A search for the Higgs boson decay to two electrons with the Compact Muon Solenoid experiment

Joseph William Davies

Imperial College London
Department of Physics

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

A search is presented for the Higgs boson decay to a pair of electrons in proton-proton collisions at a centre-of-mass energy of 13 TeV. The data set was collected with the Compact Muon Solenoid experiment at the Large Hadron Collider between 2016 and 2018, corresponding to an integrated luminosity of 138 fb^{-1}. The analysis develops event categories targeting Higgs boson production via gluon fusion and vector boson fusion, defined by selection on dedicated machine learning-based classifiers. An upper limit on the Higgs boson branching fraction to an electron pair is determined as 3.0×10^{-4} at the 95% confidence level, which is the most sensitive limit to date.

To Mum, Dad,
and my sister, Liv

Declaration

I, the author, declare that all work presented in this thesis is my own. The studies have been conducted in collaboration with several individuals, as members of the Compact Muon Solenoid (CMS) Collaboration. All figures produced by the author, or in collaboration with the author, are labelled as "CMS" if taken directly from a CMS publication, "CMS Preliminary" if taken from a CMS preliminary public document, and figures which have not explicitly been made public by the CMS Collaboration attain the label "CMS Work in Progress", or are simply left unlabelled. The figures and studies that are taken from external sources, and are not produced directly by the author, or in collaboration with the author, are appropriately referenced throughout.

Joseph William Davies

Acknowledgements

There are so many people who have helped make the last four years an enjoyable and fulfilling experience. This thesis has been greatly shaped by their influence, and although I cannot name each and every person here, I am extremely grateful to you all.

Firstly, I would like to thank the HEP department at Imperial College for allowing me to study my PhD here. The group is full of warm and witty personalities, each with their own valuable expertise to share. I could not think of a better place to study, and will miss it greatly when I leave. Thank you also to the STFC for funding my work, without which none of this would have been possible.

Nick, thank you for your guidance throughout my studies. Despite your many commitments, you have always found time to answer my (often naive) questions, and these conversations have proved invaluable. Ed, you have supported me throughout almost the entirety of my PhD, all the way from TwoStep, to submission of this thesis. I've learned so much from your mentorship, I think both academically and otherwise, and will always carry these lessons with me. Jon, you were one of the first people to convince me to study at Imperial. For that, and for so much more throughout the last few years, I am sincerely grateful. It has been a real privilege to work with you and Ed, and to say that this thesis would not have been possible without you both is an understatement.

To my other friends at Imperial, both present and past, thank you for keeping me sane and always being up for a laugh. My time here was made so much more enjoyable by your company. Thank you also to my friends at home who have provided humour, support, and beverages, at much needed times.

To my family — thank you for everything. Your love has been an enduring constant in my life, and for that I am forever grateful. Dad, although physics is perhaps/absolutely not your own interest, your support throughout my studies has allowed me to pursue it as mine. You are calm, patient, caring, and I could not ask for a better role model. Mum, I could have filled this entire page writing about what you have done for me and Liv. Your love and compassion is the bedrock of my accomplishments, and thus, this thesis is your achievement as much as it is mine. Susie, I cannot tell you how important your support for me has been over the last four years. Were it not for your patience and encouragement,

there would be no thesis to write. I am so grateful to have you in my life. Finally, Liv — I feel incredibly lucky to have grown up with you and shared almost all of our lives together. You're my best friend and I'm so proud of everything you do. It is one of the great privileges of my life to call you my twin.

I love you all very much, and promise not to make you read all of this.

Contents

List of Figures

List of Tables

Preface

This thesis presents a complete description of the search for Higgs boson decays to two electrons published in Ref [1]. I am responsible for all aspects of the analysis workflow, from the development of event categorisation procedures to extraction of the final results. In addition to this search, I made significant contributions to the $H \to \gamma\gamma$ Simplified Template Cross Section analysis, published in Ref [2]. For brevity, this work will not be discussed in this thesis. A summary and relevant accreditation for each chapter is given below.

Chapter 1 introduces this work, with reference to the current landscape of Higgs boson coupling measurements.

Chapter 2 describes the theory underpinning the standard model of particle physics, with emphasis on the Brout-Englert-Higgs mechanism. This is entirely the work of others, summarised in my own words.

Chapter 3 provides an overview of key concepts in the field of machine learning, pertinent to later chapters of this thesis. This is the work of others summarised in my own words.

Chapter 4 briefly describes the Large Hadron Collider, before focusing on the Compact Muon Solenoid detector, which was designed, built, and operated by others. This chapter concludes with studies of photon and electron identification techniques in the Level-1 Trigger for the Phase-2 upgrade to the CMS calorimeter endcaps. These studies include a baseline description of the models used, which were developed by myself and Jonathon Langford, as well as memory and latency optimisation for model deployment, which is my own work.

Chapter 5 describes the techniques used to reconstruct particles in the CMS detector, which is the work of others, although the validation of the electron energy scale corrections for the $H \to ee$ analysis was checked by myself. This chapter also outlines the simulated samples and data used to perform the $H \to ee$ search.

Chapter 6 describes the categorisation of $H \to ee$ events to target the gluon fusion and vector boson fusion Higgs boson production mechanisms. The content presented in this chapter is entirely my own work, including the construction, interpretation, and

validation of Boosted Decision Tree classifiers used in the nominal approach, and studies into deep learning techniques for categorisation of VBF events.

Chapter 7 describes the development of signal and background models, as well as the systematic uncertainties included in this analysis, and potential bias within the envelope method. This work is my own, although the original techniques, which were modified by myself for this analysis, were developed by others.

Chapter 8 reports the results of this analysis. This is my own work, although the methodology and tools were developed by others.

Chapter 9 gives a summary of this analysis in my own words.

"One day I will find the right words,
and they will be simple."

Jack Kerouac

Chapter 1

Introduction

The standard model (SM) of particle physics describes the interactions of the known elementary particles through three of the four fundamental forces of nature [3–5]. It is widely regarded as one of the most successful physical theories of the 20^{th} century, with its predictions withstanding the scrutiny of a myriad of experimental measurements. One cornerstone in the construction of the SM was the postulation of the Brout-Englert-Higgs (BEH) mechanism [6–10], which breaks the symmetry of the electroweak interaction. The BEH mechanism not only provides an explanation for how the fundamental particles acquire mass, but also predicts the existence of a new, fundamental scalar particle: the Higgs boson. The subsequent discovery of the Higgs boson by the CMS [11] and ATLAS [12] Collaborations at the Large Hadron Collider (LHC) [13] in 2012 provided confirmation of the BEH mechanism [14–16], concluding the experimental observation of all particles predicted by the SM.

Despite its successes, the SM is not considered complete, for the reason that it still leaves many observed phenomena unexplained. For example, the SM provides no explanation for neutrino masses, which have been verified experimentally to be non-zero [17]. In addition, it offers no description of the gravitational interactions between particles. The SM also fails to explain cosmological observations such as the baryon-antibaryon asymmetry of the universe, and does not currently provide an accepted candidate for dark matter. To resolve its shortcomings, various extensions beyond the SM (BSM) have been proposed, with many postulating the existence of new particles that would modify the predictions of the SM, and alter the dynamics of the particles it describes. Testing the predictions of such BSM theories is a key objective for particle physics experiments.

As the only fundamental scalar SM particle, the Higgs boson provides a unique tool with which to probe the SM, motivating precision tests of its properties, as well as searches for rare or forbidden decay channels. Many such analyses have already been performed by experiments at the LHC, including measurements of Higgs boson interactions with

the SM particles. In particular, interactions with the electroweak gauge bosons and third generation charged fermions have been observed [2, 18–30] with coupling strengths, shown in Figure 1.1, found to be in agreement with the SM predictions. More recently, the CMS Collaboration has also reported the first evidence for the decay of the Higgs boson to second generation fermions, in a search for the Higgs boson decaying to two muons [31]. Such measurements involving Higgs boson couplings to the lighter fermions are experimentally challenging, given that the SM predicts the Higgs boson coupling strength to fermions to be proportional to the fermion masses [3–5]. In addition, these rare decays are often buried amongst competing background processes that mimic the final state of signal events, yet have a significantly higher rate. Consequently, the Higgs boson couplings to the first generation fermions are yet to be confirmed experimentally.

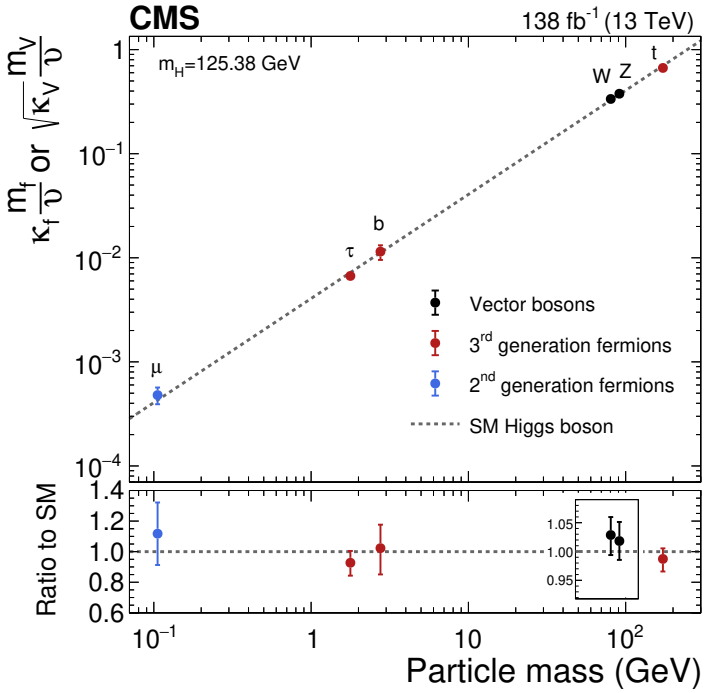

Figure 1.1: A summary of the measured couplings of the Higgs boson to the vector bosons (black markers), third generation fermions (red), and muons (blue). The results are expressed in the coupling modifier κ-framework [32], which introduces multiplicative factors at Higgs boson coupling vertices that parameterise deviations from SM behaviour. The error bars represent the 68% confidence level intervals on the measured parameters. The results are compatible with the SM predictions (dashed line), over four orders of magnitude. Figure taken from Ref [26].

This thesis studies the extremely rare decay of the Higgs boson to an electron-positron pair (hereafter referred to simply as an *electron pair*), $\mathrm{H} \rightarrow \mathrm{ee}$. The SM predicted branching fraction for this process is extremely small, at around $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee}) \approx 5.2 \times 10^{-9}$; how-

ever, a search for H $\rightarrow$ ee decays provides the only direct probe of the Higgs boson coupling to the electron[1]. Previously, both the CMS and ATLAS Collaborations have performed searches for H $\rightarrow$ ee decays. The most sensitive search from CMS was performed using data collected from proton-proton (pp) collisions at a centre-of-mass energy ($\sqrt{s}$) of 8 TeV, with an integrated luminosity of 19.7 fb^{-1}. A 95% confidence level (CL) upper limit on the H $\rightarrow$ ee branching fraction of approximately 1.9×10^{-3} (3.7×10^{5} times the SM prediction) was determined [34]. The most sensitive limit on $\mathcal{B}$(H $\rightarrow$ ee) from the ATLAS Collaboration was performed using 139 fb^{-1} of pp data, collected at $\sqrt{s} = 13$ TeV, where a 95% CL upper limit on the branching fraction was set at 3.6×10^{-4} (6.9×10^{4} times the SM prediction) [35].

The analysis described in this thesis was performed at the CMS experiment using data collected from pp collisions at $\sqrt{s} = 13$ TeV, and is published in Ref [1]. The data were collected during the Run 2 campaign of the LHC between 2016 and 2018, and correspond to an integrated luminosity of 138 fb^{-1}. In order to improve the analysis sensitivity, event categories are formed to target multiple Higgs boson production modes. Each category is defined using a dedicated machine learning (ML) discriminant trained to separate Higgs boson signal events from background processes. Finally, a simultaneous fit to the dielectron invariant mass distribution (m_{ee}) in each analysis category is used to extract a potential H $\rightarrow$ ee signal. Given the extremely small expected branching fraction, any observation of the signal process with the current dataset available from the LHC would constitute evidence for physics beyond the SM.

The remainder of this thesis is structured as follows. Chapter 2 describes the structure of the SM as a gauge field theory, with a particular focus on the BEH mechanism. The phenomenology of the Higgs boson at the LHC is also discussed. Chapter 3 provides an overview in the field of ML. The aim of this chapter is to contextualise the methods and models explored in other chapters of this thesis. The components of the LHC and CMS detector are described in chapter 4. This includes the CMS calorimeter endcap upgrade project (HGCAL), where emphasis is placed on ML models used within the first level of event triggering. Chapter 5 describes how events recorded with the CMS detector are reconstructed. The particle candidates used in the H $\rightarrow$ ee search are presented, with emphasis placed on the reconstruction and identification of electrons. The categorisation targeting H $\rightarrow$ ee events is presented in chapter 6, with a focus on the ML techniques used to reject background events. Chapter 7 outlines the signal and background models used to describe the dielectron mass distribution in each analysis category. Systematic uncertainties considered in the analysis are also discussed. Chapter 8

[1]Indirect probes of the electron Yukawa coupling also exist, such as via measurements of the electron electric dipole moment [33].

describes the statistical methods used to extract the results of the analysis. The expected and observed upper limits on the H $\rightarrow$ ee branching fraction are presented. Finally, chapter 9 summarises the H $\rightarrow$ ee analysis, and offers a brief outlook on future searches.

Chapter 2

Theory

2.1 Introduction

The standard model of particle physics is a quantum field theory (QFT) describing the fundamental particles and their interactions through the strong, electromagnetic (EM), and weak forces. In particular, the SM is a type of gauge field theory, described in the Lagrangian formalism, where the symmetries of a physical system manifest as invariances under transformations to particle wavefunctions. The framework allows for interactions between particles to be introduced naturally, as a required consequence of gauge invariance. The SM has been an extremely successful theory of nature, with its predictions verified by many experimental results. Most recently, this included the discovery of the Higgs boson by the CMS and ATLAS experiments [14–16] at the LHC, which provided confirmation of the BEH mechanism postulated almost fifty years prior.

This chapter firstly outlines the particles described in the SM, including the quarks, leptons, and gauge bosons. The development of the SM as a gauge field theory is then discussed, with reference to the EM and strong forces. Electroweak unification is then presented, before introducing symmetry breaking via the BEH mechanism [6–10]. Finally, the phenomenology of the Higgs boson at the LHC is discussed, including its production and decay channels, and their relative sensitivity.

2.2 SM particles and forces

Particles in the SM can be organised into spin-$\frac{1}{2}$ fermions, which are the fundamental constituents of matter, and integer spin gauge bosons, which mediate the three forces described in the SM. Fermions can be further divided into six quarks and six leptons, each of which belongs to one of three generations. Each generation of quark consists of an *up-type* and *down-type* quark, with fractional electric charge of $+2/3$ and $-1/3$

(in units of the elementary charge, $|e|$) respectively. Quarks are also distinguished from leptons by their colour charge, which allows them to interact via the strong force. The lepton generations consist of one charged particle and its corresponding neutrino. Charged leptons interact through either the EM or weak force, whereas neutrinos posses no electric or colour charge and are thus only permitted to interact via the weak force. All quarks and leptons have anti-particle versions of themselves with opposite charge[1], resulting in twelve additional anti-fermions.

The interactions between fermions are mediated by the gauge bosons, with each force having at least one associated carrier. The EM force is mediated by the massless photon, resulting in an effectively infinite range. The weak force is mediated by two charged massive bosons ($W^{\pm}$) and one neutral boson (Z), while the strong force is carried by massless gluons.

The final particle in the SM is the spin-0 Higgs boson, which arises as a consequence of the BEH mechanism. The mechanism is responsible for explaining how the fermions and weak gauge bosons acquire mass, while the photon remains massless.

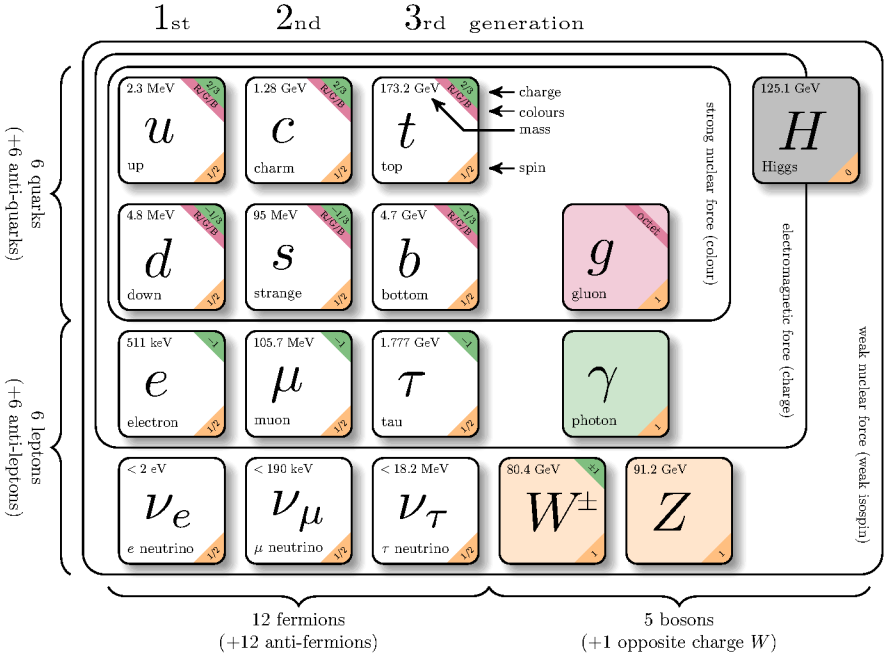

Figure 2.1: An overview of the particle content of the SM. Where relevant, particles are labelled with quantum numbers describing their spin, colour, and charge, along with their measured mass. Fermions are grouped according to their interactions through the weak, strong, and EM forces. Figure taken from Ref [36].

[1] With the exception of the three neutrinos, where the corresponding anti-particle also has no charge.

2.3 Gauge fields and symmetries

The SM is a gauge field theory, described using the Lagrangian formalism. The dynamics of such a theory can be derived from the corresponding Lagrangian using the Euler-Lagrange equations, similarly to the treatment for a classical system. The Lagrangian is typically constructed to obey an underlying symmetry of the physical system being described. According to Noether's theorem [37], for every such symmetry of the Lagrangian, there exists a corresponding conserved quantity. For example, invariance of the Lagrangian under a spatial translation implies the conservation of momentum. In gauge theory, the Lagrangian is required to be invariant under *local gauge transformations*, which apply a space-time dependent shift to the (unobservable) phase of the fields being described. The following sections describe the local gauge symmetries of the SM Lagrangian, and how they give rise to the fundamental forces.

2.3.1 Electromagnetic interactions

The QFT for electromagnetism is quantum electrodynamics (QED), describing charged fermions and their interactions with the photon. Terms in the corresponding QED Lagrangian that describe the spin-$\frac{1}{2}$ fermion fields, can be grouped into the so-called Dirac Lagrangian, $\mathcal{L}$. This Lagrangian will be used to illustrate how imposing local gauge invariance leads naturally to the introduction of additional gauge fields.

The Dirac Lagrangian [38] has the form

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi, \tag{2.1}$$

where ψ is the Dirac spinor, and $\bar{\psi}$ is the adjoint spinor defined as $\bar{\psi} = \psi^{\dagger}\gamma^{0}$. The terms γ^{μ} are the γ-matrices which obey the anticommutation relation $\{\gamma^{\mu}, \gamma^{\nu}\} = 2\eta^{\mu\nu}$, where $\eta^{\mu\nu}$ is the Minkowski metric, and $\mu, \nu \in \{0, 1, 2, 3\}$. Finally, m is the mass of the fermion described by ψ.

The Dirac Langrangian is clearly invariant under a *global* gauge transformation to the fermion fields. This transformation belongs to the U(1) symmetry group, and thus can be described by a single real parameter, θ, as

$$\psi \to e^{ig\theta}\psi, \tag{2.2}$$

where θ and g are constant, real numbers. Now consider elevating this symmetry to a *local* gauge transformation, of the kind

$$\psi \to e^{ig\theta(x)}\psi, \tag{2.3}$$

where $\theta(x)$ can now be different at any point in space-time, x. Unlike for global transformations, where the four-derivative commuted with the position independent value of θ, the corresponding Lagrangian is no longer invariant, acquiring the additional term

$$\mathcal{L} \rightarrow \mathcal{L} - g\bar{\psi}\gamma^{\mu}(\partial_{\mu}\theta(x))\psi. \tag{2.4}$$

To remove the residual term and restore gauge invariance, the derivative ∂_{μ} is replaced by the covariant derivative, D_{μ}, defined as:

$$D_{\mu} \equiv \partial_{\mu} + igA_{\mu}, \tag{2.5}$$

where A_{μ} is an additional field that has been added to the Dirac Lagrangian. The desired cancellation of the $g\bar{\psi}\gamma^{\mu}(\partial_{\mu}\theta(x))\psi$ term in Eqn 2.4 is achieved, provided the new field transforms as

$$A_{\mu} \rightarrow A_{\mu} - \partial_{\mu}\theta(x), \tag{2.6}$$

exhibiting the same gauge invariance observed in classical electromagnetism. In QED, the new boson is associated with the photon propagator, with g then assuming the electron charge. The modified Dirac Lagrangian, which no longer describes free particle fields, and is now locally gauge invariant, can be written as:

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}D_{\mu}\psi - m\bar{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{2.7}$$

where $F^{\mu\nu}$ is the electromagnetic field strength tensor, which is also invariant under local gauge transformations.

Compared with the original Lagrangian, the term $-g\bar{\psi}\gamma^{\mu}A_{\mu}\psi$ has been added, corresponding to the basic QED interaction vertex between the fermion and photon fields, shown by the Feynman diagram in Figure 2.2. The photon and its interactions thus arise naturally from requiring a locally gauge invariant description of fermionic interactions. Note that mass terms for the new field, which would take the form $-\frac{1}{2}m_A^2 A_{\mu}A^{\mu}$, are not included, since the term transforms as

$$\frac{1}{2}m_A^2 A_{\mu}A^{\mu} \rightarrow \frac{1}{2}m_A^2 \left(A_{\mu} - \partial_{\mu}\theta(x)\right)\left(A^{\mu} - \partial^{\mu}\theta(x)\right) \neq \frac{1}{2}m_A^2 A_{\mu}A^{\mu}, \tag{2.8}$$

breaking the gauge symmetry for values of $m_A \neq 0$. This feature will become important when discussing electroweak symmetry breaking, and the non-zero weak gauge boson masses.

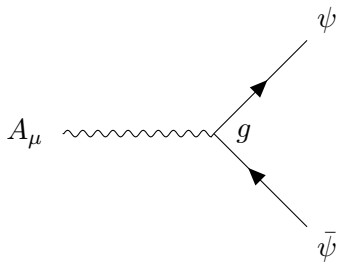

Figure 2.2: The fundamental QED vertex, where the photon field, A_μ, interacts with a charged fermion field described by ψ.

2.3.2 Strong interactions

The theory of strong interactions is known as quantum chromodynamics (QCD), describing quarks, gluons, and their interaction. The corresponding Lagrangian is again formed by requiring a physical symmetry to be respected; for QCD, this symmetry is invariance under a local SU(3) gauge transformation, which has form

$$\psi \rightarrow e^{ig_s\theta(x)^a \frac{\lambda^a}{2}}\psi, \tag{2.9}$$

where λ^a are the eight generators of the SU(3) group ($a \in \{1, 2, ..., 8\}$), represented by the 3×3 the Gell-Mann matrices, and $\theta(x)^a$ are eight functions of the spacetime co-ordinate, x [38]. The SU(3) symmetry of the QCD interaction implies a conserved quantity, which in the case of quark and gluon fields, is the colour charge. Particles (anti-particles) with colour charge exist in one of three states, labelled red (anti-red), blue (anti-blue), and green (anti-green).

In order to maintain invariance, the four-derivative is modified to the covariant derivative,

$$D_\mu \equiv \partial_\mu + ig_s\frac{\lambda^a}{2}G^a_\mu. \tag{2.10}$$

This necessitates the introduction of eight new fields, G^a_μ, which correspond to massless gluons, the mediators of the strong interaction. The value of g_s, therefore, is the strong force coupling strength.

The QCD lagragian remains invariant, provided the new fields transform as

$$G^c_\mu \rightarrow G^c_\mu - \partial_\mu\theta^c(x) - g_s f^{abc}\theta^a(x)G^b_\mu, \tag{2.11}$$

where f^{abc} are the completely antisymmetric structure constants, defined by $\left[\lambda^a, \lambda^b\right] = if^{abc}\lambda^c$, which arise due to the non-abelian nature of the SU(3) group. Unlike in the case of QED, a non-abelian gauge theory permits self-interaction of the accompanying gauge

bosons. This can be revealed by the required form of the field strength tensor, which for QCD is

$$F^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\mu, \tag{2.12}$$

resulting in trilinear and quartic terms in the QCD Lagrangian. Collection of the previous results together yields the final Lagrangian,

$$\mathcal{L}_{\text{QCD}} = \sum_f i\bar{\psi}_f(\gamma^\mu D_\mu - m_f)\psi_f - \frac{1}{4}F^{a\mu\nu}F^a_{\mu\nu}, \tag{2.13}$$

where f indexes the six quark flavours, with corresponding mass m_f.

An additional difference between QED and QCD is the relationship between the strength of the QCD interaction, and the interaction energy scale. The strong force interaction strength becomes larger at smaller momentum transfers, a property which results in the confinement of low energy quarks to colourless baryons and mesons. Conversely, at large energies such as those probed at the Large Hadron Collider, quarks and gluons behave as quasi-free particles, in an effect known as *asymptotic freedom*.

2.3.3 Electroweak unification

The unification of the EM and weak forces, originally conceived by Glashow, Weinberg and Salam (GWS) [3–5], consolidated the SM into its highly successful modern form. At a glance, the two forces already appear intimately linked; the coupling strengths are similar, and two of the weak bosons possess the charge of the EM interaction. The GWS model formalises these ideas into the modern gauge invariant theory.

It is useful to firstly consider the weak interaction alone, which is invariant under SU(2) transformations of the form

$$\psi \rightarrow e^{ig\theta(x)^i \frac{\sigma^i}{2}}\psi, \tag{2.14}$$

where the generators of the group have been identified as the three Pauli matrices, σ^i, and g is the corresponding charge known as *weak isospin*. Similarly to the previously discussed gauge groups, in order to respect the SU(2) symmetry, the four-derivative must be modified to its covariant form

$$D_\mu \equiv \partial_\mu + ig\frac{\sigma^i}{2}W^i_\mu, \tag{2.15}$$

which has resulted in the introduction of three new fields, W^i_μ. Given that the group generators are 2×2 matrices, the particle states that interact through the weak force

should be represented by two component vectors, which are termed *weak isospin doublets*.

Unlike the EM and strong forces, the weak interaction does not conserve parity [39], meaning it is not invariant under inversion of the spatial co-ordinates. The weak interaction structure has a vector minus axial-vector form

$$i\frac{g}{2\sqrt{2}}\gamma^{\mu}(1-\gamma^{5}) = i\frac{g}{\sqrt{2}}\gamma^{\mu}P_{L}, \tag{2.16}$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, and P_L has been identified as the left handed chiral projection operator. Given that the action of P_L on right handed particle spinors produces zero, an immediate consequence of parity violation is that only left (right) handed chiral particle (anti-particle) states may participate in weak interactions. To achieve this, only left handed particle and right handed anti-particle states are placed in weak isospin doublets, while right handed particle and left handed anti-particle states become isospin singlets, unaffected by SU(2) transformations. For this reason, the SU(2) symmetry group of the weak interactions is often referred to as $\mathrm{SU(2)_L}$.

In its current form, the weak interaction gauge group has three associated gauge bosons, all of which violate parity. This is, however, in contradiction with experiment, where the Z boson associated with weak neutral current interactions couples to both left handed and right handed chiral particles. The GWS model provides the resolution, where the two neutral bosons of the theory are postulated to arise from combinations of fields generated by an $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ symmetry. To maintain invariance under such a transformation, the covariant derivative associated with the weak interaction is updated with an additional term corresponding to the U(1) symmetry, such that overall

$$D_{\mu} \equiv \partial_{\mu} + ig\frac{\sigma^{i}}{2}W_{\mu}^{i} + ig'YB_{\mu}, \tag{2.17}$$

where Y is known as the *weak hypercharge*, and g' is the associated coupling. In the GWS model, the charged weak bosons are given by combinations of the W_μ^1 and W_μ^2 fields,

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}}(W_{\mu}^{1} \mp iW_{\mu}^{2}). \tag{2.18}$$

The physical neutral fields of the theory, the Z boson (Z_μ) and the photon, are then expressed as linear combinations of the remaining W_μ^3 boson, and the new B_μ field. This can be expressed succinctly as

$$\begin{pmatrix} Z_{\mu} \\ A_{\mu} \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_{\mu}^{3} \\ B_{\mu} \end{pmatrix}, \tag{2.19}$$

where θ_W is called the Weinberg angle, which controls the mixing. The charges associated with each transformation can be identified by considering the electroweak Lagrangian, using the definition of the covariant derivative in Eqn 2.17. Terms involving right handed fields, ψ_{R}, include

$$\mathcal{L} \supset -\bar{\psi}_{\mathrm{R}} g' \cos\theta_W Y \gamma^\mu A_\mu \psi_{\mathrm{R}}. \tag{2.20}$$

By comparison with the analogous term in the QED lagrangian, the relationship between $g' \cos\theta_W = |e|$ can be inferred, and the weak hypercharge, Y, is identified as the electric charge of the particle. Similarly, action of the covariant derivative on left handed fields, ψ_{L}, yields terms such as,

$$\mathcal{L} \supset -\bar{\psi_{\mathrm{L}}}(g \sin\theta_W \frac{\sigma^3}{2} + g' \cos\theta_W Y)\gamma^\mu A_\mu \psi_{\mathrm{L}}. \tag{2.21}$$

In order to assert that the two states in each weak doublet differ by a factor of $|e|$, the relation $g \sin\theta_W = |e|$ must hold. Finally, the observation that EM interactions do not depend on the handed-ness of the fermion, yields the relation between the generators of the neutral fields

$$Q = \sigma^3 + \frac{Y}{2}\mathbb{1}_2. \tag{2.22}$$

The unification has thus allowed for a photon interaction that is invariant under a parity transformation, and the weak current interaction which is not.

Whilst requiring local gauge invariance provides an elegant method to introduce interactions, as well as unify the weak and EM forces, it is, at this point, unable to permit masses for the weak gauge bosons. This prediction is in clear contradiction with the experimental observation of the massive W and Z gauge bosons. To resolve this contradiction, the BEH mechanism is introduced.

2.4 Spontaneous symmetry breaking

In the BEH mechanism, masses of the fundamental particles are generated through spontaneous breaking of a Lagrangian symmetry. In the SM, the BEH mechanism is realised by introduction of a complex scalar doublet, with a non-zero expectation value in the vacuum state. As a result of the new field, although the Lagrangian still respects the $\mathrm{SU(2)_L} \times$ U(1) symmetry associated with the weak and EM gauge groups, the ground state does [6–10].

2.4.1 Symmetry breaking for a complex scalar field

To introduce the BEH mechanism, it is instructive to first consider the simpler case of U(1) symmetry breaking, generated by the introduction of a complex scalar field, $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$. Similarly to the QED formalism, the U(1) symmetry requires the covariant derivative to be introduced, defined as [38]

$$D_\mu \equiv \partial_\mu + igB_\mu, \tag{2.23}$$

where B_μ is a new massless field with its own transformation properties. The Lagrangian describing the complex scalar field therefore has the form [38]

$$\begin{aligned}\mathcal{L} &= (D_\mu\phi)^*(D^\mu\phi) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - V(\phi) \\ &= (D_\mu\phi)^*(D^\mu\phi) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \mu^2(\phi^*\phi) - \lambda(\phi^*\phi)^2,\end{aligned} \tag{2.24}$$

where $V(\phi)$ is the potential energy of the field which is chosen to have form $V(\phi) = \mu^2(\phi^*\phi)+\lambda(\phi^*\phi)^2$. For the potential to have a finite minimum, the value of λ is constrained to positive values; however, the value of μ^2, is not. If μ^2 is chosen to be positive, the resulting potential has a singe minimum at $\phi = (\phi_1, \phi_2) = (0, 0)$, representing a scalar particle with mass μ. However, if the value of μ^2 is chosen to be negative, the associated term can no longer be interpreted as a particle mass, and the potential possesses a ring of minima satisfying $\phi^*\phi = \phi_1^2 + \phi_2^2 = -\mu^2/\lambda = v^2$. The expectation value of the Higgs potential in the vacuum state is defined as v, indicated by the dashed line of minima in Figure 2.3. Since the minima of the potential do not occur at $\phi = 0$ in the case of $\mu^2 < 0$, the field is said to have a *non-zero vacuum expectation value*. It is the actual choice of minimum for the vacuum state that spontaneously breaks the local gauge symmetry of the Lagrangian. Without loss of generality, the vacuum state can be chosen to lie entirely on the real axis, such that $\phi = (v, 0)$. The field can then be represented as a expansion around the vacuum state

$$\phi = \frac{1}{\sqrt{2}}(v + \eta + i\xi), \tag{2.25}$$

which when substituted into the Lagrangian of Eqn 2.24, yields terms [38]

$$\mathcal{L} \supset \frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \lambda v^2\eta^2 + \frac{1}{2}(\partial_\mu\xi)(\partial^\mu\xi) + \frac{1}{2}g^2v^2B_\mu B^\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{2.26}$$

The Lagrangian describes a massive scalar field, η, and massless boson ξ, which is termed

the *Goldstone* boson. In addition, the previously massless field B_μ associated with the local U(1) gauge symmetry has now acquired a mass term. The BEH mechanism therefore appears a viable method to introduce the massess of the SM vector bosons. However, the Lagrangian also contains terms such as $gvB_\mu\partial^\mu\xi$ which would appear to represent a direct coupling between the spin-0 Goldstone boson and spin-1 gauge field. In order to eliminate the Goldstone field from the Lagrangian, the *unitary gauge* is adopted. This corresponds to fixing the phase of the U(1) transformation such that the scalar field can be expressed as:

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta) \equiv \frac{1}{\sqrt{2}}(v + h), \tag{2.27}$$

where the field η has been written using h to emphasise that this is the physical field of the theory. Finally, ignoring constants, Eqn 2.24 can be written as [38]

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda v^2 h^2 + \frac{1}{2}g^2 v^2 B_\mu B^\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \\ &+ g^2 v B_\mu B^\mu h + \frac{1}{2}g^2 B_\mu B^\mu h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4.\end{aligned} \tag{2.28}$$

The Lagrangian now describes a massive gauge boson associated with the B_μ field of the U(1) symmetry, and a massive scalar field, h, known as the Higgs field. The mass of the gauge boson can be identified as $m_B = gv$, while the Higgs boson mass is given by $m_h \equiv m_H = \sqrt{2\lambda}v$. The Lagrangian also contains interaction terms between the two fields, as well as triple and quartic self-interactions of the Higgs boson. The spontaneous breaking of a U(1) local gauge invariance has therefore allowed for a previously massless field to obtain mass through introduction of a scalar field, the quanta of which is a massive boson.

2.4.2 The SM Higgs boson

In the SM, the Higgs field consists of two complex scalar fields, which transform as a doublet under $\mathrm{SU(2)_L}$ transformations. The field has four degrees of freedom, and can be written as

$$H = \begin{pmatrix}\phi_+ \\ \phi_0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2 \\ \phi_3 + i\phi_4\end{pmatrix}, \tag{2.29}$$

where the upper and lower components differ by one unit of charge. The corresponding Lagrangian therefore has terms

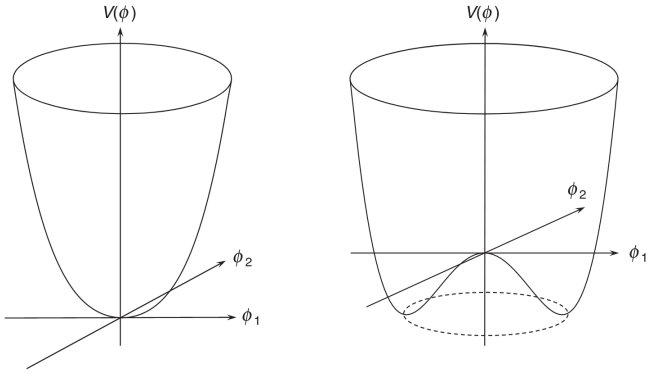

Figure 2.3: The potential $V(\phi) = \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2$ for a complex scalar field. Two cases are considered: $\mu^2 > 0$ (left), where a single minima is present, and $\mu^2 < 0$ (right), where the field acquires a non-zero vacuum expectation value, with minima at $|\phi| = -\mu^2/\lambda = v^2$. Figure taken from Ref [38].

$$\mathcal{L} \supset (D_\mu H)^\dagger (D^\mu H) + V(H) \tag{2.30}$$

$$= (D_\mu H)^\dagger (D^\mu H) - \mu^2 H^\dagger H - \frac{\lambda}{4}(H^\dagger H)^2, \tag{2.31}$$

where D_μ is defined for an $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ symmetry in Eqn 2.17. The SM Higgs potential, $V(H)$, has vacuum states with non-zero expectation value that satisfy

$$H^\dagger H = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}. \tag{2.32}$$

The vacuum state, which breaks the $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ symmetry of the Lagrangian, is chosen to correspond to the neutral complex field. Expanding around this choice of minimum yields

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \tag{2.33}$$

where, again, the unitary gauge has been adopted. Inserting this into the Lagrangian of Eqn 2.30 yields terms

$$\mathcal{L} \supset \frac{1}{4} g^2 v^2 W_\mu^+ W^{\mu-} + \frac{1}{8}(g^2 + g'^2) v^2 Z_\mu Z^\mu \tag{2.34}$$

$$+ \frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda v^2 h^2, \tag{2.35}$$

where the rotation to the physical fields has been applied using Eqn 2.19. The mass for the charged W bosons can be identified as $m_W = gv/2$, while the neutral Z (Goldstone) boson mass is $m_Z = v\sqrt{g^2 + g'^2}/2$, both of which depend on the Higgs boson vacuum expectation value. The quanta of mass for the Higgs field can also be identified as $m_H = \sqrt{2\lambda}v$. No such mass terms are present for the A_μ field, corresponding to the photon; thus, the BEH mechanism has provided mass terms for the weak gauge bosons, while preserving the massless photon. Finally, similarly to Eqn 2.28, the Lagrangian also contains triple and quartic interaction terms between one or two Higgs bosons and the weak gauge bosons.

2.4.3 Yukawa interactions

In the SM, the BEH mechanism is also used to generate the fermion masses [38]. Due to the different transformation properties of left and right handed fermions, typical mass terms of the form $-m\bar{\psi}\psi = -m(\bar{\psi}_{\mathrm{R}}\psi_{\mathrm{L}} + \bar{\psi}_{\mathrm{L}}\psi_{\mathrm{R}})$ do not respect an $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ symmetry. However, since the Higgs field behaves as a doublet under SU(2) transformations, mass terms of the form $\bar{L}H$ respect an $\mathrm{SU(2)_L}$ symmetry, where L represents a doublet of left handed fermions. When combined with a right handed fermion singlet, R, the resulting mass term $\bar{L}HR$ respects the required $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ gauge symmetry. Hence, fermion mass terms are identified in the Lagrangian as $-g_f(\bar{L}HR + \bar{R}H^\dagger L)$, where g_f is a coupling strength between the Higgs boson and fermion, known as the *Yukawa* coupling.

To illustrate the generation of fermion massess, consider the $\mathrm{SU(2)_L}\times\mathrm{U(1)}$ invariant Lagrangian mass terms for the left handed doublet containing the electron, and right handed electron singlet

$$\mathcal{L} = -g_e \left[\begin{pmatrix} \bar{\nu}_e & \bar{e} \end{pmatrix}_{\mathrm{L}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R + \bar{e}_R \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix}_{\mathrm{L}} \right]. \tag{2.36}$$

Inserting the SM Higgs doublet in the unitary gauge yields

$$\mathrm{L} = -\frac{g_e}{\sqrt{2}} v(\bar{e}_{\mathrm{L}} e_{\mathrm{R}} + \bar{e}_{\mathrm{R}} e_{\mathrm{L}}) - \frac{g_e}{\sqrt{2}} h(\bar{e}_{\mathrm{L}} e_{\mathrm{R}} + \bar{e}_{\mathrm{R}} e_{\mathrm{L}}). \tag{2.37}$$

The first term has the required form for fermion masses, which can be identified as $m_e = g_e v/\sqrt{2}$. Note that conversely, the Yukawa coupling to fermions is proportional to the fermion mass. The final form of Eqn 2.36 is therefore

$$L_e = -m_e \bar{e}e - \frac{m_e}{v} \bar{e}eh. \tag{2.38}$$

Alongside the mass term, the second term gives rise to a coupling between the electron

and Higgs boson, as shown in the Feynman diagram of Figure 2.4.

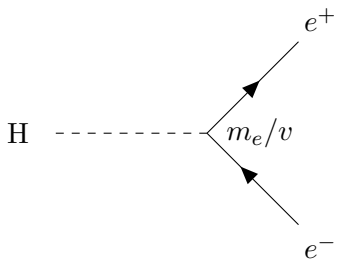

Figure 2.4: Interaction vertex between the Higgs boson and electron pair.

2.5 Higgs boson phenomenology at the LHC

With the discovery of the Higgs boson [14–16] by the CMS and ATLAS experiments at the LHC, focus has shifted to performing a detailed characterisation of the new particle. This includes an extensive programme of precision measurements, as well as searches for rare and exotic decay modes. At a hadron collider such as the LHC, the Higgs boson has a rich phenomenology, allowing for measurements across various combinations of production processes and decay channels. This section describes the prominent production mechanisms and decay channels used in many SM Higgs boson analyses at the LHC.

2.5.1 Higgs boson production modes

At the LHC, where protons are brought into collision at a centre-of-mass energy of $\sqrt{s} =$ 13 TeV, the Higgs boson is produced through a variety of different mechanisms, the cross sections of which are given in Table 2.1. The most common production process at the LHC is gluon fusion, where the Higgs boson is produced via a heavy quark loop. This is followed by vector boson fusion, where the Higgs boson is produced in association with two quarks, which hadronise to form sprays of particles known as (hadronic) jets. Production in association with a vector boson (VH), and with a pair of top ($t\bar{t}H$) or bottom ($b\bar{b}H$) quarks are the next most common modes, followed by production in association with a single top quark. Figure 2.5 gives example leading order Feynman diagrams for the four Higgs boson production modes with the highest cross section at the LHC, all of which have now been observed experimentally [26, 30].

