$ model.

two hidden layers ($L = 2$), the first containing 30 neurons and the second 20 neurons ($N = (30, 20)$). The performance gain of the ANN mostly revealed itself in the discrimination between `non-H bkg` and `H bkg` events and in the discrimination between `non-H bkg` and `H+c sens.` events, as can be seen in figure 4.6.

### Best model

To summarize, the ANN that achieved the best performance contains two hidden layers, with respectively 30 and 20 neurons. The outputs of the ANN can be interpreted as the probabilities $P(\texttt{H+c sens.})$, $P(\texttt{H+c insens.})$, $P(\texttt{H bkg})$, $P(\texttt{non-H bkg})$ for an event to belong to a certain category. The input layer contains 33 neurons, corresponding with the following variables:

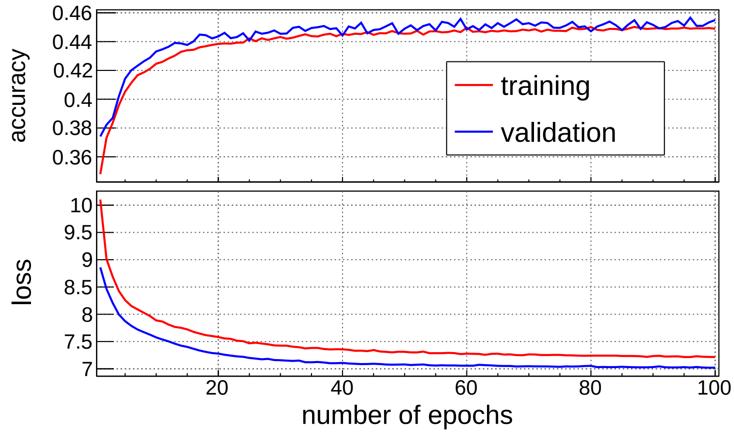



Figure 4.7: The loss and accuracy curves as a function of the training epoch of the artificial neural network described in section 4.3.2.

- The four-muon invariant mass $m_{4\mu}$.
- The leading and sub-leading jet $p_T$, $\eta$ and CvsL discriminator value, for both the $p_T$ ranking as the CvsL ranking.
- The $\Delta R$ between each of the selected muons and the leading jet of both rankings.
- The $\Delta R$ between the pairs of muons that form the $Z$ candidates.
- The $p_T$ and $\eta$ of each of the selected muons.
- The invariant mass of each of the $Z$ candidates.

The ANN was trained for 100 epochs. The loss and accuracy of the model on the training and validation set is given in figure 4.7. Both the loss and accuracy curves reach a plateau at around 100 epochs, and remain stable for both the training and validation set. Therefore, it is safe to assume that the ANN is not overtrained.

### 4.3.3 Discriminator definitions

As was mentioned earlier, the goal of the ANN that was trained in the previous sections was to obtain a `Hc` discriminator and a `ZZnoH` discriminator. The former aids in the identification of $H + c$ events that are sensitive to $y_c$, while the latter will be used to distinguish between the events that contain a Z boson pair from Higgs boson decay and the events that contain a Z boson pair formed via another interaction. Using the ANN outputs, these discriminators can be defined as

$$\text{ZZnoH discriminator} = 1 - \text{P(non-H bkg)}, \tag{4.3}$$

$$\text{Hc discriminator} = \frac{\text{P(H+c sens.)}}{\text{P(H+c sens.)} + \text{P(H+c insens.)} + \text{P(H bkg)}}. \tag{4.4}$$

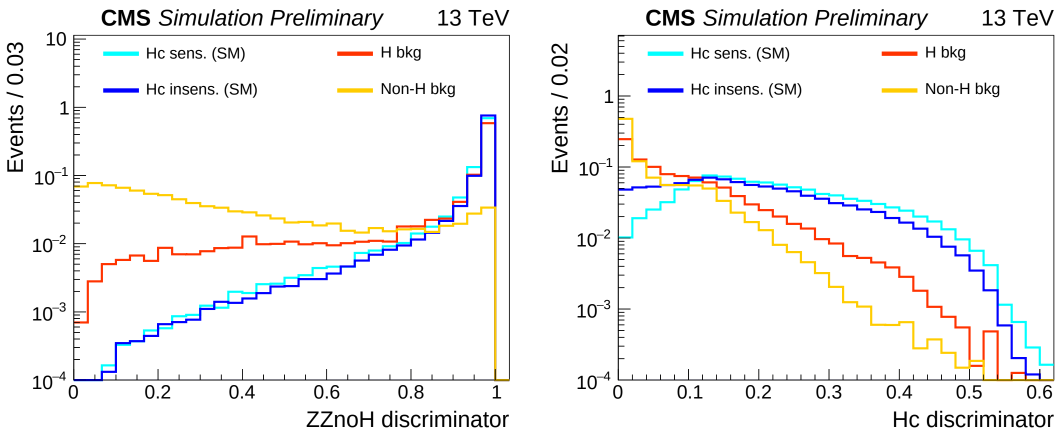



Figure 4.8: The normalised distributions of the ZZnoH and Hc discriminators defined in section 4.3.3. The former discriminates between events containing a Higgs boson, labelled $H$ events, and events that do not contain a Higgs boson, labelled non-$H$ events. The latter distinguishes events sensitive to $y_c$ from the insensitive events.

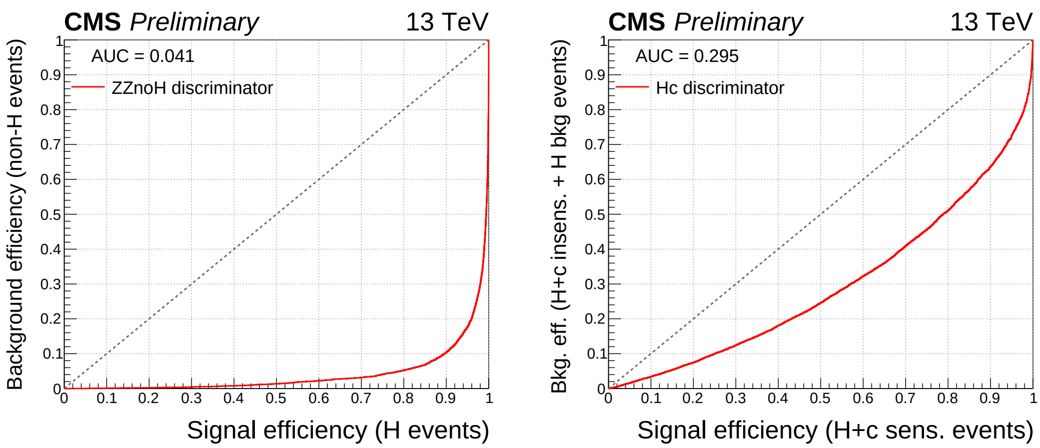



Figure 4.9: The ROC curves of the ZZnoH (left) and Hc (right) discriminators, defined in section 4.3.3.

The distributions and ROC curves of these discriminators are respectively given in figure 4.8 and 4.9. The distribution of the `ZZnoH` discriminator for the H events and non-H events are clearly separated. This separation can be exploited to further reduce the background from non-H events, and hence obtain an improved event selection. After including the ANN in the event selection, and providing it with the required input variables, the output of the ANN can be transformed into the `ZZnoH` discriminator value, which can then be used as an additional selection criterium. From the `ZZnoH` distribution in figure 4.8, most of the non-H events can be cut by requiring `ZZnoH` $> 0.8$. Since this requirement is met by most of the H events, there is no risk of throwing away a significant amount of signal events.

### Improved selection results

In table 4.4, the results of the improved selection are shown. As a result of applying the cut on the `ZZnoH` discriminator, the background from the processes that involved the production of a Z boson pair through the interactions $gg \to ZZ$ and $q\bar{q} \to ZZ$ has been drastically reduced. The number of expected events from the signal process $qg \to Hc$ on the other hand remains roughly the same. The background from $Z + X$ events appears to be dominant, judging by the number of expected events in table 4.4. However, one needs to take into account the large uncertainty that is associated with this expected event yield. From the roughly 25 million events contained in the sample, only 14 events were selected. It is therefore safe to argue that the estimate made in table 4.4 is not the one to use in a final analysis, and the $Z+X$ background needs to be estimated from data instead. In a similar analysis [45], this background is measured from control regions in data, and is found to contribute only half as much as the `ZZnoH` background category. Therefore, the contribution from $Z + X$ events to the overall background is expected to be insignificant, and will accordingly not be included in the fit presented in the next section.

Table 4.4: The number of expected events from the background and signal samples after applying the improved event selection, using an additional selection requirement involving the ZZnoH discriminator.

| ***Simulated events*** | | | | |
|---|---|---|---|---|
| Process | Sample size | Events left after initial selection | Events left after improved selection | Events expected (300 fb$^{-1}$) |
| $H \to ZZ$ (GF) | 965,198 | 22,663 | 19,716 | 77.83 |
| $H \to ZZ$ (VBF) | 928,272 | 25,871 | 23,956 | 7.63 |
| $gg \to ZZ$ | 971,500 | 15,888 | 2,137 | 1.09 |
| $q\bar{q} \to ZZ$ | 16,075,000 | 6,016 | 580 | 0.14 |
| $Z + X$ | 25,757,729 | 38 | 14 | 1,050 |
| $ttH$ | 492,273 | 8,142 | 7,786 | 1.86 |
| $W^-H$ | 200,000 | 5,143 | 4,627 | 0.96 |
| $W^+H$ | 299,710 | 6,939 | 6,233 | 1.36 |
| $ZH$ | 494,864 | 6,162 | 4,551 | 1.84 |
| $qg \to Hc$ | 465,916 | 107,625 | 100,920 | 0.39 |

## 4.4 Binned likelihood fit

### 4.4.1 Uncertainties

In the following paragraphs, constraints on $\kappa_c$ will be obtained by means of a binned maximum likelihood fit to the `Hc` discriminator, leading jet $p_T$ and Higgs boson $p_T$ distributions. Here, the Higgs boson transverse momentum is defined as the transverse momentum of the four-vector sum of the four muons associated to the ZZ candidate. For this sensitivity study, only statistical uncertainties will be used. This however does not imply that no significant systematic uncertainties are expected. On the contrary, there are multiple sources of systematic uncertainty that could be important. Some of these include:

- Uncertainty on the muon identification and reconstruction efficiency. In an analysis of the $H \to ZZ^* \to 4\ell$ process [45], this uncertainty is estimated to be 2.5% on the overall event yield for the $4\mu$ final state.
- Uncertainty on the jet energy scale and resolution.
- Uncertainty related to the c-tagging efficiency of heavy- and light-flavour jets. Recently, new calibration methods were developed to control this uncertainty [49], which can be applied to future analyses.
- Theoretical uncertainties on the cross sections of the H and non-H signal and background processes. The background cross sections can also be fit by defining control regions using the ANN outputs. The `ZZnoH` discriminator can for example be used to define a region of events to which the non-H background can be fitted. This approach could significantly reduce the uncertainties related to the cross sections of the backgrounds.

It is important that these systematic uncertainties are taken into account in future studies of the H+c process, especially if the aim is to obtain state-of-the-art limits on $\kappa_c$. Instead, in this thesis, the viability of the use of the H+c channel and the `Hc` discriminator to obtain better limits on $\kappa_c$ is explored as a proof-of-concept. To this extent, the use of only statistical uncertainties is warranted.

### 4.4.2 Asimov dataset

By combining the relevant background samples with the SM H+c expectation, an Asimov dataset [50] is obtained. Bin-per-bin, this dataset represents the expected number of events from the simulated distributions assuming that $\kappa_c = 1$ (i.e. the SM scenario). Likewise, a dataset can be constructed out of the same background samples, but with the addition of an H+c sample where $\kappa_c \neq 1$. By means of a binned maximum likelihood fit, the sensitivity to a possible excess of signal events can be quantified by quoting the resulting 95% CL interval on $\kappa_c$, which will be described further in the next section. Figure 4.10 shows the `Hc` discriminator, Higgs boson transverse momentum $p_T$ and leading jet