The experimental sensitivity to each production mode depends not only on the cross section, but also on the presence of competing processes known as background events. Many production modes benefit from the presence of characteristic final state objects, which can be used to more easily identify the process. For example, VBF Higgs boson

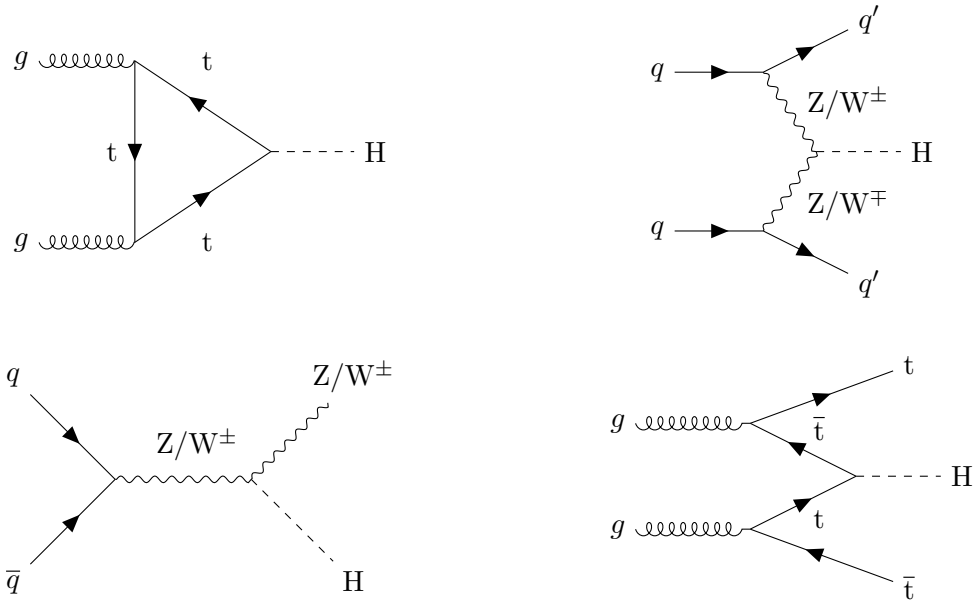

Figure 2.5: Example Feynman diagrams for the four dominant Higgs boson production mechanisms at the LHC. On the top row are the ggH (left) and VBF (right) processes, while the bottom row shows the VH (left) and $t\bar{t}H$ (right) processes.

production is characterised by two forward jets with large invariant mass, while $t\bar{t}H$ production typically results in high multiplicity final states involving B hadrons. Utilising these features results in significant suppression of background events, allowing for more sensitive measurements to be performed.

Table 2.1: Cross sections for the leading Higgs boson production modes at $\sqrt{s} = 13$ TeV, for a nominal Higgs boson mass of $m_H = 125.0$ GeV. The VH mode is split into contributions involving the Z and W gauge bosons. All values are taken from Ref [40].

Production mode	ggH	VBF	WH	ZH	$t\bar{t}H$	$b\bar{b}H$	tH
Cross section (pb)	48.58	3.78	1.37	0.76	0.51	0.49	0.09

2.5.2 Higgs boson decay channels

The Higgs boson is unstable, surviving for an extremely short time before decaying to lighter particles. These decay products are used to reconstruct the properties of the original Higgs boson. Since the Higgs boson coupling strength is proportional to the mass of the coupled particle(s), decay channels involving heavy particles are favoured. Consequently, almost 60% of all Higgs bosons decay to two b quarks. It might therefore be concluded that measurements in the $H \to b\bar{b}$ channel are the most sensitive; however, at the LHC, this considerable branching fraction is offset by the large background re-

sulting from QCD processes, which produce a signature similar to the two jets found in the $H \to b\bar{b}$ final state. In contrast, although the branching fraction for channels such as $H \to \gamma\gamma$ is comparatively small at ≈0.23%, the two final state photons produce a particularly clean signature which is more easily resolvable against background processes. The diphoton channel, which proceeds by a heavy particle loop, was therefore observed to have the highest sensitivity of all measured decay channels at the time of the Higgs boson discovery [14–16], while the $H \to b\bar{b}$ process was only observed later in Run 2 of the LHC operation [20, 21]. The branching fractions for a range of Higgs boson decay channels are given in Table 2.2, with selected Feynman diagrams shown in Figure 2.6.

Although unsuitable for precision measurements of Higgs boson properties, rarer decay channels may also be used to study the SM predictions. These decays typically involve light particles, including channels such as $H \to \mu\mu$, $H \to c\bar{c}$, and decays to the lightest charged leptons, $H \to ee$. The branching fraction for $H \to ee$ decays predicted by the SM is extremely small, at $\mathcal{B}(H \to ee) \approx 5.2 \times 10^{-9}$. Clearly, a decay channel of this rarity is inaccessible with the data collected at the LHC during Run 2 operation; however, enhancements to $\mathcal{B}(H \to ee)$ are predicted under several BSM scenarios. The simplest of these are models postulating additional Higgs doublets (2HDM) [41, 42]. Models that allow for an enhancement in κ_e are restricted to non-flavour conserving 2HDM scenarios, given current signal strength measurements in the $H \to \mu\mu$ and $H \to \tau\tau$ channels [22, 23, 31]. Other extensions to the SM include the addition of new higher order operators to the SM Lagrangian, including dimension 10 operators that could modify the electron Yukawa coupling by a factor of ≈10 [43].

Table 2.2: Branching fractions for a selection of Higgs boson decay modes, assuming $m_H = 125.0$ GeV. All values are taken from Ref [40].

Decay channel	$b\bar{b}$	WW^*	gg	$\tau\tau$	$c\bar{c}$	ZZ^*	$\gamma\gamma$	$\mu\mu$	ee
Branching fraction (%)	58.2	21.4	8.2	6.3	2.9	2.6	0.23	0.022	5.2×10^{-9}

2.6 Conclusion

The standard model of particle physics describes the fundamental particles and their interactions, and has been extraordinarily successful in explaining many experimental observations. The SM is a gauge field theory, built on an underlying $SU(3) \times SU(2)_L \times U(1)$ symmetry of the associated Lagrangian. The formalism allows for interactions between particles to be introduced as a required consequence of this invariance, including self-interactions between the gauge bosons. The BEH mechanism provides a tool for both bosons and fermions to acquire mass, through the introduction of a complex scalar doublet

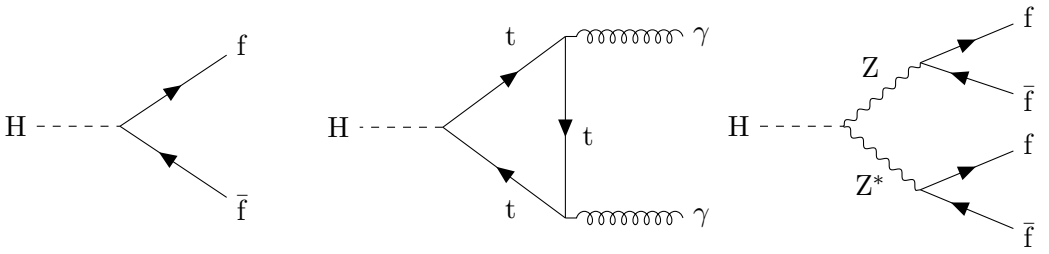

Figure 2.6: Example Feynman diagrams for three Higgs boson decay channels: decay to two fermions (left), decay to a photon pair (centre), and decay to two Z bosons, which subsequently decay to fermions (right). Note that when labelling all particles, it is implicit that the invariant mass of the decay products does not exceed that of the decaying particle, with the notable exception of the virtual Z^* in the $H \to ZZ^*$ decay channel.

with non-zero expectation value in the vacuum state. The quanta of this field, the Higgs boson, was discovered in 2012 at the Large Hadron Collider, where it is produced through a variety of production modes. Properties of the Higgs boson are inferred from its decay products, with particularly sensitive channels including $H \to \gamma\gamma$ and $H \to ZZ^*$ decays. This thesis describes a search for the extremely rare decay of the Higgs boson to an electron pair.

Chapter 3

Machine learning

3.1 Introduction

The field of machine learning is concerned with algorithms that "learn" from previous experience. These algorithms are presented with an observed dataset and tasked with connecting descriptions of the data with some underlying quantities of interest. Machine learning algorithms typically have a high degree of flexibility, giving a considerable advantage over more traditional rule-based algorithms, particularly when the learning task is complex. Subsequently, this class of algorithms has already been widely adopted in many areas of high energy physics, with a comprehensive review provided in Ref [44].

Machine learning can broadly divided into three areas: supervised learning, unsupervised learning, and reinforcement learning. A supervised machine learning algorithm, which describes all algorithms developed in this thesis, is one in which the algorithm is presented with a set of labelled examples to learn from. The algorithm aims to tune its parameters using information from given features of the observed examples. The goal of a supervised learning algorithm is to accurately predict the label, which may be categorical or continuous, of a datum, when later presented with unseen (and potentially unlabelled) data.

In contrast, unsupervised algorithms are provided with unlabelled data, and consequently the learning objective is typically less well defined. In general, these algorithms are tasked with learning a structure or groupings underlying the data. A common unsupervised learning task is the clustering of unlabelled data into discrete groups, usually based on some similarity metric between the population.

In the final category, reinforcement learning algorithms, the algorithm is again not provided with examples of labelled data. Instead, it learns from taking exploratory actions within some environment, which are either rewarded or penalised by some external mechanism, in process similar to trial-and-error. The goal of the algorithm is to maximise

the overall reward achieved by its actions.

Since neither unsupervised or reinforcement learning techniques are used within this thesis, the remainder of this chapter pertains only to supervised learning tasks. Applications of note include the classifiers trained in the triggering system under development for the High Granularity Calorimeter CMS upgrade project (chapter 4), the reconstruction and correction techniques for particles used in the H $\rightarrow$ ee analysis (chapter 5), and in the development of analysis categories targeting both ggH and VBF H $\rightarrow$ ee events (chapter 6). This chapter gives an overview of the core concepts in machine learning and discusses algorithms pertinent to later chapters of this thesis.

3.2 Concepts in machine learning

3.2.1 Problem statement

A typical supervised machine learning problem organises the data to learn from into a vector space $X = \mathbb{R}^m$, where each dimension is some observable quantity often referred to as a *feature*. Each element of the input feature space corresponds to a single feature vector, $\vec{x_i} \in X$, describing the properties of the corresponding datum. The goal of a machine learning algorithm is to construct a model, f,

$$f(\vec{x}, \vec{w}) \rightarrow Y,$$

that maps the input vector, $\vec{x}$, to a property of interest, Y, given a set of model parameters, $\vec{w}$ [45]. The property to be determined, often referred to as the label or target vector, can be continuous or categorical. This defines two types of machine learning task:

- *Classification tasks*, where the aim of the classifier is to predict one of k possible classes, $f(\vec{x}, \vec{w}) \rightarrow y$, where $y \in \{1, ..., k\}$. The simplest example is the case of a two-class prediction problem, also termed a binary classification problem, where $k = 2$. Examples of binary classifiers are presented in the search for H $\rightarrow$ ee decays, where signal events resulting from true Higgs boson events form one class, and background events entering the analysis selection resulting from processes other than the signal, form the second. The number of classes is not, however, limited to two, with so-called *multi-class* models developed to predict one of many possible outcomes.

- *Regression tasks*, where the aim of the classifier is to predict some continuously varying feature value of the data, $f(\vec{x}, \vec{w}) \rightarrow y$, where $y \in \mathbb{R}$. An example of a machine learning-based regression is presented in Section 5.4.1, where the task is to

predict the true value of the electron energy, from a list of features describing the reconstructed electron.

3.2.2 Model training

The development of a typical machine learning model can be divided into two phases, the first of which is *training*. During the training phase, the algorithm is presented with labelled examples to learn from. The learning process involves tuning the free parameters of the model according to a *loss function*, L. The loss function quantifies the performance of the model during training and depends both on the predicted outcomes and the corresponding true labels, aggregated in some way over the training data,

$$L(y, f(\vec{x}, \vec{w})) \to \mathbb{R}.$$

The exact choice of loss function is informed by both the task to be learned and desired model properties. For regression tasks, the mean squared error is often used. For classification tasks, a typical choice is the so-called cross entropy, defined for a binary classification problem as:

$$L = -\sum_i [y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)], \tag{3.1}$$

where p_i is the predicted probability for event i. The objective during model training is to minimise the loss with respect to the model parameters. This is typically achieved using a gradient descent-based optimisation where the loss hypersurface is descended with respect to $\vec{w}$, until reaching a global minimum of L,

$$\vec{\nabla}_{\vec{w}} L = 0.$$

In practice, evaluating this expression over the entire dataset is computationally costly, particularly when the number of training examples and dimensionality of the input feature space are large. Therefore, a more efficient version of the optimisation procedure, termed *stochastic gradient descent* (SGD) [46], is often preferred. This method makes sequential updates to the weight vector by computing the gradient on a subset of the entire training dataset. The model weights are updated as

$$\vec{w} \to \vec{w} - \eta \vec{\nabla}_{\vec{w}} L,$$

where η is the *learning rate* and controls the magnitude of weight updates. This update procedure is repeated either by cycling through the entire data in sequential batches, or by selecting batches at random, with replacement. Since the gradient computed in each

batch will only approximate the true gradient evaluated over all training examples, the direction of descent down the loss surface will not always be optimal. However, with enough iterations, the procedure typically still converges towards the global minimum in L, at a much faster rate than the nominal gradient descent algorithm.

Up to now, the procedure could be viewed in the context of any gradient-based objective minimisation procedure. However, we also require the model to be capable of generalising beyond the observed training data. If the descriptive power of the model, also known at the model capacity [47], is too large for the complexity of the problem, the model parameters will be tuned to the training data arbitrarily well, in a process known *overfitting* or *overtraining*. This leads to large variance in predictions if the training dataset is modified, an undesirable effect for model generalisation. In contrast, if the descriptive power of the model is too small, it is again subject to generalisation error due to poor approximation of the underlying mapping function. In this case, the resulting model is said to be *underfit*.

To illustrate this concept, Figure 3.1 shows a set of dummy observations, y, drawn from an underlying sinusoidal function $y = \sin(x)$, with the addition of some amount of Gaussian noise. The twenty observations are divided equally into a training set, used to tune the model parameters, and a testing set, where the performance of the model will be evaluated. The task is to fit an order-N polynomial to the training data, that also describes well the unseen test data. The loss function to be minimised for this simple fit is the mean squared error, which is also used to quantify how well the order-N polynomial fits both the training and testing data. Figure 3.1 demonstrates that lower order functions, for example an order-1 polynomial, lack sufficient complexity to capture the underlying relationship between x and y, resulting in a considerable loss on both the training and testing sets. Conversely, functions with high order, for example an order-8 polynomial, have enough flexibility to fit the training data arbitrarily well, yet generalise poorly on the unseen test data. The loss is therefore very small when evaluated on the training set, but largest of all models when computed on the test data. A balance between the two previous functions is achieved when using the medium complexity order-3 polynomial, which has enough freedom to model the underlying distribution, while remaining conservative enough to avoiding overfitting to statistical fluctuations in the training data.

To reduce the generalisation error of a model, a suitable balance between overfitting and underfitting must be drawn, commonly known as the *bias-variance trade-off*. One approach to maximising the trade-off is to add additional penalties to the loss function to prevent overly-complex mapping functions being learned. These penalties, known as regularisation terms [47], can be scaled during training to modify the strength of the regularisation. Although many types of regularisation exist, common forms add a penalty

proportional to the magnitude of the model weights summed. This particular approach prevents individual weights obtaining overly large values during the training phase, a common feature of overtrained models.

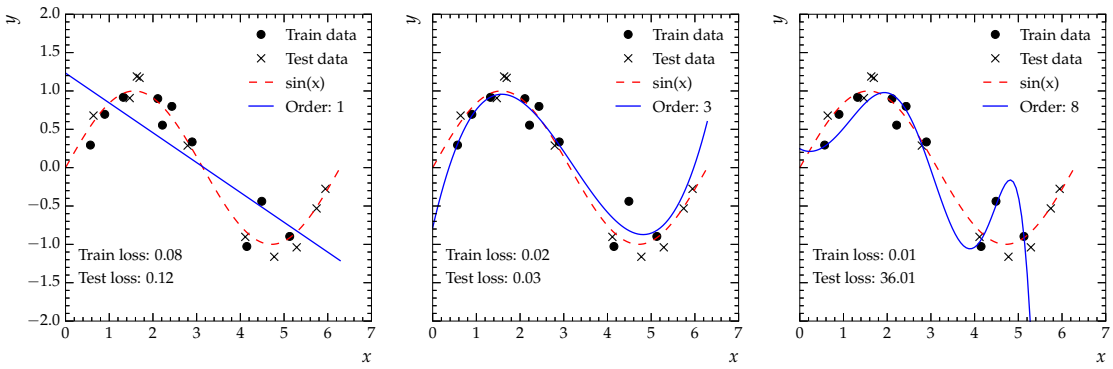

Figure 3.1: An illustrative example of underfitting and overfitting of a simple linear regression. A sample of toy data (black markers) is drawn from an underlying sinusoidal distribution, with a small amount of Gaussian noise added. The data are divided into a training set and testing set. Three polynomial fits are performed using data from the training sample: an order-1 polynomial (left), order-3 polynomial (centre), and an order-8 polynomial (right). The most simple model lacks sufficient capacity to describe the underlying distribution, resulting in a large loss (mean squared error) when evaluated on both the training and testing data. The overly-complex order-8 polynomial overfits to the Gaussian fluctuations in the training data and thus is unable to generalise well on the testing set. The order-3 polynomial provides a suitable balance between underfitting and overfitting; it describes the underlying distribution accurately, while also generalising well on the unobserved data.

3.2.3 Performance evaluation

The second phase in developing a machine learning model is the evaluation step, which can be divided into two parts: the performance evaluated on a validation set, and on a test set. The validation set comprises some fraction of the dataset that is withheld from training, and instead used to tune unlearned parameters (hyperparameters) of the model. These could include regularisation coefficients, the learning rate for SGD described above, or a host of other tuneable parameters. The test set, which was already introduced in the example of Figure 3.1, comprises a fraction of the data, usually between 10 and 30%, that remains unseen by the model during training. It is chosen to be representative of the entire distribution of the dataset. Following the training phase, the performance is evaluated on the test set to quantify both the absolute performance and the model generalisation power. A common test of generalisation is to compare the performance evaluated on both the training and testing set. For a model with good generalisation power, these should be similar; for an overtrained model, the performance on the training set is typically greater

than on the unseen data.

The exact metric with which to evaluate a model is often dictated by the learning task. For classification, a standard choice is the area under the receiver operating characteristic curve [48] (ROC AUC). Each point on a ROC curve is populated by evaluating the true positive rate (TPR) and false positive rate (FPR) at a given threshold on the classifier output. The TPR is defined as the fraction of correctly labelled signal events, while the FPR is the fraction of background events that were incorrectly labelled as signal, at the selected threshold value. The integral of this curve, which provides a single value to quantify performance, is 0.5 for a model that predicts a class at random, and approaches unity for a classifier performing perfect separation. Figure 3.2 demonstrates for a simple binary classifier, how a ROC curve may be built from the distribution of model output values (or *scores*) for each class. The ROC AUC metric can also be adapted to classification tasks with more than two classes. In such cases, all but one classes are merged, and a ROC curve is generated to quantify the separation power of the merged class against the single remaining class. In this way, for a N-class classification problem, N ROC curves are produced. Simpler metrics for evaluating the performance of multiclass models may also suffice, with a common choice being the fraction of correctly classified events, also known as the *accuracy*. For regression tasks, common evaluation metrics include the root mean squared error, or the mean absolute error, both of which offer an intuitive relation between the model predictions and associated loss.

3.3 Machine learning algorithms

The landscape of possible machine learning algorithms that may be used to achieve a learning task is extremely broad. The exact choice of algorithm is conditional on many factors including the structure and size of the input data, the desired learning outcome, and practical constraints such as the use of computational resources. This section presents a subset of algorithms pertinent to the remainder of this thesis, namely gradient boosted decision trees (BDTs) and deep neural networks (NNs).

3.3.1 Decision trees

To introduce gradient boosted decision trees, it is first useful to describe the construction of simple decision trees. Decision trees [49] perform recursive binary partitioning of the input feature space into orthogonal regions. The tree is composed of many such partitions, or nodes, each isolating a new region of input space. Nodes which terminate a branch of the tree, are labelled as leaf-nodes. Each leaf node has a value assigned to the region it isolates, which for a singular tree, corresponds to an element from the set of model

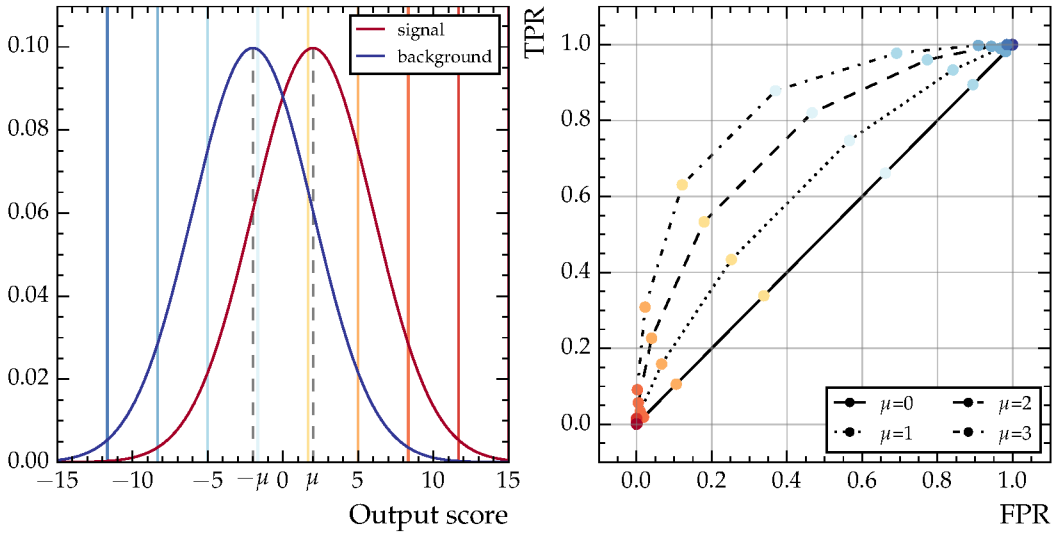

Figure 3.2: Construction of the ROC curve for a binary signal-background separation task, using the output score of a hypothetical model. Left: toy distributions of the output score are shown for the signal (red) and background (blue) classes, which are, for illustration purposes, chosen to resemble a Gaussian function symmetric about output scores of zero. The relative separation of the distributions can be parameterised using the Gaussian mean, μ, which is identical in magnitude for both distributions and opposite in sign. The corresponding ROC curves are built from applying a selection on the output score (vertical lines) and computing the TPR and FPR rates. Right: resulting ROC curves, generated for different values of the Gaussian mean(s). A larger area under the curve corresponds to a larger separation between the output score distributions. Although the TPR and FPR corresponding to ten selection thresholds are drawn, a typical ROC curve is built from applying a more granular set of selections, resulting in smoother curves.

predictions. For a regression task, this region holds a continuous number, whereas for classification, the region will assign probabilities to the possible class types. Figure 3.3 gives an illustration of a simple decision tree that partitions a two dimensional input feature space into five mutually exclusive regions. Once the tree is formed, any new input, $\vec{x} = \{x^{(1)}, x^{(2)}\}$, traverses the chain of partitions, starting at the root note and ending in a single leaf node.

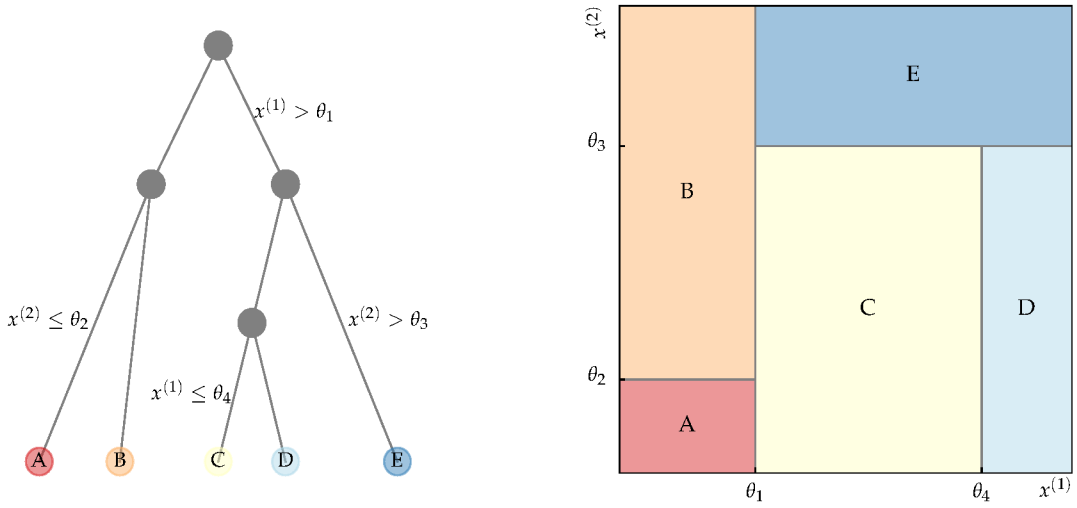

Figure 3.3: Left: an example decision tree using two input features, $x^{(1)}$ and $x^{(2)}$. The tree makes four selections on the input features, dividing the space into five regions. Right: the partitions of the input feature space resulting from the example decision tree. Each region is mutually exclusive and typically holds some associated predicted value.

The structure for a given decision tree, including which input variable is chosen at each node and the associated partition threshold, is determined during model training. Due to the computational complexity of considering all possible structures of input variables and corresponding decision thresholds, a "greedy" optimisation is performed when growing the tree. This optimisation considers one set of partitions at a time, performing an exhaustive search over the possible variables and selection thresholds [50]. The final choice of boundaries is chosen to maximise some measure of purity in the classification case, or minimise a measure of prediction error for regression tasks.

When terminating the growth of a tree, common approaches are to stop adding partitions when the number of data points landing in a leaf node reaches a critical threshold, or when a predefined maximum number of partitions is reached. This typically results in large decision trees, which are then "pruned" to prevent overfitting [50]. Pruning is a form of regularisation, specific to decision trees, which can be achieved in many ways. One frequently used technique is to remove branches that contribute little to the overall

performance of the model.

Decision tree algorithms have many advantages. They are simple, easy to interpret, and require much less computational time to train and evaluate in comparison to algorithms that result in more complex model architectures. However, the resulting models struggle to capture the mapping function for tasks where the optimal decision boundary is not somewhat aligned with the dimensions of the feature space. In addition, decision trees are prone to overfitting, if unregularised.

3.3.2 Gradient boosted decision trees

A gradient boosted tree combines many individual decision trees in order to generate strong final predictions. Each individual tree is trained sequentially in order to correct the mistakes of previous ones. This is one example of model *ensembling*, where multiple base models are combined in some way to produce a committee of classifiers (or regressors), whose performance is significantly better than that of each constituent learner [50, 51].

Suppose that a model is to be built on a dataset $\{\vec{x}_i, y_i\}$, with $i = 1, 2, ..., n$, indexing the number of data points to learn from. The process to construct a gradient boosted decision tree, F_m, where $m = 0, 1, ..., M$, labels the iteration of the boosting algorithm, is summarised as follows:

1. Train a decision tree to predict the *pseudo-residuals*, defined as the set of $y_i - F(\vec{x}_i)_{m-1}$. For a regression task, where the loss might be the mean squared error, this results in minimising a quantity proportional to the negative of the classical residuals (hence the name pseudo-residual). This makes clear that training to predict the negative residuals is analogous to gradient descent of the loss surface, a key insight underpinning the boosting algorithm.

2. For each of the j leaf nodes in the resulting tree, R_{jm}, assign a single score, γ_{jm}, that satisfies

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{\vec{x}_i \in R_{jm}} L(y_i, F_{m-1}(\vec{x}_i) + \gamma). \tag{3.2}$$

   The leaf node scores account for the previous predictions $F_{m-1}(\vec{x}_i)$ and are computed using only $\vec{x}_i \in R_{jm}$. The value of γ can be found using a line search strategy, or in some cases, solved for analytically.

3. Update the predictions using

$$F_m(\vec{x_i}) = F_{m-1}(\vec{x_i}) + \eta \sum_j \gamma_{jm} I(\vec{x}_i \in R_{jm}). \tag{3.3}$$

This step amounts to summing the predictions for each event from the current tree, with the previous predictions. The value for η weights the predictions from all decision trees equally when summing, and can be viewed analogously to the step size, or learning rate, in gradient descent.

The boosting process is terminated when reaching a pre-defined maximum number of decision trees, or when successive trees fail to improve the performance by a certain threshold.

In practice, gradient boosted trees offer a significant improvement in performance over a singular decision tree, and are still relatively quick to train. The structure of each component tree is interpretable, allowing the model development to remain tractable. However, the ensemble is sensitive to the exact details of the training data, with small changes often resulting in a significantly different set of splits [45]. In addition, given the increased model capacity, gradient boosted trees are prone to overfitting and must be carefully regularised.

3.3.3 Deep neural networks

So-called *deep learning* algorithms are classes of machine learning algorithms that use artificial neural networks. Their design is partly inspired by information processing structures in biological systems. Deep networks are characterised by a large number of learnable parameters and possible configurations, resulting in complex and powerful overall structures.

The most simple of deep learning models is a multi-layer perceptron (MLP). A MLP comprises a collection of nodes, organised into layers, where each node is connected by weights. The vector of all such weights, $\vec{w}$, are learnable parameters fit during training. Each node receives a weighted collection of input values, which are added with one learnable bias parameter, b, particular to the given node. This sum is then passed through a non-linear *activation function*, $f(\cdot)$, with the resulting output fed forward to the next set of nodes. The output, or activation, at each node is therefore:

$$o = f\left(\sum_i w_i x^{(i)} + b\right), \tag{3.4}$$

where i enumerates the number of connections arriving from nodes in the previous layer. Common choices for the activation function include the rectified linear unit (ReLU),

$f(\phi) = \max(0, \phi)$, the sigmoid activation function, $f(\phi) = 1/(1 + e^{-\phi})$, and hyperbolic tangent function. In the case of the sigmoid function, the node output is identical to a logistic regression hypothesis.

Figure 3.4 shows a network diagram for a typical MLP, which can be divided into three regions: the input, hidden, and output layers. For each training example, the input layer is simply the feature vector, with each $x^{(i)}$ connected to all nodes in the subsequent layer. The hidden layers, of which there may be many, comprise sets of fully connected (FC) internal nodes. These layers form abstract representations of the input features in a process known as *automatic feature engineering*. The activations from the final hidden layer are then fed into the output layer. For regression tasks, this comprises a single node, which holds the predicted target variable. For a generic classification problem, the softmax function is commonly used at the output nodes. Outputs from softmax are given by $y_k = e^{a_k}/\sum e^{a_j}$, where a_k is the input to node k, and a_j is the activation for the j^{th} element of the output vector. The softmax function bounds node values between zero and one, allowing each output to be interpreted as a class probability.

To calculate the gradient vector, $\vec{\nabla}_{\vec{w}} = \partial L/\partial \vec{w}$, where the biases are implicitly included in $\vec{w}$, an algorithm called backpropagation [46] is used. At the first iteration of the algorithm, all model parameters are initialised to some starting value. Each subsequent iteration of the algorithm can be then divided into three stages:

1. *Forward pass*: predictions are generated for a given set of examples, referred to as a batch, by propagating each example through the network connections. The resulting loss, summed over all batch training instances, is then evaluated.

2. *Backward pass*: the gradient of the loss with respect to each learnable parameter is computed by repeated application of the chain rule. Inputs for each node, m, in given layer, n, are a function of the bias at node m, the activation of all weights in node $n-1$, and the weights joining to node m from all nodes in layer $n-1$. This is of course true for each layer with $n > 0$ in the network, and hence the gradient is propagated backwards through each one.

3. *Weight update*: after the backward pass, the gradient vector is used to update model weights using a gradient descent process.

One complete iteration of the backpropagation algorithm, where the gradient vector has been computed and updated using all examples from the training set, is called an *epoch*. Training a deep neural network typically involves many hundreds of epochs in order to reach optimal performance.

In deep learning models, the exact network structure has many non-learnable hyperparameters that can be tuned during training. These parameters often control the capacity

for the network to learn the underlying mapping of input features to target vector. For example, increasing the number of hidden layers or the number of nodes in each hidden layer typically increases the model capacity. However, like any machine learning application, models with high capacity are likely to overfit; therefore, regularisation techniques specific to NNs are usually imposed. One highly effective method is *dropout* [52], where a random fraction of nodes are eliminated during training, as illustrated in Figure 3.4. This prevents the model from over-using specific neurons, and leads to *effective ensembling*, where the predictions of many separate networks are aggregated during inference.

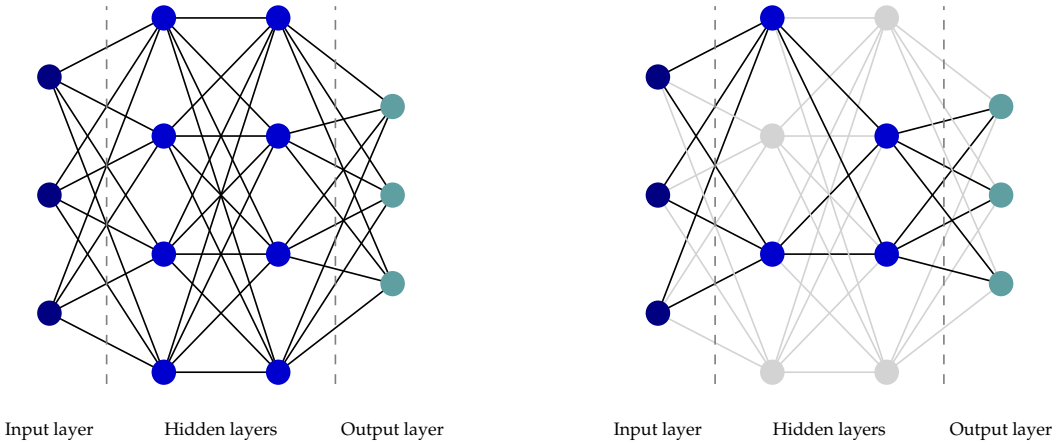

Figure 3.4: Left: schematic of a simple feed forward NN. Nodes in the input layer hold values of the input features for each training example. These feed into layers of hidden nodes, where meaningful representations of the inputs are abstracted. The above example network is relatively shallow; deep learning models are typically characterised by having many more hidden layers. For classification tasks, the final set of nodes hold the class probabilities. For regression tasks, the network output is a single node holding the predicted value. Right: an illustration of "dropout", a technique used to regularise NNs where a fraction of nodes and associated connections are dropped during training.

Deep learning techniques typically produce very powerful models, with a large learning capacity. There is also less need to construct features with known predictive power (*engineered* features), since this process occurs automatically within the network's hidden layers. However, deep networks typically have millions of learnable parameters that must be fit during training time. Not only does this require enormous computational power, it also obscures the learned relationship between input features and output predictions, making interpretation of the final model difficult.

Long short-term memory networks

Although, thus far, only the simple case of a feed forward NN has been presented, many other variants on this structure exist. When deciding which ML algorithm to choose for a given learning task, one may appeal to the so-called *inductive bias* of an algorithm [53], which is the set of assumptions used in the model's construction that determine how it makes predictions. For example, a linear regression assumes that the predicted variable is linearly related to the input feature values. More complex architectures make additional assumptions. For example, a convolutional neural network (CNN), which hierarchically combines input features through many filters, is well suited to classification of image data, since the assumption of correlations between neighbouring pixels in an image works well with the inductive bias of the algorithm.

Another widely used set of ML models are recurrent neural networks (RNNs). A RNN makes assumptions similar to a CNN over a *sequence* of data, motivating the organisation of inputs into list-like structures. The key feature of such a network is the repeated or recurrent propagation of previously learned information from earlier elements of the sequence, to the current element. With this view, the architecture of a RNN can be likened to a series of identical connected networks. The input to each network consists of the output from the previous network, alongside some element of the input feature sequence, as shown in Figure 3.5. The contribution of each input is controlled through learnable weights and biases, which are tuned using the standard backpropagation algorithm on the "unrolled" network. A typical RNN may have many recurrent layers, the last of which is fed into a feed forward network that takes the current hidden state vector to the final network output. Since a RNN is capable of propagating temporal information, typical learning tasks include speech recognition or language translation, where the input text is naturally formed from a sequence of words. However, the simple RNN implementation described above is rarely used for such tasks, since such networks often fail to capture long-range dependencies; information from earlier features of the sequence are not propagated effectively to latter parts of the network. This issue is known as the *vanishing gradient problem* [54] and arises in cases where small weight vectors recurrently force previous networks weights to be close to zero. Since the resulting gradient vector also becomes small, the network is unable to escape this issue by updating the weight vector significantly.

In order to solve the vanishing gradient problem, many variations on traditional RNNs have been proposed, including the long short-term memory (LSTM) network [55]. As shown in Figure 3.6, an LSTM contains repeating cells with interacting components. To illustrate the function of each unit, consider an input sequence with elements, $\vec{x}_t$. The key property of each cell is the *state* vector, $\vec{C}_t$, which for a cell observing input $\vec{x}_t$, can

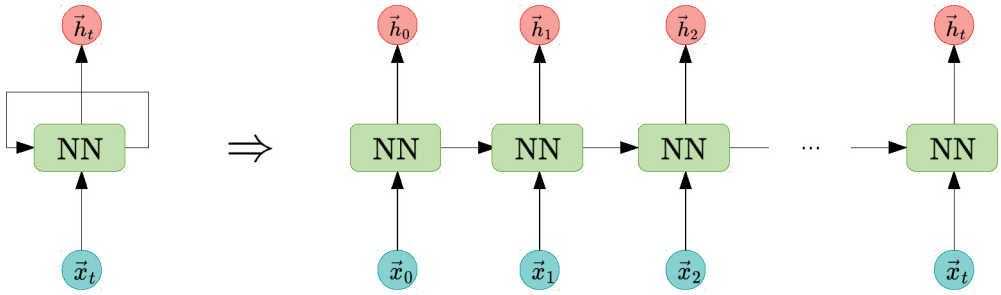

Figure 3.5: A schematic of an "unrolled" recurrent neural network, where the recurrent architecture is viewed as a series of connected networks (cells) propagating the hidden state forward. Inputs to each cell are an element of the input sequence, $\vec{x}_t$, and the previous hidden state, $\vec{h}_{t-1}$.

be thought of as holding learned information from all previous elements, $\{\vec{x}_0, \vec{x}_1, ..., \vec{x}_t\}$. Each unit in the LSTM can modify the cell state using three gate mechanisms:

- *Forget gate:* this gate learns which information will be discarded from the previous cell. Inputs to this gate are $\vec{x}_t$, and the previous hidden state, $\vec{h}_{t-1}$. These are fed into a feed-forward NN with learnable weights, $\vec{w}^f$, and biases, typically using a sigmoid activation function. The resulting output can be therefore be expressed as: $\vec{o}_t^f = \sigma\left(\vec{w}^f \cdot [\vec{h}_{t-1}, \vec{x}_t]\right)$, where biases have been absorbed into $\vec{w}^f$, and the σ function is applied to each element in the output vector. Elements in the previous cell state are updated through the point-wise multiplication $\vec{o}_t^f \cdot \vec{C}_{t-1}$. This can be viewed as a learnable filter for the previous state — elements multiplied by values close to zero are forgotten, while values close to one are remembered and propagated onward.

- *Input gate:* the aim of this gate is to control the information added to the current cell state. This section is composed of two units, each taking as input $[\vec{h}_{t-1}, \vec{x}_t]$. The first unit learns to filter which values of the current cell state should be updated, passing inputs through a feed forward NN with sigmoid activation, to give output: $\vec{o}_t^i = \sigma\left(\vec{w}^i \cdot [\vec{h}_{t-1}, \vec{x}_t]\right)$. The second unit creates a vector of new candidate values, $\tilde{\vec{C}}_t$, used to update the state, using a similar network with $\tanh(\cdot)$ activation, yielding the output: $\tilde{\vec{C}}_t = \tanh\left(\vec{w}^C \cdot [\vec{h}_{t-1}, \vec{x}_t]\right)$. Combining with the *forget gate*, this results in the updated cell state: $\vec{C}_t = \vec{o}_t^f \cdot \vec{C}_{t-1} + \vec{o}_t^i \cdot \tilde{\vec{C}}_t$

- *Output gate:* finally, the hidden state, which is also passed to the next cell, must be updated. Firstly, a NN with sigmoid activation is used to filter the previous hidden

state information, to be passed on. This is combined with $\tanh(\vec{C}_t)$ in pointwise multiplication, where the hyperbolic tangent ensures each element in the current cell state spans the range $\{-1, 1\}$. Along with $\vec{C}_t$, this hidden state is passed to cell $t+1$, which has an identical structure of gates formed from independent learnable weights and biases.

Many variants on the gated cell structure exist, including most notably the gated recurrent unit model, which combines the *input* and *forget* gate into a single update step [56]. However, for many leaning tasks, the resulting performance is typically unaffected by the exact choice of recurrent algorithm [57].

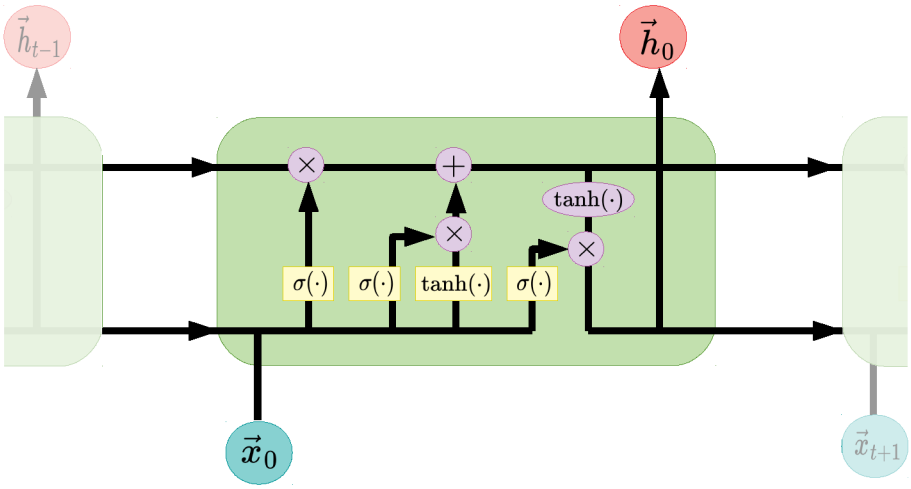

Figure 3.6: Schematic of the repeating module inside an LSTM deep neural network, a variant on the traditional RNN. Each cell is composed of a *forget gate*, *input gate*, and *output gate*, which regulate the flow of information to and from the overall cell state. Inputs to each cell are an element of the input sequence, $\vec{x}_t$, and the previous hidden state, $\vec{h}_{t-1}$. Yellow boxes indicate neural network layers, where all output nodes are transformed either by a sigmoid or hyperbolic tangent function, although hidden layer activations may differ. Purple circles indicate point-wise addition or multiplication.

3.4 Summary

A supervised machine learning algorithm is one that learns to perform a desired task through observation of many labelled examples. The parameters of the model are ad-

justed with the aim of jointly maximising the performance on the learning task and the generalisation capabilities on unobserved data. Although these models may require dedicated tuning of their unlearned parameters, they typically offer a considerable improvement in performance over rule-based approaches. As a result, these models have been widely applied in particle physics [44]. The machine learning models trained in this thesis include shallow learning architectures, which are interpretable and lightweight to train, such as the boosted decision tree performing e/γ identification in the CMS Level-1 Trigger (L1T) phase-2 upgrade (chapter 4). Deep learning models are also explored, including a LSTM NN used in categorisation of VBF H $\rightarrow$ ee events (chapter 6). In contrast to shallower architectures, these models have a large number of learnable parameters resulting in resource-heavy trainings. Their implementation, optimisation details, and performance comparisons are left to later chapters.