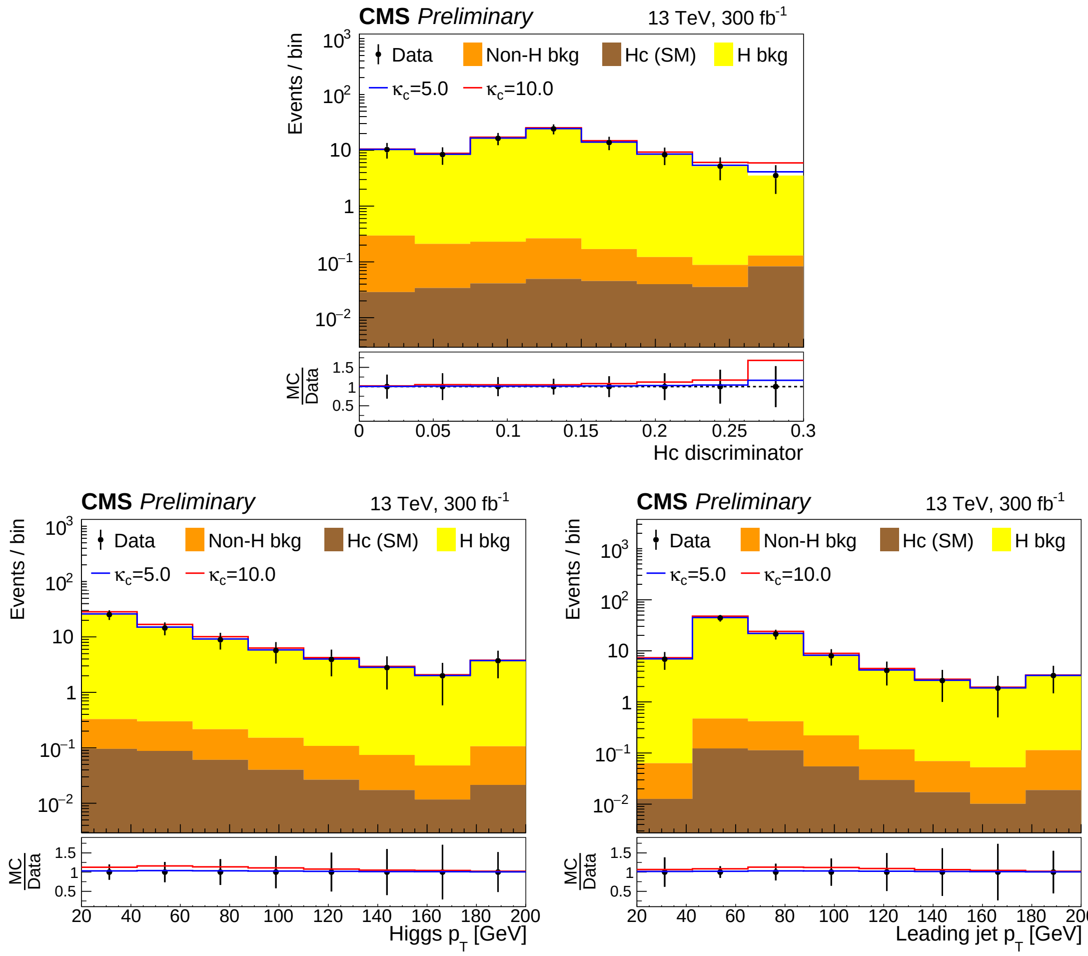



Figure 4.10: The Asimov dataset (black) constructed out of the background samples and the SM H+c sample, compared to the datasets constructed similarly but with a modified $\kappa_c$ value (blue and red). The distributions of these datasets are shown for the Hc discriminator, the leading jet $p_T$ (in the $p_T$ ranking) and the Higgs boson $p_T$. The shown uncertainties on the Asimov dataset are statistical uncertainties only.

$p_T$ (in the $p_T$ ranking) distributions for the Asimov dataset and the dataset corresponding with a modified $\kappa_c$ with respect to the SM value. In the `Hc` discriminator distribution, the last few bins seem to be most sensitive to changes in $\kappa_c$. In the last bin, this leads to an increase in the expected number of events of roughly 50% in the $\kappa_c = 10$ case. However, few events are expected in these bins for an integrated luminosity of 300 $fb^{-1}$. It is therefore expected that the availability of more data will lead to an increased event yield in these bins, which will result in a higher sensitivity to $\kappa_c$. The Higgs boson $p_T$ distribution on the other hand seems to be less sensitive to the modulation of $\kappa_c$. Here, the first few bins seem to be most sensitive, but the excess is less significant compared to the `Hc` discriminator distribution. The same is true for the leading jet $p_T$ distribution.

### 4.4.3 Maximum likelihood method

In the previous section, an Asimov dataset was constructed, along with similar datasets where the coupling $\kappa_c$ is modified with respect to the SM prediction, $\kappa_c = 1$. The total expected number of events $\nu_i$ is given bin-per-bin by this Asimov dataset. The other datasets, where $\kappa_c$ is modified, will generally contain a different value of expected events in each bin. These datasets act analogous to the result of an observation, in the sense that they will be used for comparison with the distributions of the Asimov dataset similarly to how real data is compared with an expected distribution. The *observed* number of events $n_i$ in bin $i$ is expected to follow a Poisson distribution, i.e. $n_i \sim \text{Pois}(\nu_i)$, where $\nu_i$ is the expectation value of the Poisson distribution. In other words, the probability for bin $i$ to contain $n_i$ events, given the expectation $\nu_i$, is given by

$$P(n_i, \nu_i) = \frac{\nu_i^{n_i} \mathrm{e}^{-\nu_i}}{n_i!}. \tag{4.5}$$

This expectation value is dependent on $\kappa_c$, since different values of $\kappa_c$ predict different bin-per-bin event yields for the variable distributions in figure 4.10. From these distributions, a likelihood function $\mathcal{L}$ can be constructed. By writing $\mathbf{n}$ and $\mathbf{v}$ respectively as the vectors containing the bin-per-bin *observed* number of events and bin-per-bin expected number of events, the likelihood function can be expressed as

$$\mathcal{L}(\mathbf{n}|\mathbf{v}) = \prod_{i=1}^{N} \frac{\nu_i^{n_i} \mathrm{e}^{-\nu_i}}{n_i!}, \tag{4.6}$$

where $N$ is the number of bins. The above quantity quickly reaches extremely small values, therefore one generally works with the logarithm of the likelihood function, or the log likelihood for short. This has the effect of transforming the product in equation (4.6) into a sum. The (negative) log likelihood is then given by

$$-\ln\left(\mathcal{L}(\mathbf{n}|\mathbf{v})\right) = \sum_{i=1}^{N} \left(-n_i \ln v_i + \ln(n_i!) + \nu_i\right). \tag{4.7}$$

### 4.4.4 Results

The log likelihood in equation (4.7) can be calculated for each of the distributions given in figure 4.10 by taking $\mathbf{n}$ from the Asimov dataset and $\mathbf{v}$ from each of the datasets containing modified $\kappa_c$ values. As a result, a curve can be fitted through the points $(\kappa_c, -\ln\left(\mathcal{L}(\mathbf{n}|\mathbf{v})\right))$ to obtain the log likelihood curve, which will reach a minimum in the trivial case where $\kappa_c = 1$ or $y_c = y_c^{\text{SM}}$. In this fit, the background yields remain fixed at their simulated predictions. In figure 4.11, the negative log likelihood curves are given for the Hc discriminator, leading jet $p_T$ (in the $p_T$ ranking) and Higgs boson $p_T$ distributions for different integrated luminosities of 35.9, 300 and 3000 fb$^{-1}$. The integrated luminosity of 35.9 fb$^{-1}$ corresponds with the 2016 dataset of CMS. In figure 4.11, the minimum of the log likelihood curve is shifted to zero, i.e. $-\Delta\ln(L) = -\left[\ln(L) - \ln(L_{\text{min}})\right]$

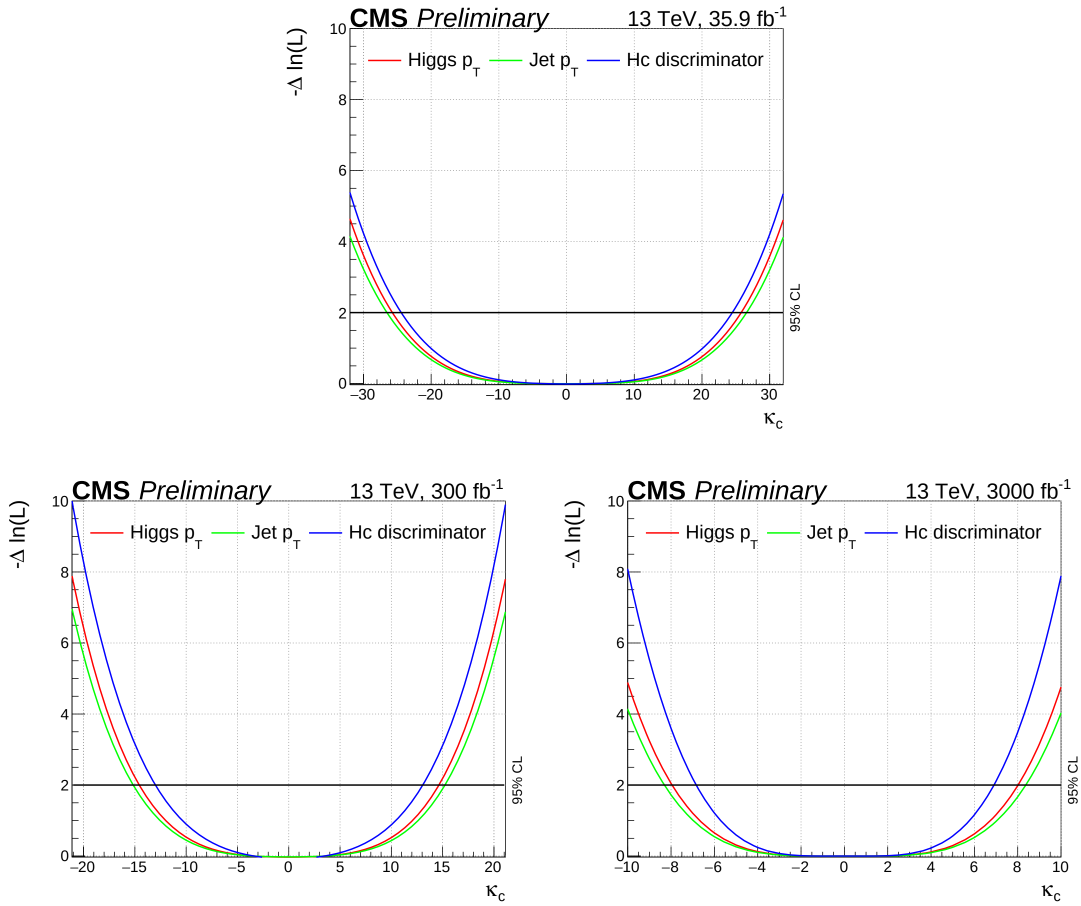



Figure 4.11: Results of the binned maximum likelihood for the Hc discriminator, leading jet $p_T$ (in the $p_T$ ranking) and Higgs boson $p_T$ distributions, for an integrated luminosity of 35.9 fb$^{-1}$, 300 fb$^{-1}$ and 3000 fb$^{-1}$.

is shown on the y-axis instead. The intersection of the log likelihood curve with the horizontal line $-\Delta \ln(L) = 2$ define the 95% confidence level (CL) intervals for $\kappa_c$. In each of these figures, the log likelihood curve is narrower for the fit to the `Hc` discriminator distribution compared to the curves associated with the other variables and as a consequence, a smaller 95% CL interval is obtained. The resulting CL intervals are given in table 4.5. As was mentioned previously, the `Hc` discriminator showed a higher sensitivity to $\kappa_c$ especially in the last few bins, where less events are expected. The leading jet $p_T$ and Higgs boson $p_T$ are most sensitive to $y_c$ in the bins which typically have a larger amount of events compared to the `Hc` discriminator distribution. Therefore, the 95% CL intervals `Hc` discriminator benefit the most from the additional data that becomes available for integrated luminosities of 300 and 3000 fb$^{-1}$, as can be seen in figure 4.11. For the latter scenario, a relative improvement in the limits on $\kappa_c$ of about 18% and 15% is

Table 4.5: The 95% confidence level (CL) intervals from the binned likelihood fit to the Hc discriminator, leading jet $p_T$ (in the $p_T$ ranking) and Higgs boson $p_T$ distributions.

| ***95% CL intervals for*** $\kappa_{\mathbf{c}}$ *(simulation)* | | | |
|---|---|---|---|
| Variable | 35.9 fb$^{-1}$ | 300 fb$^{-1}$ | 3000 fb$^{-1}$ |
| Hc discriminator | [-24.4, 24.5] | [-13.0, 13.1] | [-6.8, 6.9] |
| Leading jet $p_T$ | [-26.5, 26.6] | [-15.2, 15.2] | [-8.3, 8.4] |
| Higgs $p_T$ | [-25.7, 25.8] | [-14.6, 14.6] | [-8.0, 8.0] |

obtained with respect to the limits resulting from the fit to respectively the leading jet and Higgs boson $p_T$ distributions. For an integrated luminosity of 35.9 and 300 fb$^{-1}$, a relative increase in the sensitivity of $\kappa_c$ of respectively up to 8% and 14% is expected. Variations of $\kappa_c$ have a larger shape effect on the distribution of the `Hc` discriminator compared to the leading jet $p_T$ and Higgs boson $p_T$ distributions, where the variations for the largest part result in a yield shift in the bulk of the distributions. For this reason, the sensitivity gain becomes more apparent with higher integrated luminosity.