Chapter 4

The Compact Muon Solenoid detector

4.1 Introduction

The Large Hadron Collider [13] at CERN is the world's longest and most powerful particle accelerator, with a centre-of-mass energy of 13 TeV. It is designed to collide particles at high energy, and in sufficient number, to be used in tests of the SM. To this end, the LHC ring is equipped with four main detectors aiming to measure the collision products: CMS [11], ATLAS [12], LHCb [58], and ALICE [59]. Each detector is designed with a particular set of physics objectives in mind. For example, the forward geometry and particle identification capabilities of the LHCb detector facilitate a rich B hadron physics programme, while the ALICE experiment is optimised to probe the strong interaction using data collected during heavy ion runs of the LHC. The CMS and ATLAS experiments are more general-purpose detectors, capable of measuring a wide range of processes up to the TeV energy scale. Their physics programmes include precision measurements of SM processes, including detailed exploration of the electroweak sector, as well as direct and indirect searches for processes beyond the SM.

This chapter briefly describes the LHC and associated accelerator complex, before focusing on the CMS detector used to collect all data used in this thesis. The CMS upgrade project in preparation for the High Luminosity LHC [60] is also discussed, with a particular emphasis on electron and photon identification in the calorimeter endcaps, as part of the CMS trigger system.

4.2 The Large Hadron Collider

Located approximately 100 m below ground, crossing the French-Swiss border, the Large Hadron Collider is a sychrotron-type hadron accelerator and collider. The experiment is housed in a 27 km tunnel, previously servicing the LEP [61] collider, where two hadron beams are counter-circulated at energies up to 6.5 TeV before being brought into collision at one of four interaction points (IPs). To achieve this beam energy, the LHC is fed by a series of accelerator circuits known as the CERN accelerator complex, shown in Figure 4.1, which boost the particle energies in several stages. Although the CERN complex also has a heavy ion programme, only the acceleration and collision of proton beams will be discussed here.

Firstly, to obtain protons, hydrogen atoms are ionised using a strong electric field. The resulting protons are fed into LINAC2, a radio-frequency (RF) cavity-based accelerator, which brings the proton energy up to 50 MeV. From here, the beam is transferred into the PS booster, composed of four superimposed synchrotron rings that accelerate protons up to an energy of 1.4 GeV, while also increasing the beam intensity. Protons are then delivered to the PS, a synchrotron ring that brings the beam energy up to 25 GeV. At this stage, the beam is segmented into bunches containing several billion protons, each separated by a 25 ns time interval, resulting in a 40 MHz event rate once inside the LHC ring. Beams are then injected into the SPS, the penultimate accelerator measuring nearly 7 km in circumference, which injects 2808 proton bunches directly into the LHC at energies of 450 GeV.

The LHC ring itself consists of 16 RF cavities, oscillating at a frequency of 400 MHz, that bring the energy per beam up to its final value of 6.5 TeV. To maintain a stable orbit, 1232 superconducting dipole magnets cooled to as low as 1.9 K are used to guide the beam around the LHC ring. The strength of the magnetic field is synchronised with increasing beam energy; at its peak, the maximum magnetic field is just over 8.3 T. An additional 392 quadrapole magnets are used both to compress the beams in the vertical and horizontal planes, and to bring them into collision at interaction points. After many hours of circulation, the beam is deflected towards a dump consisting of a water cooled carbon absorber.

4.2.1 Luminosity

The LHC has operated in two main phases: Run 1, where data were collected at a centre-of-mass energy of both 7 and 8 TeV, over the period spanning 2009–2013, and Run 2, where data were taken between[1] 2016 and 2018 at an increased centre-of-mass energy of

[1] A limited dataset was also collected during 2015, amounting to approximately 2 fb^{-1}

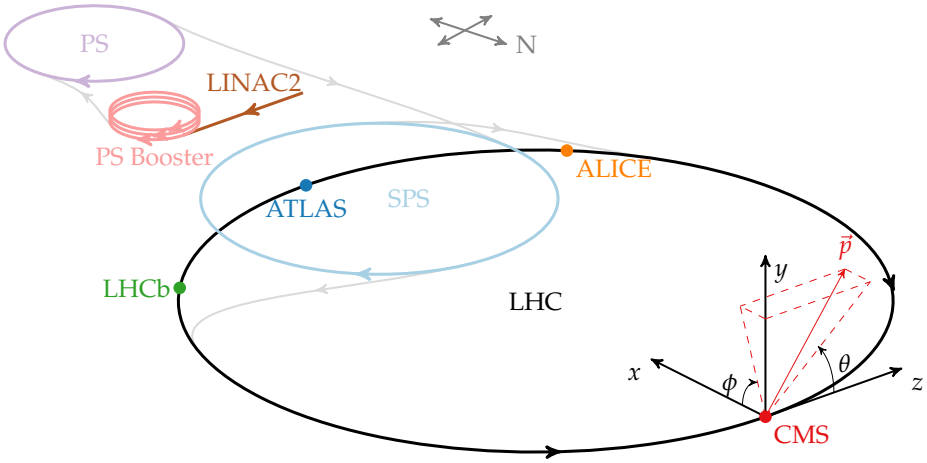

Figure 4.1: A schematic view of the accelerator chain that services the LHC, known as the CERN accelerator complex. Proton beams are accelerated sequentially by LINAC2, the PS booster, PS, and the SPS, before finally entering the LHC ring where the beam energy reaches 6.5 TeV. The four primary LHC experiments located at interaction points are also shown. The co-ordinate system used by the CMS detector is shown in both cartesian and polar co-ordinates, illustrated for a particle of momentum $\vec{p}$. Note that the size of each component is not to scale, and that the positions serve only as a guide to the real physical location. Figure taken from Ref [36].

13 TeV. During the Run 2 phase, the LHC surpassed its design instantaneous luminosity, $\mathcal{L}$, reaching $\mathcal{L} = 2 \times 10^{34}$ cm^{-2}s^{-1} throughout much of 2018. Operating at a high $\mathcal{L}$ benefits the statistical power of resulting analyses, since the rate of a particular physics process, n, is given by the relation

$$n = \sigma\mathcal{L}, \tag{4.1}$$

where σ is the cross section for the process. A high centre-of-mass energy can also benefit analyses of rare physics processes, including Higgs boson production, where the cross section scales with increasing $\sqrt{s}$. Integrating Eqn. 4.1 over the data taking period gives the total number of events for the given physics process, $N = \sigma L$, where L is the integrated luminosity, a measure of the dataset size that is dependent only on the LHC beam parameters and operating time. Figure 4.2 shows the cumulative integrated luminosity delivered to the CMS experiment by the LHC for each year of the Run 2 data taking period, amounting to 159.3 fb^{-1} in total. Due to inefficiencies and quality requirements during data collection, the size of the dataset recorded at CMS is approximately 15% smaller than that delivered by the LHC.

Although an increased instantaneous luminosity allows for collection of larger datasets,

it poses challenges elsewhere, including an increased rate of inelastic pp collisions known as *pileup* (PU). Pileup events occur concurrently with rarer processes of interest, and must therefore be separated during many physics analyses, particularly those using jets. Figure 4.2 shows the distribution of PU interactions per bunch crossing for each year of the Run 2 data taking period. As the instantaneous luminosity was raised, the average number of PU interactions per bunch crossing increased from 27 in 2016, to 38 and 37 in 2017 and 2018, respectively.

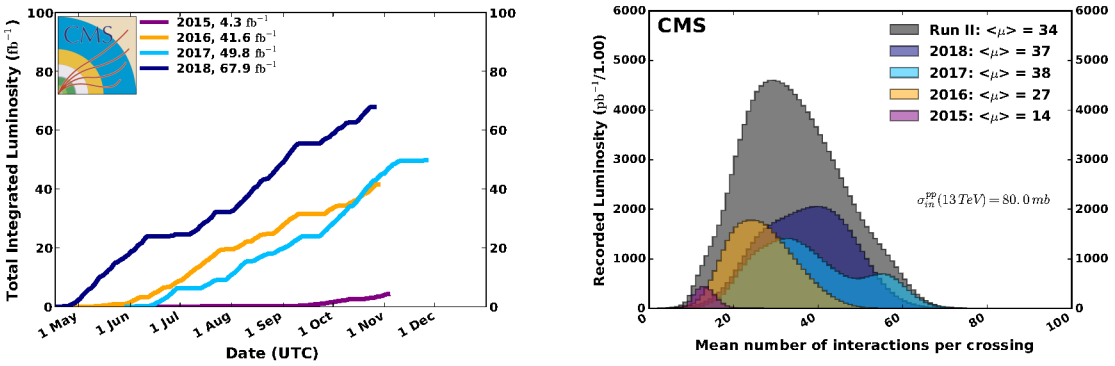

Figure 4.2: Left: the cumulative integrated luminosity delivered to the CMS experiment per month, shown for the years 2015–2018. Right: the distribution of the number of PU interactions per bunch crossing, between the years 2015 and 2018. The average number of PU interactions increases with the instantaneous luminosity of the LHC; for the total Run 2 period (2016–2018), the average number is 34. Figures are taken from Ref [62].

4.3 The CMS detector

The CMS detector is one of two general-purpose physics apparatus on the LHC ring, designed to measure the products of particle collisions. It has a cylindrical shape, measuring over 28 m in length and 15.0 m in diameter, and weighs approximately 14,000 tonnes. The detector is composed of many subsystems, each responsible for measuring different aspects of collisions. The component closest to the beamline is the silicon tracker, followed by the electromagnetic and hadronic calorimeter (ECAL and HCAL, respectively) systems. These are enclosed by the superconducting solenoid, a central feature of the detector which drives the overall design. The most exterior components are the muon detection systems which are interleaved with the return yoke of the magnet.

Each component is required to operate within the high luminosity LHC environment, posing significant constraints on many aspects of the detector design. Materials must be radiation tolerant in order to withstand high particle fluences, while also providing a readout fast enough to contend with the 25 ns collision window. In addition, the detector

instrumentation must be granular enough to provide fine spatial and temporal measurements within the high occupancy environment. Further to the operating constraints, the CMS detector should facilitate a diverse physics programme, requiring sensitivity to a wide range of particles and other objects used in physics analyses (*physics objects*). This includes making precise tests of the SM, such as measurements of the Higgs boson and its interactions with other particles, as well as performing searches for dark matter and other processes not predicted in the SM. To achieve these physics objectives, CMS was designed with following key considerations:

- *Muon identification and momentum resolution:* over a wide range of momenta and angles, muons should be identified with high efficiency and be well measured. This includes good invariant mass resolution for dimuon pairs, and a requirement to unambiguously determine the muon charge. These considerations are particularly important for Higgs boson measurements in the $H \to ZZ^*$ decay channels.

- *Charged particle momentum resolution:* charged particles must be reconstructed with a high efficiency, and their momentum should be well measured by the tracker components. The efficiency for triggering on, and subsequent offline tagging of, τ leptons and jets originating from b quarks should be good.

- *Energy resolution for EM objects*: a wide coverage that should provide good invariant mass resolution for diphoton and dielectron objects. These requirements directly impact the $H \to ee$ decay channel studied within this thesis.

- *Missing transverse energy resolution*: missing energy is an expected signature in many BSM theories, including supersymmetric extensions to the SM. Measuring the missing energy with high resolution requires hadronic calorimeters with large geometric coverage and granular instrumentation.

A schematic of the CMS detector, including a brief description of each component, is shown in Figure 4.3.

4.3.1 Co-ordinate system

The co-ordinate system adopted by CMS has the origin centred at the collision point. The z-axis points in the longitudinal direction along the beamline, the x-axis points radially inward to the centre of the LHC, and the y-axis points vertically upward. For convenience, quantities are also expressed in a cylindrical co-ordinate system, with particle directions described by the quantities η and ϕ. The quantity η, referred to as the pseudorapidity, is defined as $\eta = -\ln[\tan(\theta/2)]$, and used as a measure of the polar angle relative to the

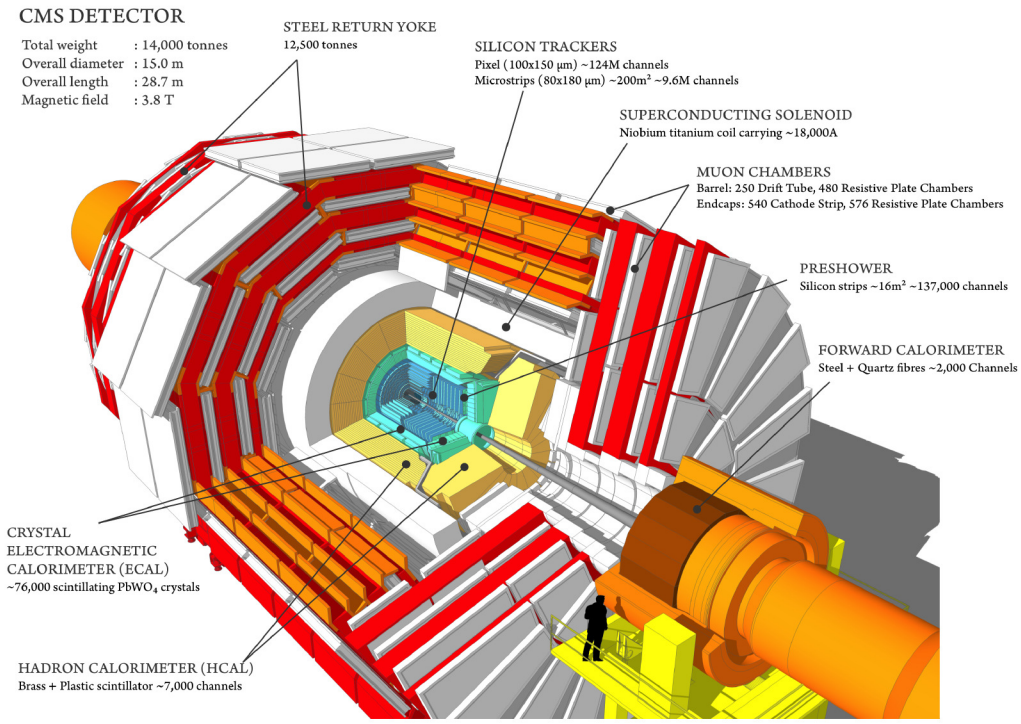

Figure 4.3: A schematic of the CMS detector, where a section is cut away in order to expose each subsystem. Figure taken from Ref [63].

beam axis, θ. Particles with high η values are described as *forward*. The quantity ϕ is the azimuthal angle in the x-y plane. The distance measure in η-ϕ space, ΔR, is also used to describe, for example, the separation of two particles, defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Other common quantities include the component of momentum (energy) in the direction transverse to the beamline, p_T (E_T), and the magnitude of the negative vector sum of particle momentum in the same transverse plane, E_T^{miss}. The CMS co-ordinate system is summarised in Figure 4.1.

4.3.2 Tracker

The tracking system is the innermost component of the CMS detector, aiming to precisely determine the trajectories of charged particles passing through its coverage. In addition, the tracker must accurately reconstruct the primary vertex of the hard scattering process, as well as locating any decay vertices of secondary particles, such as those resulting from B hadron decays. To perform these tasks at the high instantaneous luminosity and bunch crossing frequency of the LHC, the tracker components must be instrumented granularly, with extremely fast readout capabilities. Such requirements imply a high density of detector electronics with associated cooling systems, significantly increasing the material budget. However, a larger mass of instrumented material results in undesirable interactions that may adversely affect the physics performance of the detector. These include multiple scattering events and bremsstrahlung radiation, which degrade the tracker momentum resolution. Proximity to the beam pipe also necessitates the tracker to be built using radiation hard materials. As a compromise between the above constraints, the tracker is instrumented solely with silicon-based technologies, which provide good spatial and temporal resolution at a low material budget, while also tolerating a high radiation environment.

The CMS tracker has cylindrical geometry and provides tracking capabilities up to $|\eta| < 2.5$. The system provides measurements along the charged particle trajectory, known as *hits*, which seed the track reconstruction algorithms presented in chapter 6. It is composed of two distinct components: a silicon pixel detector, which extends to a radial distance of 20 cm, and a silicon strip detector which covers the remaining volume out to a radius of 1.1 m. The original pixel detector, which was designed to operate for ten years, was instrumented with 66 million sensors, each measuring 100 μm $\times$ 150 μm. Closest to the interaction point, the sensors were organised into three cylindrical layers situated at radii of 4.4, 7.3, and 10.2 cm. Two additional pixel endcap detectors arranged in disks sat either side of this central module at longitudinal distances of $\pm$34.5 and $\pm$46.5 cm, extending the η coverage. During the technical pause scheduled at the end of 2016, the pixel tracking system was upgraded, in expectation of greater particle fluence during the

remainder of Run 2 [64]. The upgrade aimed to mitigate radiation damage associated with the increased luminosity environment which can reduce the charge collection efficiency of modules, leading to reduced spatial resolution, particularly for components closest to the beamline. The associated increase in number of pileup interactions is also detrimental to the physics performance, resulting in lower tracking efficiencies. The upgrade improved the granularity of readout components, with the number of cylindrical layers increased by one. This allows for four spatial measurements along the charged particle trajectory in the barrel region, at radii of 2.9, 6.8, 10.1, and 10.6 cm. An additional silicon disk was also added to each endcap section, such that the endcaps provide measurements at longitudinal distances of ±29.1, ±39.6, and ±51.6 cm. These modifications result in an excellent spatial resolution of around 13 and 20 μm in the transverse and longitudinal directions respectively. The new components are lower mass in order to maintain a reasonable material budget, whilst also more radiation hard than their predecessors. A schematic of the upgraded tracker layout is shown in Figure 4.4.

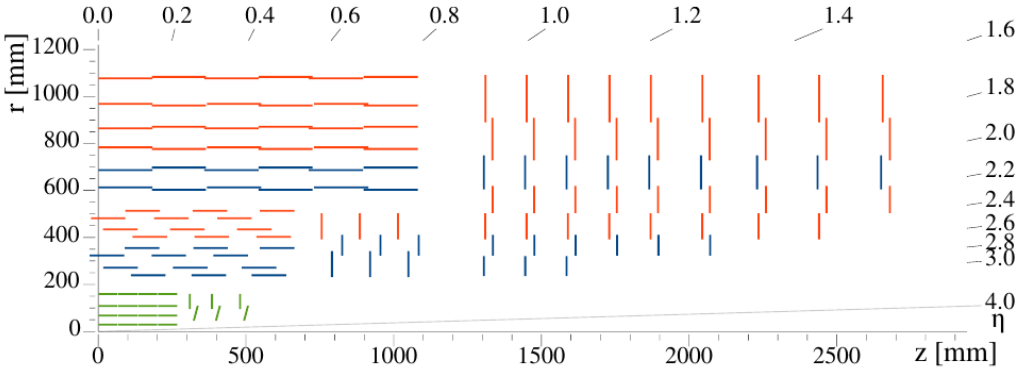

Figure 4.4: One quarter of the CMS tracking system, shown in the r-z view. Components of the silicon pixel detector are shown in green, while the strip detectors are coloured in both orange and blue, denoting single-sided and double-sided modules respectively. The silicon pixel detector was upgraded during the technical pause between 2016 and 2017 to account for the expected increase in particle fluence. Figure taken from Ref [65].

The outer tracker uses silicon strip technologies, and can be divided into three components. The innermost components are the tracker inner barrel (TIB) and associated endcaps (TIDs), which extend to a radial distance of 55 cm. The TIB consists of four layers of silicon strip sensors, arranged parallel to the beam axis, orthogonal to the three disk layers that comprise each TID. Together, the TIB and TID can perform up to four position measurements along a particle trajectory, with a spatial resolution between approximately 20 and 35 μm. Exterior to these components is the tracker outer barrel (TOB), the outermost component, extending to a radius of 116 cm. The TOB consists of

six layers, each of which provides an additional measurement along the charged particle trajectory. Note that the first two layers of the TIB, TIDs, and TOB are also mounted back-to-back with a second silicon strip module, which provides an additional z (ϕ) measurement in the barrel (endcap) region. Finally, the η coverage is extended by two tracker endcaps (TECs), positioned symmetrically about the interaction point, between z coordinates of $\pm$124 cm and $\pm$282 cm. Each TEC is instrumented with 9 disks, carrying up to seven concentric rings of silicon strip sensors. The first, second, and fifth rings of the TEC are mounted with second silicon sensors, providing further 2D measurements.

With all components combined, the tracker geometry ensures a minimum of 9 hits over the full η range, with at least four of them providing two-dimensional measurements. The momentum resolution is $< 1\%$ for particles produced centrally ($|\eta| < 1$) with $p_T \approx$ 10 GeV, which is dominated by multiple scattering affects. At higher energies, these effects are significantly reduced; however, the associated tracks experience less curvature in the CMS magnetic field, resulting in a worsening in p_T resolution to $< 2\%$ for particles with $p_T \approx 100$ GeV and $|\eta| < 1$.

4.3.3 Electromagnetic calorimeter

The CMS electromagnetic calorimeter is a homogeneous calorimeter designed to measure the energy deposits of particles that interact electromagnetically. Alongside the tracker, it is an important component in reconstructing and identifying electrons produced in the $H \to ee$ decay.

The ECAL, shown in Figure 4.5, is comprised of a cylindrical barrel section (EB) which spans the pseudorapidity range $|\eta| < 1.48$, enclosed by two endcaps (EE) that increase the angular coverage to $|\eta| < 3.0$. Both the EB and EE are instrumented with lead tungstate ($PbWO_4$) crystals that produce scintillation light when incident with photons or electrons. The light is collected by photodetectors that convert the response to a voltage, proportional to the energy of the incident particle. Each crystal in the EB (EE) has a frontal area of 22×22 mm^2 (28.6×28.6 mm^2), and is 230 mm (220 mm) in length. The transverse size of each crystal is comparable to the typical width of an EM shower deposited in $PbWO_4$, facilitating photon and electron identification using properties of the (resolvable) shower shape. The barrel region consists of 61,200 such crystals, which are grouped into 36 supermodules that each cover half the EB length and 20° in ϕ. The axis of crystals inside a supermodule is rotated by a 3° angle in both η and ϕ with respect to the interaction point, in order to avoid particles passing entirely through a gap region. The endcaps, which are divided into two semi-circular halves, or *dee's*, each hold a further 3662 crystals arranged into 5×5 groups of so-called *supercrystals*. Finally, preshower detectors (ES) are placed in front of each EE region, covering the range $1.65 < |\eta| < 2.6$. Each

ES contains two layers of lead absorber designed to initiate EM showers, instrumented with granular silicon strip modules used to improve the misidentification of photons from $\pi^0 \to \gamma\gamma$ decays.

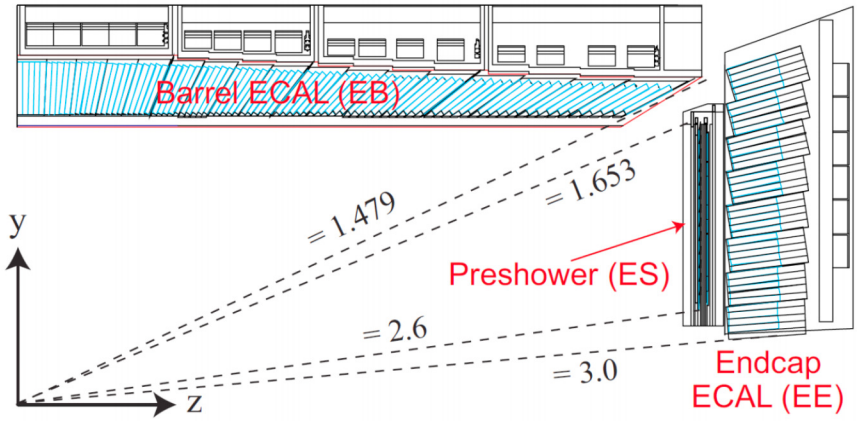

Figure 4.5: One quarter of the CMS ECAL, annotated with the pseudorapidity coverage of the EB, EE, and ES. Figure taken from Ref [66].

The choice of lead tungstate crystal is driven by the material's high density (8.28 g/cm^3) and short radiation length[2], X_0, (0.89 cm) which allow for a compact design that provides good containment of EM showers over the full 25.8 X_0 longitudinal length. A small Moliére radius[3] (2.2 cm) also inhibits shower spreading in the transverse plane. The LHC operating conditions pose further requirements on the crystal properties, including a response time compatible with the 40 MHz bunch crossing frequency, and adequate radiation hardness. The $PbWO_4$ crystals emit 80% of light within the required 25 ns collision window, and are reasonably tolerant to radiation damage; however, over time, the exposure to high particle fluences causes the crystals to transmit less light. This effect can be at the level of ≈10% for crystals positioned at high η, and must therefore be monitored and corrected for by measuring the optical transparency using laser light.

The light yield from $PbWO_4$ crystals is relatively low and must be amplified by on-detector electronics. To this end, each crystal in the barrel is mounted with an avalanche photodiode, responsible for both collecting and amplifying scintillation light. In the endcap region, where the particle fluence is significantly higher, crystals are instrumented with vacuum phototriodes, capable of operating within the high axial magnetic field. The photodetectors produce approximately 4,500 photoelectrons per GeV of energy deposited,

[2] A radiation length is defined as the mean distance over which a high energy electron will lose all but $1/e$ of its initial energy, due primarily to bremsstrahlung radiation.

[3] A Moliére radius characterises the lateral shower containment, and is defined as the radius of a cylinder containing on average 90% of the deposited shower energy.

which are converted to digital format using analog-to-digital (ADC) converters. The digitised amplitudes are buffered while awaiting trigger decision, before being transferred to off-detector electronics.

The energy resolution of the CMS ECAL is particularly important for the H $\rightarrow$ ee analysis, where resolving the narrow signal peak in the dielectron mass distribution increases the signal-to-background ratio. For the majority of particle energies at the LHC, the intrinsic energy resolution can be parameterised by

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{4.2}$$

where S is the stochastic term, N is the noise term, and C is the constant term. Each coefficient is fit to data collected during test beams; for energies in GeV, characteristic values of S=2.8%, N=12%, and C=0.3% are obtained [11].

Showers resulting from both electrons and photons are typically spread over several ECAL crystals. Therefore, in order to build candidate EM objects, these energy deposits must be combined into so-called *superclusters* (SC), from which an energy can be estimated. The reconstructed energy [66] of an electron or photon SC can be described using

$$E_{e/\gamma} = F_{e/\gamma} \cdot \left[G \sum_i S_i(t) C_i A_i + E_{\mathrm{ES}} \right], \tag{4.3}$$

where, firstly, the digitised signal amplitudes (A_i) are summed over each of the i channels in the SC, weighted by coefficients that correct for time ($S_i(t)$) and channel (C_i) dependent changes in response. The resulting amplitude is multiplied by the ADC-to-GeV conversion factor (G), and summed with any associated energy deposited in the preshower plates (E_{ES}). Lastly, any residual effects from imperfect clustering, upstream material, and detector geometry are accounted for by a global correction factor ($F_{e/\gamma}$), estimated using a series of multivariate regressions. The procedure to obtain $F_{e/\gamma}$ is described in chapter 6, where it is also validated for the H $\rightarrow$ ee analysis.

4.3.4 Hadronic calorimeter

The Hadronic calorimeter is a sampling calorimeter, placed between the ECAL and inner extent of the magnet coil [11, 67]. It is designed primarily to measure the energy deposits of charged and neutral hadrons, but also plays an important role in measurements of E_T^{miss}. Figure 4.6 shows the layout of each sub-component in the HCAL. It is comprised of a barrel region (HB) which extends to a radius of 2.95 m, two endcaps (HE), and a forward region (HF) that increase the angular coverage up to large pseudorapidities.

Modules in the HB are arranged into two half barrels consisting of 18 azimuthal wedges, placed either side of the interaction point, and covering the range $|\eta| < 1.3$. The wedges are constructed using brass absorber plates designed to initiate hadronic showers. Brass is chosen for its large nuclear interaction length[4], λ_I, (16.4 cm) which allows for good shower containment within the HCAL volume, and for its tolerance to a high radiation environment. In total, the absorber contains between 5.8 and 10.6 λ_I of material, depending on the η angle. Each absorber wedge is interleaved with trays of silicon tiles, divided into 16 sectors based on η. This results in approximately 70,000 tiles in total, with a segmentation of 0.087 in both η and ϕ. The tiles produce scintillation light which is read out by wavelength-shifting plastic fibres and used to infer the energy of the incident hadron.

Beyond the solenoid sits the most exterior component of the HCAL: the outer hadronic calorimeter (HO). The HO uses the solenoid and iron return yoke as the absorber material, extending the number of nuclear interaction lengths by 2–3. Without this additional material, space constraints imposed by the magnet would result in an inadequate amount of stopping power within the central region, impacting the energy resolution for late-starting hadronic showers and E_T^{miss} measurements. The HO uses the same plastic scintillator technology readout, positioned within the rings of the magnet return yoke, with similar segmentation to the HB.

The HCAL endcaps cover a substantial portion of the pseudorapidity range, extending out to $|\eta| < 3$. Given the proximity to the beamline, the HE must be tolerant to large particle fluences, and contain as many nuclear interaction lengths as possible in order to contain hadronic showers. These considerations lead to brass also being chosen as the absorber, resulting in approximately 10 λ_I of material when including the EE. Silicon scintillator tiles instrument the absorber with a granularity equivalent to the HB for $|\eta| < 1.6$. For angles closer to the beamline, this segmentation reduces to 0.17 in both η and ϕ.

The final components of the HCAL system are the forward calorimeters, positioned at $z = \pm 11.2$ m downstream of the interaction point. These components experience extremely high particle fluxes; on average, the energy deposition in the two HF detectors is almost 800 GeV per interaction, compared to around 100 GeV for the rest of the detector. This environment poses considerable challenges to calorimetery and is the driving design requirement for the HF. To meet these challenges, the HF uses a steel absorber material, with grooves instrumented by radiation hard quartz fibres for readout. The fibres produce Cerenkov light which is collected by photomultiplier tubes. Over 1000 km of fibres are

[4] A nuclear interaction length is defined as the mean distance travelled by a hadron before undergoing an inelastic nuclear interaction.

arranged parallel to the absorbers, and are divided into two lengths — a longer fibre runs the full length of the HF, while shorter fibres start at a depth of 22 cm from the front face. This arrangement allows for separation of EM and hadronic showers, which penetrate the material to differing depths. The HF calorimeters play an important role in reconstructing the VBF H $\rightarrow$ ee production mode, where the Higgs boson is produced in association with two forward jets.

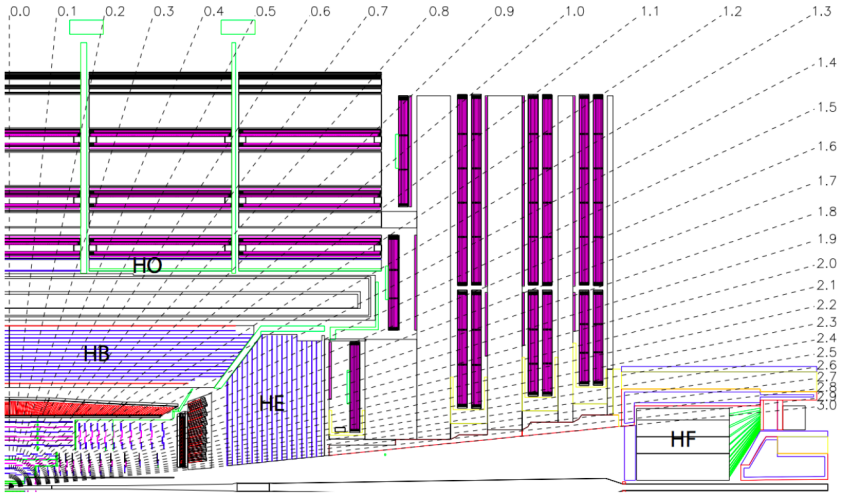

Figure 4.6: A quadrant of the CMS hadronic calorimeter, with the primary four regions labelled (HB, HE, HO, HF). Various pseudorapidity angles are also shown by the dashed lines. Figure taken from Ref [11].

4.3.5 Solenoid

The central feature of the CMS detector is a 13 m long, 6 m (inner) diameter superconducting solenoid, cooled to a temperature of 4.5 K, and storing an energy of 2.6 GJ at full operating current. The solenoid produces a strong magnetic field of 3.8 T, which greatly benefits the charge assignment efficiency and measurements of the momentum of highly energetic charged particles. These particles bend under the field's influence, allowing the momentum and charge to be determined from the radius and direction of track curvature, respectively. The magnetic flux is confined to the volume of the detector by the steel return yoke, which at 10,000 tonnes, comprises the majority of the CMS detector mass.

4.3.6 Muon chambers

Precise muon reconstruction has been a key physics objective for CMS from its earliest design stages, since muonic final states are expected for many physics processes of interest, including the H $\rightarrow$ ZZ* $\rightarrow 4\mu$ Higgs boson decay channel. Muons traverse the CMS

apparatus relatively unimpeded, leaving deposits in the tracker and calorimetry systems only through ionisation processes. A dedicated muon tracking system is built into the return yoke for the CMS magnet, designed to extend tracking measurements several metres beyond the interaction point. This system aims to reconstruct the muon charge and momentum over the entire kinematic profile expected at the LHC. It consists of a barrel (MB) and two endcap regions (ME), instrumented with three gaseous detector types. In a typical gas detector, as a charged particle passes through the active volume, atomic nuclei are ionised, creating electron-ion pairs along the particle path. A strong electric field is applied across the chamber, causing a drift of the liberated electrons towards a wire anode. During this drift, an avalanche of subsequent gas ionisations take place, resulting in a measurable charge reaching the anode. The charge pulse across each each anode can then be used to infer the position of the incident particle. The exact choice of gaseous technology varies with η, motivated by the expected muon flux, the neutron-induced background, and the local properties of the magnetic field.

In the MB, drift tube (DT) chambers provide tracking capabilities over the range $|\eta| < 1.2$. The DTs use a mixture of Ar and CO_2 gases, and are organised into four overlapping stations interspersed throughout the return plates. Each station contains four chambers that measure muon co-ordinates in the ϕ direction, and four that provide a measurement of z co-ordinates. In the two endcap regions, the muon system uses cathode strip chambers (CSCs) which provide a fast response time and radiation resistance suitable for the increased particle fluence. Four stations of CSCs instrument each endcap, arranged perpendicular to the beamline, and provide coverage for muons produced between $0.9 < |\eta| < 2.4$. Both the MB and ME regions are also instrumented with resistive plate chambers (RPCs) that operate independently. The RPCs produce a fast response allowing for improved timing resolution that is used to inform the trigger decision. The additional position measurements also help to resolve ambiguities when attempting to form tracks from multiple hits in a single chamber.

Overall, the efficiency to reconstruct and identify muons is greater than 96% over the full η range.

4.3.7 Trigger

Given that each 25 ns LHC beam crossing generates approximately 1 Mb of data, recording information from every event would result in insurmountable data processing and storage challenges. The CMS detector therefore uses a two-level trigger system [68, 69] to reduce the recorded event rate to a tolerable level. The first Level-1 Trigger consists of custom electronics, which use coarse data from a subset of detectors in order to reduce the recorded event rate to 100 kHz. The second system is the High-Level Trigger (HLT), implemented

in software using a farm of commercial processors. In contrast to the L1T, the HLT has access to the full granularity of the detector readout, which is used to form more complex selection criteria that further reduce the event rate to around 1 kHz.

Level-1 Trigger

The L1T has to perform a factor of 400 decrease in event rate within a latency of 3.2 μs. Inputs to the decision algorithms are therefore restricted to basic information from the muon and calorimeter systems only, with no access to the highly segmented readout from the tracker. At this stage, no effort is made to distinguish photon and electron objects, which produce similar signatures in the ECAL. The coarse information is used to form a list of algorithms, known as *seeds*, which check each event against a set of predetermined criteria known as the trigger *menu*. The menu covers the broad physics programme at CMS, maintaining a high efficiency for signals of potential interest. The seed selection typically applies thresholds to the p_T and η of physics objects, which are adjusted such that the total menu rate is within 100 kHz. For the Run 2 H $\rightarrow$ ee search, the final state contains multiple e/γ candidates, a signature which is allocated 6.4% of the total L1T rate [68].

The full event resolution information, including the tracker readout, is buffered until a L1T accept signal is generated, where it is then passed on to the HLT.

High level Trigger

Filtering by the HLT uses the full precision of data from the detector in order to reduce the event rate by a factor of 100. The event selection is similar to that used in offline algorithms; physics objects including electrons, muons, and jets, are reconstructed, including information from the tracker, and compared against selection criteria that informs whether an event should be accepted. For the H $\rightarrow$ ee search, the HLT selection comprises asymmetric thresholds on electron p_T, as well as loose isolation requirements which limit energy deposits nearby the electron candidates. Events passing selection are sent to a separate computing farm for further reconstruction offline, and storage.

4.4 The High Granularity Calorimeter Phase-2 upgrade

4.4.1 The High Luminosity LHC

With Run 2 of the LHC operation concluded, many upgrade projects for both accelerator and detector components are underway. These upgrades are typically scheduled within so-called *long shutdown* (LS) periods, where the LHC beam is inactive. Amongst other

projects, the LS2 period, which has been underway since the end of Run 2 operation in 2018, has prepared the LHC to run at a higher centre-of-mass energy of 13.6 TeV. Such upgrades are in anticipation of Run 3, which will span over mid-2022 to 2024, and bring the total integrated luminosity collected at the CMS detector to approximately 300 fb^{-1}. Beyond Run 3, operation of the LHC with the current beam parameters is of limited interest for physics analyses; in order to improve the statistical precision on measurements by a factor of two, ten more years of operation with the Run 3 conditions would be required. Instead, a major upgrade to the LHC is planned, aiming to achieve a substantially increased instantaneous luminosity of up to 5×10^{34} $\text{cm}^{-2}\text{s}^{-1}$. The upgraded LHC, known as the High Luminosity LHC (HL-LHC) [70], will run until the mid-2030's, collecting approximately 3000 fb^{-1} of data. The considerable increase in luminosity benefits many physics analyses, from precision measurements, to searches for rare and exotic processes. In the Higgs sector, rare interactions will be measured with significantly improved sensitivity, including couplings to the light fermions and the Higgs boson self interaction. The expected precision of the Yukawa couplings to various SM particles projected at the end of the HL-LHC operation is shown in Figure 4.7. Measurements of Higgs boson production cross sections, such as those defined in the Simplified Template Cross Section scheme [40], will also be facilitated with unprecedented resolution, allowing for precise tests of the SM predictions.

The benefits of operating at a higher instantaneous luminosity are, however, accompanied by difficulties elsewhere. The average number of pileup interactions per bunch crossing is expected to increase to around 140, requiring highly segmented detector readouts and sophisticated subtraction techniques to prevent confusion with objects of interest. Furthermore, detector components must be sufficiently resistant to the high radiation levels, given that the expected dose in year one of the HL-LHC operation will be equivalent to that absorbed during the entire LHC operation up to Run 3. In order to preserve the performance of the CMS detector despite the challenging environment, many of the detector subsystems will require upgrading. These Phase-2 upgrades are extensive, including a new highly segmented tracker [65] that will also inform L1T decisions, upgraded and unified ECAL and HCAL endcap components [71], and additional instrumentation of the muon system.

In this section, the calorimeter endcap upgrade project, known as the High Granularity Calorimeter, will be discussed, with a particular focus on its use within L1T algorithms used to identify electromagnetic objects.

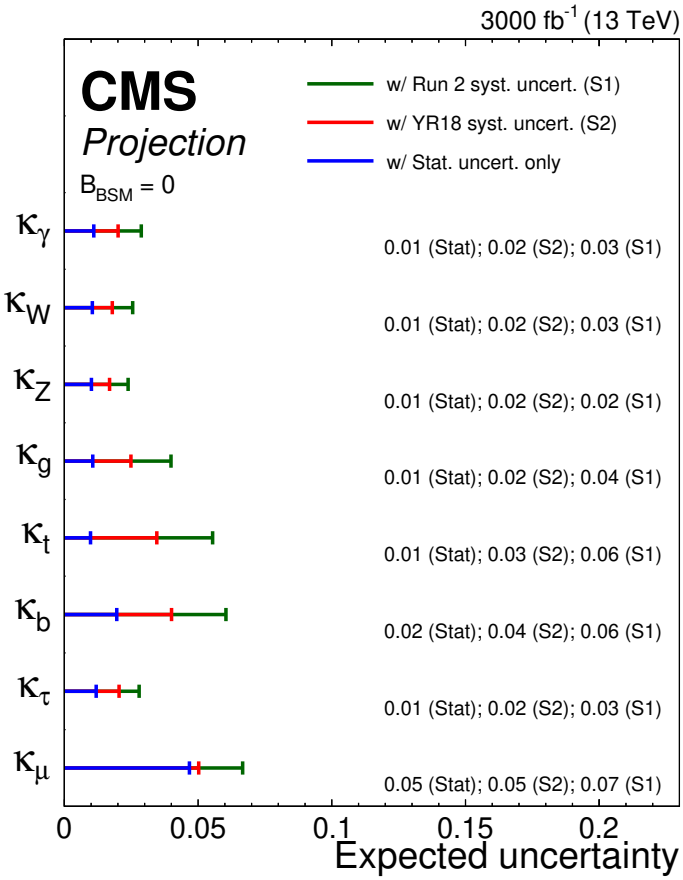

Figure 4.7: The expected precision of Higgs boson coupling measurements projected by the end of the HL-LHC operation. The $\pm 1\sigma$ systematic uncertainty bands are broken into two regimes: S1, where Run 2 systematic uncertainties are assumed, and S2, where the (improved) systematic uncertainties from 2018 only are assumed. The statistical-only component of the uncertainty is also shown (blue). Figure taken from Ref [72].

4.4.2 The High Granularity Calorimeter

The current technologies used in the CMS endcap calorimeters were designed to operate up to an integrated luminosity of ≈500 fb^{-1}, and are thus unable to withstand the considerably higher particle fluxes expected at the HL-LHC. The requirement of radiation tolerance drives the CMS endcap upgrade project, which will see both the EE and HE subsystems replaced by a single, unified sampling calorimeter, comprising both electromagnetic (CE-E) and hadronic (CE-H) sections. Figure 4.8 shows the calorimeter design, which covers the range $1.5 < |\eta| < 3$.

Silicon sensors are chosen as the active material for the bulk of the detector upgrade, since they have been demonstrated capable of withstanding the expected radiation doses throughout the entire Phase-2 operation. Their fast response also provides timing capabilities for energy deposits, allowing for a significant reduction in out-of-time PU. The CE-E portion consists of 28 layers of tungsten absorber interleaved with silicon sensors of size 0.5–1cm^2, corresponding to around 26 X_0 of material. The same silicon sensors instrument the entire front section of the CE-H, which contains 12 brass and copper absorber plates, providing excellent energy containment. The rear CE-H component, which extends the calorimeter depth to around 10 λ_I, is instrumented using a combination of silicon sensors and plastic scintillator tiles. The scintillator-technology is positioned at radii further from the beampipe, such that the radiation induced light-loss is minimised.

Both the electromagnetic and hadronic components of the HGCAL feature an unprecedented readout granularity with over 6 million channels [71], providing considerable benefits for the particle flow (PF)-based reconstruction of physics objects (see Section 5.3). In the lateral direction, fine segmentation allows for separation of neighbouring showers, improving jet resolution and discrimination against PU clusters. The longitudinal segmentation enables frequent sampling of the shower development, providing good energy resolution for electromagnetic objects, including photons from $\text{H} \rightarrow \gamma\gamma$ decays. Moreover, the high density of the absorber materials results in laterally compact showers that are well contained, improving the resolution of physics signatures involving boosted jet topologies, such as VBF Higgs boson production.

4.4.3 The HGCAL L1T

Alongside detector upgrades, a complete replacement of the Level-1 and High-Level Trigger is planned for Phase-2, designed to maintain the trigger efficiency despite the higher luminosity environment. Improvements to front-end electronics and the data acquisition system are expected to increase the L1T latency to 12.5 μs, with an allowed total event rate of 750 kHz [73]. As a result, more granular information may be exploited to inform L1T decisions, including, for the first time, readout from the tracker system.

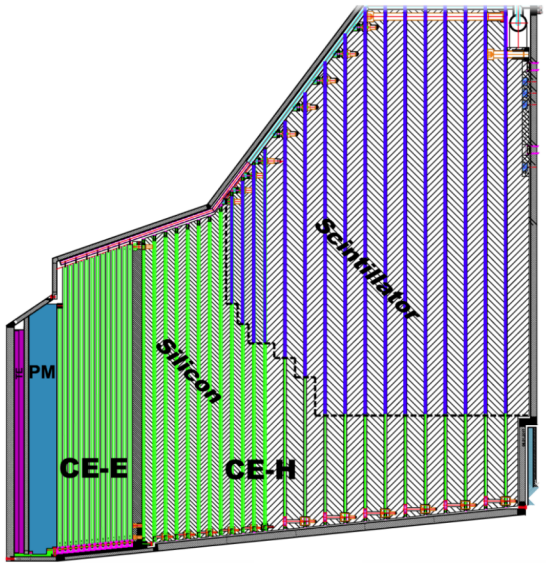

Figure 4.8: A cut-away section showing the structure of the HGCAL sampling calorimeter. The HGCAL is divided into an initial electromagnetic section, instrumented with a tungsten absorber and silicon readout , and a hadronic section, which uses brass absorbers instrumented with both silicon sensors and plastic scintillator tiles. These components are positioned behind an additional moderator (PM), which reduces the neutron flux penetrating the electromagnetic section. Figure adapted from Ref [71].

Tracking information in particular will be crucial in preventing unacceptably high trigger rates, while also allowing for particle flow capabilities at Level-1. An increased latency also enables reconstruction and identification of higher-level objects using more complex selection, including machine learning techniques that could considerably improve trigger efficiencies. These algorithms may be realised through customised, re-programmable electronics known as Field Programmable Gate Arrays, which are expected to bear much of the computational load at Level-1.

Although the L1T system is not fully defined, the envisaged menu will reconstruct a wide variety of objects in order to support a diverse physics programme. This includes triggers targeting single Higgs and promising di-Higgs channels including, for example, $HH \rightarrow bb\gamma\gamma$; paths supporting B-physics signals that will utilise soft lepton triggers; and triggers targeting large E_T^{miss} signatures typically associated with searches for supersymmetric particles [73]. Many trigger paths at the L1T will require excellent identification of EM showers produced by single electrons or photons (e/γ), despite the high pileup conditions. The e/γ ID algorithms are currently informed by calorimeter information, although studies incorporating track information are also ongoing.

To produce quantities that describe calorimeter deposits for use in the trigger, neighbouring silicon sensors are first grouped into so-called trigger cells (TCs), with a granular-

ity of around 4 cm^2. To reduce computational resources, only half of the CE-E layers are used to construct TCs; this is, however, 500 times more granular than the corresponding system used throughout Run 2. The TCs are then sent to backend electronics, which rely on FPGA technologies, and amongst other duties, perform basic clustering of TCs. The clustering is similar to that described in Section 5.3, and can be divided into two steps. The first step defines *seeds* around which 2D clusters can be built. Seeds are defined as TCs with energy above some local, position-dependent maxima. The second step attaches TCs to seeds, provided they are within some geometric neighbourhood. Information from each 2D cluster is then combined depth-wise across layers to form 3D clusters. The position of a cluster is defined as the energy-weighted centre of all TC's in the final cluster, while the energy is defined using TCs within some limited distance from the core, as a compromise between shower containment and contamination from PU. Several other cluster properties describing the shower profile are computed, serving as input to various identification algorithms, including the e/γ ID deployed in the HGCAL.

4.4.4 Electron and photon identification at the L1T

The nominal strategy for e/γ ID within the HGCAL uses BDT-based classifiers trained to separate clusters resulting from e/γ deposits against those resulting from PU. Since observables in electron and photon identification evolve rapidly with η, two decision trees are trained: one pertaining to a lower η region, $1.5 < |\eta| \leq 2.7$, the other to a higher eta region, $2.7 < |\eta| \leq 3.0$. Given that the shower properties in the HGCAL are similar between photons and electrons, in practice, BDTs are trained only on electron clusters, simulated within a 200 PU event environment. The benchmark pileup per bunch crossing of 200 is chosen instead of the expected value of 140, since the HL-LHC has the ability to deliver 50% higher values of instantaneous luminosity. Signal clusters used for training are defined as those consistent with originating from a generator-level electron, by requiring each cluster to be within a cone of radius $\Delta R < 0.2$ relative to the generator-level electron. Electron clusters must also pass a minimum p_{T} threshold of 10 GeV. The simulated background comprises all clusters with $p_{\mathrm{T}} > 20$ GeV which are not matched to a generator-level electron.

The BDT is implemented using the XGBoost [74] python library, a gradient boosting package optimised for large datasets. The model is trained on 70% of the total dataset, with the remaining 30% withheld for performance evaluation. Nine input features describing both the lateral and longitudinal cluster profile are considered in training, including:

- the ratio between the cluster energy deposited in the CE-H to that deposited in the CE-E, H/E;

- the first layer in the cluster to contain an energy deposit above a minimum threshold;
- the length of the cluster, defined as the difference between the first and last layers that contain energy deposits above threshold;
- the core cluster length, defined as the maximum number of layers with consecutive energy deposits;
- the energy weighted RMS of the ρ, η, ϕ, and z co-ordinates of the cluster TCs, $\sigma_{\rho\rho}$, $\sigma_{\eta\eta}$, $\sigma_{\phi\phi}$, and σ_{zz} respectively, defined for a generic trigger cell co-ordinate, x, as

$$\text{Weighted RMS}(x) = \sqrt{\frac{1}{E_{\text{tot}}} \sum_{i}^{N_{\text{TC}}} E_i (x_i - \langle x \rangle)^2}, \tag{4.4}$$

  where i enumerates trigger cells, of which there are N_{TC}, each with energy E_i and co-ordinate x_i. The quantity $\langle x \rangle$ is the energy weighted mean of x over all TCs, and E_{tot} is the total energy obtained from the sum over all TCs; and
- the cluster centre, in z-coorindates, $\langle z \rangle$.

The majority of inputs inputs to both classifiers show good separation between signal and background clusters. In particular, Figure 4.9 (left) shows the distribution of $\sigma_{\eta\eta}$, describing the lateral shower spread in η, which is typically narrower for e/γ clusters. Also shown in Figure 4.9 (right) is the ratio H/E. Events associated with signal clusters typically deposit the majority of their energy in the CE-E, resulting in small values of H/E in comparison to PU clusters, where the energy deposited extends out to later layers of the CE-H.

The performance of each BDT is evaluated using the area under the associated ROC curve. The output score distributions used to generate performance curves are shown for each classifier in Figure 4.10. The distributions for signal and pileup events are extremely well separated, resulting in ROC curves with large areas shown in Figure 4.11. The performance is summarised for each classifier using thresholds on the output score, known as *working points*, which correspond to various signal efficiency benchmarks. For the classifier trained in the lower η region, a benchmark signal efficiency of 97.5% is chosen, while for the higher η region, where the number of pileup interactions per bunch crossing is higher, a value of 90.0% is preferred. These correspond to high background rejection rates of 99.36% and 99.69% respectively, where background rejection is defined as one minus the background efficiency. Note that the current working points are chosen only to provide benchmarks of the background rejection rate; the final optimised choice of working point will consider a host of additional factors, such as the trigger rate.

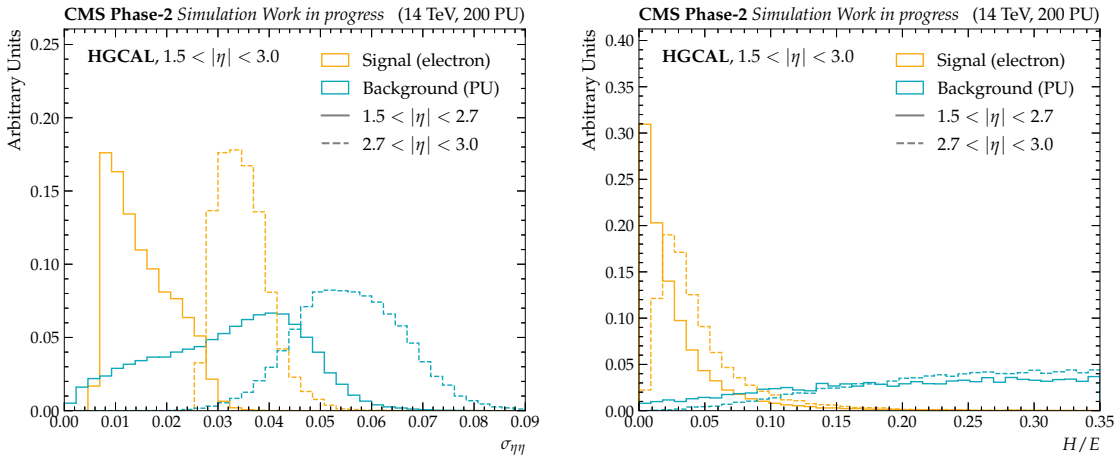

Figure 4.9: Selected inputs to the HGCAL e/γ ID BDTs, shown for simulated single electron (signal) and pileup (background) clusters. The outcome of the classifier will inform decisions in the CMS Phase-2 L1T. Events are separated into the two η regions in which a model is trained, and normalised such that the integral of each histogram is unity. Left: the energy weighted RMS of the η co-ordinate of the cluster TCs. Right: the cluster energy deposited in the CE-H divided by that deposited in the CE-E.

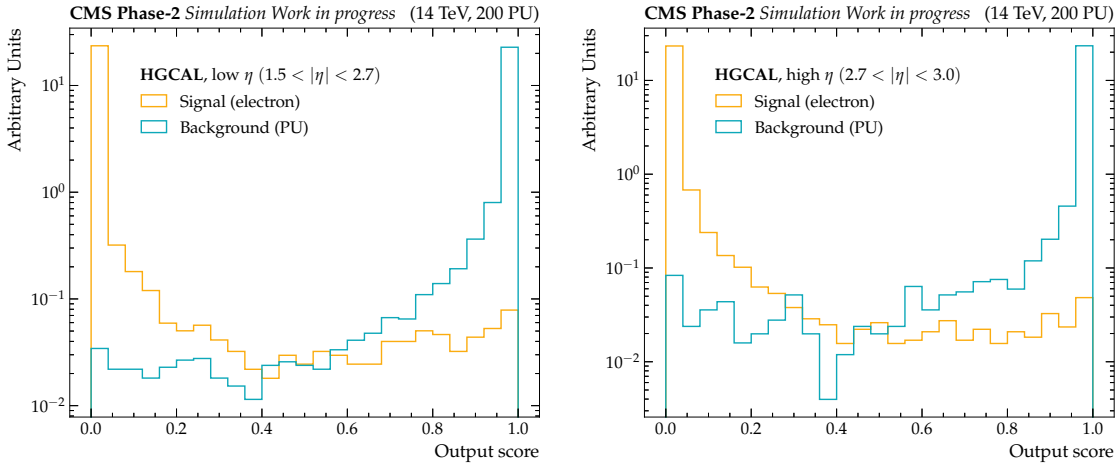

Figure 4.10: Output score distributions for the HGCAL e/γ ID BDTs for simulated electron (signal) and pileup (background) clusters. The BDT trained in the lower η region ($1.5 < |\eta| \leq 2.7$) is shown on the left, while the BDT trained in the higher η region ($2.7 < |\eta| \leq 3$) is shown on the right. Output scores are shown for the test set, containing 30% of the total events. The output score distributions are well separated between signal and background clusters.

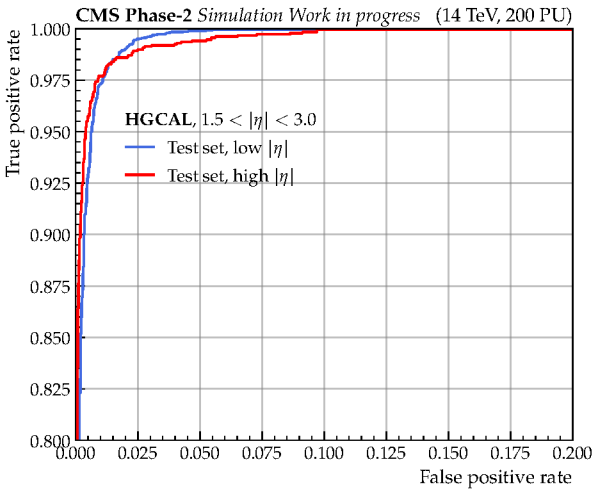

Figure 4.11: The ROC curves for the two e/γ identification BDTs used in the HGCAL L1T, separated by η region. The curve is produced using events from the test set, comprising 30% of the total events. For both classifiers, the TPR remains high even at low FPR values, resulting in large areas under the curve.

All clusters passing the e/γ ID requirements are then promoted to calorimeter-only e/γ candidates, with the energy determined using the CE-E layers. Clusters within the tracker acceptance are considered for track-matching, using track-finder trigger primitives with $p_T > 10$ GeV that have been propagated to the calorimeter surface. These electron-like objects are also corrected for bremsstrahlung losses by merging neighbouring low energy deposits passing the e/γ ID into the highest energy cluster. Photon deposits are identified as showers without a matching track. Finally, the resulting objects are compared against the L1T trigger menu, with events receiving an "accept" signal sent to detector read-out electronics.

4.4.5 Firmware constraints and feature selection

The e/γ ID algorithms for the HGCAL will ultimately be realised in firmware, and must operate in real time within the constraints of the L1T system. The entire HCAL output should be received by the central L1T system within 5 μs, which requires the latency for clustering and property evaluation to be less than 1 μs. Although the exact cluster properties to be evaluated are not completely defined, it is expected that they should require between 128 and 416 bits of memory, in order to remain within the total HGCAL bandwidth allocation of 60,000 bits per endcap, in each bunch crossing [71]. In order to be compatible with these constraints, the e/γ ID must use as few computational resources as possible; importantly, it must be both quick to evaluate and memory efficient.

One strategy to reduce the computational complexity is to truncate the number of

input features provided to the classifier. Boosted decision trees trained with fewer features typically result in more simple base learners, resulting in quicker evaluation during online inference. In addition, removing quantities that do not significantly contribute to performance prevents redundant variables being constructed in firmware, reducing memory requirements. Naturally, removing many features will eventually degrade the performance of the classifier; the objective, therefore, is to remove a maximum number of features, whilst still maintaining reasonable discrimination power against PU clusters.

Typical feature reduction algorithms aim to remove inputs that are deemed to contribute little predictive power to the overall model. To quantify the importance of a feature, the notion of Shapley values is used [75, 76]. A Shapley value measures the contribution of a feature value, for a given event, to the difference between the prediction for that event and the mean prediction over all events[5]. Values with large magnitude are associated with predictions that strongly support a given class. Since a Shapley value is computed at the per-event level, to obtain a single measure of importance for each feature, the mean of the absolute Shapley values is taken. Figure 4.12 shows the fractional breakdown of mean absolute Shapley values for each variable, for both e/γ ID BDTs. In both models, around 75% of the total importance is attributed to just two inputs, namely H/E and $\sigma_{\rho\rho}$. For the lower η classifier, the remaining feature importance values are distributed semi-evenly, with the exception of the three least important features which contribute very little. For the higher η classifier, the $\sigma_{\phi\phi}$ quantity contains the majority of the remaining importance, meaning around 85% can be attributed to only three inputs. Although these asymmetric importance distributions highlight an opportunity for variable pruning, it should be noted that the relative importance can also shift between updates to the feature set.

One common feature reduction technique, known as *sequential backward selection* (SBS), removes one feature per iteration, starting from the full set of (nine) features. At each training step, the feature with the smallest Shapley value is dropped from the set. The process is repeated until a termination criteria is reached, which is typically a lower performance threshold. The model is re-trained between updates to the feature set, to account for shifts in importance between models. The resulting performance when applied to each e/γ ID BDT, measured using the area under the ROC curve, is summarised for each step in the algorithm in Figure 4.13. To obtain an estimate of the uncertainty, the model is evaluated many thousands of times on bootstrapped samples of the test set, where events are drawn at random and with replacement. The average ROC AUC is taken as the central value, with the standard deviation of the scores used to estimate the

[5]For a feature value x_i, the Shapley value is x_i's marginal contribution to the predicted value for event i minus the mean prediction over all events, averaged over all possible permutations (or *coalitions*) of the features under consideration.

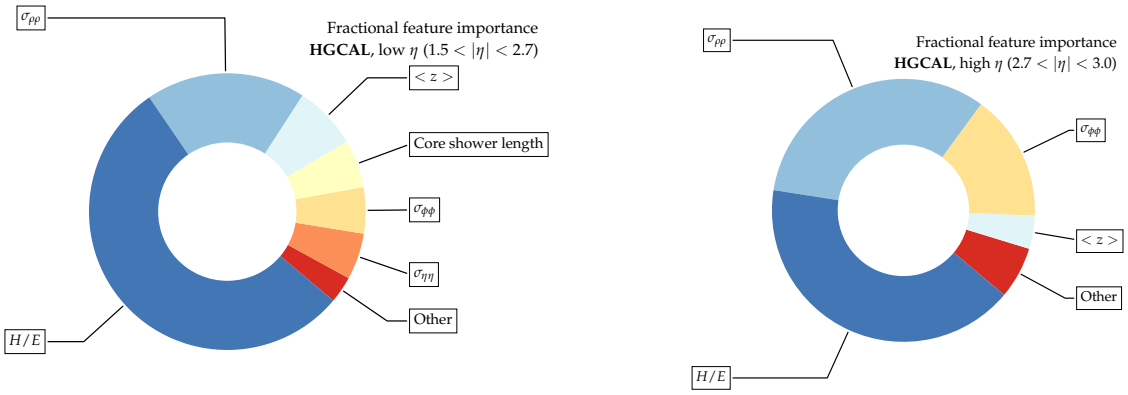

Figure 4.12: The fractional mean absolute Shapley values for each input feature to the HGCAL L1T e/γ ID BDTs. The model trained in the lower η region is shown on the left, while the model trained in the higher η region is shown on the right. For both models, the quantities H/E and $\sigma_{\rho\rho}$ are responsible for approximately 75% of the total feature importance scores. Variables with small importance are shown inclusively by the segment labelled "other".

uncertainty on the mean.

For the classifier trained in the low η region, the performance is unaffected up to the removal of three variables. This could be expected from the importance distribution in Figure 4.12, where the three least important variables contribute very little to the overall model importance. A small degradation is seen when removing additional variables, although the loss in background rejection efficiency is modest; removing five variables, which would decrease the memory required for cluster properties by more than one half, still maintains a background rejection efficiency of 99.22%, at the usual 97.5% signal efficiency benchmark. Results for the BDT trained in the higher η region are also encouraging, with the original performance maintained up to the removal of six variables. This is again expected from Figure 4.12, where the majority of the feature importance can be attributed to only a handful of inputs. At the benchmark signal efficiency of 90.0%, removing six variables maintains a background rejection rate of 99.67%, comparable to the nominal model trained with all features.

Although promising, feature reduction through the SBS algorithm may not yield optimal feature sets at each iteration. The algorithm is *greedy*, given that it can only assess the most important feature in the current training, and cannot look ahead to future iterations where the distribution of importances may change. Therefore, extensions to this work may consider alternative approaches using more holistic algorithms. Studies using more sophisticated methods, for example the Boruta feature selection algorithm [77], have shown similar results to that of SBS. Other approaches using evolutionary algorithms could also be applied with success. In addition, the complication of training two

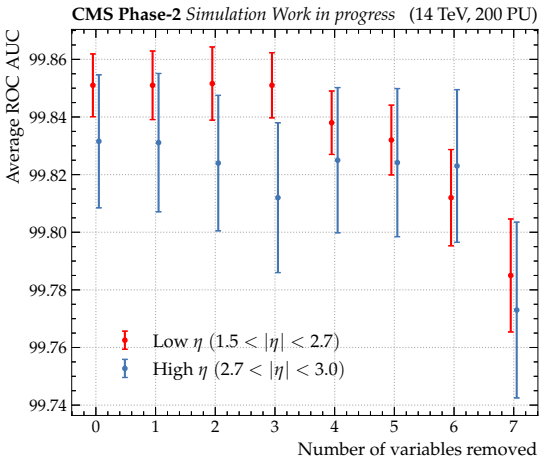

Figure 4.13: The evolution of performance with decreasing number of variables for the HGCAL L1T e/γ ID BDTs. Performance is measured using the area under the ROC curve, averaged across many bootstrapped samples of the test set, with the standard deviation used to estimate the uncertainty on the mean. The performance of the low η BDT, shown in red, degrades slightly upon removal of more than three variables, and noticeably worsens once six and seven variables have been removed. Performance of the model trained in the high η region, shown in blue, is maintained up to the removal of six input variables. A small amount of overtraining results in larger uncertainties on the average ROC score when compared with the corresponding low η model.

separate BDTs, one pertaining to each region in η, has effectively been ignored, given that the top ranking features are similar between models. With the possible advent of new cluster properties in future L1T design iterations, the feature importance rankings could differ significantly between models. This may require features that are redundant in one classifier still being computed for use in the other, resulting in larger memory requirements. Overall, however, feature reduction in the e/γ ID appears a viable method to reduce the computational requirements associated with online model implementation and inference in the HGCAL L1T. The technique is applicable not only to BDTs, but to any such machine learning model that may eventually be deployed. Thus far, studies have aimed at removing a maximum number of features while maintaining close to the original performance; the final realistic constraints that will ultimately inform the choice of model will only be available once the L1T architecture for Phase-2 operation is decided upon.

4.5 Summary

The CMS detector detector is one of two general-purpose detectors situated on the LHC ring, aiming to measure the properties of hadron collisions. Its design centres around a 3.8 T solenoidal magnet and comprises an inner tracker, electromagnetic and hadronic calorimeters, and muon system. Together, these components allow for reconstruction of a wide range of physics objects with high efficiency, including electrons, photons, muons, and jets, over a range of momenta and angles. In addition, the two-level trigger system allows for events of interest to be selected with high efficiency, while maintaining a tolerable readout rate. Upgrade projects for the CMS detector in preparation for the high luminosity Phase-2 operation of the LHC include a new endcap calorimeter, which will inform the L1T decision. A key component of the trigger menu will be identification of e/γ objects against pileup clusters, to which end, dedicated BDT-based classifiers are trained. The classifiers will be evaluated online, and must therefore adhere to the latency and bandwidth constraints associated with the L1T architecture. Various techniques to reduce the computational resources of the e/γ ID are being considered, including reducing the input feature set provided to online models. These feature selection algorithms have been demonstrated to reduce the cluster memory requirements, while also maintaining excellent rejection of pileup clusters.

Chapter 5

Particle reconstruction

5.1 Introduction

To extract a potential H $\rightarrow$ ee signal, a fit is performed to the dielectron invariant mass distribution in each constructed analysis category. For H $\rightarrow$ ee signal events, the m_{ee} distribution forms a narrow peak centred close to the value of the Higgs boson mass, sitting on top of a smoothly falling background from SM processes other than the H $\rightarrow$ ee decay. For events with two electrons, e_1 and e_2, under the approximation of negligible electron mass, the dielectron invariant mass is given by the expression:

$$m_{ee} = \sqrt{2E_{e_1}E_{e_2}(1-\cos\theta)}, \tag{5.1}$$

where E_{e_1} and E_{e_1} are the energies of the leading and subleading-E_T electron respectively, and θ is the opening angle between them. In order to resolve the signal peak resulting from H $\rightarrow$ ee decays over the background continuum, Eqn 5.1 demonstrates that both the electron energies and directions, as well as the position of the primary interaction vertex, must be well measured. Since the analysis sensitivity depends on the signal-to-background ratio, S/B, electrons from the H $\rightarrow$ ee decay must also be correctly identified against those resulting from background processes. The improvement in sensitivity as a result of improved m_{ee} resolution and S/B ratio is shown graphically in Figure 5.1.

This chapter firstly outlines the simulated samples and data used within the event categorisation and to produce the analysis' results. The remainder of the chapter describes the reconstruction of physics objects at CMS, many of which are used within the H $\rightarrow$ ee search, and the corrections applied to them. A particular focus is placed on the reconstruction of electrons, including the per-electron identification techniques. Additional suppression of background events at the per-event level is also discussed in chapter 6.

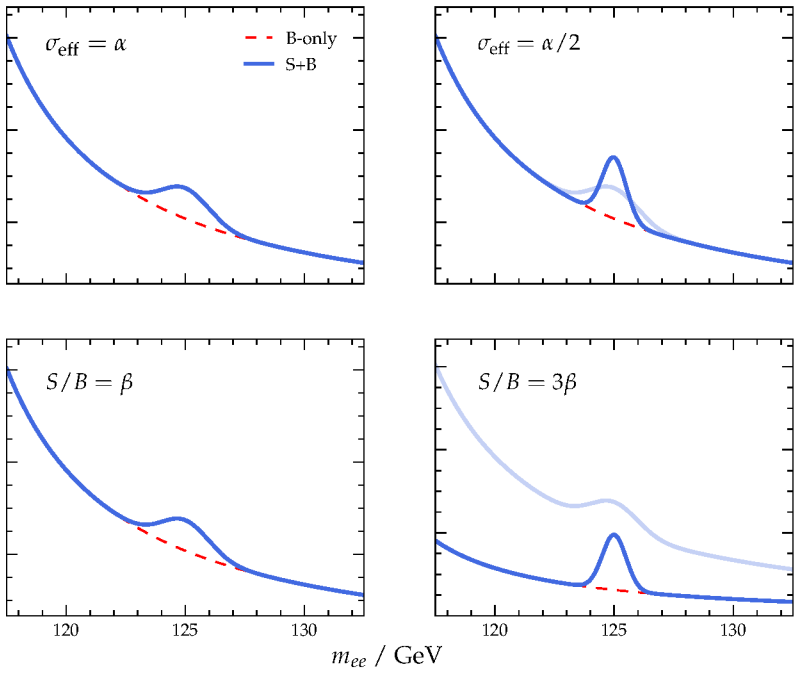

Figure 5.1: An illustration of the means of improving the sensitivity for the H $\rightarrow$ ee search. Top: the effect of improving the m_{ee} resolution, resulting in a narrower signal peak. Bottom: The effect of improving the S/B ratio, due to better discrimination of signal and background events.

5.2 Samples

5.2.1 Data

This search uses 138 fb^{-1} of pp collision data collected by the CMS experiment at $\sqrt{s} =$ 13 TeV, of which 36.3, 41.5, and 59.8 fb^{-1} were collected during 2016, 2017, and 2018, respectively. Data is selected using the two-tiered trigger system at CMS. The analysis uses a dedicated dielectron HLT path, which requires each electron to be seeded by at least one EM candidate in the L1T. For data taken in 2016, seeds were required to have a minimum transverse energy of 23 and 10 GeV for the leading and subleading electron respectively. For 2017 and 2018, the requirements were increased to 25 and 14 GeV respectively [78]. In the HLT, energy deposits associated with the L1T seeds are clustered using a procedure similar to the offline algorithm. Electron candidates are then selected based on properties of the resulting supercluster. For triggers targeting dielectron objects, this selection is similar to that of the L1T, and includes thresholds of 23 GeV and 12 GeV on the leading and subleading electron respectively. Requirements are also placed on the shower shape and isolation properties of the cluster. This includes selection on the root-mean-square of the shower width in η, which describes the lateral shower profile. The isolation requirement limits contributions from energy deposits outside of a cone with axis centred on the EM candidate, with radius $\Delta R = 0.3$, to be less than 20% of the cluster E_T.

The efficiency of the trigger is evaluated in Drell-Yan (DY) $\text{Z} \rightarrow \text{ee}$ events in a control region around the Z boson mass. Drell-Yan events are chosen since a relatively pure sample of electrons that are kinematically similar to those from the $\text{H} \rightarrow \text{ee}$ decay can be obtained. Trigger efficiencies are measured using the tag-and-probe [79] method. This method selects one electron, named the *tag*, which satisfies tight electron identification and isolation requirements. A second electron, named the *probe*, is then chosen to pass a criteria specific to the efficiency being measured — in this case, the dielectron HLT requirements. The efficiency is measured in both data and simulated samples, with the ratio being used to correct simulation to data. The set of corrections are derived differentially in bins of supercluster p_T and η, independently for each era of data taking. Figure 5.2 shows the resulting scale factor distributions, where for all eras, the corrections are close to unity.

5.2.2 Simulation

In this analysis, simulated samples produced using Monte Carlo (MC) techniques are used to model signal events and train MVA-based event classifiers. Simulated signal samples are produced for the four Higgs boson production modes with the highest cross

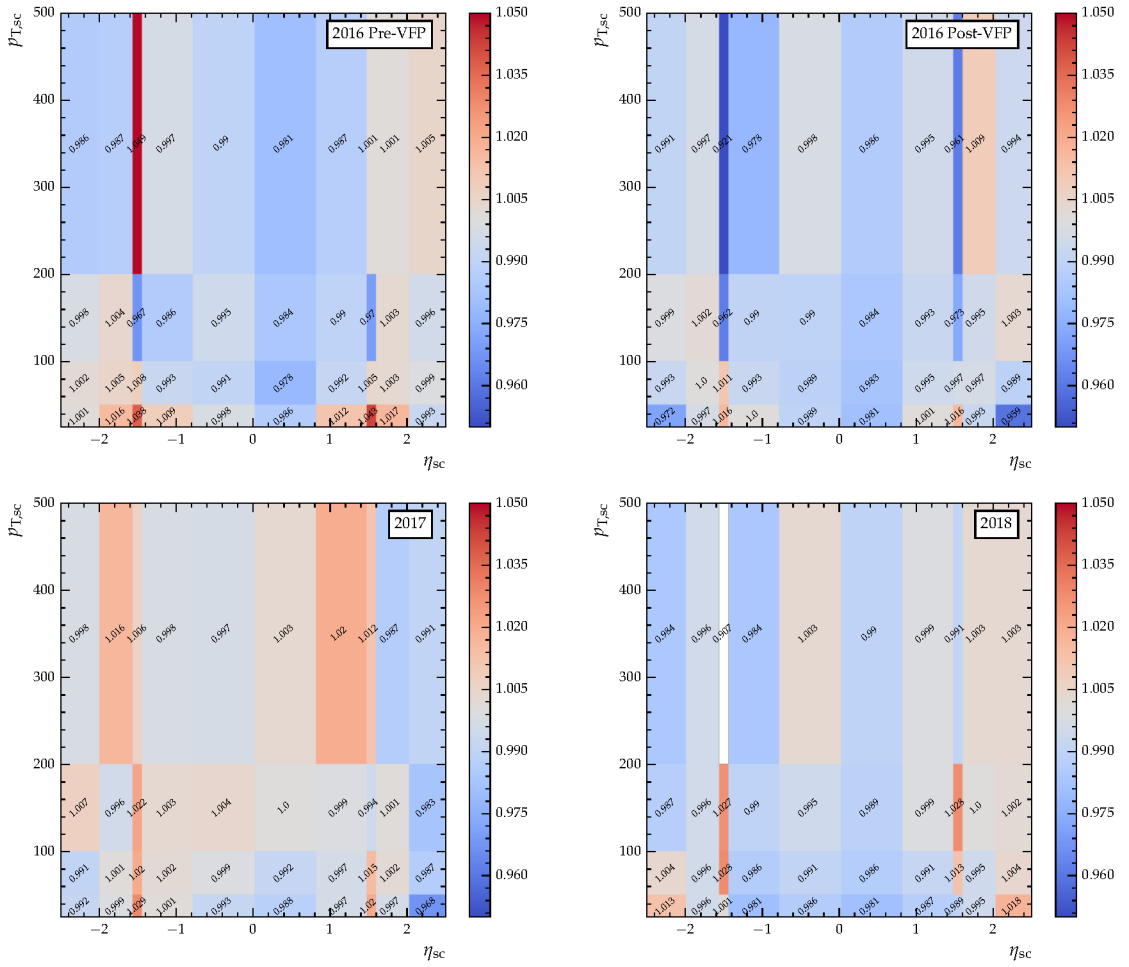

Figure 5.2: Ratio of trigger efficiencies for data and simulation as a joint function of supercluster p_T and η, for 2016, 2017, and 2018. Corrections are evaluated in a sample of $Z \to ee$ events using the tag-and-probe method, before being applied to simulation. Note that two independent sets of corrections are derived for samples in 2016, to account for changing conditions in the mitigation of highly ionising particles.

sections at the LHC, including ggH, VBF, VH, and ttH events. The matrix element (ME) for each process is calculated using the MADGRAPH5_aMC@NLO v2.6.5 [80] package at next-to-leading order accuracy in perturbative QCD. The resulting parton-level samples are then interfaced with PYTHIA v8.230 [81] which simulates the parton showering and hadronisation processes, using the CP5 underlying event tune [82] to calibrate soft QCD processes. The yield of each signal sample is normalised by the corresponding Higgs boson production cross section [40], the H $\rightarrow$ ee branching fraction, and the acceptance of the analysis preselection.

The largest fraction of background events originate from Drell-Yan processes, where the Z boson decays to two electrons. Smaller contributions from $t\bar{t}$ processes are also present, where each top quark in the pair decay via a W boson. These decays are divided into a dilepton mode, where both W bosons decay to a lepton and associated neutrino, and a semi-leptonic mode where one W boson decays hadronically, with the corresponding jet misidentified as an electron. Since the electron misidentifaction rate is low, semi-leptonic $t\bar{t}$ decays account only for a very small fraction of background events. Finally, processes where a Z($\rightarrow$ ee) boson is produced via vector boson fusion are also considered. The MEs for DY and electroweak Z $\rightarrow$ ee processes are calculated using MADGRAPH5_aMC@NLO, while for $t\bar{t}$ events, the POWHEG v2.0 [83–86] generator is used. The simulation of hadronisation and parton showering processes for all background samples is performed with the PYTHIA v8.230 package. Note that simulated background events are used only to train event classifiers; the final background models used to produce the results of the analysis are instead constructed using a data-driven approach.

Finally, the response of the CMS detector is simulated using the GEANT4 [87] package, which models the interactions of particles with the detector material. This includes simulation of pileup interactions, which are re-weighted to reproduce the average number observed in data, and mixed with the hard scattering event.

5.3 Particle flow

The global event description at CMS combines measurements from multiple detector components to reconstruct the properties of candidate objects. This holistic approach, known as particle flow reconstruction [88], makes use of the fine spatial granularity of the CMS apparatus, allowing for efficient identification of all object types, and energy measurements driven by the subdetector with the best resolution. Although the particle flow reconstruction is applied offline, a simplified version optimised for processing speed is run within the HLT. Inputs to PF reconstruction are tracks from charged particles in the inner tracker and muon stations, as well as calorimeter energy deposits.

5.3.1 Tracking

For the inner tracker, a generic set of PF tracks are reconstructed using a multi-step iterative procedure [89], referred to as the Combinatorial Track Finder (CTF):

1. Seeds are generated from up to two or three hits in the pixel tracker material, which define the initial parameters of the particle trajectory, including their uncertainties. A seed must pass a minimum p_T threshold and be consistent with originating from the IP.

2. The seed trajectories are extrapolated along their expected flight path. The objective of this step is to find additional hits that can be assigned to the track candidate. A Kalman filter [90] is then used within a track-fitting module to better estimate the parameters of the particle trajectory. This takes into account caveats that are ignored during the previous track finding step, including the non-uniform magnetic field and Coulomb scattering.

3. A selection is applied aiming to remove tracks faking charged particles. This consists of track quality metrics including the number of hits, the χ^2/n_{dof} for the track fit which provides a measure of the goodness-of-fit (where n_{dof} is the number of degrees of freedom), and compatibility with the primary vertex, determined from the track impact parameters.

Several repetitions of the above sequence are performed. Initial iterations search for tracks that are easiest to find, typically with large p_T and produced near the hard scattering vertex. After each iteration, hits associated with the (higher purity) reconstructed tracks are removed, reducing the combinatorial complexity of the algorithm. This allows for new seeds to be formed using relaxed quality criterea, increasing the track reconstruction efficiency without degrading the purity. The latter CTF iterations, seeded by the lower quality hits, search for more challenging classes of tracks, such as those produced with low p_T or with only two pixel seeds.

Primary vertex location

Location of the primary interaction vertex corresponding to the hard scattering event is determined using information from the tracker [91]. The assignment of this vertex is critical in mitigating the impact of pileup interactions, which typically produce many vertices with low associated p_T. For each candidate vertex, all compatible tracks are clustered with the anti-k_t [92] algorithm into pseudo-jets, as described in Section 5.4.4. The sum of the squared transverse energy, $\sum p_T^2$, is then evaluated for each pseudo-jet

and combined with the p_T^2 of any remaining associated single tracks. Finally, the squared missing transverse momentum associated with the vertex is added to the sum, to account for contributions from neutral particles. The candidate vertex with the largest $\sum p_T^2$ is chosen as the primary vertex.

In general, if the location of the primary vertex is correctly determined within 1 cm of the true vertex, the dielectron mass resolution is dominated by the electron energy resolution. For electrons considered in the search for H $\rightarrow$ ee decays, the primary vertex resolution in the z-direction varies between approximately 16 and 100 μm in 2016, and between 12 and 70 μm in 2017 and 2018. Therefore, the uncertainty on the electron energy dominates the overall uncertainty on m_{ee}. The efficiency of reconstructing the primary vertex for H $\rightarrow$ ee decays is close to unity.

5.3.2 Calorimeter clustering

In the calorimeters, inputs to the PF algorithm use energy deposits collected together to form clusters. This is particularly important for electrons and photons, which have a significant probability of showering when traversing the tracker. The resulting ECAL deposits must be combined into a single (super)cluster that captures the energy of the original electron/photon. The clustering algorithm is performed separately in each calorimeter subdetector, excluding the HF. The procedure for clustering in the ECAL [78, 88] is described here, although the HCAL procedure is similar:

1. Seeds for clustering are identified as calorimeter cells with a minimum energy threshold of 0.18 and 1 GeV for seeds in the endcap and barrel regions respectively. Seed cells are also required to be larger in energy than neighbouring cells (those that share an edge with the seed).

2. So-called *topological clusters* are built iteratively by aggregating cells that share a corner or edge in common with the growing cluster, provided the candidate cell has energy greater than twice the noise level. For crystals that could belong to multiple clusters, the energy is shared assuming a Gaussian shower profile. A seed cluster is then defined as the one containing most of the energy deposited.

3. Topological clusters within a geometric window around the seed cluster are grouped into so-called superclusters. The supercluster energy is defined previously in Eqn 4.3, while its position is given by the p_T weighted values of its constituent clusters.

The resulting SCs are then assessed for compatibility with reconstructed tracks.

5.4 Physics objects

The PF elements generated in each subdetector are connected to form physics objects for analysis using the so-called *link algorithm*. The algorithm generates links between the basic PF elements, defined using some distance metric that quantifies the link quality. For example, a track in the inner tracker may be linked to a cluster if its extrapolated position is within the cluster area. Here, the quality of the matching is determined by the impact parameter between the cluster position and extrapolated track. Other links are also developed, including cluster-to-cluster associations, connections between bremsstrahlung photons and ECAL energy clusters, and links between tracks in the muon and inner tracking systems. The resulting physics objects are described below.

5.4.1 Electrons

Electron candidates comprise reconstructed clusters linked to tracks in the inner tracker. However, dedicated modifications to the basic PF inputs are required to account for bremsstrahlung energy losses and subsequent photon conversions to electrons. In addition, a suite of corrections designed to calibrate electron energy measurements and improve their resolution are derived.

Tracking

Although CTF tracks can be built for any charged particle trajectory, electrons that loose large amounts of energy through bremsstrahlung are dealt with poorly by the algorithm, resulting in a degradation in tracking efficiency and resolution. To remedy this issue, an approach using a modified version of the Kalman filter, known as the Gaussian-sum filter (GSF), is used [93]. This approach accommodates large and abrupt changes in momentum along the track trajectory associated with bremsstrahlung losses. Similarly to the standard algorithm, a final set of track quality selection are applied, including thresholds on the number of hits, the χ^2 for the GSF track, and the energy lost along the trajectory. The GSF tracks provide new estimates of electron track parameters which are extrapolated towards the ECAL for association with clusters.

To associate tracks with clusters, a BDT-based classifier is used. Inputs to the BDT include kinematic and quality-related information from the GSF track, SC descriptions including lateral and longitudinal shower shape information, and track-cluster spatial compatibility measures such as the η and ϕ co-ordinates of the extrapolated track at the position of closest approach to the SC.

Bremsstrahlung recovery and photon conversions

Electron reconstruction must collect and associate any bremsstrahlung photons and electrons from subsequent photon conversions to the incident electron, in order to correctly estimate the particle energy. To this end, a dedicated energy recovery algorithm searches for photons produced tangentially to GSF tracks, that could be associated with ECAL clusters. The bremsstrahlung photon is associated with the electron object provided the extrapolated tangents are within a spatial margin of $\eta < 0.5$ from the cluster. To identify photon conversions, a dedicated BDT-based classifier [94] is employed, the inputs to which include missing track hits (which are not expected for prompt electrons), the radius of the first track hit, and the track impact parameter with respect to the interaction vertex.

Electron identification

Prompt electrons can be faked by various background sources that do not originate from the primary vertex. These include photon conversions, misidentified hadrons, and secondary electrons produced from leptonic decays of heavy quarks. To mitigate the incorrect identification of prompt electrons, a dedicated electron ID BDT is used, trained on simulated events containing $Z \rightarrow ee$ decays and hadronic jets. As input, the electron ID uses descriptions of the energy deposits in the ECAL, including shower shape and isolation quantities, tracker information such as the fraction of energy lost through bremsstrahlung, and information on the compatibility of measurements from the tracker and ECAL. Several BDTs are trained in independent bins of electron η and E_T to account for varying material budget and background composition respectively. For electrons in the barrel region, a background efficiency of approximately 2.5% is achieved for a signal efficiency of 95%; for electrons in the ECAL endcaps, the background efficiency is slightly worse, at $\approx$6% [78].

Electron energy scale and resolution corrections

The energy of an electron is obtained using information from both the tracker and ECAL, which are subject to losses from several sources. These include:

- Loss through bremsstrahlung photons that are not reconstructed in the ECAL SC, resulting in an underestimation of the electron energy.
- Leakage of SC energy into inactive regions of the detector, including gaps between modules and dead ECAL crystals.
- Poor shower containment for electrons that begin showering towards the back of the ECAL, resulting in energy leakage into the inner HCAL layers.

These losses result in systematic variations in the measured energy, which in turn degrade the energy resolution for EM objects. To mitigate these effects, multivariate techniques are applied to correct estimates of the energy scale and resolution in both simulation and data.