# Conclusion and outlook

In the 1960s, Robert Brout, Francois Englert [4] and Peter Higgs [5] independently predicted the existence of the scalar Brout-Englert-Higgs (BEH) boson. The BEH boson, or Higgs boson in short, is an essential component of the Standard Model (SM) of particle physics. Through its interactions with fermions and massive vector bosons, mass terms are generated for these particles. Nearly 50 years later, the elusive Higgs boson was finally discovered by the CMS [7] and ATLAS [6] Collaborations at the Large Hadron Collider at CERN. A linear relationship is predicted by the SM between the mass of a fermion and its coupling to the Higgs boson. In other words, the SM predicts that heavier particles couple more strongly to the Higgs boson. Recent findings of the ATLAS and CMS Collaborations indicate that the couplings of the third-generation fermions with the Higgs boson are consistent with the relationship predicted by the SM [19–22]. Recently, the CMS Collaboration has also found the muon coupling to be compatible with the SM prediction [16]. The next particle to have their coupling measured is the charm quark, the heavier of the second-generation quarks.

Previous attempts at constraining the charm quark Yukawa coupling $y_c$ relied on the direct detection of the Higgs boson decay channel $H \to c\bar{c}$ to a charm quark-antiquark pair [26]. Due to the large QCD multijet background and relatively low branching rate, this process is challenging to observe. On top of this, the Higgs boson needs to be reconstructed from two charm jets, which need to be identified by heavy-flavour taggers over a large background of bottom jets from the Higgs decay $H \to b\bar{b}$. Instead, a new method is proposed where the charm quark will be probed in the process $gc \to Hc$. In this process, the sensitivity to $y_c$ enters in the production of a Higgs boson instead of in the Higgs decay. As a consequence, the Higgs boson can be reconstructed from one of its clean decay signatures. Additionally, this process relies on the identification of only one charm jet instead of two. In this thesis, the H+c production channel will be studied by reconstructing the Higgs from the four-muon final state, $H \to ZZ^* \to 4\mu$.

The H+c process was simulated at leading order using the `MadGraph5_aMC@NLO` matrix element generator supplied with the `Higgs Effective Couplings` model, which allows to include variations of $\kappa_c = y_c/y_c^{\mathrm{SM}}$ directly in the simulation by using event weights. The interactions of the simulated physics objects with the CMS detector were simulated by `Geant4`. Finally, the result of this simulation is passed through the entire

CMS reconstruction chain to obtain a H+c sample ready for analysis. The kinematical properties of the event were augmented with information on the jet flavour using state-of-the-art charm tagging algorithms, and were subsequently used to train an artificial neural network (ANN). The outputs of this ANN were them transformed into a `Hc` and a `ZZnoH` discriminator. The former was constructed to distinguish events from the H+c process that are sensitive to $y_c$ from the H+c and H background events that do not directly depend on $y_c$. This way, a variable was obtained with improved sensitivity to $y_c$. The `ZZnoH` discriminator on the other hand, attempts to identify events that contain a pair of Z bosons from a Higgs boson decay from events that contain a pair of Z bosons that were formed through other mechanisms. The `ZZnoH` discriminator was then used to further reduce the background from non-H events.

The simulated H+c sample with $\kappa_c = 1$ was combined with the background samples to obtain an Asimov dataset. By means of a binned maximum likelihood fit, a possible excess of signal events resulting from a deviating charm quark Yukawa coupling with respect to the SM prediction is transformed into a 95% confidence level (CL) interval. Compared to the fit using only the leading jet or Higgs boson transverse momentum $p_T$ distributions, more strict limits are obtained from the fit to the `Hc` discriminator distribution. For an integrated luminosity of 300 fb$^{-1}$, the 95% CL constraints on $\kappa_c$ obtained from the fit to the leading jet and Higgs boson $p_T$ distributions are respectively $\kappa_c \in [-15.2, 15.2]$ and $\kappa_c \in [-14.6, 14.6]$, while the fit to the `Hc` discriminator spectrum yielded a 95% CL of $\kappa_c \in [-13.0, 13.1]$, which is an improvement of respectively 14% and 11%. This improvement with respect to the other variables, which were used in previous analyses [23, 28] to constrain $y_c$, only grows when more data becomes available. The shape difference in the Hc discriminator distribution becomes more dominant at higher integrated luminosity than simply the increase in cross section of the $gc \to Hc$ process. As a result, the `Hc` discriminator has an improved sensitivity compared to the leading jet and Higgs boson $p_T$, and a binned likelihood fit to its distribution seems to lead to stricter 95% CL intervals on $\kappa_c$.

This study however has some limitations. For one, no systematic uncertainties were included in this analysis. Especially the uncertainties related to c-tagging might be significant, though these can be controlled using calibration methods as outlined in [49]. Additionally, the simulation of the H+c was only at leading order. The advantage of this was that the H+c events insensitive to $y_c$ coming from the effective diagram in figure 1.8 could be separated from the sensitive H+c events in a straightforward way. For next-to-leading order (NLO) simulations, this distinction will be less clear. In the analysis presented in this thesis, the backgrounds remain fixed during the likelihood fit. Instead, one could consider a two-dimensional fit of both the Hc and ZZnoH discriminator distributions of the backgrounds together with the signal, shown in figure 4.12. A downside of this approach is however the increased number of bins in the 2D phase-space compared to the 1D case, which leads to a decreased event yield in each bin. This problem can however be taken care of by including additional clean decay channels of