Firstly, the energy scale in simulation is corrected using the sequential application of three BDT-based regressions. The initial regression targets the ratio between the true electron energy, E_{true}, and reconstructed energy from the SC, E_{SC}. The regression parameterises the ratio $E_{\text{true}}/E_{\text{SC}}$ using a double-sided crystal ball function (DCB) that has a Gaussian core with power law tails on each side. The aim of the regression is to estimate the DCB parameters using a set of input features that describe the cluster profile. These include quantities such as the reconstructed energy and position of the SC, which provide information on the variation in energy containment as a function of detector geometry. The so-called R_9 variable, defined as the energy sum of the 3×3 crystal grid centred on the most energetic crystal in the EM cluster, divided by the total cluster energy, is also provided as a measure of local shower containment. Variables sensitive to clusters resulting from pileup interactions are also included, such as the energy density of the event, which help to mitigate overestimation of the SC energy. The correction applied to the SC energy is the most probable value of the $E_{\text{true}}/E_{\text{SC}}$ probability distribution returned by the regression.

A similar approach is used to estimate the per-electron energy resolution. This correction uses a second BDT-based regression using identical features, with the exception of the E_{SC} which is replaced by E_{corr}, the corrected electron SC energy, to predict the parameters of the $E_{\text{true}}/E_{\text{corr}}$ probability distribution.

This approach to correcting the SC energy and resolution is common to both electron and photon objects; however, for electrons, the final correction to the energy uses a third regression aiming to correct the electron energy estimate obtained from both the ECAL and the tracker. The target of this regression is the ratio between E_{true}, and a weighted combination of E_{corr} and p_{trk}, where p_{trk} is the momentum measurement of the track associated with the SC. Input features include the corrected SC energy, alongside measurements from the tracker such as the fractional energy loss, and p_{trk}. The most probable value of the predicted distribution is used as the correction to the electron energy.

After these corrections, small differences between data and simulation remain in the energy scale and resolution, for which further sets of corrections are derived. Firstly, the electron energy scale in data is shifted to correct for changes in the detector response due to degradation over time. These corrections are derived with a sample of DY $\text{Z} \rightarrow \text{ee}$ decays using the well measured Z boson mass, in bins roughly equivalent to one LHC

fill. Following this, a final set of granular corrections to the energy scale in data are also derived in bins of η and R_9. Using η information enables the correction of position dependent radiation damage, while including R_9 allows for isolation of electrons that undergo conversion in material prior to the ECAL. Lastly, the energy resolution in simulation is increased to match that in data using a Gaussian smearing. Corrections are derived in the same granular η and R_9 binning, and range from 0.1 to 1.5%. A similar approach is used to calibrate the energy scale and resolution of photons, where electrons from the Z boson decay are instead reconstructed without tracking information.

To validate corrections to the energy scale and resolution in the context of the $H \rightarrow ee$ analysis, the agreement of the dielectron mass distribution between data and simulation is compared for $Z \rightarrow ee$ events within a dedicated analysis control region. The selection defining this region is designed to be as similar to the analysis preselection (Section 6.2) as possible, whilst still remaining orthogonal. The requirements are therefore identical, with the exception of the dielectron mass selection, which is chosen to centre on the Z boson mass. The resulting dielectron mass distributions are shown in Figure 5.3, where the agreement between data and simulation is within the combined statistical and systematic uncertainties for all three years.

Final electron energy and resolution

The final electron energy is obtained from a combination of the ECAL energy, the GSF track momentum, and all associated energy deposits from bremsstrahlung photons and electron-positron pairs identified by the conversion finding algorithm. The energy resolution is measured using electrons from $Z \rightarrow ee$ decays. Appendix A shows the cumulative impact of the electron reconstruction process, including application of the electron energy scale and resolution corrections. The performance as a function of electron $|\eta|$ is shown for each year in the Run 2 campaign in Figure 5.4 [78, 95]. The resolution is extracted for all electrons inclusively, and independently for low bremsstrahlung electrons, defined by a minimum requirement on the R_9 quantity. In the latter regime, the momentum resolution for electrons is around 1.6% for those in the barrel, rising to values of around 5% in the endcaps, where the effect of an increased upstream material budget is more evident. Inclusively, the energy resolution varies between 2 to 4% in the barrel, and worsens to around 4 to 5% in the endcaps. The radiation damage sustained to ECAL crystals during the LHC operation results in a small worsening of the resolution with year; however, the overall performance during Run 2 is stable, despite the increased luminosity and material degradation.

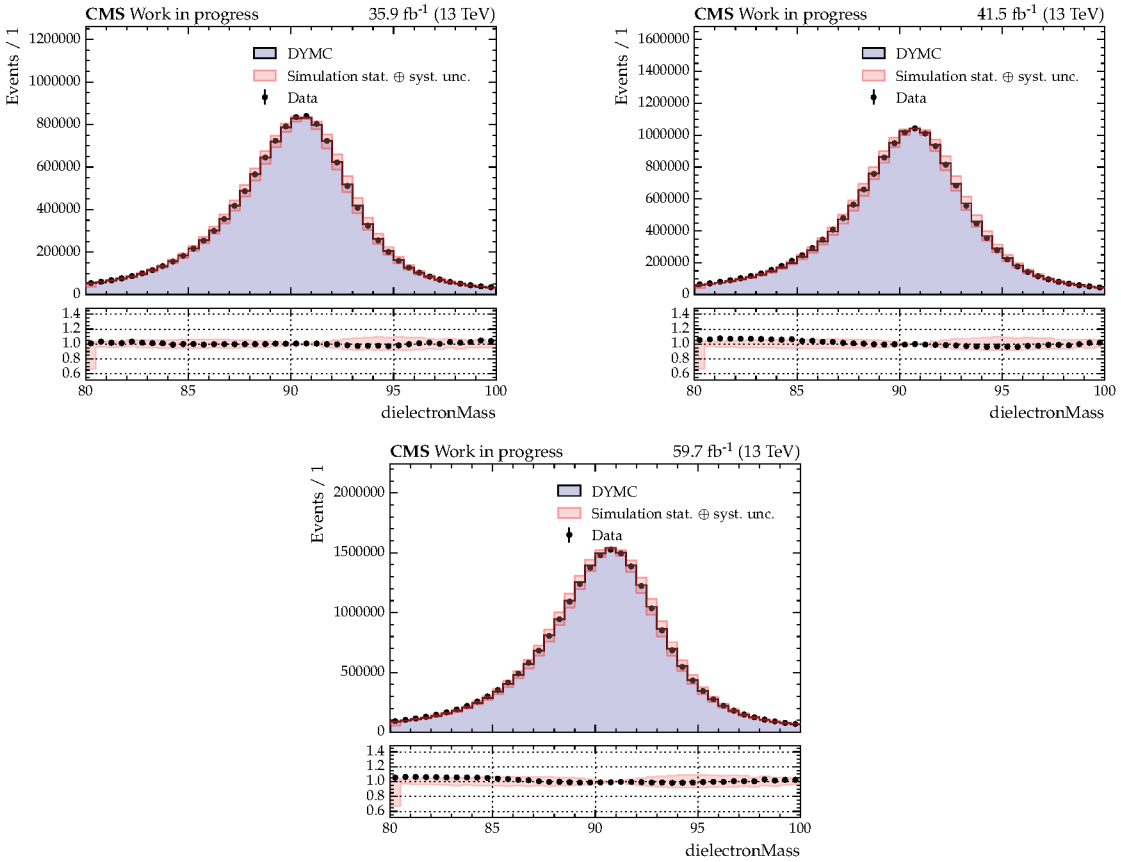

Figure 5.3: The dielectron invariant mass distribution for Z → ee events in the H → ee control region, compared between data and simulation following the electron energy scale and resolution corrections. The control region is defined identically to the analysis preselection, with the exception of the dielectron mass requirements, which are shifted to $80 < m_{ee} < 100$ GeV. The red band indicates the combination of statistical and systematic uncertainties for simulated Z → ee events, where the systematic component comprises the uncertainty on the electron energy scale corrections. The agreement between data and simulation is within the combined uncertainty from statistical and systematic sources, for each year independently.

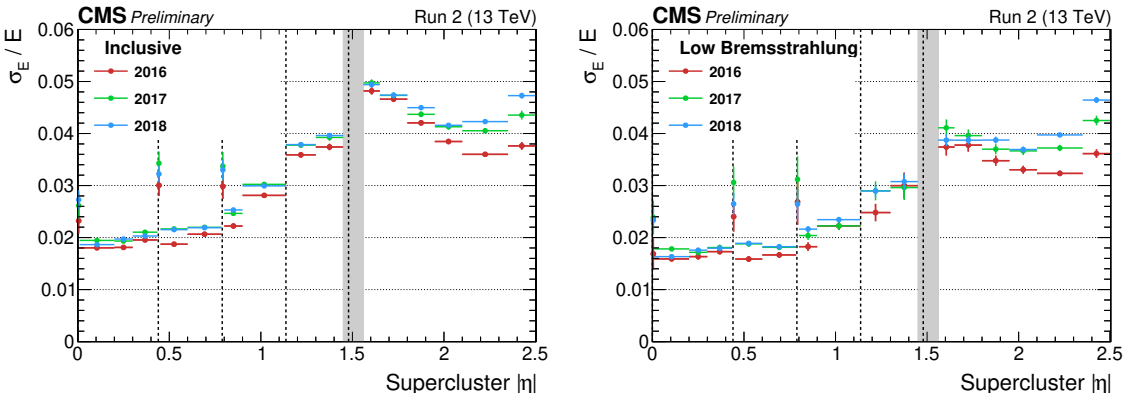

Figure 5.4: The relative energy resolution for electrons from Z → ee decays, for each year of data taking during Run 2 operation, shown as a function of electron $|\eta|$. The transition region between the ECAL barrel and endcap ($1.44 < |\eta| < 1.57$) is shaded grey. Left: the inclusive performance shown for all electrons. Right: the performance evaluated in a low electron bremsstrahlung regime, defined by $R_9 > 0.94$. The resolution is extracted from a likelihood fit to the m_{ee} distribution, using a Breit-Wigner convolved with a Gaussian function as the Z → ee signal model. Figures taken from Ref [95].

5.4.2 Photons

Photon reconstruction is performed separately for isolated and non-isolated photons. Since the basic properties and energy deposition patterns of isolated photons are similar to electrons, their reconstruction within the PF framework is performed together. These photons are formed from ECAL clusters with E_T above 10 GeV that are well isolated from tracks and calorimeter clusters, and have no link to a GSF track. A selection is also placed on quantities that describe photon energy deposit patterns, such as the ratio of energy clustered in the HCAL to that in the ECAL.

Non-isolated photons are typically found within hadronic jets, most notably from the $\pi^0 \to \gamma\gamma$ meson decay. These are formed using ECAL clusters not linked to tracks, provided the photon was reconstructed within the tracker acceptance. The preference for identifying ECAL deposits with photons over neutral hadrons is justified by the observation that in hadronic jets, 25% of the energy is carried by photons, while neutral hadrons only deposit 3% of their energy. Beyond the tracker acceptance however, the presence of charged hadrons adds additional ambiguity, and the preference given to photons is no longer justified. Therefore, reconstruction in this region asserts that ECAL clusters resulting from photon interactions are not linked to any HCAL clusters.

The final energy measurement is taken from the SC deposits in the calorimeter systems. This includes the energy of all associated electron pairs resulting from conversions (and associated bremsstrahlung photons), which is particularly important since the frac-

tion of conversions occurring before the last tracker layer can be as high as 60% for regions with the largest amount of tracker material in front of the ECAL [78]. Once the energy scale and resolution are corrected in the procedure described above, the energy resolution of photons is around 1–2.5% in the barrel, which degrades to approximately 2.5%–4% in the endcaps [94, 95].

5.4.3 Muons

Tracks resulting from muons are reconstructed independently in the inner tracker and muon systems [96]. Hits within DT or CSC detector systems are used to form track segments. The segments seed track fits which also use information from the RPCs, forming *standalone* muon tracks. Association of standalone tracks to those in the inner tracker results in one of two different muon types being reconstructed:

- *Tracker muon:* unique inner tracks are extrapolated to the muon system, and are compatible with one standalone muon track segment.
- *Global muon:* standalone muon tracks are matched to tracks in the inner tracker.

At low p_T, the reconstruction is driven by tracker muons, since such muons rarely penetrate multiple components in the muon system. Conversely, global muons drive the reconstruction of muons with high p_T, where the tracker-only fit is poor. The momentum resolution for muons is around 1% (3%) in the barrel (endcaps) for muons with $p_T < 100$ GeV. This resolution degrades significantly for high p_T muons; in the barrel region, muons with p_T up to 1 TeV achieve a momentum resolution of around 7%.

5.4.4 Hadrons

At this stage, only charged hadrons including $\pi^{\pm}$, $K^{\pm}$, and protons, and neutral hadrons such a K^0 and neutrons, remain to be identified. In practise, these are considered in the same processing step as non-isolated photons, described earlier. Within the tracker acceptance, all HCAL reconstructed clusters not linked to tracks are identified as neutral hadrons. The situation is complicated outside of the tracker acceptance where charged and neutral hadrons cannot be separated. Instead, ECAL clusters that are also linked to clusters in the HCAL are assigned as hadron showers with unknown charge, while those with ECAL clusters without this link are considered as photons.

Jets

Particles produced from the hadronisation of quarks or gluons are collected into collimated sprays known as jets. The clustering at CMS is performed using the anti-k_t [92] jet

finding algorithm, which groups PF candidates into conical clusters. Typical jet finding algorithms define a distance metric between constituent particles; in the anti-k_t algorithm, this metric is formed in an abstract space of momentum, rapidity, and azimuthal angle. Clustering starts with the two objects that are nearest in distance and repeats iteratively until the largest distance is between an object and the beampipe. At this point, the resulting jet is removed from consideration and the process may begin again.

The PF candidates considered during jet finding are subject to a charged hadron subtraction procedure (CHS), designed to reduce jet contributions resulting from pileup. The CHS procedure identifies the vertex for each charged PF candidate, removing those which are unambiguously associated with pileup vertices. Since the CHS procedure requires tracking information to identify PU vertices, only jets within $|\eta| < 2.5$ are subject to this correction. For jets within the tracker acceptance, CHS removes approximately 50% of pileup clusters.

Although CHS removes a significant fraction of PU, additional corrections are applied to specifically reduce PU contributions to jet energy and momenta. The correction factors are differential in quantities sensitive to pileup, including η, p_T, average event energy density, and jet area. Following these corrections, the jet energy scale and resolution is corrected in both data and simulation [97]. These corrections are derived in bins of p_T and η, resulting in agreement within 0.5% of the generator-level jet energy over a wide range of momenta. Finally, any residual differences between data and simulation are corrected using dijet and single jet events in samples of $Z(\rightarrow \mu\mu)$+jet, $Z(\rightarrow ee)$+jet, γ+jet, and multi-jet events. These corrections exploit the transverse momentum balance between the jet to be calibrated and a well-measured reference object, to correct for jet energy scales deviating from unity. Following these sets of corrections, the jet energy resolution ranges between approximately 15 to 20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [97].

5.4.5 Missing transverse momentum

The presence of particles that do not interact with the detector material, notably neutrinos, can be inferred from a p_T imbalance in an event, referred to as missing transverse momentum. This is defined as the negative vector sum of the transverse momenta of all PF candidates in an event, p_T^{miss}. The reconstruction of p_T^{miss} relies heavily on the hermeticity of the detector, and maintaining low misidentification and misreconstruction efficiencies for physics objects that could fake p_T^{miss}.

5.5 Summary

Particle reconstruction at the CMS detector optimally combines information from each subdetector, in a technique known as particle flow. This results in a set of physics objects available for analysis, including electrons, photons, hadrons, and muons. The reconstruction of many objects includes a set of dedicated corrections; for electrons, these are designed to improve agreement between simulation and data, and the dielectron mass (resolution), which is later fit to extract the analysis' results. These corrections, which are similar for electron and photon objects, proceed via a multi-step regression which uses descriptions of the supercluster profile and track-related quantities as inputs. The reconstruction of electrons also uses an identification criterion designed to reduce contamination from objects that could fake prompt electrons. Further selection is applied to other objects used within the $H \rightarrow ee$ search, including jets, where the requirements aim to reduce contamination from pileup clusters.

Chapter 6

Categorisation of H $\rightarrow$ ee events

6.1 Introduction

All events considered within the H $\rightarrow$ ee analysis are subject to categorisation in order to construct high purity analysis regions. Separate categories are developed to target both ggH and VBF Higgs boson production, the dominant modes at the LHC. Rarer production processes, such as in association with top quarks or a vector boson, are also considered in the analysis, although no dedicated event categories are constructed to target them since their contribution to the overall sensitivity of the search is small.

With an inclusive cross section of 48.6 pb, ggH production comprises the majority of SM Higgs boson events produced at the LHC, and thus is a natural target for a search with small expected signal yield. Although the cross section for VBF production is roughly an order of magnitude smaller than for ggH, the final state provides a unique signature in the detector whereby the dielectron from the Higgs boson decay is produced in association with two forward jets. These jets typically have a high combined invariant mass and are separated by a large pseudorapidity difference. These distinctive event properties enable a significant suppression of background events, allowing the sensitivity of categories targeting VBF events to become similar to those targeting ggH.

The categorisation strategy uses many quantities that characterise both VBF and ggH production as inputs to dedicated machine learning algorithms designed to reject background events (where the two electrons are produced by other SM processes) while keeping as many signal events (where two electrons originate from the Higgs boson decay) as possible. The chosen classification algorithm for both production modes are boosted decision trees, although studies using deep learning techniques are also presented. The goal of each classifier is to assign high scores to signal-like events, where score is interpreted as the probability for an event to be from a particular signal process. Analysis categories are then defined by a tight selection upon the output score of each classifier.

Since the search targets multiple Higgs boson production modes, a priority sequence must also be defined to ensure that all analysis categories are mutually exclusive. The sequence prioritises rarer production modes, and therefore would preferentially assign an event to VBF categories over ggH, if satisfying the requirements for both. A summary of the priority sequence for all analysis categories constructed is given in Section 6.7.

This chapter describes the event categorisation techniques, including the training, optimisation, evaluation, and interpretation of the machine learning-based classifiers. The modelling of signal events by the final classifiers is also validated.

6.2 Event preselection

A loose preselection defining the analysis signal region is applied to events to ensure they are consistent with a Higgs boson decaying to two electrons. All events must contain two electrons with opposite charge, fulfilling the following requirements:

- $p_T > 35$ (25) GeV for the leading (subleading) electron,
- 90% signal efficiency working point on the electron ID BDT,
- pseudorapidity within the ECAL acceptance ($|\eta| < 2.5$), and not in the barrel-endcap transition region ($1.44 < |\eta| < 1.57$), and
- dielectron mass between 110 and 150 GeV, chosen to ensure the m_{ee} sideband regions have sufficient events to constrain the background expectation in the signal region, and to limit contributions from Z → ee decays.

Jets entering the analysis are clustered using the anti-k_t [92] clustering algorithm with a distance parameter of 0.4. All jets are then subject to a selection designed to reduce contamination by jets from processes other than the hard scattering. Firstly, jets are required to have $p_T > 25$ GeV, $|\eta| < 4.7$, and to pass a tight PU identification criteria [98] which uses the jet shape, the number of charged and neutral jet constituents, and information on any associated tracks, to reject jets resulting from PU. A tight requirement is also placed on a second jet ID, which rejects spurious jets resulting from detector noise. Finally, additional selection is applied to suppress the observed noise in the ECAL endcaps for low p_T jets in 2017 data exceptionally. This selection vetoes jets with $p_T < 50$ GeV within the problematic region defined by $2.7 < |\eta| < 3.1$.

6.3 Categories targeting ggH events

The categorisation of ggH events is based on a boosted decision tree, referred to as the ggH BDT, trained to discriminate events produced via ggH production from all background processes combined. The largest contribution to the background in the ggH phase space consists of electron pairs from Drell-Yan production, the kinematics of which are typically similar to the ggH signal. The BDT output score is then used to define categories with differing S/B, in order to increase the sensitivity to ggH events.

6.3.1 Input features

The classification task for the ggH BDT is non-trivial and thus many features are leveraged to improve the separation power. These include kinematic properties of each electron from the Higgs boson decay, as well as properties of the composite dielectron object. In addition, descriptions of up to two jets are provided as inputs, since in approximately 40% of events passing the analysis preselection, the dielectron pair is produced in association with jets. Including descriptions of these jets allows the classifier to exploit possible differences in jet kinematics between the ggH signal and DY background, while also improving the selection of any residual VBF events entering ggH categories. A description of each of the 27 input features to the ggH BDT is given as follows:

- the transverse momentum of the dielectron system, $p_{T,ee}$;
- cosine of the angle between the two electrons in the transverse plane, $\cos(\Delta\phi_{ee})$;
- transverse momentum for the leading (subleading) electron, scaled by the dielectron invariant mass, $p_{T,e_1(e_2)}/m_{ee}$;
- pseudorapidity of each electron, $\eta_{e_1(e_2)}$;
- absolute value of the difference in pseudorapidity between the two leading jets, $|\Delta\eta(jj)|$;
- difference in angle between the two leading jets in the transverse plane, $\Delta\phi(jj)$;
- smallest ΔR between a single electron in the dielectron pair, and the dijet system, $\min(\Delta R(e, jj))$;
- invariant mass of the leading two jets, known as the dijet mass, m_{jj};
- absolute value of the difference in pseudorapidity between the dielectron and dijet systems, $|\Delta\eta(jj, ee)|$;

- difference in azimuthal angle between the dielectron and dijet systems, $\Delta\phi(jj, ee)$;
- the centrality variable [99], defined as: $\exp(-4(Z/|\eta_{j_1} - \eta_{j_2}|)^2)$, where Z is the Zeppenfeld variable, defined as $Z = |\eta_{ee} - \frac{1}{2}(\eta_{j_1} + \eta_{j_2})|$;
- difference in angle between the dielectron and leading (subleading) jet in the transverse plane, $\Delta\phi(j_{1,(2)}, ee)$;
- difference in pseudorapidity angle between the dielectron and leading (subleading) jet, $\Delta\eta(j_{1,(2)}, ee)$;
- the four vector of the leading and subleading jets, $(E_{j_1(j_2)}, p_{T,j_1(j_2)}, \eta_{j_1(j_2)}, \phi_{j_1(j_2)})$;
- the quark-gluon likelihood (QGL) ID score for the leading and subleading jets [100]. The ID uses quantities such as the p_T and multiplicity of the PF candidates reconstructed within a jet, in a likelihood discriminant to separate gluon and quark initiated jets.

The distributions for a selection of features are shown for simulated background, ggH signal, and data in Figure 6.1 (left). The agreement between data and simulated background events is reasonable; however, since this analysis uses data-driven background models, any non-closure in the input feature distributions, or BDT score, cannot bias the background modelling. Figure 6.1 (right) shows the same inputs normalised to unit area, allowing for comparison of the shape of each variable between simulated ggH signal and background events. Distributions which present large separation between the two classes are typically good predictors for the task. The normalised distributions show the dielectron p_T to be an important discriminating observable, where the p_T spectrum in ggH events is harder than the Drell-Yan background. A handful of inputs describing single jet and dijet kinematics also display reasonable separation power. Overall, however, the distributions are similar between the H → ee signal and background events, indicating a difficult separation task for the classifier to perform. Distributions for all inputs to the ggH BDT are provided in Appendix C.

6.3.2 Training and optimisation

To train the ggH BDT, simulated ggH events from all years passing the basic preselection are used. Simulated background events include Drell-Yan processes which account for approximately 92% of all background events passing preselection, alongside smaller contributions from $t\bar{t}$ decays. For both signal and background, events are weighted initially in accordance with their SM production cross section. The full sample of events are divided randomly into a training set, which comprises 70% of the original sample used to

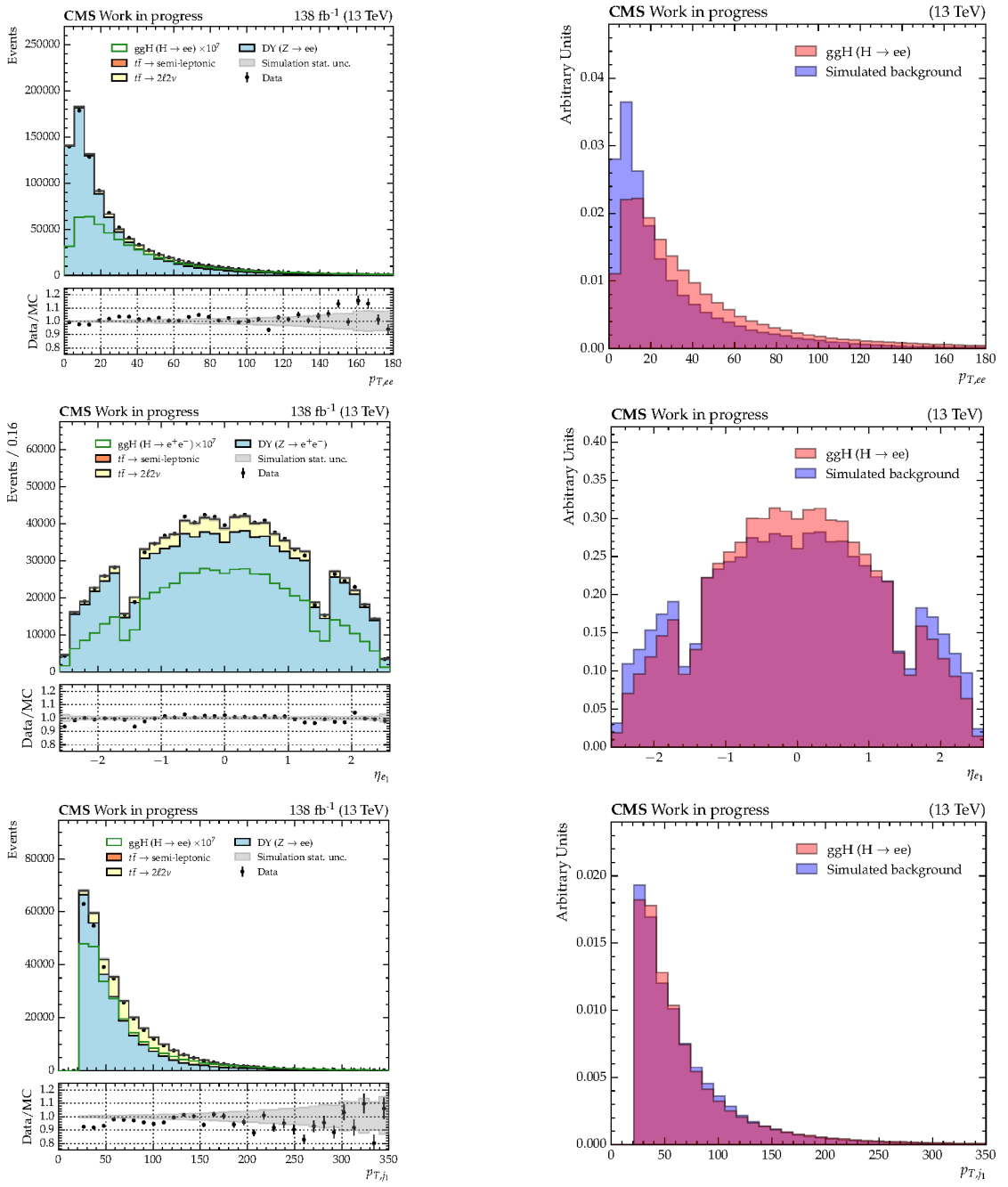

Figure 6.1: Distributions of selected input variables to the ggH BDT. From top to bottom row, these include the p_T of the dielectron object, the η position of the leading electron, and p_T of the leading jet. Left: feature distributions for simulated background processes (bold face), stacked for comparison with data (black points). Simulated ggH signal is shown in green, with the overall normalisation scaled such that it is visible. The ratio of data to simulated background is shown in the lower panel. Reasonable agreement is observed, with respect to the statistical uncertainty (grey band). Right: the same input variables, normalised to unit area, with ggH signal shown in red and simulated background, integrated over each processes, shown in blue. The majority of input features show similarity between the signal and background classes, with the exception of the dielectron p_T distribution which is harder for simulated signal, and the η position of the leading electron, which tends to be more central.

train the BDT, and a testing set, which contains the remaining 30% of events withheld to evaluate the performance of the final model. Separating the data in this way ensures that any predictions made on the test set remain unbiased to events used in the classifier training and optimisation processes, as discussed in Section 3.2.2.

The BDT is implemented using the XGBoost [74] python library. The implementation has multiple hyperparameters that may be tuned to improve both performance and convergence time during training. For the ggH BDT, the main hyperparameters that govern the model complexity are optimised, including the learning rate and maximum depth of ensembled learners. Larger values of the learning rate result in quicker descent of the loss manifold, but produce models that may never converge to the global minimum. The maximum depth refers to the number of successive splits permitted for all decision trees in the ensemble. Increasing the tree depth allows the algorithm to learn a more complex function of the input feature set, which may offer increased separation power. However, increasing the depth arbitrarily may result in overfitting to statistical fluctuations in the training set.

Since the underlying function mapping hyperparameters to performance is unknown, gradient-based methods cannot be used to find optimal configurations. Therefore, this analysis uses a 3-fold cross-validation technique [50] to draw conclusions on the best model hyperparameters. This is a derivative-free method where each configuration is sampled exhaustively from a grid of all possibilities. A typical k-fold cross validation strategy partitions the training set into k intervals, or "folds". The model is trained on all but one of the folds, with the remaining fold used to quantify the model performance. The process is then repeated, with a different fold being held for testing purposes and the remaining used for training. The final performance of the model is quantified by averaging the performance over all k trainings. The hyperparameter search space considered for all BDTs trained in this search is given Table 6.1.

Table 6.1: The hyperparameter search space for both the ggH and VBF BDT. The final choice of parameters are chosen to maximise the model performance on a withheld validation set. The performance is largely similar across the possible hyperparameter configurations; hence the nominal values, highlighted in bold font, are used in the final models.

Model hyperparameter	Search space
Learning rate	0.01, 0.05, 0.1, **0.3**
maximum tree depth	3, 4, 5, **6**, 7, 8, 9, 10
regularisation coefficient (γ)	**0**–5
events subsample for learners	0.5, 0.8, **1.0**
number of learners	**100**, 200, 300, 400, 500

Another parameter that can be changed is the relative weight of the signal and background samples. With the default sample weights, which correspond to the expected

number of events, the two classes are extremely imbalanced, resulting in the classifier predicting the background class for almost all events. To mitigate this, the BDT is trained in a scenario where the signal event weights are increased by a uniform factor such that the total sum of weights for signal and background processes is equal. This serves purely as a technical change to the training, designed to improve the learning outcomes of the classifier — when evaluating the performance, the nominal weights are used. The reweighting forces the algorithm to consider each class as equally important and results in dramatically improved performance. In addition, given that the analysis sensitivity depends on the dielectron mass resolution, a second reweighting is applied to ggH events during training to make the classifier aware of this fact. Events are weighted by $1/\sigma(m_{ee})$, where $\sigma(m_{ee})$ is the per-event mass resolution. This strategy is preferred to simply adding $\sigma(m_{ee})$ as an input feature, since this distribution itself holds little separation power between signal and background processes.

Finally, since the dataset analysed is composed of three individual years of data taking, an alternative strategy where a single classifier is trained for each year independently is explored. This strategy could allow the classifier to take into account any per-year differences between simulated samples. However, splitting the training data more finely reduces the size of the dataset available to learn from, potentially worsening the overall performances. Two studies are performed to look for possible improvement by training independently: firstly, it is verified that including a quantity encoding the year of simulation to the input feature set does not add any discrimination power. Secondly, it is confirmed that the performance when training three separate classifiers is consistent between years. Since the results of these studies indicate little difference between simulation in each year, the classifier is trained on the combination of all simulated samples.

6.3.3 Performance evaluation

The performance of the classifier is evaluated using the area under the ROC curve. The curve is generated by placing sequentially tighter selection on the ggH BDT output score, which is shown for both data and simulation in Figure 6.2 (left), and computing the associated true and false positive rates. The output score distributions for ggH signal and simulated background are not well separated, which is expected from the similarity of the event kinematics. This is reflected in the ROC curve, shown in Figure 6.2 (right), where the associated AUC score is relatively small, evaluating on the test set to 0.675. The metric is also computed on the training set and compared to confirm no significant overfitting.

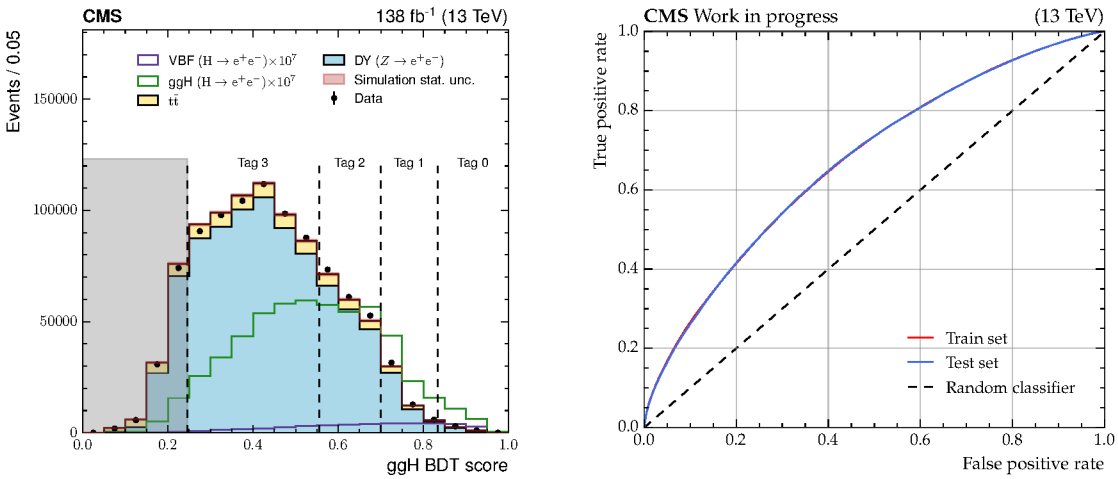

Figure 6.2: Left: output score of the ggH BDT for simulated signal events (green and purple histograms), background processes (bold face), and data (black markers). The distribution is similar between ggH signal events and the Drell-Yan background, reflective of the common features between event topologies. Category boundaries targeting ggH production are denoted with dashed lines. Events with scores in the low S/B grey shaded region are discarded from the analysis. Right: receiver operating characteristic curves for the ggH BDT, evaluated on the training and testing datasets, with only ggH events considered as signal. The area under the curves are similar for both sets, indicating no significant overtraining. The dashed line (black) provides a benchmark for a model assigning events to classes at random.

6.3.4 Model interpretation

Given that a typical gradient boosted decision tree may ensemble thousands of individual learners, each with their own set of decision nodes, it is often difficult to explain how a model arrives at a prediction. Therefore, in order to improve the so-called *explainability* of the ggH BDT, additional studies are performed to understand how each input feature is connected to the overall prediction.

Firstly, in order to understand how the ggH BDT score is correlated with each input, Figure 6.3 shows, for simulated ggH signal events, the distribution of selected features as a function of the BDT output score. Red and orange colours indicate events occupying a lower BDT score interval, associated with background-like predictions, while blue and green colours correspond to signal-like events. The BDT assigns high scores to centrally produced electrons forming a high $p_{T,ee}$ pair, features which are characteristic of the ggH event topology. In addition, a significant fraction of events (with two or more jets) receiving a high BDT score are characterised by a large ΔR separation between the dijet system the closest electron in the dielectron pair. The classifier is also aware of the dielectron mass resolution, placing events with smaller $\sigma(m_{ee})$ at higher scores, as seen in Figure 6.4 (left). This effect is partially generated by the reweighting of signal events applied in training by $1/\sigma(m_{ee})$. Lastly, it is checked that the classifier cannot infer the value of the Higgs boson mass from the set of input features, which leads to undesirable sculpting of the m_{ee} distribution. To illustrate this check, Figure 6.4 (centre) shows the m_{ee} spectrum for simulated background events as a function of the BDT output score. The distribution remains smoothly falling in each BDT score interval, and thus is highly uncorrelated with the dielectron mass.

It is also useful to understand which features are good predictors for the classification task, or in other words, how important each one is when generating a prediction. The importance of each input to the ggH BDT is ranked using Shapley values, introduced in Section 4.4.5, with the mean of the absolute Shapley values taken as the figure of merit. Figure 6.5 shows the Shapley values for all simulated events used to train the classifier, for a selection of the most predictive features. It can be seen that events with smaller dielectron p_T values cause a large shift in predictions towards smaller (background-like) values. When evaluating the figure of merit, the dielectron p_T scores highly, as expected, alongside the leading jet QGL ID score and subleading jet p_T.

The high importance of jet descriptions may not be forseen when considering, for example, the individual jet p_T distributions, given that these alone hold little separation power. It is therefore hypothesised that these quantities are used by the BDT to construct some other useful feature. For example, the jet multiplicity could be inferred from the set of default values for jet p_T when a jet is not present, which are set far from the nominal

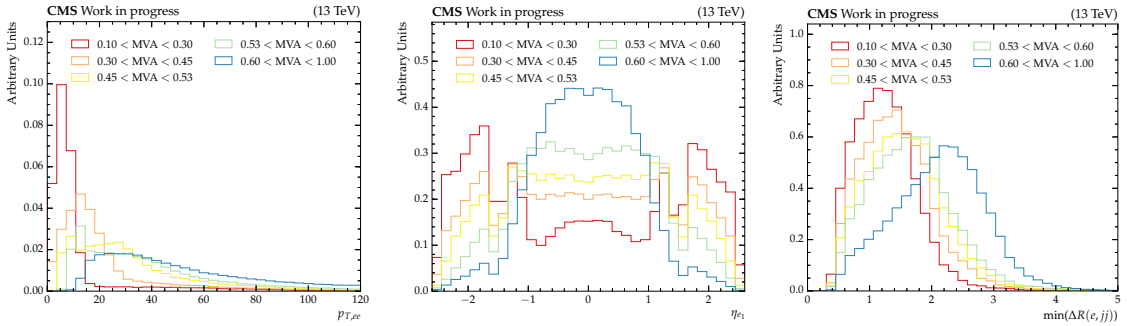

Figure 6.3: Distributions for selected input variables to the ggH BDT for simulated ggH signal events, as a function of the ggH BDT output score. From left to right, these include the dielectron p_T distribution, the η position for the leading electron, and the smallest ΔR between a single electron in the dielectron pair and the dijet system. Each histogram is normalised to unit area. Blue and green colours indicate signal-like predictions, which are characterised by an electron pair with high p_T, produced centrally in the detector. Events with two jets are also assigned high scores if the dijet pair is well separated in R from the closest electron in the dielectron object.

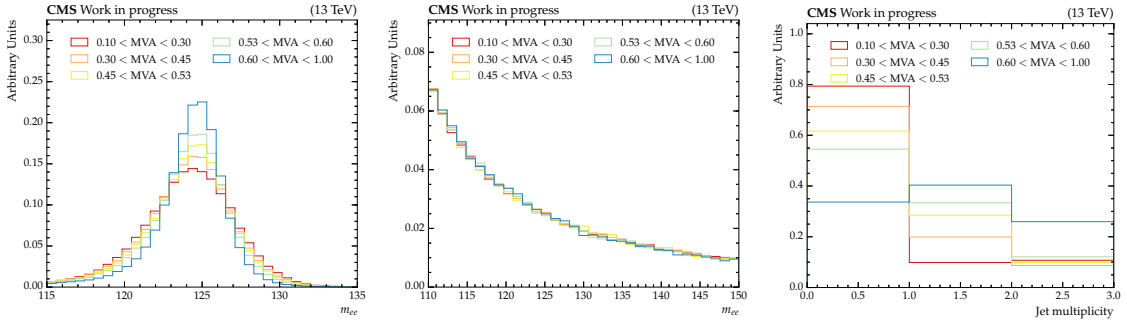

Figure 6.4: Left: the dielectron mass distribution for simulated ggH signal events as a function of ggH BDT score, normalised to unit area. The classifier assigns higher scores to events with a good mass resolution. Centre: the dielectron mass distribution in simulated background events, as a function of ggH BDT score, normalised to unit area. The distribution is clearly uncorrelated with the BDT output score. Right: the jet multiplicity in various intervals of the BDT score, for simulated ggH signal events, normalised to unit area. This variable is not explicitly included in the feature set, yet the distribution evolves as a function of the output score. The BDT has therefore learned the distribution implicitly from other jet descriptions.

range. To verify this, it is checked that the jet multiplicity for simulated signal events changes with increasing ggH BDT score, which is confirmed in Figure 6.4 (right). It is also verified that adding the jet multiplicity to the set of input features brings no improvement in performance, since the classifier already has access to this information implicitly. This partly illustrates the benefit of using lower-level event information when constructing a feature set, a concept discussed in more detail in Section 6.5.

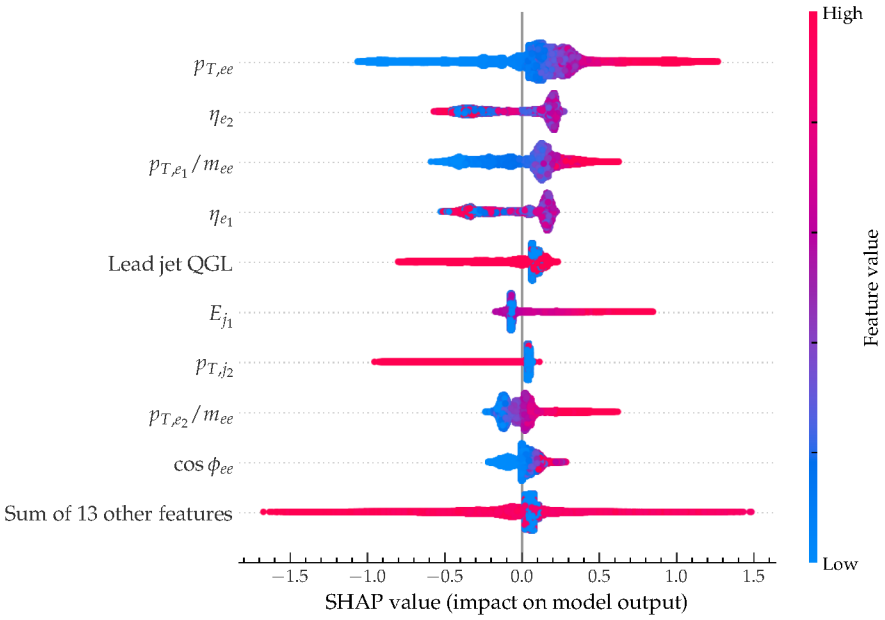

Figure 6.5: The distribution of Shapley values (filled circles) for selected input features to the ggH BDT, shown at the per-event level. Positive Shapley values indicate a preference towards signal-like predictions, while negative values are associated with background like predictions. Larger values of the input variable are coloured in red hues, while smaller values are shown by blue hues. Features are ordered by descending importance, measured by taking the mean over the magnitude of all Shapley values. Electron kinematics score highly, including the dielectron p_T, alongside various jet descriptions.

6.3.5 Category definitions

Categories targeting ggH events are defined using the output score of the ggH BDT. To construct these categories, an optimisation procedure is performed where the score distribution is divided into multiple regions, ordered by S/B ratio. The Approximate Median Significance [101] (AMS) is used as the figure of merit for the optimisation, defined as:

$$\text{AMS} = \sqrt{2\left((S+B)\ln\left(1+\frac{S}{B}\right)-S\right)}, \tag{6.1}$$

where S is the number of signal events within the window $m_H \pm \sigma_{\text{eff}}$, with σ_{eff} defined as the smallest interval containing 68.3% of the signal distribution, and B is the background yield. The value of B is calculated by first performing an exponential fit to the m_{ee} distribution in background, and then summing the number of events in the same integration region as the signal. Both yields are determined only for events passing the BDT selection being tested. The value of the AMS metric corresponds to the expected significance of a signal from the likelihood ratio test statistic for a simple counting experiment, neglecting the impact of systematic uncertainties. The expression reduces to the familiar $S/\sqrt{S+B}$ in the limit of small S/B.