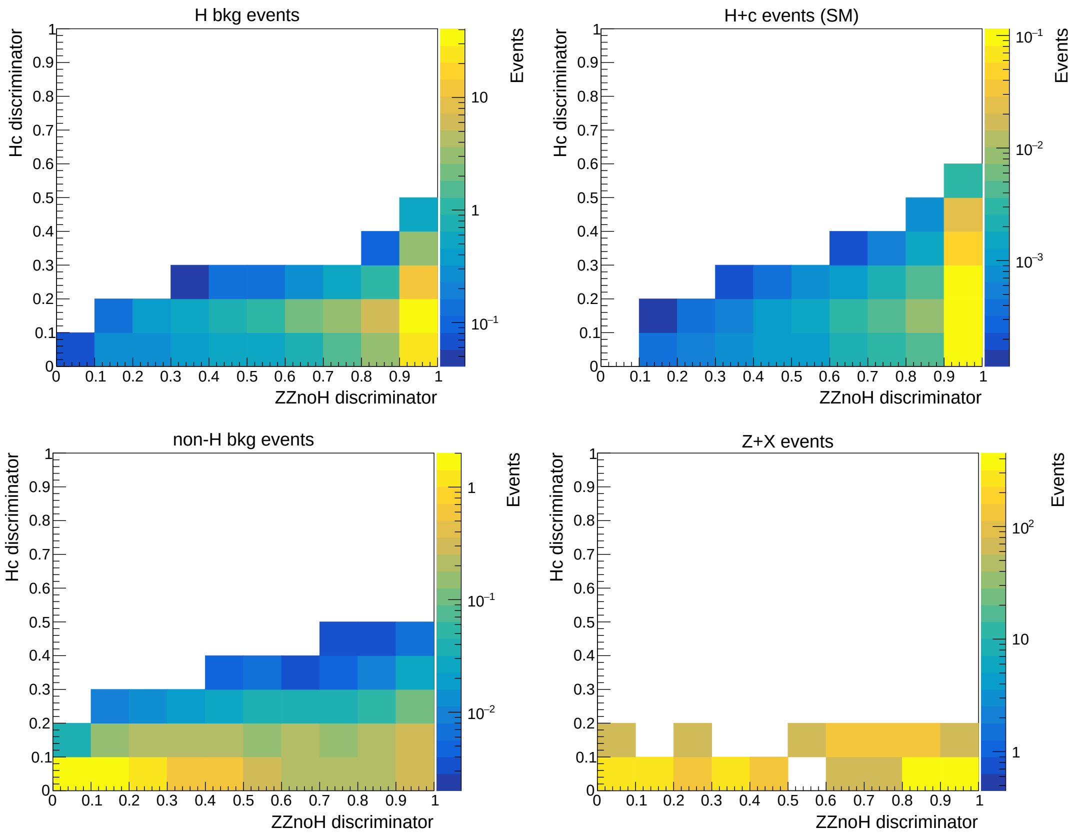



Figure 4.12: Two-dimensional distributions of the ZZnoH and Hc discriminators for the H background (top left), H+c (top right), non-H background (bottom left) and Z + X (bottom right) samples.

the Higgs boson, for example the $2e2\mu$ final state of the $H \to ZZ^*$ decay or the $H \to 2\gamma$ decay channel. Including these contributions on top of the four-muon final state that was used in this analysis will lead to an increased event yield, and thus an increased sensitivity of the Hc discriminator to $\kappa_c$.

In closing, a new method was developed in this thesis to constrain the charm quark Yukawa coupling using a Hc discriminator constructed from the outputs of an artificial neural network, which combined the kinematical properties of events with the charm-tagging information of the jets. A fit to the distribution of this discriminator resulted in more sensitive 95% CL intervals on $\kappa_c$, owing to the improved sensitivity of the Hc discriminator compared to the leading jet or Higgs boson $p_T$. The method in which $y_c$ is probed in the production of a Higgs boson $gc \to Hc$ has for the first time been studied in an experimental context, and is complementary to the $H \to c\bar{c}$ analysis. Accordingly, the method is an added value in the LHC programme for the measurement of $y_c$.

# Contributions by the author

This section serves to list the original contributions that I made during the development of this project. These achievements are listed below.

- I have created a dedicated simulation of the $gc \to Hc$ process, from the matrix element generation with `MadGraph5` up to the reconstruction of physics objects at the CMS detector.
- I have developed and implemented an event selection to select $H + c$ events in the $H \to ZZ^* \to 4\mu$ decay channel.
- I have trained an artificial neural network which combines charm tagging information on the jets with variables related to the kinematics of an event to define two discriminators. I have also increased the performance of this neural network by optimising the architecture and the input variables.
- I have extracted 95% confidence level intervals on $\kappa_c$ using a binned maximum likelihood fit on the distributions of the `Hc` discriminator and the Higgs and leading jet transverse momentum.

# Acknowledgements

This project has been an incredibly educational experience for me. I definitely would not have been able to bring this work to a conclusion if it wasn't for the contributions of various people. First of all, I would like to express my deep appreciation to my promotor, Professor Dr. Jorgen D'Hondt, for providing me with the opportunity to work on this project. Due to the protective measures related to the COVID-19 pandemic, he arranged for me to sit in on virtual meetings which allowed me to collaborate with the wonderful people of the IIHE. The valuable and constructive feedback from these meetings were especially helpful to the development of this thesis.

My special thanks are extended to my copromotor, Dr. Seth Moortgat, who guided me and closely followed my progress throughout the academic year. Even on the busiest of days, whether it be between meetings or after work hours, he was always available to support me and answer any of my questions. His patience when navigating me through a maze of account setups and permission queries, or his genuine excitement when a particular result came out right; I truly could not have wished for a better supervisor, for which I am very grateful.

I currently find myself at the end of an online academic year. Without the group of people with which I have shared my entire academic journey, this year would have been far more difficult than it already was. Our virtual nights spent chatting, gaming or even watching movies provided me with the moral support I so direly needed. For this I thank you, my friends at the VUB.

Finally, I should mention that without the support of my family, none of this would have been possible. Staying at home for such a long time made me realise even more how lucky I am to have these fantastic people in my life.

# Appendices

# Appendix A

# Backpropagation in more detail

Calculating the gradient in equation (2.10) is certainly a challenge. One approach which is widely used in training is backpropagation. The general idea behind backpropagation is to introduce a small change in the weights of the network layer-by-layer, observe the effect on the result of the loss function, and then update the weights accordingly. The activation $z_j^\ell$ of a neuron in a hidden layer $l$ can generally be expressed as

$$z_j^\ell = f\left(\sum_{i=1}^{M} w_{ji} z_i^{\ell-1} + w_{j0}\right) = f(s_j^\ell), \tag{A.1}$$

where $z_i^{\ell-1}$ is the activation of the $i$-th hidden neuron in the previous layer $\ell - 1$. Here, $s_j^\ell$ is defined as the argument of the activation function. The dependence of the loss function $L$ on $w_{jk}$, which is one of the weights used in the calculation of $z_j^\ell$, can be expressed using the chain rule,

$$\frac{\partial L}{\partial w_{jk}} = \sum_{i=1}^{M} \frac{\partial L}{\partial z_i^\ell} \frac{\partial z_i^\ell}{\partial w_{jk}} = \frac{\partial L}{\partial z_j^\ell} \frac{\partial z_j^\ell}{\partial w_{jk}} \tag{A.2}$$

where the sum goes over the activations in layer $\ell$. Since tweaking the weight $w_{jk}$ will only have an effect on the activation $z_j^\ell$, the derivative $\partial z_i^\ell / \partial w_{jk}$ is only non-zero for $i = j$. The chain rule can be applied again to $z_j^\ell$ such that

$$\frac{\partial L}{\partial w_{jk}} = \frac{\partial L}{\partial z_j^\ell} \frac{\partial z_j^\ell}{\partial s_j^\ell} \frac{\partial s_j^\ell}{\partial w_{jk}} = \frac{\partial L}{\partial z_j^\ell} f'(s_j^\ell) z_k^{\ell-1}, \tag{A.3}$$