The placement of each category boundary is optimised using a random search over the parameter space. This strategy is found to be more efficient than an exhaustive grid search, particularly when the number of categories is high. The final set of boundaries are chosen such that the AMS, combined in quadrature over categories, is maximised. It is found that four categories targeting ggH events provides optimal sensitivity; no significant improvement is found when using more categories. The exact placement of each boundary is shown by the dashed lines on the output score in Figure 6.2 (left). Table 6.2 details the expected signal and background yields, the fractional contribution from each signal process, and expected significance, for each ggH category. Finally, the observed dielectron mass distribution in each ggH analysis category is shown for simulated ggH signal and background events, and data, in Figure 6.6.

Table 6.2: Table showing the total expected number of signal events for $m_H = 125.38$ GeV, the current best measurement of the Higgs boson mass [102], in analysis categories targeting ggH and VBF events, for an integrated luminosity of 138 fb^{-1}. The fractional contribution from each production mode to each category is also shown. The σ_{eff}, defined as the smallest interval containing 68.3% of the m_{ee} distribution, is listed for each analysis category. The final column shows the expected ratio of signal to signal-plus-background, S/(S+B), where S and B are the numbers of expected signal and background events in a $\pm\ \sigma_{\text{eff}}$ window centred on m_H.

Analysis category	Signal yield ($\times 10^{-4}$)	Production mode fractions				σ_{eff} (GeV)	Bkg per GeV	S/(S+B)
		ggH	VBF	VH	ttH			
ggH Tag 0	4.7	81%	12.1%	5.0%	1.9%	1.65	250	7.7×10^{-7}
ggH Tag 1	18	86.5%	8.4%	4.2%	0.9%	1.84	2340	2.8×10^{-7}
ggH Tag 2	35	91.7%	4.8%	3.1%	0.5%	2.02	8760	1.4×10^{-7}
ggH Tag 3	61	92.6%	3.5%	3.2%	0.7%	2.52	30500	5.4×10^{-8}
VBF Tag 0	2.2	19.8%	80.0%	0.1%	0.1%	2.04	25.9	2.8×10^{-6}
VBF Tag 1	1.7	40.3%	58.5%	0.8%	0.4%	2.26	81.7	6.3×10^{-7}

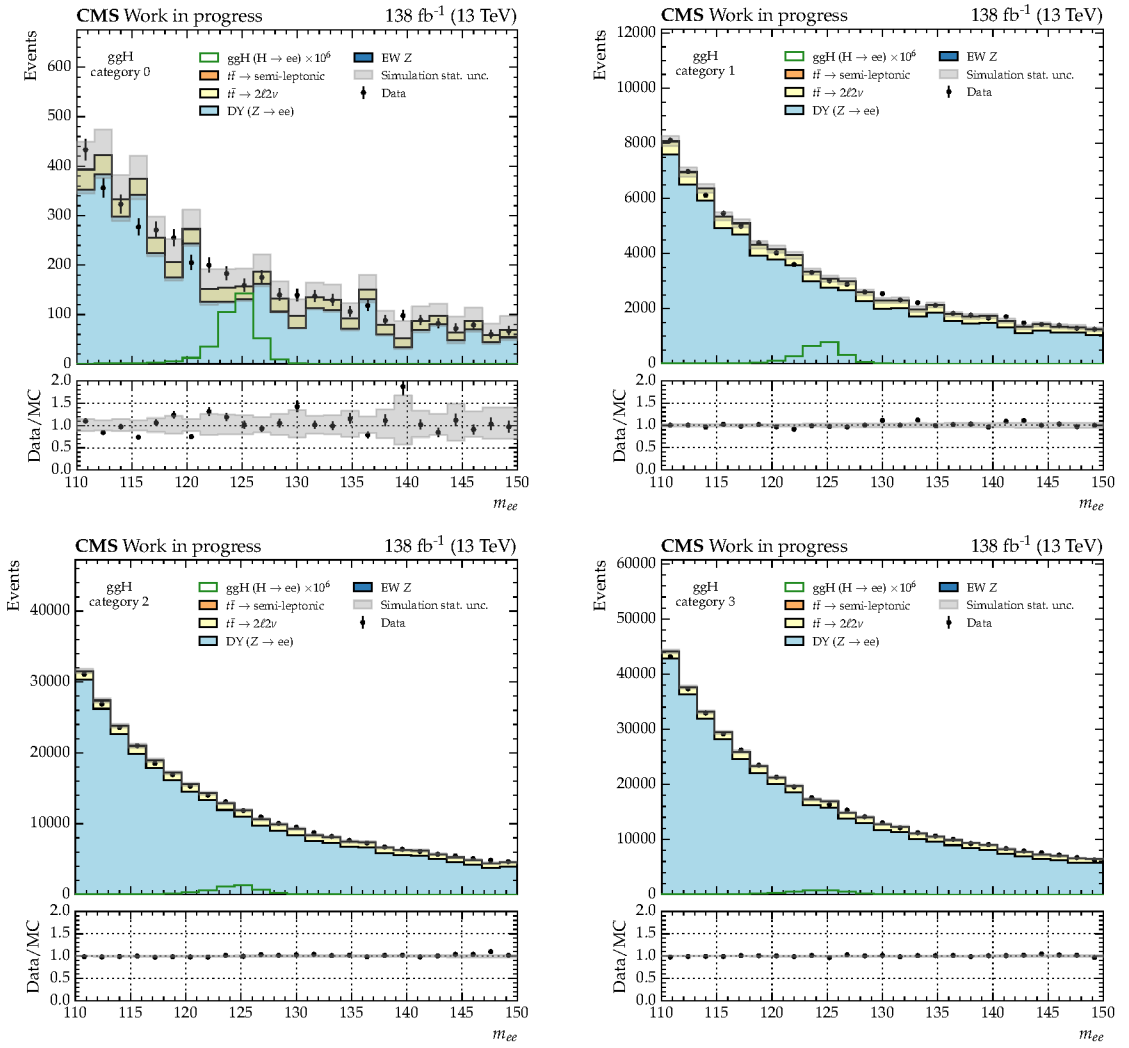

Figure 6.6: Observed m_{ee} distributions for the analysis categories targeting ggH events, shown for simulated background (bold face) and data (black points). The statistical uncertainty on the simulated background is shown by the grey shaded band. Simulated ggH signal events are also shown (green), scaled such that they are visible. The ratio of data to the total simulated background is shown in the lower panels.

6.4 Categories targeting VBF events

The nominal categorisation strategy used to target VBF events is similar to that used for ggH. Analysis regions are defined using the output score of a BDT trained to discriminate between VBF signal and background events, referred to as the VBF BDT. Alternative categorisation strategies using deep learning techniques are also presented in Section 6.5.

The background composition in this channel is significantly different than for the ggH BDT, comprising roughly equal numbers of Drell-Yan and $t\bar{t}$ events. Additionally, electroweak Z boson production, where the two electrons resulting from the Z decay are produced in association with two quark-initiated jets, are also considered. Although these events account for less than 2% of the total background passing VBF preselection, their kinematics and event topology are similar to the VBF signal, making them more difficult to separate compared with the dominant backgrounds.

The VBF categories are defined firstly by applying additional requirements on top of the nominal analysis preselection, to favour events with VBF-like topology. This is referred to as the VBF preselection and comprises the following requirements:

- transverse momentum of the leading (subleading) jet > 40 (25) GeV, and
- dijet invariant mass greater than 350 GeV.

All simulated signal and background events passing the VBF preselection are used to train the VBF BDT. It is checked that including residual ggH events that pass the VBF preselection does not improve the resulting sensitivity of the VBF analysis categories (and vice-verse when training the ggH BDT).

6.4.1 Input features

The set of input features provided to the VBF BDT include kinematics of the individual electrons, dielectron system, and hadronic jets. The feature set is similar to the ggH BDT, with the exception that descriptions of a third leading jet are also included. This results in the addition of five new variables: the four vector and QGL ID score for the third leading jet. The distribution of selected input features to the VBF BDT with good prediction power are shown in Figure 6.7 (left) for simulated VBF signal and background events, and data. Also shown in Figure 6.7 (right) are the simulated distributions normalised to unit area, allowing for shape comparison between the signal and background events. In comparison to the ggH BDT, many inputs display good separation power; VBF events typically include a dijet pair with high invariant mass and large opening angle in η between the two constituent jets, features which allow for significant suppression of background events. Additionally, considerable separation of events are observed in the centrality

distribution, where the VBF signal typically peaks at values close to one, while background events occupy values nearer to zero. Distributions for all inputs to the VBF BDT are provided in Appendix C.

6.4.2 Training and optimisation

The training and optimisation procedure for the VBF BDT is identical to the ggH BDT and hence will not be repeated in detail. The main hyperparameters of the model are optimised through a 3-fold cross-validation, with approximately ten thousand possible combinations considered. Simulated signal samples are reweighted by $1/\sigma(m_{ee})$ to encourage high resolution events to occupy signal-like BDT scores. Finally, the sum of signal and background event weights are equalised during training in order to mitigate the large class imbalance.

6.4.3 Performance evaluation

Similarly to the ggH BDT, the performance of the VBF BDT is evaluated using the area under the ROC curve. This metric is computed on both the training and testing sets, and compared as an overtraining check. The output score of the VBF BDT used to construct the ROC curve is shown for simulated VBF signal, background, and data events in Figure 6.8 (left). Unlike the ggH BDT output score distribution, good separation between the signal and background classes is observed. In general, the classifier is able to reject DY and $t\bar{t}$ events, which comprise the majority of the background yield, with high efficiency. Conversely, background events resulting from electroweak Z boson production, which exhibit signal-like topology, are harder to discriminate. These events account for up to 20% of the total background in the most signal-like VBF BDT score bins. The ROC curves for the training and testing sets are shown in Figure 6.8 (right), with corresponding AUC values of 0.928 and 0.925 respectively.

6.4.4 Model interpretation

As discussed in Section 6.3, for any complex classifier architecture, it is useful to investigate how the feature set is connected with the output predictions. To understand how the inputs to the VBF BDT are correlated with the output score distribution, Figure 6.9 shows a selection of the most predictive input features as a function of BDT score. Events associated with background-like predictions are shown in red and orange hues, while signal-like predictions comprising events predicted with high score are shown in green and blue. As expected, the VBF BDT assigns high scores to events characterised by a dijet pair with large invariant mass and angular separation of jets. Also shown is the

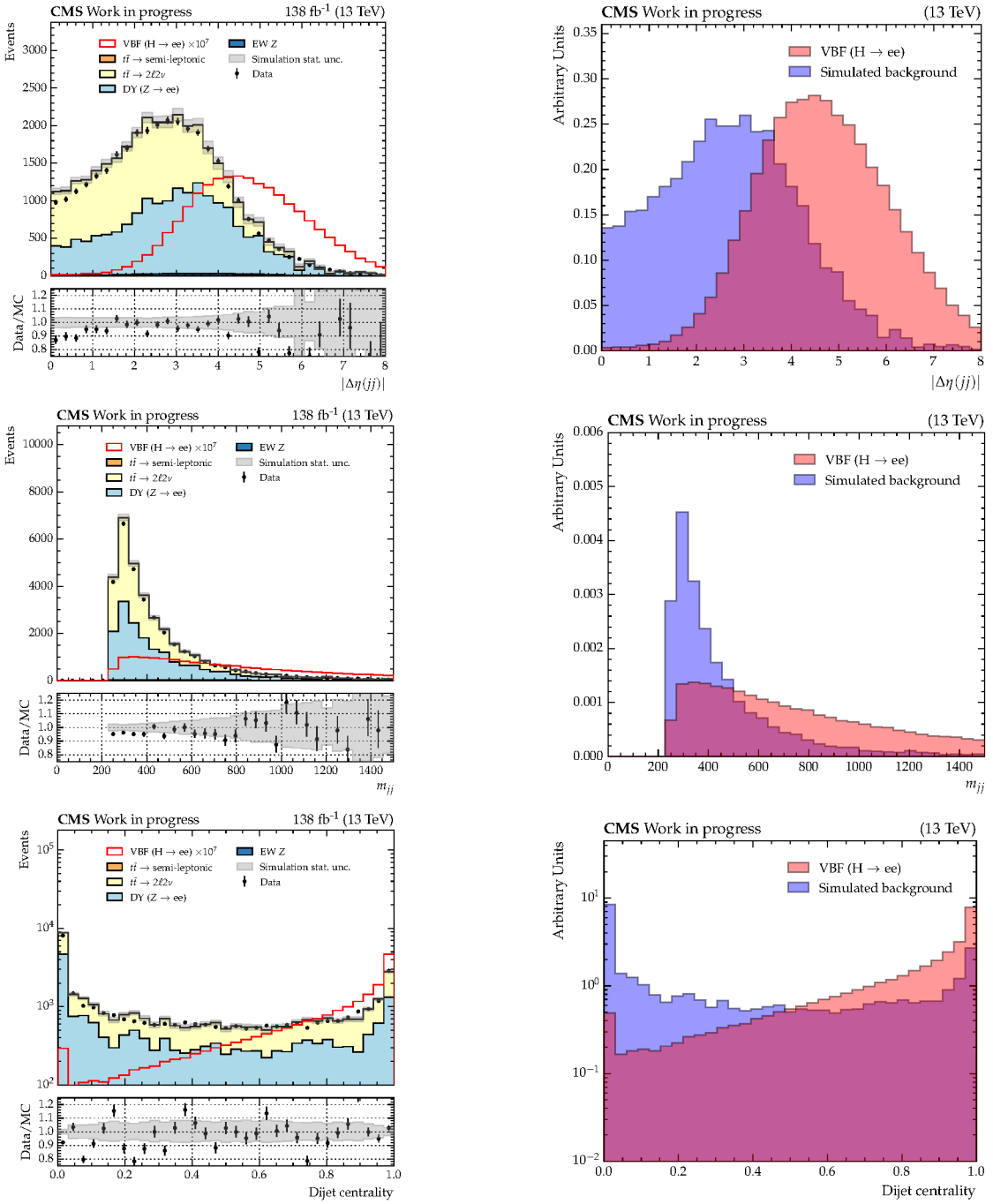

Figure 6.7: The distributions for selected input variables to the VBF BDT. From top to bottom row, these include the magnitude of the difference in η position between the leading two jets, the invariant mass of the system defined by the leading and subleading jets, and the centrality variable defined by the η positions of the two leading jets and dielectron system. Left: feature distributions for simulated background processes (bold face), stacked for comparison with data (black points). Simulated VBF signal is also shown (red histogram), with the overall normalisation scaled such that it is visible. Reasonable agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band). Right: the same input variables, normalised to unit area, with VBF signal shown in red, and simulated background integrated over each processes shown in blue. In general, the set of VBF input features display good separation between signal and background events.

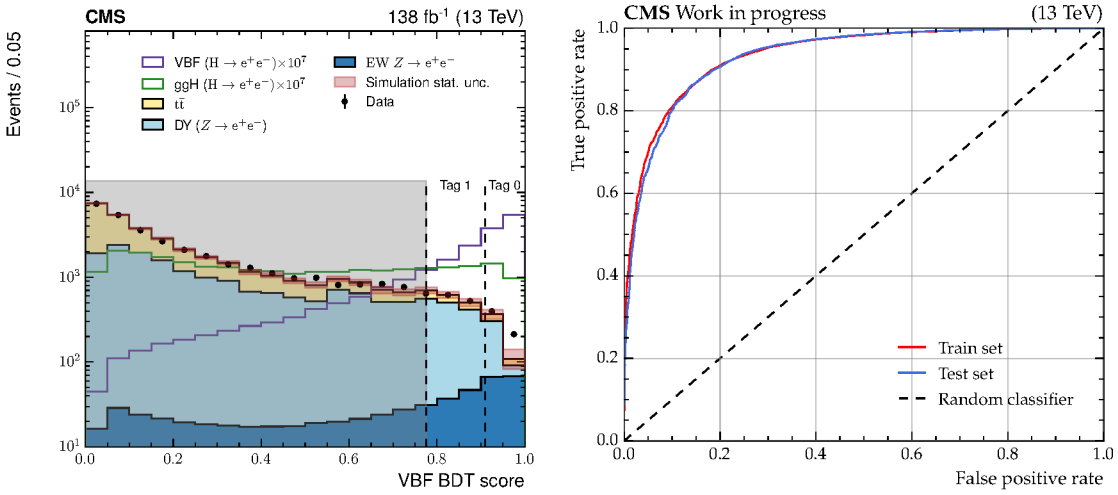

Figure 6.8: Left: the output score of the VBF BDT in simulated signal events (green and purple histograms), background (bold face), and data (black markers) events. Good separation between the signal, and Drell-Yan and $\mathrm{t\bar{t}}$ backgrounds is observed. Category boundaries targeting VBF production are denoted with dashed lines. Events with scores in the low S/B grey shaded region are not considered for VBF categories, but may populate categories targeting ggH production. Right: receiver operating characteristic curves for the VBF BDT, evaluated separately on the training and testing datasets, with only VBF events considered as signal. The area under the curves, which measure classifier performance, are comparable for both sets indicating negligible overtraining. The dashed line (black) provides a benchmark for a model that assigns events to classes at random.

QGL ID value for the leading jet, which is expected a priori to hold reasonable rejection power against DY events, where jets are typically produced via radiated gluons. This is confirmed in Figure 6.9, where lower QGL scores are associated with background-like predictions. Finally, similar distributions of the m_{ee} spectrum in simulated signal and background events are produced to check both that the classifier assigns high scores to events with good mass resolution, and cannot learn the value of the Higgs boson mass.

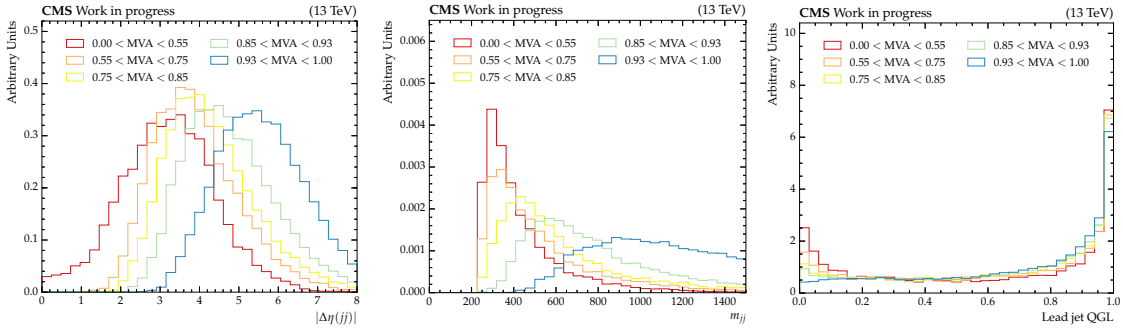

Figure 6.9: The distribution of selected inputs to the VBF BDT as a function of the VBF BDT score, in simulated VBF signal events. From left to right, these include the magnitude of the difference in η position between the two leading jets, the invariant mass of the system defined by the leading and subleading jets, and the QGL ID score of the leading jet. Events obtaining a signal-like prediction are drawn in blue and green colours, while background-like events are shown in red and orange. Events with high dijet mass and angular separation of the leading two jets are associated with higher scores. Background-like scores are assigned to events with low (quark-like) QGL ID score.

Similarly to the ggH BDT, it is useful to generate Shapley values for the VBF BDT input feature set to understand which are the most important. Figure 6.10 shows the distribution of Shapley values at the per-event level, for a selection of the most predictive inputs. Quantities that encode jet kinematics typically score highly, including the dijet centrality and other angular descriptions. The directional pulls for each feature are also intuitive — for example, higher values of centrality are typically associated with positive, signal-like predictions, which is expected from the distributions shown in Figure 6.7.

6.4.5 Category definitions

To construct the final VBF categories, a boundary optimisation procedure identical to that presented in Section 6.3 is followed, yielding two VBF categories. The resulting signal and background yields, fractional signal composition, and expected sensitivity are shown for each category in Table 6.2. The m_{ee} distributions for simulated background events and data are shown for both VBF categories in Figure 6.11. In this phase space, the statistical uncertainty for simulated Drell-Yan events becomes large, partially motivating

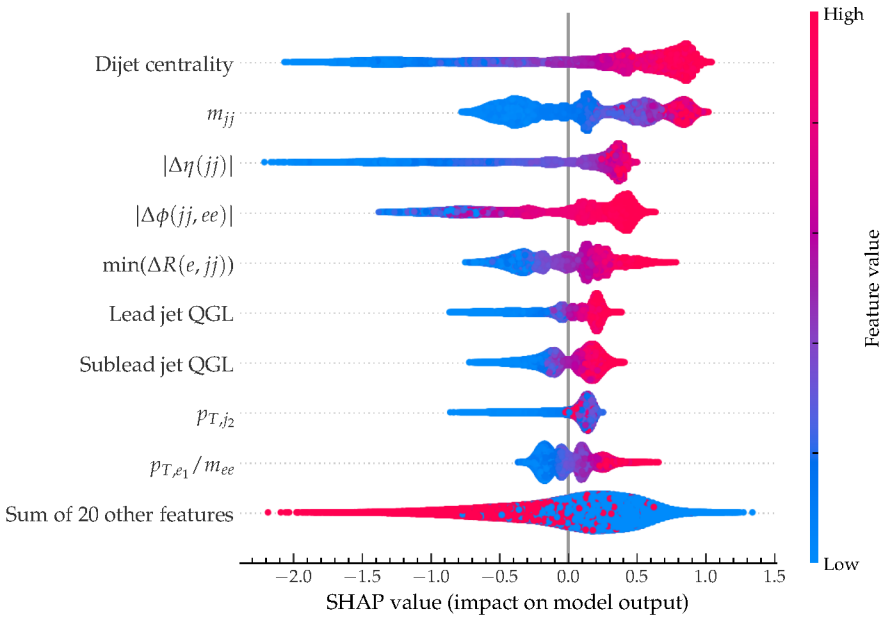

Figure 6.10: The distribution of Shapley values (filled circles) for selected input features to the VBF BDT, shown at the per-event level. Features are ordered by descending importance, measured by the mean of the absolute Shapley values over all events. Larger values of the input variable are coloured with red hues, while smaller values are shown in blue hues. The dijet centrality and invariant mass score highly, as expected from the large class separation visible in the 1D distributions. Events containing two jets separated by a large η difference are also observed to pull the model output towards signal-like predictions, as expected.

the data-driven method for constructing background models discussed in Section 7.3.

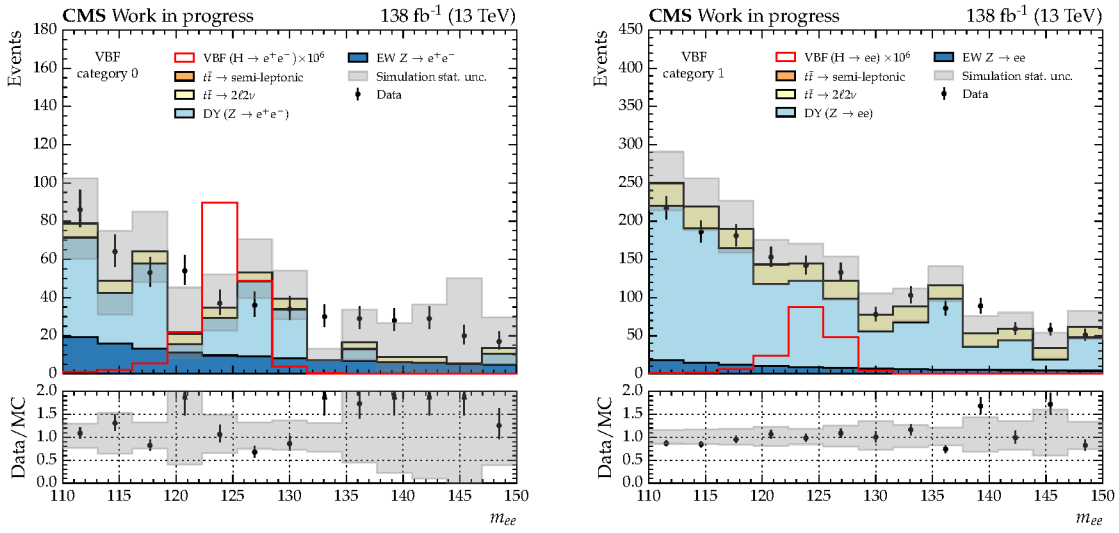

Figure 6.11: Observed m_{ee} distributions for the two VBF analysis categories, shown for simulated background (bold face) and data (black points). The distribution for VBF signal events is also shown (red), with the overall yield scaled for visibility. The ratio of data to simulation is shown in the lower panel. Arrows indicate entries outside the y-axis range. The statistical uncertainty for DY events can be large for events with unphysically high simulated event weight, particularly in the higher S/B analysis category.

6.5 Deep learning based VBF categorisation strategies

The BDT-based approach to background rejection in VBF categories makes use of features that are carefully engineered such that the distributions of signal and background events show some separation power. However, in the process of summarising event information "by hand", some of the original information in the event can be lost. This lost information may still provide discrimination power for the classification task, yet a BDT is unable to exploit it. Therefore, it could be beneficial to invoke a machine learning algorithm that can abstract its own feature representations from low-level inputs. A deep neural network is a natural choice, where feature engineering occurs automatically within the network's hidden layers.

6.5.1 Motivation and strategy

The motivation for a neural network in the context of background rejection in the VBF phase space is two-fold. Firstly, summarising low-level detector information into higher-level input features with known predictive power is more suited to decision tree algorithms,

where obvious partitions can be made in the input feature space. In comparison, a NN builds abstractions and higher-level representations from low-level inputs automatically within hidden layers, which may be more complex and indeed more useful that the original set of features chosen by hand. Secondly, the structure of the input feature set could be arranged to appeal to the inductive bias of the choice of NN architecture, potentially resulting in additional separation power.

The chosen VBF NN structure can be divided into two components. The first comprises long short-term memory layers which take low-level event features as input. These include the four vector and QGL ID score for the leading three jets in the event. The four vectors for the electrons are not provided to prevent the network from learning the value of m_{ee}. Since the LSTM structure motivates the organisation of these jets by some meaningful ordering, the jet descriptions are arranged into a 1D sequence by descending jet p_T. This treatment is partially motivated by the DeepJet [103] architecture for jet flavour tagging, where an LSTM is used to process a sequence of particle flow candidates, ordered by impact parameter.

In addition, a set of high-level input features are added such that the network can learn correlations between the low-level physics objects and the rest of the event. These are identical to those described in Section 6.4 (minus the low-level features) and include electron kinematics and high-level jet features. With this treatment, the input features are deliberately identical between models, in order to compare the NN and nominal BDT approaches fairly. The high-level features are combined with the output of the LSTM network in a series of fully connected layers. The number of nodes in each of these FC layers is designed to gradually reduce the dimensionality of the feature representations, before reaching the single output node. Hence, earlier FC layers are more dense than the latter ones. A summary of the architecture is illustrated in Figure 6.12.

6.5.2 Training and optimisation

The network is implemented using the Keras [104] python library, with the Tensorflow [105] backend. The differentiable loss function for the categorisation task is chosen to be identical to the BDT-based approach, since the classification goal is identical. Although Keras supports many gradient-descent based algorithms to minimise this loss, in practise, the Adam [106] optimiser is chosen. Adam uses an approach similar to classical SGD, with a variable learning rate that is gradually decreased as the number of steps taken by the optimiser increases.

Several hyperparameters of the network are optimised as part of the model selection process. For the LSTM portion of the network, the number of hidden layers and number of nodes in each layer are optimised. A similar strategy is employed for the fully connected

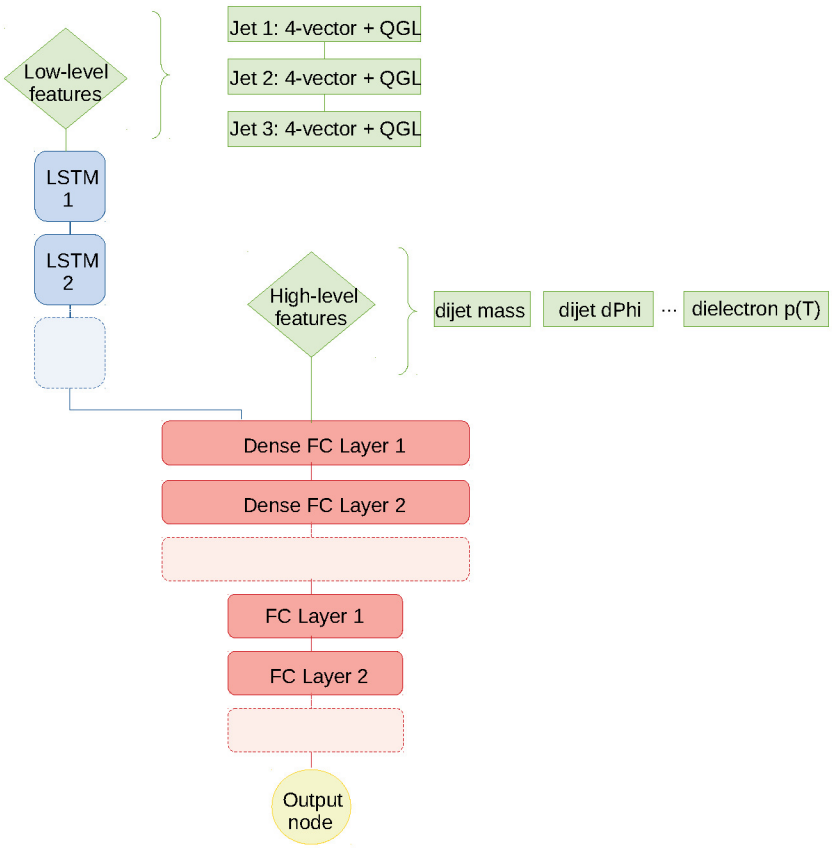

Figure 6.12: A schematic of the VBF neural network architecture. The structure comprises long short-term memory layers (blue) which take low-level descriptions of the three leading jets. The output from the final long short-term memory cell is interfaced with fully connected layers (red) which take high-level per-event features as inputs. The number of nodes in each layer is decreased gradually before reaching the terminal output node (yellow) which holds the discriminator score. Empty components indicate possible extensions to the network architecture that are explored during hyperparmeter tuning.

network layers. To regularise the network, a dropout technique is implemented [52], where the exact probability to drop each neuron is considered as a hyperparameter. Each of the above network parameters are optimised using a grid search strategy, where a range of possible hyperparameter values are specified, and a model is trained for each possible combination. This results in the training of a few thousand networks. The final choice of NN hyperparameters that maximise the model performance is summarised in Table 6.3. Note, however, that the performance is largely insensitive to the exact hyperparameter choices; most configurations return equally performant models.

Input features are also pre-processed with a "Z-score" normalisation procedure, where the mean and standard deviation are transformed to be zero and one respectively. These transformations provide a uniform scale for all features, a common technique used to speed up convergence during training and improve network performance [46].

Finally, to further prevent overfitting, the network is trained with an early-stopping procedure in which the number of training epochs is decided dynamically by the loss, computed on a validation set containing 20% of the events. During this process, the batch size is also increased while training, to encourage convergence to the global minimum [107].

Table 6.3: The optimal choice of hyperparameters for the VBF NN, chosen to maximise the network performance on a withheld validation set. A grid search strategy over the potential architecture space is performed, resulting in the training of a few thousand networks. Multiple entries indicate the configuration for successive units of the same type.

Network hyperparameter	Value(s)
Number of FC layers	3
Number of nodes in each FC layer	200, 100, 100
Number of LSTM layers	2
Number of nodes in each LSTM layer	100, 100
Dropout rate	0.2
Activation function (FC layers)	Rectified linear unit
Activation function (output layer)	Sigmoid

6.5.3 Performance comparisons

The performance of the VBF NN is compared with the nominal BDT-based approach to assess whether the deep learning architecture offers any improvement in background rejection. The figure of merit in this comparison is the area under the ROC curve for each model, which for the VBF NN is generated from the output score distribution shown in Figure 6.13 (left). The distribution of scores is similar between the VBF NN and nominal VBF BDT (Figure 6.8); the simulated signal is well separated from both the DY and $t\bar{t}$ backgrounds, while backgrounds from EW Z processes are harder to separate. The

areas under the corresponding ROC curves, shown in Figure 6.13 (right), are also similar between classifiers — when evaluated on the test set, the area under the ROC curve associated with the VBF NN (BDT) is 0.923 (0.925). It is therefore concluded that, for this learning task, a more complex approach using a deep learning-based classifier offers near identical performance to the nominal BDT-based approach.

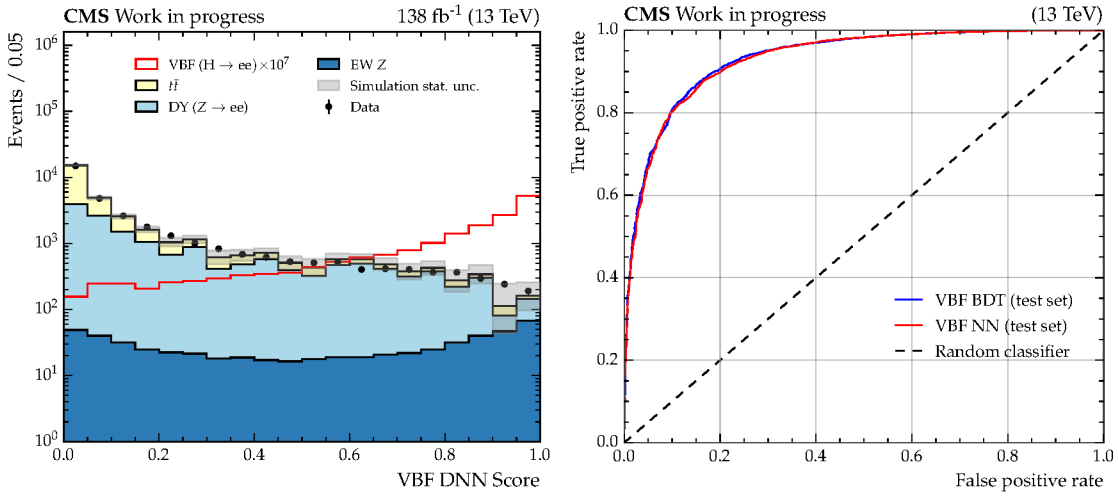

Figure 6.13: Left: the output score of the VBF NN in simulated signal (red), background (bold face), and data (black markers). The separation observed between the signal, and Drell-Yan and $t\bar{t}$ backgrounds is similar to that of the VBF BDT. Right: ROC curves for the VBF NN (red) and the VBF BDT (blue), evaluated on the test set. The area under the curves are comparable between models, indicating similar performance.

To understand why the two learning algorithms may perform similarly, it is again useful to understand how the model predictions are connected with the input features. To this end, a selection of the most predictive features are shown as a function of the VBF NN output score in Figure 6.14. The evolution of each feature is almost identical to the analogous distributions for the VBF BDT shown in Figure 6.9, suggesting that input features are being used in similar ways. Both models assign high scores to events containing a dijet pair with high invariant mass, and large separation in pseudorapidity between the constituent jets. The fact that both algorithms produce models yielding similar performances, using variables in near identical ways, suggests that the underlying function mapping input features to the target class is learned near optimally. This is also supported by the fact that the NN performance is largely insensitive to the exact choice of architecture during hyperparameter optimisation; each model returns a similar performance, regardless of its complexity.

An alternative explanation is that the representation of features, combined with the choice of learning algorithm, may not offer any additional information for the task. Al-

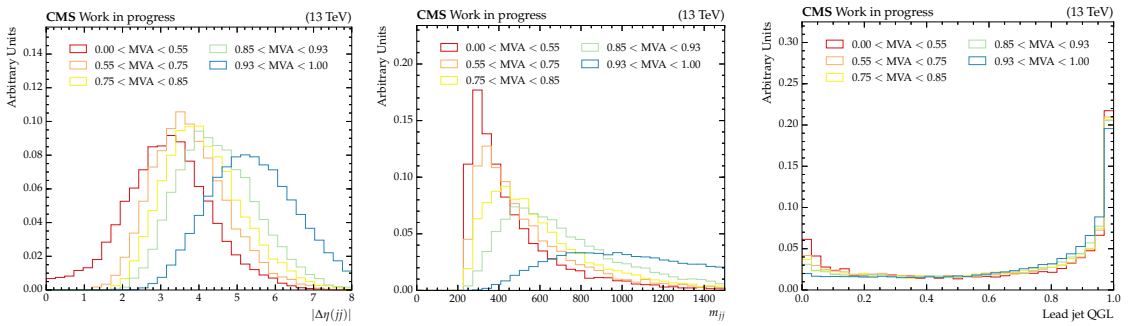

Figure 6.14: Distributions of selected inputs to the VBF NN as a function of VBF NN output score, in simulated VBF signal events. Each histogram is normalised to unit area. The variables shown have identical definitions to those in Figure 6.9. The evolution of each feature with the NN score is almost identical to the analogous plots for the VBF BDT, suggesting that variables are used in similar ways between the two classifier types.

though ordering low-level jet inputs by p_T has been shown to offer small improvements in other applications [108], LSTM input data is typically ordered temporally. This is perhaps a more natural ordering, where the connection between elements is less abstract and can be easily interpreted by the model. In the case of the VBF classification task, an alternative structuring of the input dataset may appeal more naturally to the inductive bias of alternative algorithms. For example, treating the jet-based information as an image, unfolded in the η-ϕ detector plane, and applying a convolutional NN has been shown to offer a small advantage over BDT-based approaches for VBF classification in the H $\rightarrow$ $\gamma\gamma$ decay channel [109].

Given the similarity between performances, the nominal VBF BDT-based approach is kept for VBF classification. This model is significantly quicker to train and uses less computational resources than the deep learning approach.

6.6 BDT validation

In this analysis, background models are taken directly from data, whereas signal models are derived from simulated samples. Therefore, it is necessary to ensure there is reasonable agreement between data and simulation for signal-like objects in the inputs to each BDT, which in turn control agreement in the output score. This is validated using a sample of Z $\rightarrow$ ee events from control regions in data, defined to be orthogonal to the analysis category phase space. Drell-Yan processes are chosen for this validation since a large number of events that mimic the H $\rightarrow$ ee final state are available, and the decay is relatively free from contaminating backgrounds. For the ggH BDT, the control region is defined by the nominal analysis preselection, with the exception of the dielectron mass

requirements which are shifted to $80 < m_{ee} < 100$ GeV, in order to roughly centre on the Z boson mass. For validation of the VBF BDT, the VBF preselection defined in Section 6.4 is applied in addition.

Figure 6.15 shows the output score distribution of both the ggH and VBF BDT for data and simulation. The effect of the dominant systematic uncertainties are included in both plots. For electrons, these include the uncertainty on the electron energy scale corrections, and uncertainties on the electron reconstruction and identification efficiencies. For inputs that describe jets, the uncertainties associated with the jet energy scale and resolution corrections are also included. The agreement between data and simulation in both output score distributions is within the uncertainty permitted by the combined systematic and statistical variations, over the majority of each distribution. The residual differences observed in the ggH BDT score are smaller than the ggH theoretical uncertainties included in the final fits, presented in Section 7.4. The closure between data and simulation for the input features to both BDTs is also good.

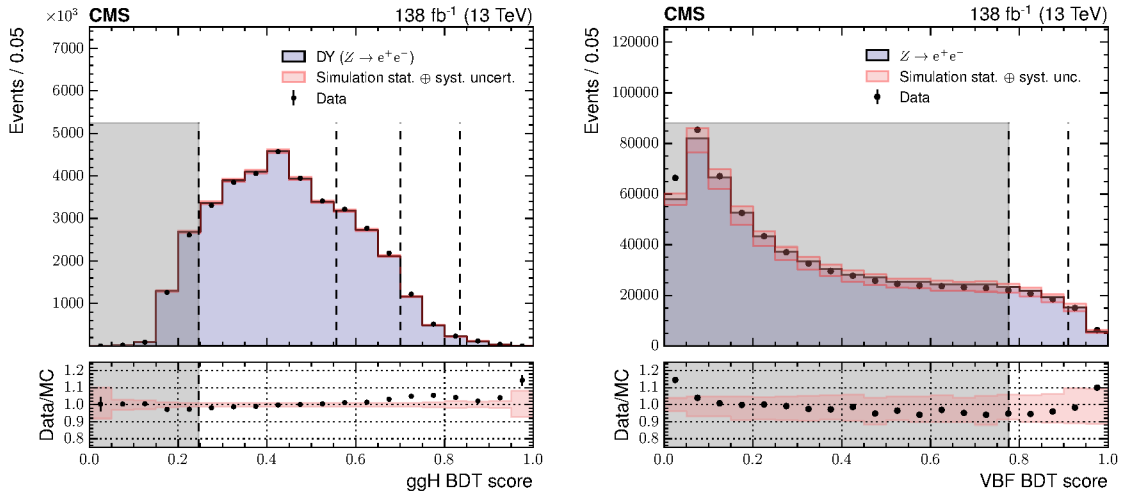

Figure 6.15: Distribution of the output score of the ggH BDT (left) and VBF BDT (right) in their respective Z → ee control regions. The combination of systematic and statistical uncertainties is shown by the red shaded band. Category boundaries targeting each Higgs boson production process are denoted with dashed lines. Events with scores in the low S/B grey shaded regions are not considered for their respective categories. Good agreement is observed between the Drell-Yan simulation (filled histograms) and data (black points), for the phase space in which analysis categories are constructed. For the ggH BDT output score, residual differences between the data and simulation are covered by theoretical uncertainties on the ggH cross section included in the final fits.

6.7 Summary

The categorisation for H $\rightarrow$ ee events targets production via both gluon fusion and vector boson fusion. Analysis categories are constructed using the output of dedicated BDTs, designed to improve the S/B ratio. Each BDT is trained on kinematic properties of the dielectron object, alongside descriptions of any associated jets. For the categorisation of VBF events, a deep learning approach using a long short-term memory neural network is also studied, yielding similar performance to the nominal BDT-based approach. Two categories are developed to target VBF events, while four categories are used for ggH. Events satisfying requirements for both VBF and ggH categories are assigned preferentially to those targeting VBF, as illustrated in Figure 6.16.

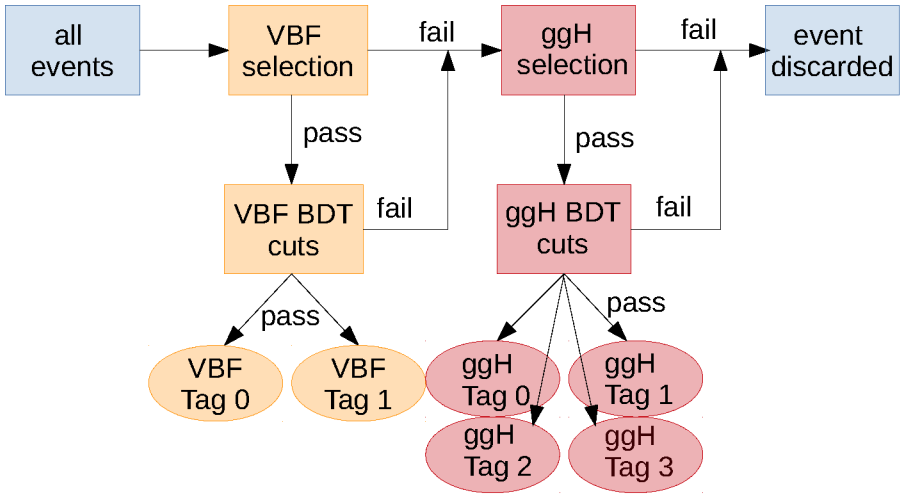

Figure 6.16: The priority sequence for H $\rightarrow$ ee analysis categories. Events are considered firstly for one of two categories targeting VBF, provided they pass the jet-based VBF preselection. Events failing the VBF preselection requirements, or receiving VBF BDT scores lower than the minimum category BDT threshold, are then considered for one of four ggH categories. Events that do not satisfy the requirements for any category, or that do not pass the analysis preselection (equivalent to the ggH selection), are discarded.