where $f'(s_j^\ell)$ denotes the derivative of the activation function evaluated at $s_j^\ell$. The last step made use of the fact that $z_j^\ell = f(s_j^\ell)$. Finally, an expression for the partial derivative of the loss function $L$ with respect to the activation $z_j^\ell$ is needed. Remark that the loss function $L$ is also dependent on the neurons of the next hidden layer $\ell + 1$. Furthermore, the activation $z_j^\ell$ is used in the calculation of the activations of each of the neurons in this next layer. As a result, the chain rule can be applied again to this term such that

$$\begin{aligned}\frac{\partial L}{\partial z_j^\ell} &= \sum_{i=1}^{M} \frac{\partial L}{\partial z_i^{\ell+1}} \frac{\partial z_i^{\ell+1}}{\partial z_j^\ell} = \sum_{i=1}^{M} \frac{\partial L}{\partial z_i^{\ell+1}} \frac{\partial z_i^{\ell+1}}{\partial s_i^{\ell+1}} \frac{\partial s_i^{\ell+1}}{\partial z_j^\ell} \\ &= \sum_{i=1}^{M} \frac{\partial L}{\partial z_i^{\ell+1}} f'(s_i^{\ell+1}) w_{ij}^{\ell+1}\end{aligned} \tag{A.4}$$

As a result, the partial derivative $\partial L/\partial z_j^\ell$ is known if the dependence of the loss function on the subsequent layer's neurons is known. At the final hidden layer, the neurons in the next layer are simply the output neurons, on which the loss function's dependence is known. Therefore, starting from the output layer, the above equations can be used to calculate the layer-by-layer gradient of the loss function. By running through the ANN in reverse, the weights can be updated layer-by-layer according to equation (2.10).

# Appendix B

# ANN bare output distributions

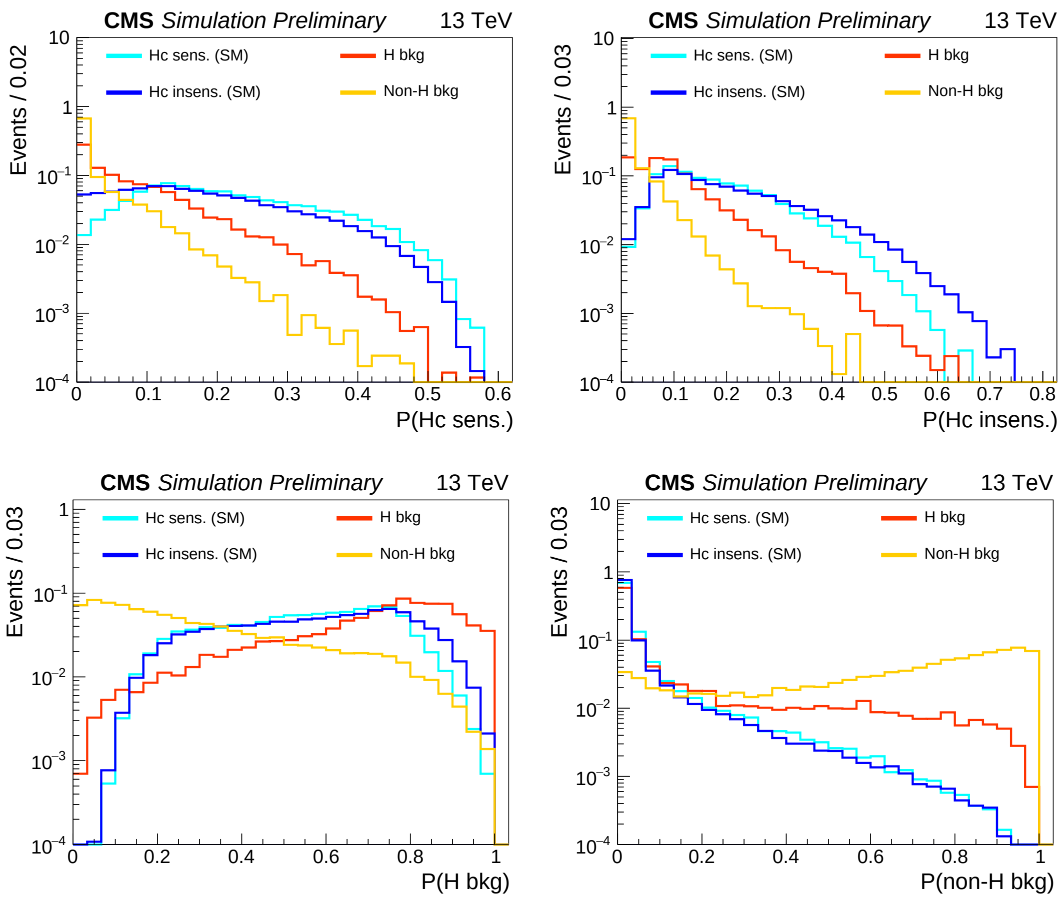



Figure B.1: The normalised distributions of the four outputs P(Hc sens.), P(Hc insens.), P(H bkg) and P(non-H bkg) of the ANN that was trained in chapter 4.

# Bibliography


[1] Istvan Bodnar. "Aristotle's Natural Philosophy". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2018. Metaphysics Research Lab, Stanford University, 2018.

[2] Q. R. Ahmad et al. "Measurement of the Rate ofe+d→p+p+eInteractions Produced byB8Solar Neutrinos at the Sudbury Neutrino Observatory". In: *Physical Review Letters* 87.7 (July 2001). ISSN: 1079-7114. DOI: 10.1103/physrevlett.87.071301. URL: http://dx.doi.org/10.1103/PhysRevLett.87.071301.

[3] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Physical Review Letters* 81.8 (Aug. 1998), pp. 1562–1567. ISSN: 1079-7114. DOI: 10.1103/physrevlett.81.1562. URL: http://dx.doi.org/10.1103/PhysRevLett.81.1562.

[4] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[5] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[6] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[7] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[8] Esma Mobs. "The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019". In: (July 2019). General Photo. URL: https://cds.cern.ch/record/2684277.

[9] Lyndon Evans. "The Large Hadron Collider". In: *Annual Review of Nuclear and Particle Science* 61.1 (2011), pp. 435–466. DOI: 10.1146/annurev-nucl-102010-130438. eprint: https://doi.org/10.1146/annurev-nucl-102010-130438. URL: https://doi.org/10.1146/annurev-nucl-102010-130438.

[10] J. T. Boyd. *LHC Run-2 and Future Prospects*. 2020. arXiv: 2001.04370 [hep-ex].

[11] Apollinari G. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*. CERN Yellow Reports: Monographs. Geneva: CERN, 2017. DOI: 10.23731/CYRM-2017-004. URL: https://cds.cern.ch/record/2284929.