Chapter 7

Signal and background modelling

7.1 Introduction

In order to extract a potential H $\rightarrow$ ee signal, a maximum likelihood fit to the observed m_{ee} distribution in data is performed. To estimate the expected mass distributions, a parameterisation of both the signal and background components of the m_{ee} spectrum is required. Once constructed, the parameters describing the models are chosen such that the value of twice the negative logarithm of the likelihood (2NLL) is minimised.

Signal models are derived from simulated samples, for a selection of Higgs boson mass hypotheses. The dependence on m_H is built into the shape parameters and overall normalisation of the signal yield. Conversely, background modelling uses a data-driven approach, where a set of candidate models are constructed by fitting to sideband regions of the m_{ee} distribution. All candidate functions fitting well are considered when extracting the final results, in a procedure known as the discrete profiling method [110]. The method is similar to that used in many previous SM H $\rightarrow \gamma\gamma$ analyses [2, 111, 112] by the CMS collaboration, where it has been extensively validated.

When extracting the analysis' results, a set of systematic uncertainties are included to account for imperfect descriptions of the analysis inputs. These uncertainties, which are divided into theoretical and experimental sources, are implemented using nuisance parameters. While not being of interest themselves, nuisance parameters are allowed to modify the extracted parameter(s) of interest during the final fit. The analysis uses both continuous and discrete nuisance parameters to implement systematic uncertainties.

This chapter describes the strategy used to construct models for signal and background events, with the formalisation of the likelihood fit left to chapter 8. A description of all theoretical and experimental systematic uncertainties is also given, including the magnitude of their effect.

7.2 Signal modelling

In this analysis, models of the m_{ee} distribution for H $\rightarrow$ ee signal processes are built using simulated samples. An independent model is constructed for each signal process contributing to each analysis category. Models are also constructed separately for each year of data taking, to account for any per-year differences in the CMS detector response and performance. In total, this results in 72 models (4 signal processes $\times$ 6 analysis categories $\times$ 3 years).

Each signal model is constructed as a sum of up to five Gaussian probability distributions. Four of the five Gaussians are used to model the core of the m_{ee} distribution, where the majority of events lie, while the fifth is included to account for the lower tail in m_{ee}, resulting from final-state radiation (FSR). The mean of this so-called FSR Gaussian is centred at lower masses with respect to the nominal set and has larger width, to model the extended tail. Alternative signal parameterisations are also tested, namely Double Crystal Ball functions [113]. The resulting signal models are similar to those produced with the nominal strategy; hence, the more simple approach of summing Gaussian probability distributions is kept. Figure 7.1 shows the shape components of the signal models for ggH signal entering the highest S/B category targeting ggH events in 2017, and similarly for VBF signal entering the purest category targeting VBF events, at a nominal Higgs boson mass of 125 GeV.

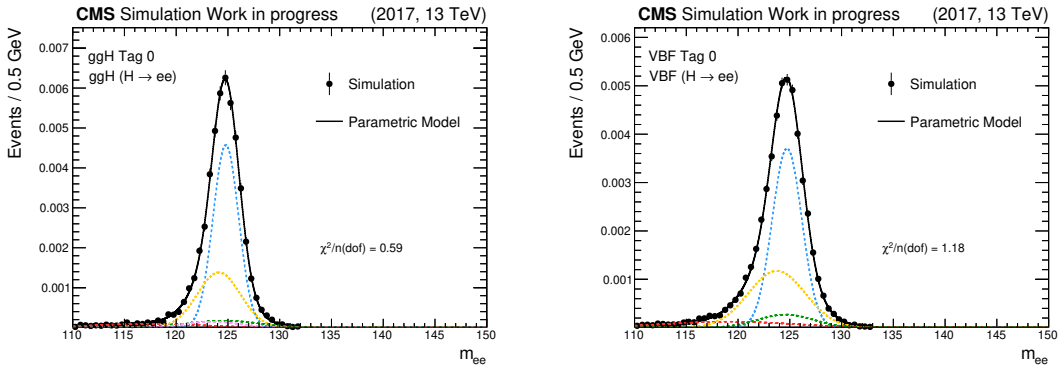

Figure 7.1: Shape components of the m_{ee} signal model for simulated ggH events in the most pure ggH analysis category (left), and simulated VBF events in the most pure analysis category targeting VBF events (right). Events are shown for 2017 only, and simulated assuming $m_H = 125$ GeV. Contributions from up to five Gaussian functions are shown by the dashed lines, with the FSR Gaussian displayed in red. The overall model, resulting from the sum of each Gaussian distribution, is shown by the solid black line.

To allow for the extraction of limits on $\mathcal{B}$(H $\rightarrow$ ee) as a function of the Higgs boson mass, signal models are constructed from a simultaneous fit to samples at $m_H = 120$,

125, and 130 GeV. Each parameter for each Gaussian is a linear function of m_H, which accounts for the dependence of the signal model shape on the Higgs boson mass. The exact number of Gaussian probability distributions used for each model depends on the shape of the m_{ee} distribution, and is chosen to maximise the χ^2/n_{dof} in the signal fit.

For each value of m_H, the total signal yield for each production process is normalised by the product of the cross section times branching fraction for the H $\rightarrow$ ee decay. For intermediate mass hypotheses, this normalisation is extracted using a spline constructed from piece-wise polynomial functions, through the three original m_H values. Each yield is also normalised to the product of the detector efficiency and analysis acceptance, defined by the ratio of the number of events entering the overall analysis category acceptance, to the total number of expected events. To illustrate the evolution of signal models with m_H, Figure 7.2 shows example models at a selection of Higgs boson mass hypotheses, for ggH signal entering in the highest S/B category targeting ggH events, in 2016 simulation. The signal normalisation is also shown, divided into its constituent components.

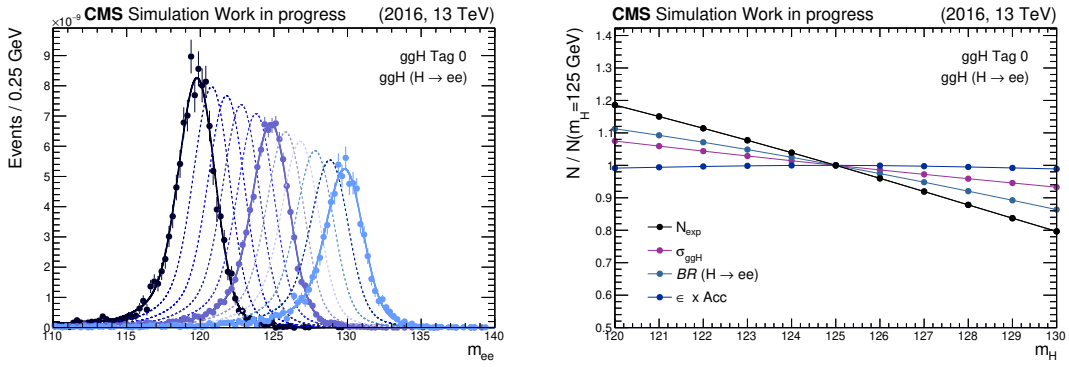

Figure 7.2: Left: the evolution of signal models with m_H, for simulated ggH signal events in ggH Tag 0, in 2016. Simulated events for Higgs boson masses of $m_H = 120$, 125, and 130 GeV are shown, along with their statistical uncertainty, by the filled markers. Dashed lines indicate signal models for intermediate Higgs boson mass hypotheses. Right: normalisation of the signal yield for the same simulated events, as a function of m_H, scaled to the nominal yield at $m_H = 125$ GeV. The overall normalisation is broken into contributions from the production cross section, H $\rightarrow$ ee branching fraction, and the combination of the detector efficiency and analysis acceptance.

Figure 7.3 shows the final signal models for the highest S/B analysis categories targeting ggH and VBF events, for a nominal Higgs boson mass of $m_H = 125$ GeV.

7.3 Background modelling

Background events in the H $\rightarrow$ ee search comprise those entering analysis categories that originate from processes other than the H $\rightarrow$ ee signal. These events, predominantly

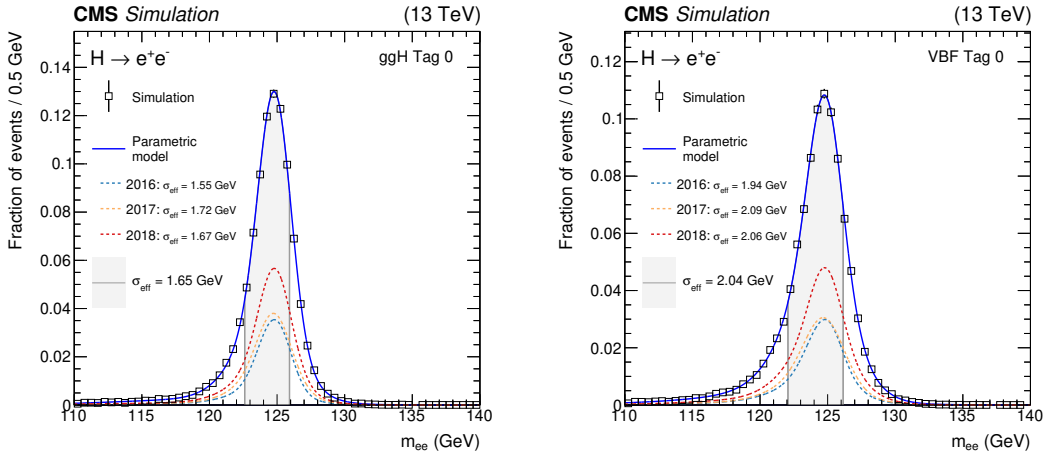

Figure 7.3: Signal models for the two highest S/B analysis categories targeting ggH (left) and VBF (right) production, integrated over Higgs boson production processes, for a nominal Higgs boson mass of $m_H = 125$ GeV. The overall yield is normalised such that the integral is equal to unity. The contribution from each of the three years is shown by the dashed lines. The resolution, parameterised using σ_{eff}, is also labelled for each model.

originating from DY decays, form a smoothly falling distribution in m_{ee}, on top of which the signal peak is situated. Unlike in the construction of signal models, an entirely data-driven approach is used to model the background. The strategy uses data in the sideband region, defined as data outside $115 < m_{ee} < 135$ GeV, to constrain functional forms for the background models. This offers the advantage that any systematic effects associated with the modelling and reconstruction of simulated background samples is avoided.

The exact parameterisation of the background distributions are not known a priori, and thus many functional forms that fit well are considered. This degeneracy in the parameterisation translates into an uncertainty on any measurements of the signal, since each choice of function results in a different number of events counted under the signal peak. To account for this uncertainty, the discrete profiling method is used, described in detail in Ref [110]. This reference includes tests of the coverage for the method, and bias on the extracted parameters of interest (POIs), the latter of which is also tested explicitly in Section 7.3.1.

To introduce the discrete profiling method, it is useful to outline the concept of nuisance parameters, and their effect on the NLL curve. In general, a nuisance parameter is any parameter of a model which affects the parameters of interest, but is not of interest itself. When fitting for the POIs, the nuisance parameter is profiled, meaning its value is allowed to vary during the minimisation of the likelihood, potentially within some constraint. This is illustrated in Figure 7.4, where the blue curve shows the negative log-

arithm of the likelihood, for some parameter of interest, μ, when the nuisance parameters are fully profiled. If the nuisances were fixed to their best fit values, as illustrated in the orange parabola, the likelihood curve would be narrower, and the uncertainty interval on μ, smaller. However, when profiling the nuisance parameters, a configuration can can be found for each point in μ-space that decreases the NLL (increases the likelihood). Smaller values for the NLL result in the likelihood curve tracing a wider parabola, naturally increasing the measured uncertainty on μ.

Although the above example describes a continuous profiled nuisance, the same curve could be built by ensembling a set of negative log-likelihood curves where nuisance parameters can take on discrete values. In the example of Figure 7.4, curves generated by arbitrary values of the nuisance parameters are shown in red. Although a discrete number of these are drawn, the entire profile likelihood curve could be constructed from the envelope enclosing all such red curves, provided the nuisance parameter space is sampled sufficiently. This equivalent formalisation demonstrates the generalisation between a discrete and continuous set of nuisance parameters.

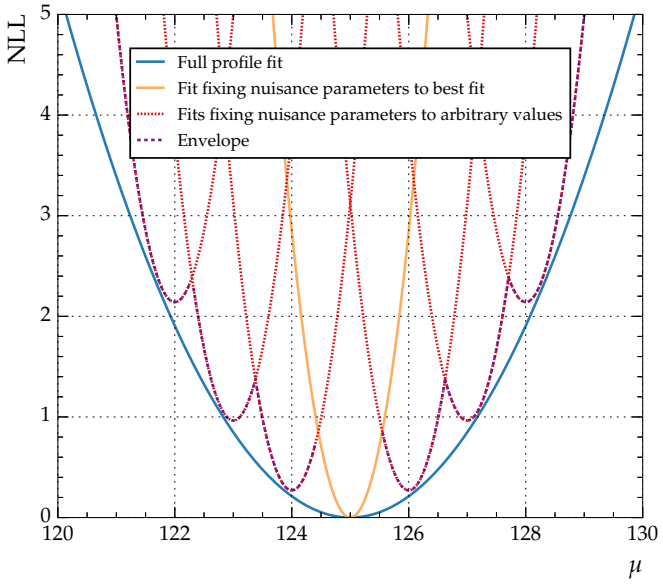

Figure 7.4: An illustration of the discrete profiling method. The negative log-likelihood curve, where the set of nuisance parameters are allowed to vary at each value of the POI, μ, is shown in the solid blue line. The fully profiled curve can be built from the envelope (dashed purple line) of curves generated at discrete configurations of the nuisance parameter space. These curves, shown in red, are produced by nuisances fixed at arbitrary values other than those at the best fit, and approximate the fully profiled fit when the number of nuisance parameter configurations sampled is large. Figure inspired by Ref [110].

In the case of modelling the background m_{ee} distribution in the H $\rightarrow$ ee analysis, the

choice of functional form with which to fit is treated as a discrete nuisance parameter. The resulting 2NLL envelope is then constructed by profiling all possible choices of background parameterisation. Although in principle all analytical functions that fit the background m_{ee} distribution well could be tested, in practise, the chosen parameterisations can be divided into two groups of functional families. The first set consists of functions that are able to describe, in principle, smoothly falling distributions, given an appropriate choice for the number of degrees of freedom. Four families are considered in this group; for a N-parameter function with parameters $p_0, p_1, ..., p_N$ to be determined, these include:

- sum of exponential functions, where

$$f_N(m_{ee}) = \sum_{i=0}^{N} p_{2i} \exp\left(p_{2i+1} \cdot m_{ee}\right),$$

- sum of power law functions, where

$$f_N(m_{ee}) = \sum_{i=0}^{N} p_{2i} \cdot m_{ee}^{-p_{2i+1}},$$

- Bernstein polynomials, where

$$f_N(m_{ee}) = \sum_{i=0}^{N} p_i \binom{N}{i} m_{ee}^{i} \left(1 - m_{ee}\right)^{N-i},$$

  and

- Laurent series, where

$$f_N(m_{ee}) = \sum_{i=0}^{N} p_i \cdot m_{ee}^{-4+L(i)},$$

  with

$$L(i) = \sum_{j=0}^{i} (-1)^j j.$$

Since the expected background in all analysis categories is composed primarily of DY events, the second set of functions are included to model this smoothly falling distribution, driven by the Breit-Wigner nature of the Z boson lineshape. Three individual functions are considered, characterised by a core Breit-Winger component, modified by exponential or power law terms. This set of so-called physics-inspired functions include:

- Breit-Wigner core modified by an exponential function of m_{ee}, where

$$f_N(m_{ee}) = \frac{\Gamma_Z \cdot \exp(p_0 \cdot m_{ee})}{(m_{ee} - m_Z)^2 + (\Gamma_Z/2)^2},$$

- Breit-Wigner Redux, where the Breit-Wigner core is modified by a further exponential term, and the exponent in the denominator becomes a free parameter, where

$$f_N(m_{ee}) = \frac{\Gamma_Z \cdot \exp(p_0 \cdot m_{ee} + p_1 \cdot m_{ee}^2)}{(m_{ee} - m_Z)^{p_2} + (\Gamma_Z/2)^{p_2}},$$

- Breit-Wigner Gamma, where a Breit Wigner, modified by a single exponential factor, is summed with a gamma function, where

$$f_N(m_{ee}) = p_2 \cdot \frac{\Gamma_Z \cdot \exp(p_0 \cdot m_{ee})}{(m_{ee} - m_Z)^2 + (\Gamma_Z/2)^2} + (1 - p_2) \cdot \frac{\exp(p_1 \cdot m_{ee})}{m_{ee}^2}.$$

For efficiency, the final envelope of candidate functions considers only a subset of the possible choices. The procedure to select functions is performed separately for each analysis category, and proceeds as follows. The lowest order functions in the family are fit, using a maximum likelihood approach, to the m_{ee} sideband regions in data. The process is repeated for subsequent orders of the function, until a minimum requirement on the goodness-of-fit is reached. This criteria ensures that only functions fitting the m_{ee} distribution well are included in the final category envelopes. For functions defined at a fixed order, namely the set of physics-inspired functions, the procedure is complete. For parameterisations that span multiple orders, an F-test [114] is then performed to quantify whether the improvement in fit quality when considering the next highest order function justifies the increase in model complexity. The test computes a p-value by determining the difference in 2NLL between fits of successive order as a test statistic. Assuming sufficient sample size, the statistic follows a χ^2 distribution with degrees of freedom equal to the difference in number of parameters between the two functions under test. If the observed p-value is below a threshold of 0.05, the higher order function is deemed to offer a worthwhile improvement in fit quality, and is added to the category envelope. The process is repeated for each successive order until the p-value requirement is failed. To further prevent overfitting, an additional penalty term is applied to penalise overly complex functions. The penalty adds one unit to the negative log-likelihood for each free parameter in the background model and is applied in the final signal-plus-background fit.

Figure 7.5 shows all functions entering the envelope for the highest S/B analysis categories targeting ggH and VBF events. The degeneracy in possible fits is well spanned by the set of candidate functions. The envelope in each category consists of roughly

equal numbers of physics-inspired and nominal smoothly falling functions; the best fitting functions in each category, defined as those with the smallest NLL value amongst possible background candidates, are also equally divided between these two sets.

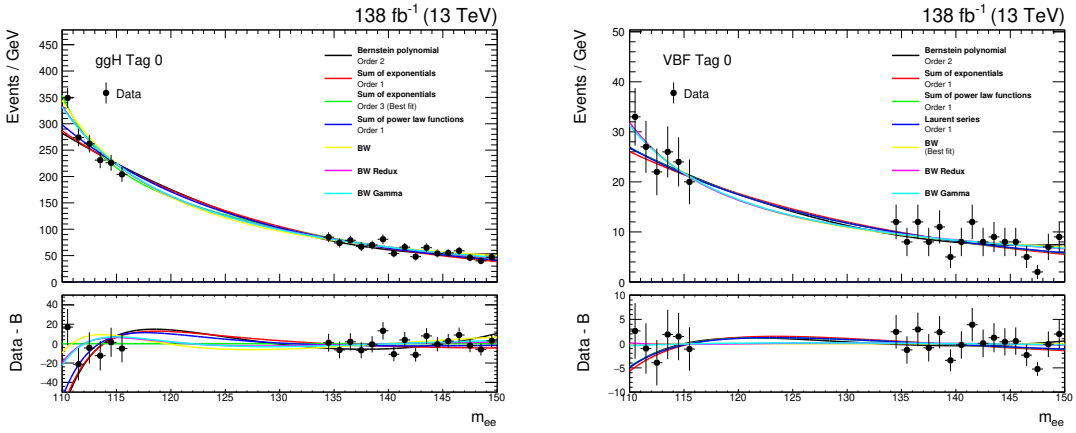

Figure 7.5: The set of candidate functions chosen to fit data in the m_{ee} sideband regions using the envelope method, for the highest S/B category targeting ggH (left) and VBF (right) events. The functional families considered include sums of exponential functions, Bernstein polynomials, sums of power law functions, and Laurent Series', as well as modified Breit-Wigner functions aiming to model the lower tail in m_{ee} resulting from Z boson decays. The best fitting function is indicated in the legend. The residual difference between each background model and data is shown in the lower panel.

Finally, the background resulting from SM H $\rightarrow \gamma\gamma$ production, where both photons are misreconstructed as electrons, is studied with simulated events. The process is found to contribute less than 0.1% to the inclusive analysis categories, for a H $\rightarrow$ ee branching fraction scaled to the expected limit obtained in chapter 8. It is therefore neglected in the rest of the analysis.

7.3.1 Bias studies

Although many parameterisations of the background m_{ee} distribution are considered for each value of the POI, in the end, only one function minimising the likelihood is chosen. Therefore, it is possible that this choice of function may inject a systematic bias on the extracted H $\rightarrow$ ee branching fraction. To check for possible bias, a large sample of pseudo-experiments, or *toys*, are generated using a Monte Carlo technique. Toy datasets are built for the m_{ee} distribution in all analysis categories from each of the candidate background functions in the envelope. The function used to generate the background m_{ee} distribution is considered as the "true" background parameterisation for the given set of toys. Each toy also is also injected with a signal proportional to the expected limit on $\mathcal{B}$(H $\rightarrow$ ee), assuming a Higgs boson mass of 125 GeV. For each true background

function, two thousand generated toys are fit using the discrete profiling method with the full envelope of functions included in the analysis category.

The bias for each toy dataset is defined using the pull of the distribution, where the pull for a given toy is

$$P(\mu, \hat{\sigma}) = \frac{\hat{\mu} - \mu}{\hat{\sigma}}, \tag{7.1}$$

where μ is the injected value of the rate parameter scaling $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})$, $\hat{\mu}$ is the value of μ which minimises the NLL for each toy, as returned from the fit, and $\hat{\sigma}$ is the uncertainty on $\hat{\mu}$. In the absence of systematic effects, the pull over all toys is expected to be Gaussian distributed, with a mean of zero and standard deviation of one. Practically, a Gaussian function is fit to the pull distribution, with the mean of the fit being used to quote the bias. Good functional forms are those characterised by an absolute bias of less than 0.2, where this threshold is chosen to ensure that the systematic uncertainty due to a potential bias on the upper limit for $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})$ does not increase the overall uncertainty by more than 2%.

The biases for the highest S/B analysis categories targeting ggH and VBF events are summarised in Figure 7.6. For almost all truth functions across each analysis category, the bias is within the accepted tolerance. It is also checked from toys that those functions which are outside the acceptable bias range are rarely the best-fit choice from the envelope in each category.

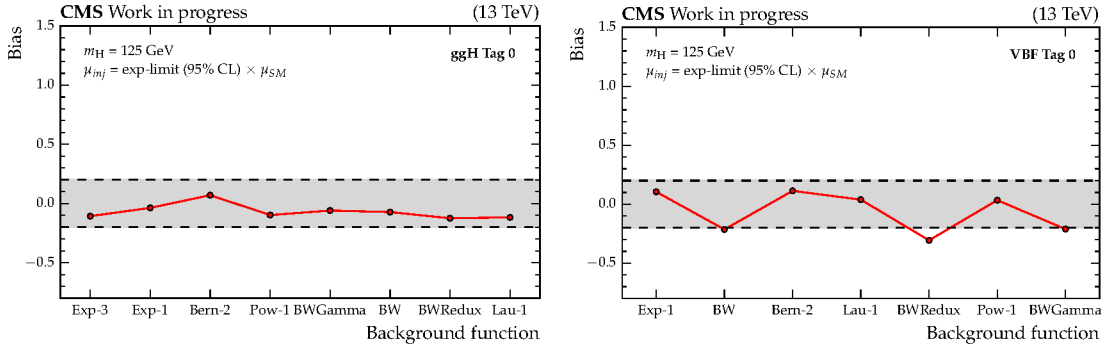

Figure 7.6: The measured bias of each "true" function inside the envelope of possible background fits used to generate toy datasets, for the highest S/B analysis categories targeting ggH (left) and VBF (right) events. Each toy dataset is injected with a signal model with yield scaled to the expected limit on $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})$, assuming a nominal Higgs boson mass of 125 GeV. The individual functions and their order are specified along the x-axis. For almost all functions, the bias is within the chosen tolerance of ± 0.2 (grey shaded-band).

7.4 Systematic uncertainties

Unlike the data-driven background models, where systematic uncertainties are handled naturally using the discrete profiling method, uncertainties affecting the signal models must be estimated directly from simulation. These uncertainties are treated in one of two ways. Uncertainties which could modify the shape of the m_{ee} distribution are incorporated as Gaussian constrained nuisance parameters that modify the mean and width of the signal distributions in each analysis category. Variations that modify the signal shape typically arise from uncertainties in the electron energy scale calibration. Uncertainties that do not affect the m_{ee} shape are treated as log-normal variations in the signal yield. These are allowed to modify the normalisation in both directions asymmetrically, although each variation is typically comparable in magnitude. Systematic uncertainties that affect only the signal yields typically arise from theoretical sources associated with the simulation of signal events, and experimental uncertainties describing, for example, the electron reconstruction and identification efficiencies.

7.4.1 Theoretical uncertainties

The sources of theoretical uncertainty considered in this analysis are as follows.

- *QCD scale uncertainty*: the uncertainty arising from variations of the renormalisation and factorisation scales used when computing expected SM cross sections and event kinematics. These account for the missing higher order terms in perturbative calculations. The recommendations provided in Ref [40] are followed, where the uncertainty in the yield is estimated using three sources: varying the renormalisation scale by a factor of two, varying the factorisation scale by a factor of two, and varying both in the same direction simultaneously. The impact on the normalisation is largest for $t\bar{t}H$ events at 5.8%.

- *PDF uncertainties*: these account for the uncertainty due to imperfect knowledge of the composition of the proton, which affects which partons are most likely to initiate high energy events. Uncertainties are computed following the PDF4LHC prescription [32, 115], with the impact on the normalisation ranging between 1.9 and 3%.

- *Uncertainty in the strong force coupling constant*: the uncertainty in the value of the strong force coupling constant is included in the treatment of the parton density function uncertainties. The impact on the yield is largest for ggH production, with a value of 2.6%.

- *Underlying event and parton shower uncertainties*: these uncertainties are obtained using dedicated simulated samples which vary the PYTHIA8 tune from that used in the nominal samples, and vary the renormalisation scale for QCD emissions in initial state and final state radiation by a factor of 2 and 0.5. These uncertainties are treated as migrations of events from a given production mode into and out of the VBF and ggH analysis categories. The largest effect comes from the parton shower uncertainty on VBF events, which can change the signal yield in the VBF analysis categories by up to 5%.

The total impact on the $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})$ measurement from theoretical systematic uncertainties is $^{+0.2}_{-0.1} \times 10^{-4}$, with dominant contributions from uncertainties affecting the ggH production cross section.

7.4.2 Experimental uncertainties

Experimental uncertainties that affect the shape of the m_{ee} distribution in signal events are listed below.

- *Electron energy scale*: the uncertainty associated with the corrections applied to the electron energy scale in simulation [78], derived using a sample of $\mathrm{Z} \rightarrow \mathrm{ee}$ tag-and-probe events [79]. Four nuisance parameters are defined for the possible combinations of electrons with low or high R_9 values, and reconstructed within the EE or EB.

- *Electron energy scale non-linearity*: an uncertainty to cover possible differences between the linearity of the electron energy scale between data and simulation, estimated on a sample of boosted tag-and-probe events. An uncertainty of 0.1% is assigned for electrons with $p_T < 80$ GeV, and 0.2% for above.

Experimental uncertainties that only modify the event yield are as follows.

- *Integrated luminosity*: the integrated luminosities for the 2016, 2017, and 2018 data-taking years have individual uncertainties of 1.2, 2.3, and 2.5% respectively [116–118], while the overall uncertainty for the 2016–2018 period is 1.6%.

- *Electron ID and reconstruction efficiencies*: uncertainties on the scale factors derived to correct for differences in simulation and data for the electron ID and reconstruction efficiencies. For both sources, the size of the uncertainty is approximately 1% in each category.

- *Jet energy scale and smearing corrections*: the energy scale of jets is measured using the p_T balance of jets with Z bosons and photons in $Z \to ee$, $Z \to \mu^+\mu^-$ and γ+jets events, as well as the p_T balance between jets in dijet and multijet events [100]. The uncertainty in the jet energy scale is a few per-cent and depends on p_T and η. The size of the jet energy scale uncertainties on event yields is evaluated by varying the jet energy corrections within their uncertainties and propagating the effect to the final result. The impact on category yields is largest for those targeting VBF, and can be as high as 15%.

- *Trigger efficiency*: the uncertainty on the efficiency of the trigger selection, measured with $Z \to ee$ events using the tag-and-probe technique. The size of the uncertainty is less than 1%. An additional uncertainty is introduced to account for a gradual shift in the timing of the inputs to the ECAL L1 Trigger in the region at $|\eta| > 2.0$, which caused a specific trigger inefficiency during 2016 and 2017 data taking [68]. Both electrons, and to a greater extent jets, can be affected by this inefficiency.

The total impact on the $\mathcal{B}(H \to ee)$ measurement from experimental systematic uncertainties is $\pm 0.2 \times 10^{-4}$, with the largest contributions arising from uncertainties affecting the electron energy scale.

7.5 Summary

The maximum likelihood fit used to extract a potential $H \to ee$ signal requires modelling of both signal and background events. In this analysis, background models are developed with a data-driven approach using the discrete profiling method. This method naturally handles the uncertainty associated with choosing a function to model the m_{ee} distribution in background events. Conversely, signal models are derived using simulated samples. The dependence on the Higgs boson mass is deliberately built into the parameters of each signal model, such that limits can be extracted as a function of m_H. Systematic uncertainties affecting signal models are handled using nuisance parameters that can modify both the shape of signal models and the overall yields.

Chapter 8

Results

8.1 Introduction

Since the sensitivity of the analysis is much lower than the SM prediction, the reported results set upper limits on the branching fraction for the H $\rightarrow$ ee decay. To do so, a likelihood fit of the signal and background models to the m_{ee} distributions is performed in all analysis categories simultaneously. The fitted parameters are then used to extract observed and expected limits at various confidence levels.

In the following chapter, the construction of the likelihood is presented, alongside the corresponding signal-plus-background fits to m_{ee} distributions. The results of the analysis comprise limits on $\mathcal{B}$(H $\rightarrow$ ee) for a selection of Higgs boson masses, as well as per analysis category.

8.2 The likelihood function

In order to extract upper limits on the H $\rightarrow$ ee branching fraction, this analysis uses a binned simultaneous maximum likelihood fit to the m_{ee} distributions in data over all analysis categories. The fit is performed over the the range $110 < m_{ee} < 150$ GeV, with bin widths of 0.25 GeV, chosen to be as small as possible to prevent information loss. In this case, the likelihood is proportional to the product of Poisson terms over count observations from all bins in the m_{ee} distribution. For a given category, the likelihood, $\mathcal{L}_c$, is expressed as:

$$\mathcal{L}_c(\text{data}|\vec{\mu}, m_H, \vec{\theta}) = \prod_i^{N_b} \text{Poisson}\left(d_i \,\middle|\, \left[s_i(\vec{\mu}, m_H, \vec{\theta_s}) + b_i(\vec{\theta_b})\right]\right), \tag{8.1}$$

where $\vec{\mu}$ are the parameters of interest. The quantities $\vec{\theta_s}$ and $\vec{\theta_b}$ are the set of nuisance

parameters affecting signal and background models respectively, which may be discrete or continuous ($\vec{\theta} = \{\vec{\theta_s}, \vec{\theta_b}\}$). For the i^{th} bin in the m_{ee} distribution, of which there are N_b total bins, the associated Poission term is composed of:

- the *observed* number of events in data, denoted by d_i, in bin i,
- the *expected* number of events, obtained from the sum of signal and background yields, s_i and b_i respectively. The expected number of signal events depends on the POIs being tested, the Higgs boson mass, and the set of nuisance parameters affecting signal models. The background expectation is simply a function of the (unconstrained) nuisance parameters affecting the functional form and overall choice of background model.

Since each analysis category contains an exclusive set of events, the final likelihood describing all events is obtained by taking the product over all per-category likelihoods

$$\mathcal{L}(\text{data}|\vec{\mu}, m_H, \vec{\theta}) = \prod_{c}^{N_c} \left[\mathcal{L}_c(\text{data}|\vec{\mu}, m_H, \vec{\theta})\right] \cdot \Lambda(\vec{\theta}), \tag{8.2}$$

where N_c is the number of analysis categories, and $\Lambda(\vec{\theta})$ is the constraint term. For background models, the constraint adds a penalty in the likelihood for background candidates with large numbers of free parameters, to prevent fitting overly complex models. For nuisances affecting signal models, the term penalises deviations from the nominal nuisance parameter values. To prevent the signal yield becoming negative, the constraint terms on $\vec{\theta_s}$ that affect the normalisation are modelled using a log-normal distribution.

During the fit, the value of twice the negative logarithm of the likelihood is minimised using a numerical minimisation tool [119]. The transformations to the likelihood are monotonic and do not change the resulting minima; they are performed to make the numerical minimisation task easier. The set of parameters which minimise the value of the 2NLL, for a given $\vec{\mu}$ being tested, are referred to as *best-fit*. Since the H $\rightarrow$ ee analysis aims to set a limit on a single POI which scales the $\mathcal{B}$(H $\rightarrow$ ee) for each value of m_H, the replacement $\vec{\mu} \rightarrow \mu$ will be made in the notation from here on.

8.3 Hypothesis testing and exclusion limits

Setting limits on parameters of interest can be interpreted in the context of hypothesis testing. In the case of the H $\rightarrow$ ee search, the set of all possible values of the POI that scale the signal yield form a set of null hypotheses. The objective is to use the observed data to reject these hypothesis at various confidence levels. The CL is specified by the

size of the test, α, which corresponds to the probability to reject the null hypothesis, even if it was true.

In order to set upper limits, a test statistic which summarises the observed data in a single value, t_μ, is used. Typical test statistics often use ratios of likelihood functions constructed under a null and alternative hypothesis for μ, since these maximise the *power* of the test for a given test size, and have useful asymptotic properties [101, 120]. The test statistic distribution, $f(t_\mu, \vec{\theta})$, is constructed for each considered value of μ. The observed value of t_μ in data, t_μ^{obs}, can then be used to compute a p-value; comparing the p-value with the chosen confidence level informs whether the null hypothesis under test can be excluded. For this analysis, the chosen test statistic is given by:

$$t_\mu = \begin{cases} -2\log\mathcal{L}(\mu, m_H, \hat{\vec{\theta}}_\mu) + 2\log\mathcal{L}(0, m_H, \hat{\vec{\theta}}_0), & \hat{\mu} < 0; \\ -2\log\mathcal{L}(\mu, m_H, \hat{\vec{\theta}}_\mu) + 2\log\mathcal{L}(\hat{\mu}, m_H, \hat{\vec{\theta}}), & \hat{\mu} < (0, \mu]; \\ 0, & \hat{\mu} > \mu, \end{cases} \tag{8.3}$$

where $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the global best fit value of μ and $\vec{\theta}$ respectively, for all values of μ tested, while $\hat{\vec{\theta}}_\mu$ are the best-fit values of $\vec{\theta}$ for the current tested POI value, μ. The test statistic has three regimes. The first case prevents unphysical values of $\hat{\mu} < 0$, while the third imposes an upper limit on μ. The second is equivalent to the ratio of log-likelihoods, constructed under the tested value of μ, the null hypothesis, and under the best fit value, $\hat{\mu}$, the alternative hypothesis. The distribution of the test statistic can be determined analytically or from pseudo experiments.

In this analysis, the CL_s procedure [121–123] is used to set exclusion limits. This method uses the distribution of the test statistic to compute a ratio of p-values, obtained from integrating t_μ under two hypotheses: the signal-plus-background ($\mu \neq 0$), and background-only ($\mu = 0$) scenarios. It is expressed as:

$$\begin{aligned} CL_s &= \frac{p_{s+b}}{1 - p_b} \\ &= \frac{\int_{t_\mu^{obs}}^{\infty} f(t_\mu|\mu, \vec{\theta} = \hat{\vec{\theta}}_\mu^{obs}) dt_\mu}{\int_{t_0^{obs}}^{\infty} f(t_\mu|0, \vec{\theta} = \hat{\vec{\theta}}_0^{obs}) dt_\mu}, \end{aligned}$$

where the nuisance parameters are set to their best-fit values for the chosen μ being tested. The CL_s ratio is designed to make the test more conservative, avoiding excluding values of μ where the data also does not favour the background-only hypothesis. The smallest value of μ which satisfies $CL_s < \alpha$ is quoted as the upper limit on μ, at the $1 - \alpha$ confidence level.

8.4 Limits on $\mathcal{B}(H \rightarrow ee)$

This analysis reports exclusion limits on the branching fraction for H $\rightarrow$ ee decays at the 95% confidence level ($\alpha = 0.05$), for a selection of Higgs boson mass hypotheses between 120 and 130 GeV. Example m_{ee} distributions, along with corresponding signal-plus-background models, are shown for the highest S/B categories targeting ggH and VBF events in Figure 8.1, for a H $\rightarrow$ ee branching fraction scaled to the observed limit (at the 95% CL), and assuming a Higgs boson mass of 125.38 GeV [102].

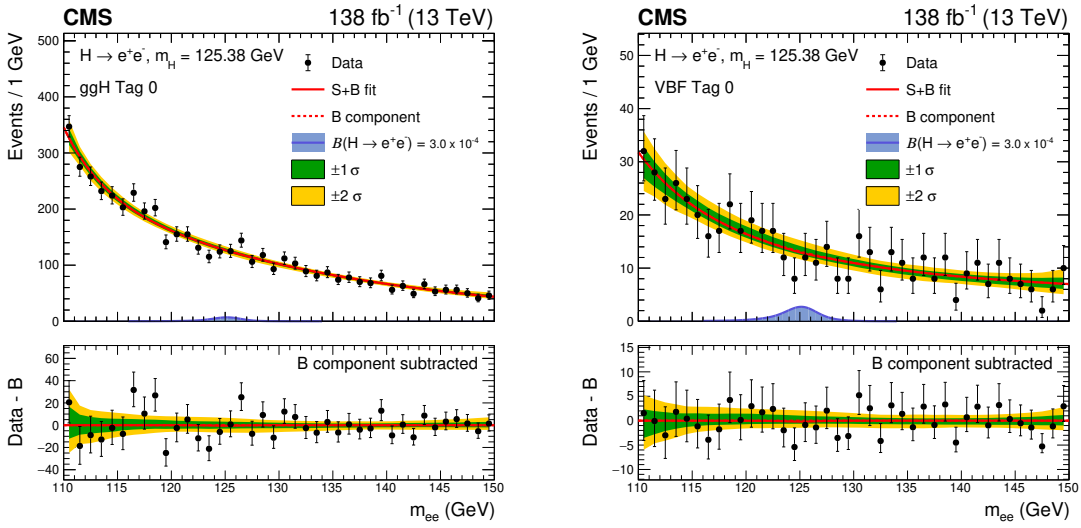

Figure 8.1: The signal-plus-background model fit (solid red line) to the m_{ee} distribution for the highest S/B analysis categories targeting the ggH (left) and VBF (right) processes. The background-only component is also shown (dashed red line). The signal model for each category is shown (blue), scaled to the observed limit at $m_H = 125.38$ GeV. The one (green) and two (yellow) standard deviation bands show the uncertainties in the background component of the fit. The lower panel shows the residual difference after subtraction of this background component.

Exclusion limits on $\mathcal{B}$(H $\rightarrow$ ee) are presented in Figure 8.2, for selected values of the Higgs boson mass. At the current best measured value of $m_H = 125.38$ GeV [102], the observed 95% CL limit is

$$\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee}) < 3.0 \times 10^{-4}, \tag{8.4}$$

while the expected limit is (also) $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee}) < 3.0 \times 10^{-4}$. The p-value for this fit is 90%, indicating good compatibility with the SM prediction. Limits at alternative Higgs boson mass points are of similar magnitude, with the observed limit consistently within the 1σ uncertainty interval of the expected value. Limits are presented for each analysis category independently in Figure 8.3, where it can be seen that the overall expected sensitivity is

driven equally by ggH and VBF categories.

Finally, the limit on the H $\rightarrow$ ee branching fraction is translated into an upper bound on the effective coupling modifier to electrons, $|\kappa_e|$. The coupling modifier is defined as the ratio of the observed electron-Yukawa coupling, to the SM prediction, $|\kappa_e| = |Y_e^{obs}|/|Y_e^{\mathrm{SM}}|$, such that deviations from $|\kappa_e| = 1$ indicate non-SM behaviour. It is related to the H $\rightarrow$ ee branching fraction by

$$\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee}) = \frac{|\kappa_e|^2 \mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})_{\mathrm{SM}}}{1 + (|\kappa_e|^2 - 1)\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})_{\mathrm{SM}}}, \tag{8.5}$$

where $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})_{\mathrm{SM}}$ is the H $\rightarrow$ ee branching fraction predicted by the SM. At a nominal Higgs boson mass of $m_H = 125.38$ GeV, an upper limit on the effective coupling modifier to electrons is observed at $|\kappa_e| < 240$.

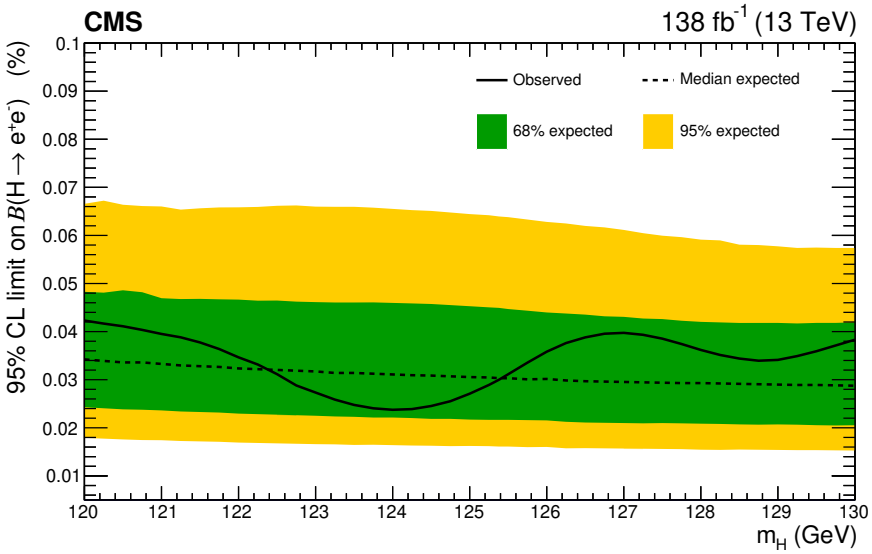

Figure 8.2: Expected and observed limits on $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})$ for Higgs boson masses between 120 and 130 GeV, along with the associated 1σ (green) and 2σ (yellow) CL intervals. The observed limits are within the 1σ interval of the expected limits for all considered values of the Higgs boson mass.