[12] CMS Detector [online]. "CMS detector schematic". In: (). [Accessed 20 May 2021]. URL: https://cms.cern/detector.

[13] S. Chatrchyan et al. "The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment". In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08004. 361 p. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://cds.cern.ch/record/1129810.

[14] David Barney. "Presentation for public - Introduction to CMS for CERN guides". In: (June 2013). URL: https://cds.cern.ch/record/2629323.

[15] A. M. Sirunyan et al. "Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$". In: *The European Physical Journal C* 79.5 (May 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6909-y. URL: http://dx.doi.org/10.1140/epjc/s10052-019-6909-y.

[16] A. M. Sirunyan et al. "Evidence for Higgs boson decay to a pair of muons". In: *Journal of High Energy Physics* 2021.1 (Jan. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep01(2021)148. URL: http://dx.doi.org/10.1007/JHEP01(2021)148.

[17] A.M. Sirunyan et al. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Physics Letters B* 805 (June 2020), p. 135425. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135425. URL: http://dx.doi.org/10.1016/j.physletb.2020.135425.

[18] P.A. Zyla et al. "Review of Particle Physics". In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[19] A.M. Sirunyan et al. "Observation of Higgs Boson Decay to Bottom Quarks". In: *Physical Review Letters* 121.12 (Sept. 2018). ISSN: 1079-7114. DOI: 10.1103/physrevlett.121.121801. URL: http://dx.doi.org/10.1103/PhysRevLett.121.121801.

[20] A.M. Sirunyan et al. "Observation of $t\bar{t}H$ Production". In: *Physical Review Letters* 120.23 (June 2018). ISSN: 1079-7114. DOI: 10.1103/physrevlett.120.231801. URL: http://dx.doi.org/10.1103/PhysRevLett.120.231801.

[21] M. Aaboud et al. "Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector". In: *Physics Letters B* 786 (Nov. 2018), pp. 59–86. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2018.09.013. URL: http://dx.doi.org/10.1016/j.physletb.2018.09.013.

[22] M. Aaboud et al. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Physics Letters B* 784 (Sept. 2018), pp. 173–191. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2018.07.035. URL: http://dx.doi.org/10.1016/j.physletb.2018.07.035.

[23] Fady Bishara et al. "Constraining Light-Quark Yukawa Couplings from Higgs Distributions". In: *Physical Review Letters* 118.12 (Mar. 2017). ISSN: 1079-7114. DOI: 10.1103/physrevlett.118.121801. URL: http://dx.doi.org/10.1103/PhysRevLett.118.121801.

[24] Gilad Perez et al. "Constraining the charm Yukawa and Higgs-quark coupling universality". In: *Physical Review D* 92.3 (Aug. 2015). ISSN: 1550-2368. DOI: 10.1103/physrevd.92.033016. URL: http://dx.doi.org/10.1103/PhysRevD.92.033016.

[25] Albert M Sirunyan et al. "Measurement and interpretation of differential cross sections for Higgs boson production at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 792 (2019), pp. 369–396. DOI: 10.1016/j.physletb.2019.03.059. arXiv: 1812.06504 [hep-ex].

[26] A. M. Sirunyan et al. "A search for the standard model Higgs boson decaying to charm quarks". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep03(2020)131. URL: http://dx.doi.org/10.1007/JHEP03(2020)131.

[27] M. Aaboud et al. "Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment". In: *Phys. Rev. Lett.* 120.21 (2018), p. 211802. DOI: 10.1103/PhysRevLett.120.211802. arXiv: 1802.04329 [hep-ex].

[28] Yotam Soreq, Hua Xing Zhu, and Jure Zupan. "Light quark Yukawa couplings from Higgs kinematics". In: *Journal of High Energy Physics* 2016.12 (Dec. 2016). ISSN: 1029-8479. DOI: 10.1007/jhep12(2016)045. URL: http://dx.doi.org/10.1007/JHEP12(2016)045.

[29] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain." In: *Psychological Review* 65.6 (1958), pp. 386–408. DOI: 10.1037/h0042519.

[30] Mark Thomson. *Modern particle physics*. Cambridge University Press, 2019.

[31] 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". In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079.

[32] Adam Alloul et al. "FeynRules 2.0— A complete toolbox for tree-level phenomenology". In: *Computer Physics Communications* 185.8 (Aug. 2014), pp. 2250–2300. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2014.04.012. URL: http://dx.doi.org/10.1016/j.cpc.2014.04.012.

[33] Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Boulder, CO: Westview, 1995.

[34] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "A brief introduction to PYTHIA 8.1". In: *Computer Physics Communications* 178.11 (June 2008), pp. 852–867. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2008.01.036. URL: http://dx.doi.org/10.1016/j.cpc.2008.01.036.

[35] Bryan R Webber. "Hadronization". In: (1994). URL: https://cds.cern.ch/record/272660.

[36] "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[37] A.M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/10/p10003. URL: http://dx.doi.org/10.1088/1748-0221/12/10/P10003.

[38] The CMS collaboration. "Performance of CMS muon reconstruction in pp collision events at s = 7TeV". In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002–P10002. ISSN: 1748-0221. DOI: 10.1088/1748-0221/7/10/p10002. URL: http://dx.doi.org/10.1088/1748-0221/7/10/P10002.

[39] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[40] A.M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011. URL: http://dx.doi.org/10.1088/1748-0221/13/05/P05011.

[41] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/p12012. URL: http://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[42] Seth Moortgat. "When charm and beauty adjoin the top. First measurement of the cross section of top quark pair production with additional charm jets with the CMS experiment." Presented 20 May 2019. May 2019. URL: https://cds.cern.ch/record/2676133.

[43] D. Claude. *Higgs Effective Theory - FeynRules*. URL: https://feynrules.irmp.ucl.ac.be/wiki/HiggsEffectiveTheory.

[44] Olivier Mattelaer. "On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy". In: *The European Physical Journal C* 76.12 (Dec. 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4533-7. URL: http://dx.doi.org/10.1140/epjc/s10052-016-4533-7.

[45] Albert M Sirunyan et al. "Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (Mar. 2021). arXiv: 2103.04956 [hep-ex].

[46] *HL-LHC High Luminosity Large Hadron Collider: The HL-LHC project.* [*Online*]. URL: https://hilumilhc.web.cern.ch/content/hl-lhc-project.

[47] François Chollet et al. *Keras*. https://github.com/fchollet/keras. 2015.

[48] Martın Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[49] CMS Collaboration. *Calibration of charm jet identification algorithms using proton-proton collision events at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2021. URL: http://cds.cern.ch/record/2758866.

[50] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.