8.5 Summary

The results of this analysis set exclusion limits on the branching fraction for H $\rightarrow$ ee decays. This is achieved using a maximum likelihood fit to the dielectron mass distribution across all analysis categories. Limits computed using the CL_s method are presented both as a function of the Higgs boson mass, and per analysis category. The contribution to the overall sensitivity from categories targeting ggH and VBF production is similar in

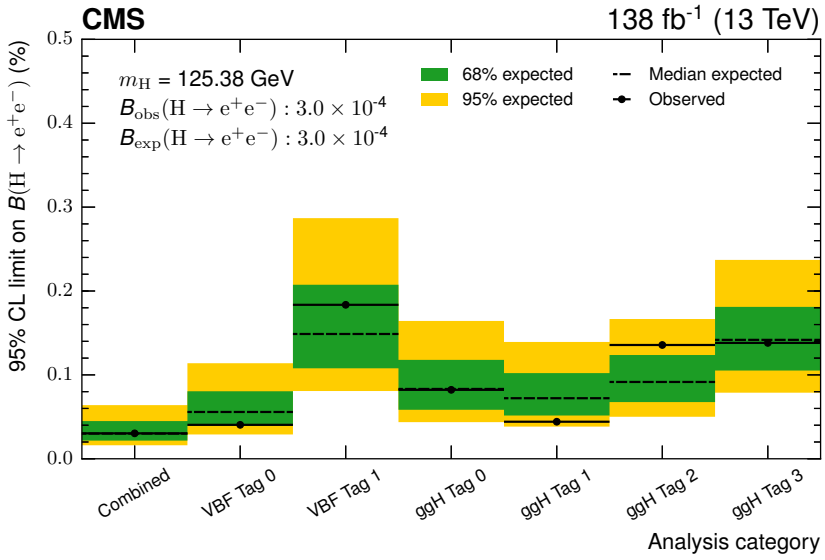

Figure 8.3: Expected and observed limits on $\mathcal{B}(\mathrm{H} \to \mathrm{ee})$ for each constructed analysis category, and all categories combined. The relative contribution to the overall sensitivity is similar between the most sensitive categories targeting ggH and VBF events. The results here assume a SM Higgs boson with mass of 125.38 GeV.

magnitude. For a nominal Higgs boson mass of 125.38 GeV, the observed (expected) limit at the 95% CL on $\mathcal{B}(\mathrm{H} \to \mathrm{ee})$ is found to be 3.0×10^{-4} (3.0×10^{-4}). Results are also translated into an upper limit on the Higgs boson effective coupling modifier to electrons, $|\kappa_e|$, which is found to be $|\kappa_e| < 240$. These results provide the most sensitive constraints on the $\mathrm{H} \to \mathrm{ee}$ decay to date.

Chapter 9

Conclusions

The standard model of particle physics provides an extremely successful description of the fundamental particles and forces, with its predictions surviving much experimental scrutiny. After the discovery of the Higgs boson at the LHC, the entire particle content of the SM has now been experimentally confirmed. Despite this, the SM fails to provide explanations for a number of observed phenomena, including neutrino masses, the presence of dark matter, and incorporation of the gravitational force. These shortcomings motivate theories beyond the SM, many of which modify the Higgs sector. A detailed characterisation of the Higgs boson is therefore a priority for particle physics experiments, such as those at the LHC.

The properties of the Higgs boson can be studied in a variety of ways, including through precision measurements in experimentally clean decay channels, or via searches for rare and forbidden interactions. This thesis presents a search for the rare decay of the Higgs boson to two electrons. The search is performed using pp collision data collected at $\sqrt{s} = 13$ TeV by the CMS experiment at the LHC between 2016 and 2018, corresponding to an integrated luminosity of 138 fb^{-1}. Analysis categories are developed to target Higgs boson production via both gluon fusion and vector boson fusion. To improve the ratio of signal-to-background events, dedicated BDTs are trained for each Higgs boson production mode. These classifiers use characteristic features of the event topology to improve rejection of background, which consists primarily of Drell-Yan $\text{Z} \rightarrow \text{ee}$ and $\text{t}\bar{\text{t}}$ decays. The modelling of signal events by the BDT is also validated using a sample of $\text{Z} \rightarrow \text{ee}$ events in a control region orthogonal to the selection defining the signal region. Since the analysis sensitivity is not close to the SM prediction, upper limits on the branching fraction for $\text{H} \rightarrow \text{ee}$ decays are extracted. Limits are determined via a maximum likelihood fit to the dielectron mass distribution in each analysis category. Models for signal events are derived using simulated samples at various Higgs boson masses, while background models are taken directly from a sideband region in data. The

observed upper limit at the 95% confidence level on the branching fraction for the H $\rightarrow$ ee decay is

$$\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee}) < 3.0 \times 10^{-4}.$$

This limit is also translated into an upper bound on the Higgs boson coupling modifier to electrons, yielding $|\kappa_e| < 240$.

These results provide the most stringent constraints on the Higgs boson to electron pair decay channel to date. However, the observed limits remain orders of magnitude greater than the corresponding SM predictions. In order to improve the precision of existing measurements, including those in the H $\rightarrow$ ee channel, upgrades to the LHC, such as those discussed in Section 4.4, are crucial. In particular, the HL-LHC upgrade project is expected to collect over twenty times the data analysed in this thesis, improving the statistical precision of measurements made with the LHC Run 2 dataset by almost a factor of five. The upgrade will also facilitate entirely new measurements, such as a detailed characterisation of the Higgs boson potential. These benefits are, however, accompanied by several experimental challenges, which will require significant modifications to the existing detector technologies and reconstruction techniques, such as those developed in the CMS HGCAL project. Many of the proposed solutions are expected to rely heavily on the application of sophisticated ML algorithms, some of which have been discussed in chapter 3. Overall, however, the expected improvement in sensitivity offered by the HL-LHC is unfortunately not near to closing the five orders of magnitude difference between the observed limit on $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})$ presented in this thesis, and the SM prediction.

Beyond the LHC, future colliders offer a unique opportunity to probe the electron-Yukawa coupling. In particular, the Future Circular Collider [124], a proposed 100 km long successor to the LHC, is expected to collide electron and positron beams as part of its precision physics programme [125]. The FCC electron-positron collider will be operated at a selection of threshold energies for various electroweak processes, allowing for tens of attobarns of data to be collected. Dedicated runs at $\sqrt{s} = m_H$ would allow for resonant Higgs boson production via the electron fusion s-channel process, ee $\rightarrow$ H, facilitating measurements of the electron-Yukawa coupling at the Higgs boson production vertex. Recent feasibility studies using 10 ab^{-1} of integrated luminosity project a sensitivity to this coupling of 1.6 times the SM prediction [126], providing unparalleled precision when compared with previous direct measurements. Future colliders are thus an exciting prospect with which to probe properties of the Higgs boson, and elucidate the nature of the Yukawa coupling to the first generation fermions. Through such tests, performed at the LHC and beyond, it is hoped that explanations for the shortcomings of the SM will be found, and our understanding of the universe, improved.

Appendix A

Electron energy scale and resolution corrections

The cumulative impact of the electron reconstruction process (summarised in Eqn 4.3), including application of the electron energy scale and resolution corrections, is shown in Figure A.1, for electrons from $Z \to ee$ decays.

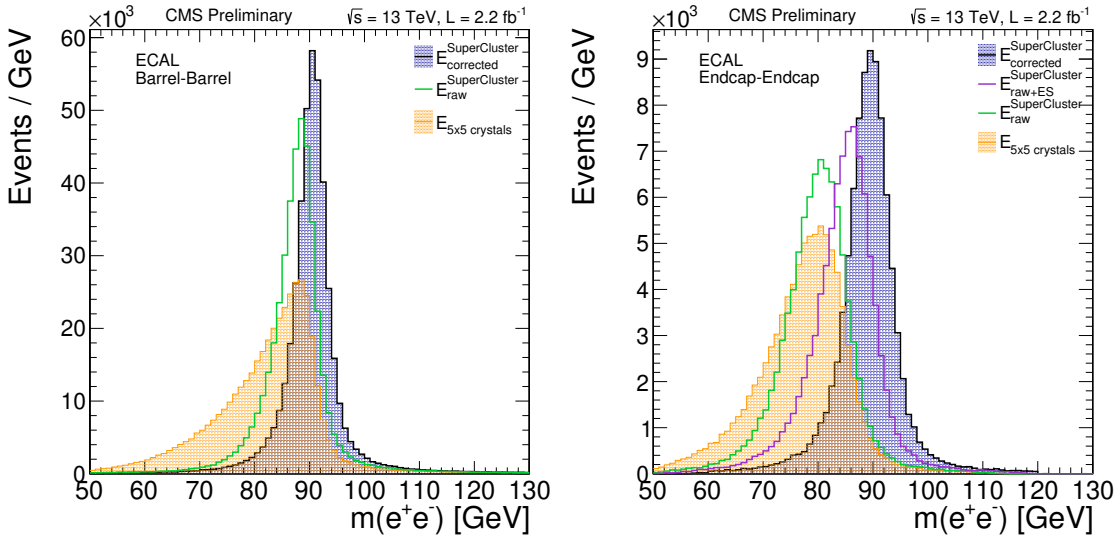

Figure A.1: The invariant mass of electrons from Z → ee decays, shown at different stages of the electron reconstruction process, including application of the electron energy scale and resolution corrections. Electrons reconstructed in the EB (EE) region are shown in the left (right) plot. The data shown were collected during the 2015 period of the LHC operation. Each stage in the reconstruction is compared to a simple energy sum of the 5×5 array of ECAL crystals centred on the electron candidate, the distribution for which is shown in yellow. Following the clustering process, the invariant mass determined using the electron supercluster properties is shown in the green histogram. For electrons reconstructed in the endcap, the invariant mass computed using the combination of the supercluster energy and the energy deposited in the ES is shown by the purple histogram. The final invariant mass distribution, following the application of the electron energy scale and resolution corrections, is shown in the blue histogram. Figure taken from Ref [127].

Appendix B

The Z boson transverse momentum correction

Dedicated corrections are applied to the dielectron p_T spectrum in simulated Drell-Yan events in order to improve the agreement between simulation and data. These aim to correct differences with respect to data at low p_T, where the emission of soft gluons and associated re-summation effects are not accounted for in simulation [128]. This leads to a softer p_T spectrum in simulation, compared to data. To correct for this effect, a simple reweighting of the simulation to data is performed. This correction acts only to improve the performance of the classifiers used in the event categorisation (that are trained on DY events), rather than to correct background models, which are taken directly from data. The reweighting is derived in bins of dielectron p_T in the analysis control region around the Z boson mass. Scale factors are derived in a binning scheme that roughly evolves with the yield in the control region; the bin widths are 1.6 GeV, 4 GeV, and 5 GeV for the regions $0 < p_{T,ee} \leq 40$ GeV, $40 < p_{T,ee} \leq 80$ GeV, $80 < p_{T,ee} \leq 180$ GeV respectively. No correction is derived for events with $p_{T,ee} > 180$ GeV, since the relative statistical uncertainty on the simulation is large. Corrections are also derived separately for each year of data taking, to account for potential per-year differences. To illustrate the impact of this correction, Figure B.1 shows the dielectron p_T distribution in the signal region, before and after the p_T reweighting is applied. Good agreement is observed between the corrected simulation and data, particularly in the low p_T regime. A doubly-differential reweighting was also attempted by further partitioning events into bins of jet multiplicity (0-jet, 1-jet, and $\geq$ 2-jets). The additional binning does not significantly improve the agreement between simulation and data; it is therefore sufficient to reweight DY samples in bins of $p_{T,ee}$ only.

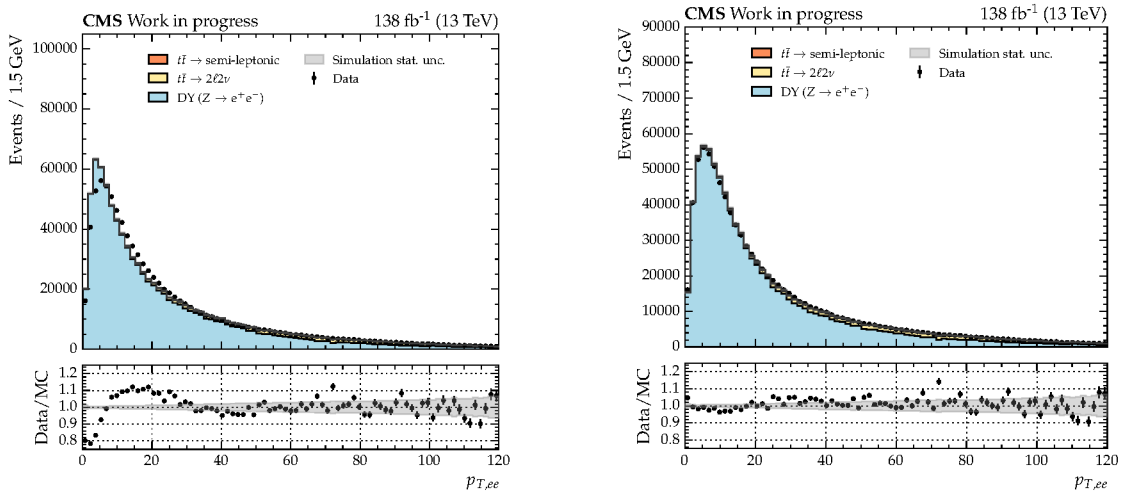

Figure B.1: Distributions of $p_{T,ee}$ for simulated Drell-Yan events, additional $\mathrm{t\bar{t}}$ backgrounds, and data in the signal region. The statistical uncertainty on the total background yield is shown in the grey shaded band. Left: nominal $p_{T,ee}$ distribution with no correction applied. Right: $p_{T,ee}$ distribution where Drell-Yan simulation is corrected by scale factors derived in a control region around the Z boson mass. A significant improvement in agreement between data and simulation is observed following these corrections, particularly in the low $p_{T,ee}$ regime.

Appendix C

Distribution of input features to the ggH and VBF BDTs

The distribution of each input feature to the ggH BDT is shown in Figures C.1–C.3 for the DY and $t\bar{t}$ backgrounds, data, and ggH signal. The agreement between data and simulated background events is reasonable; however, any non-closure cannot bias the modelling of background events in the final fit. Similarly, Figures C.4–C.8 show inputs to the VBF BDT for simulated DY, $t\bar{t}$, and electroweak Z boson background processes, data, and VBF signal. Agreement between data and simulated background events is good.

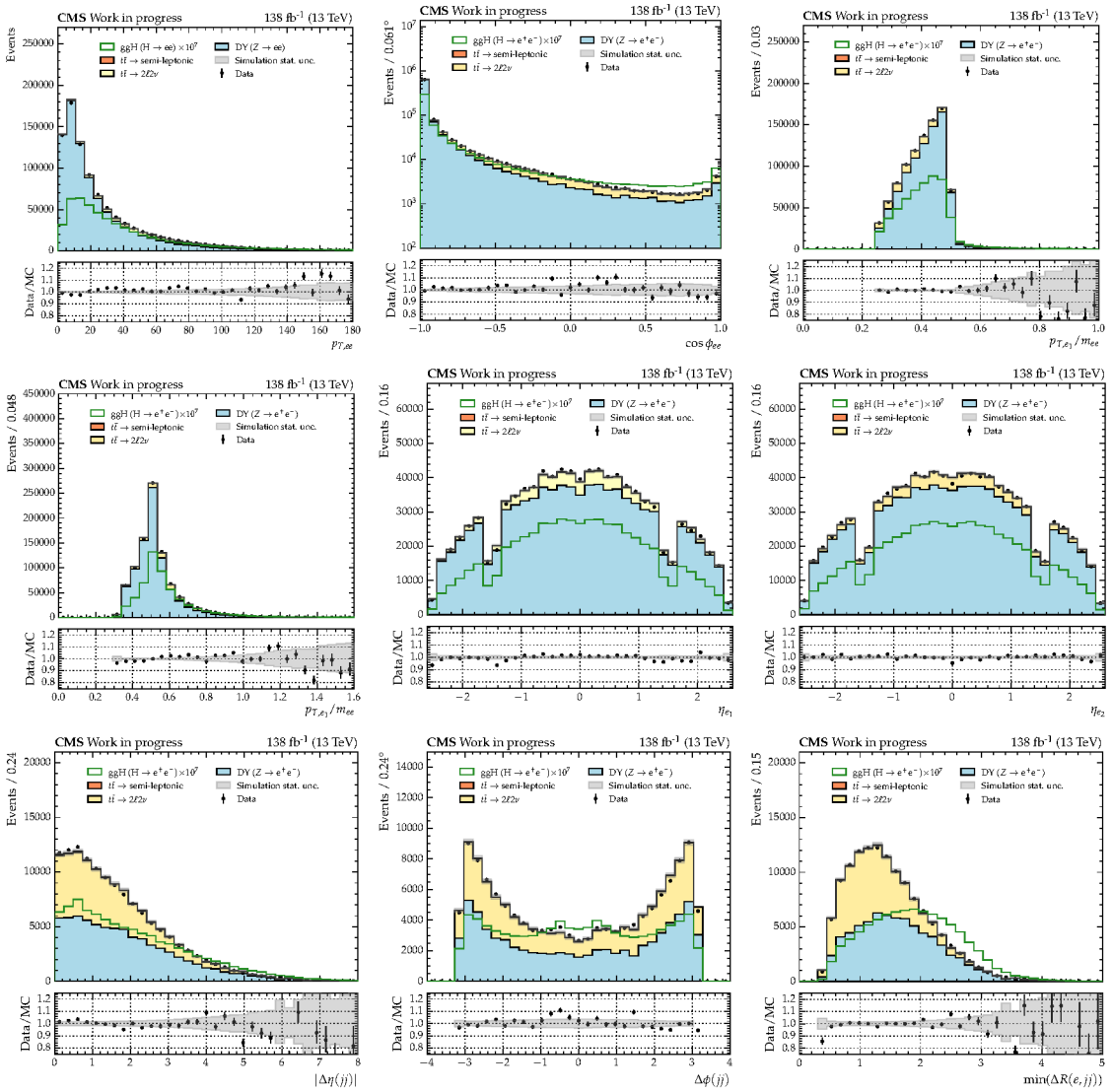

Figure C.1: Distributions for the input variables to the gluon-fusion BDT. The ggH signal is shown in green, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Reasonable agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

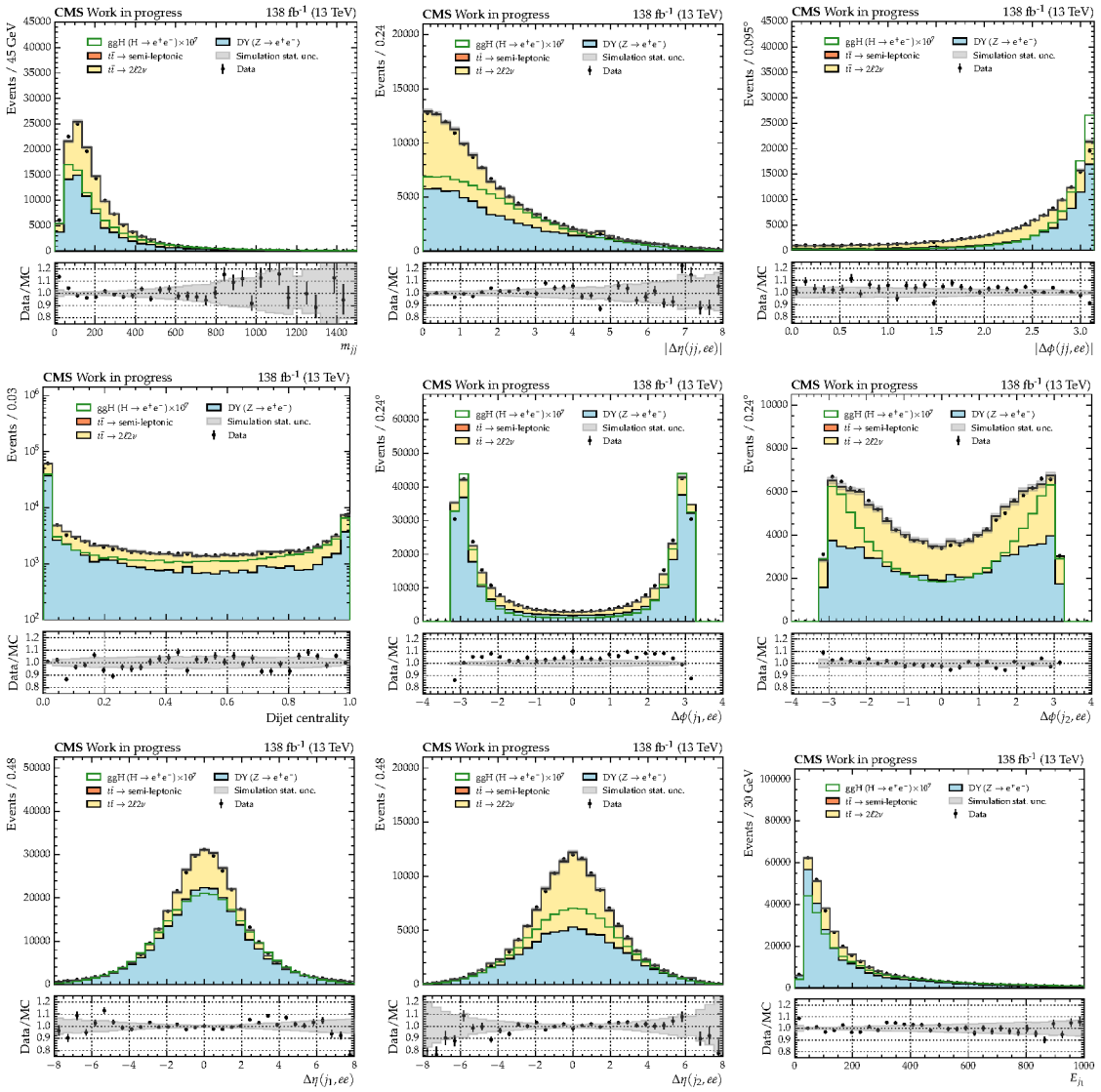

Figure C.2: Distributions for the input variables to the gluon-fusion BDT. The ggH signal is shown in green, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Reasonable agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

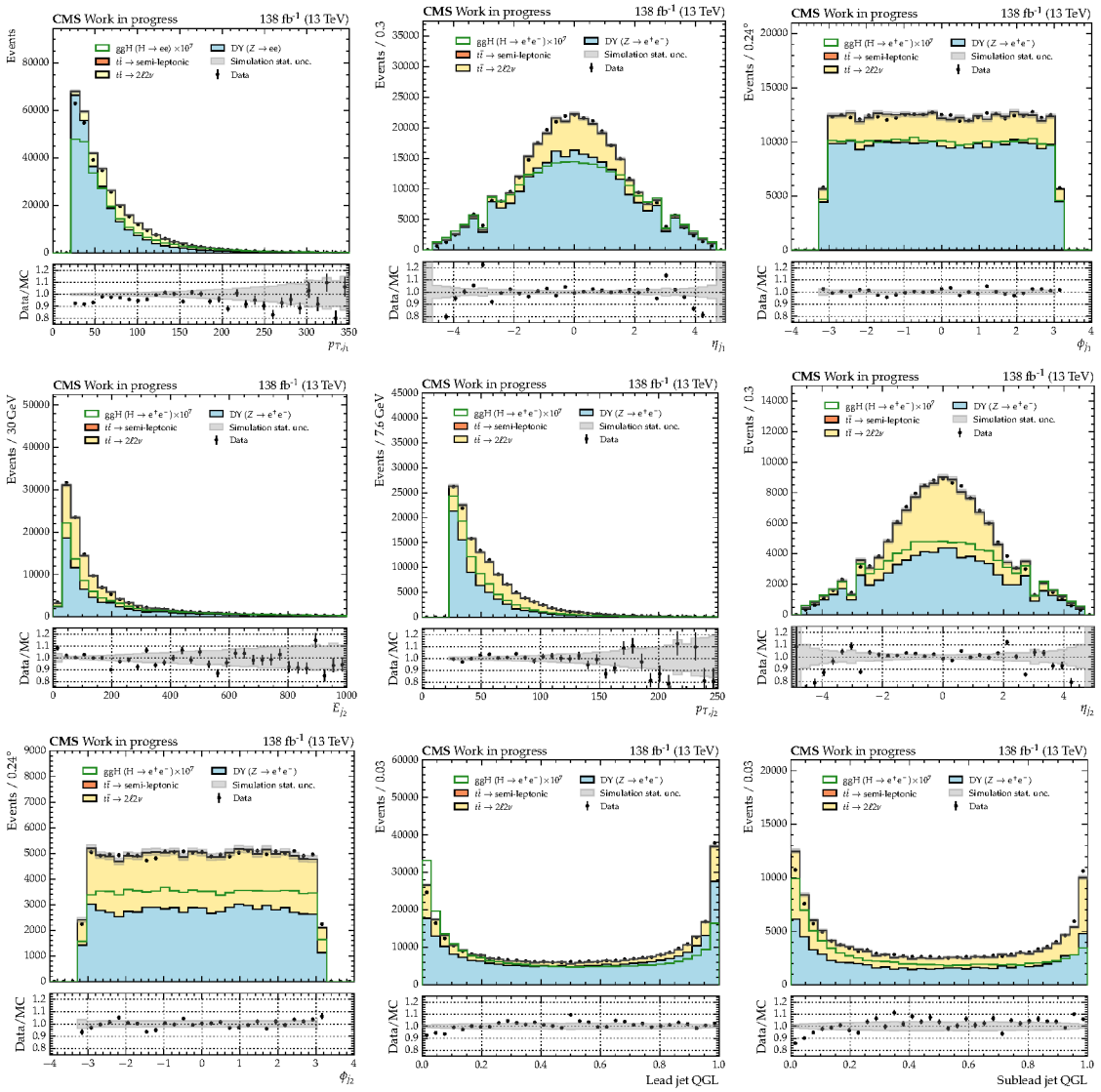

Figure C.3: Distributions for the input variables to the gluon-fusion BDT. The ggH signal is shown in green, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Reasonable agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

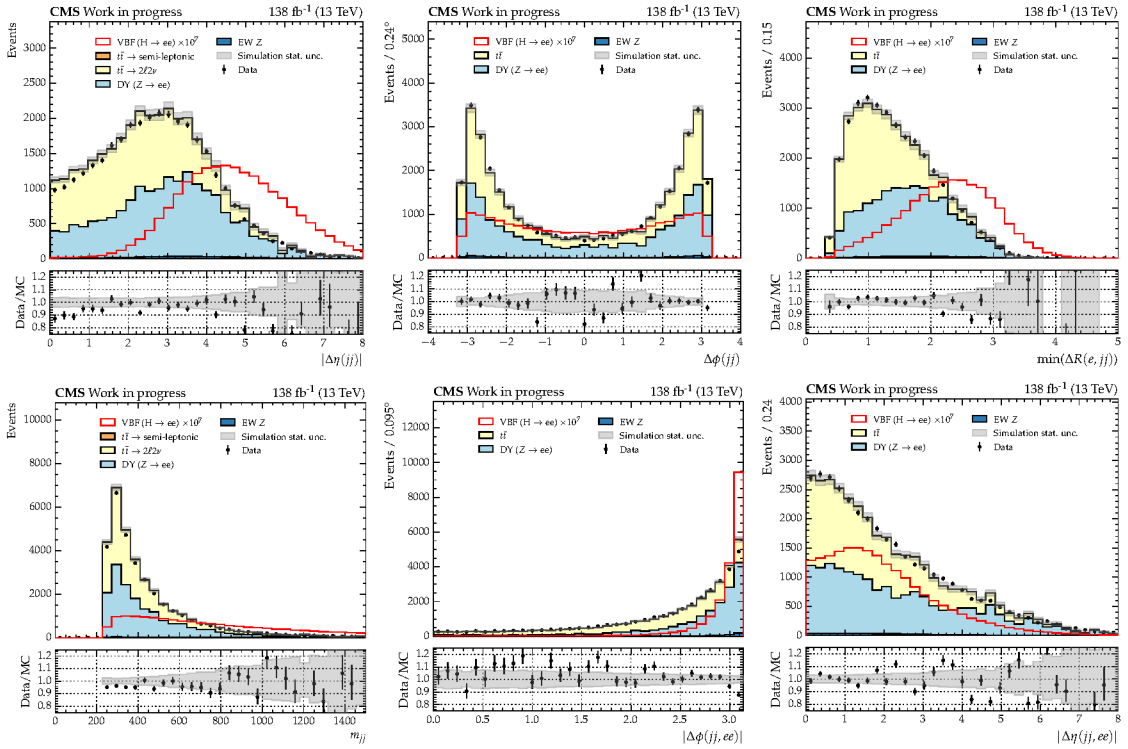

Figure C.4: Distributions for the input variables to the VBF BDT. The VBF signal is shown in red, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Good agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

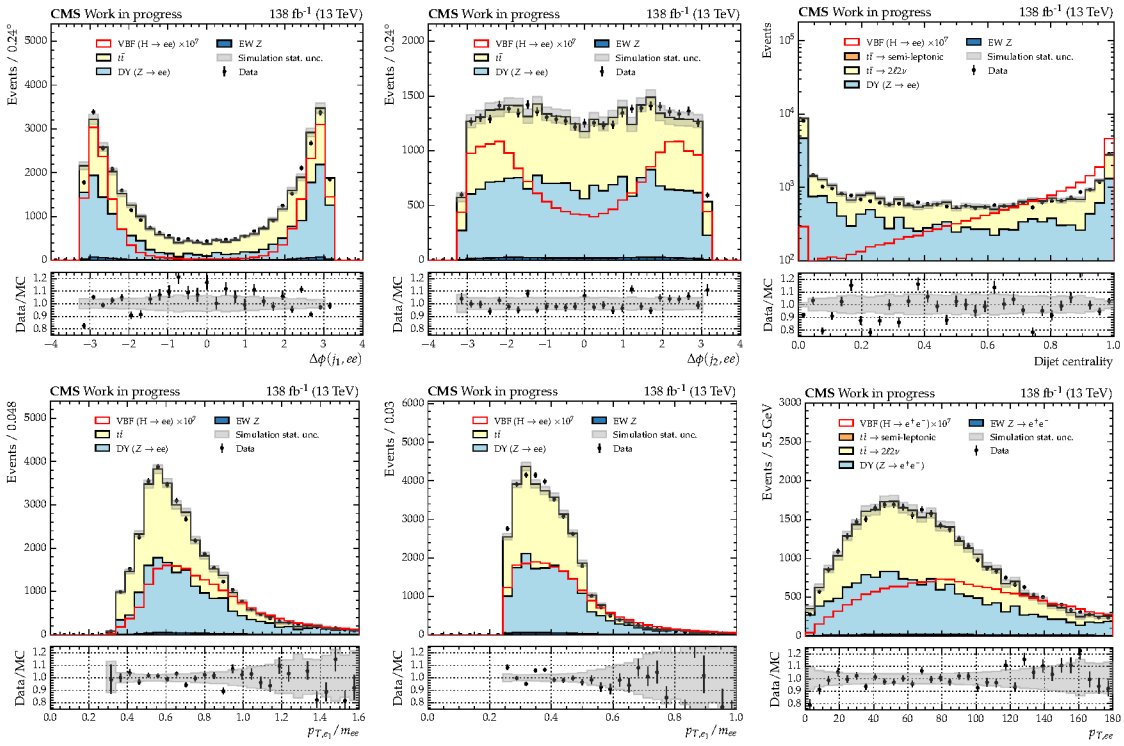

Figure C.5: Distributions for the input variables to the VBF BDT. The VBF signal is shown in red, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Good agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

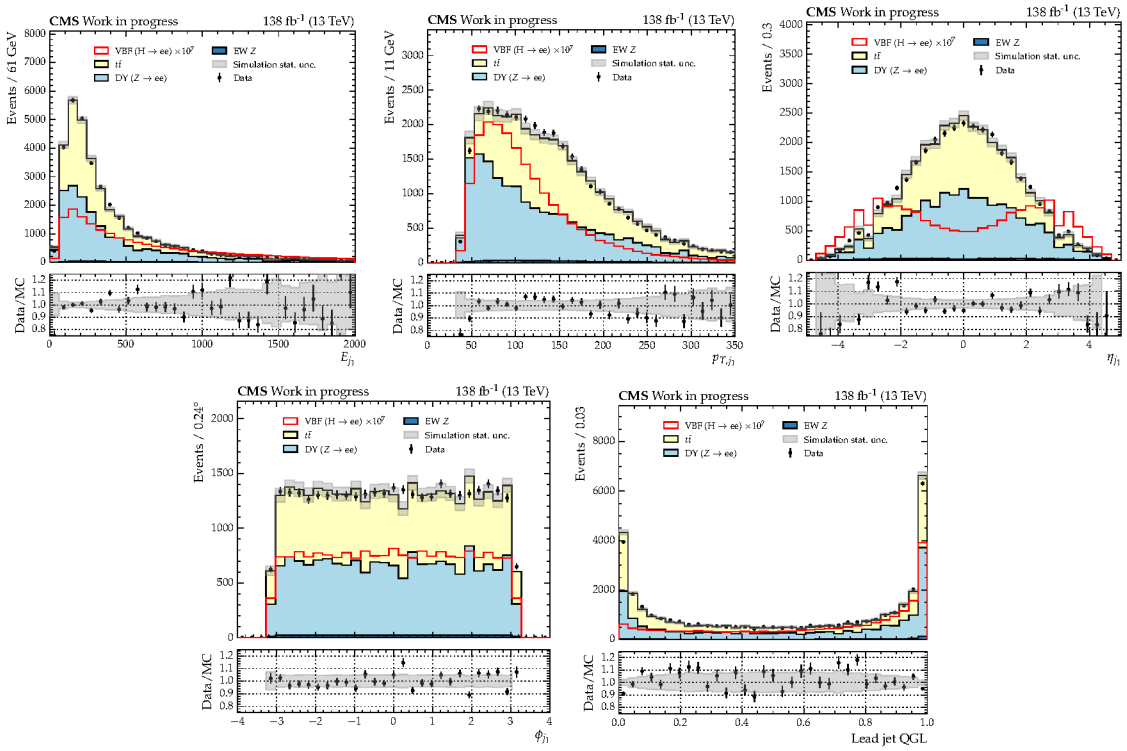

Figure C.6: Distributions for the input variables to the VBF BDT. The VBF signal is shown in red, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Good agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

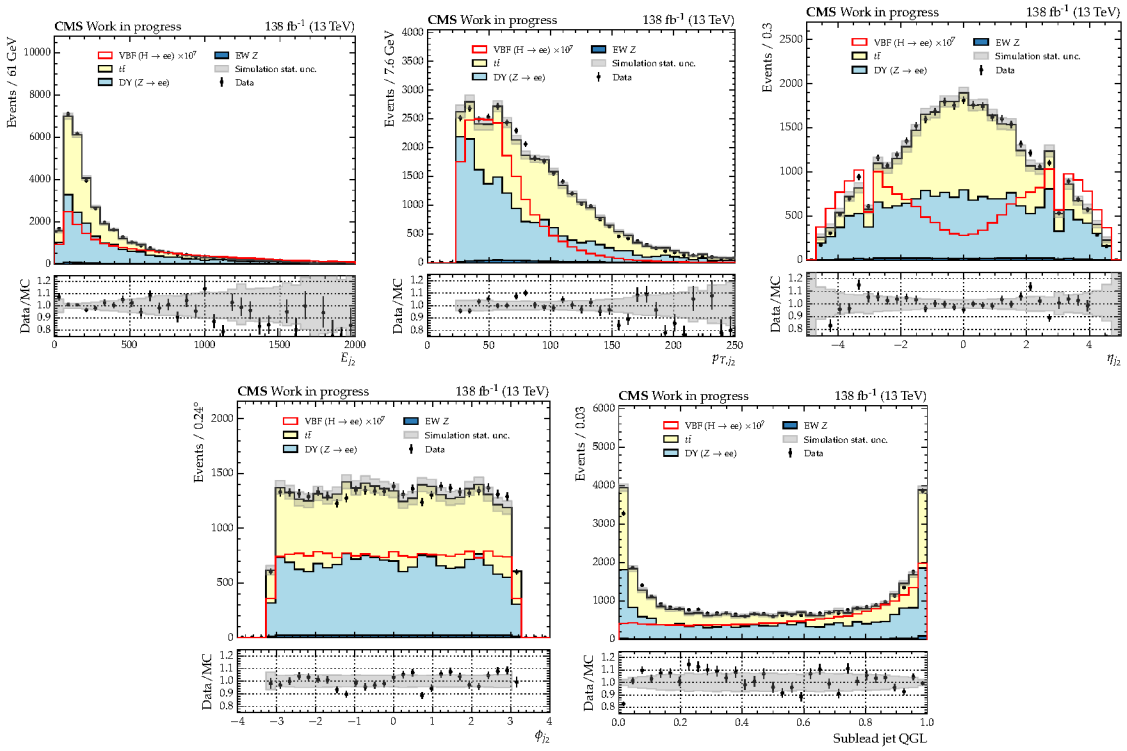

Figure C.7: Distributions for the input variables to the VBF BDT. The VBF signal is shown in red, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Good agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

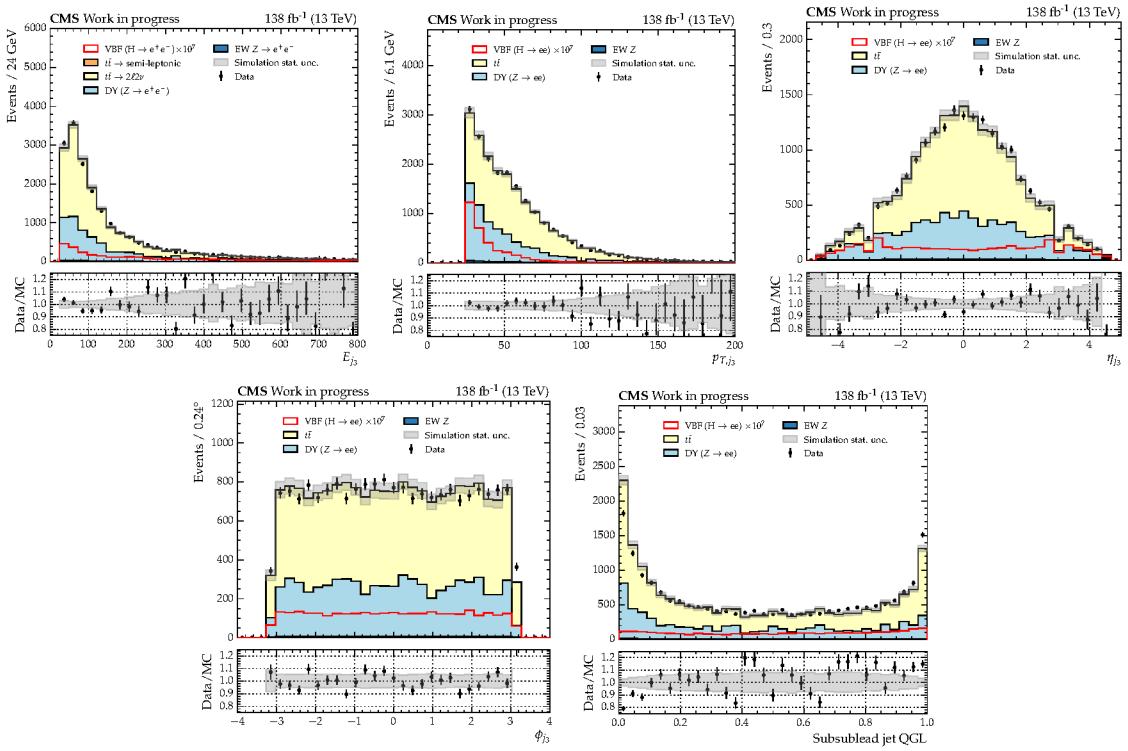

Figure C.8: Distributions for the input variables to the VBF BDT. The VBF signal is shown in red, with the overall normalisation scaled such that it is visible. The simulated background processes (bold face) are stacked for comparison with data (black points). Good agreement is observed between data and simulation, with respect to the statistical uncertainty (grey band).

Appendix D

Impact of the largest systematic uncertainties on $\mathcal{B}(\mathrm{H} \rightarrow \mathrm{ee})$

The impact of the largest systematic uncertainties on the observed H $\rightarrow$ ee branching fraction is shown in Figure D.1. The uncertainties with the largest impact are those affecting the electron energy scale, as well as those affecting the ggH production cross section. Also shown is the pull of each nuisance parameter, defined as the difference between the postfit and prefit uncertainties, divided by the prefit uncertainty. No significant pulls away from zero are observed.

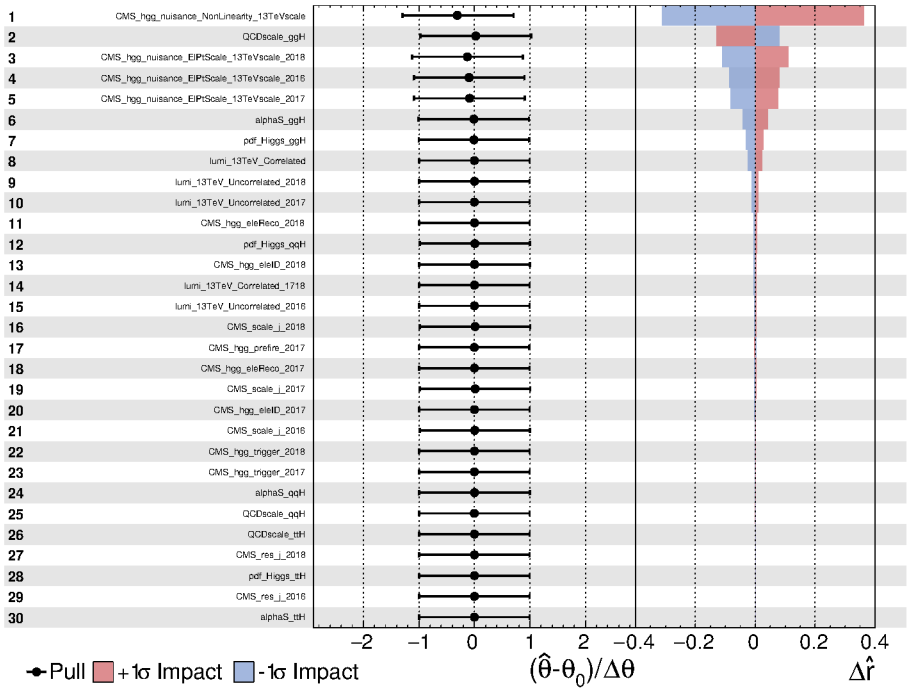

Figure D.1: Left: the pull of each nuisance parameter, defined as the difference between the postfit and prefit uncertainties, divided by the prefit uncertainty. The error bar indicates the ratio of the postfit and prefit uncertainties. No significant pulls away from zero are observed. Nuisance paramaters are also not signficantly constrained from the fit. Right: the impact of fixing nuisance parameters to their $\pm 1\sigma$ values on the observed $\mathrm{H} \to \mathrm{ee}$ branching fraction (labelled here as $\Delta\hat{r}$).

Appendix E

Observed dielectron mass distributions

The observed dielectron invariant mass distributions for the remaining analysis categories not presented in Chapter 8 are shown in Figure E.1.

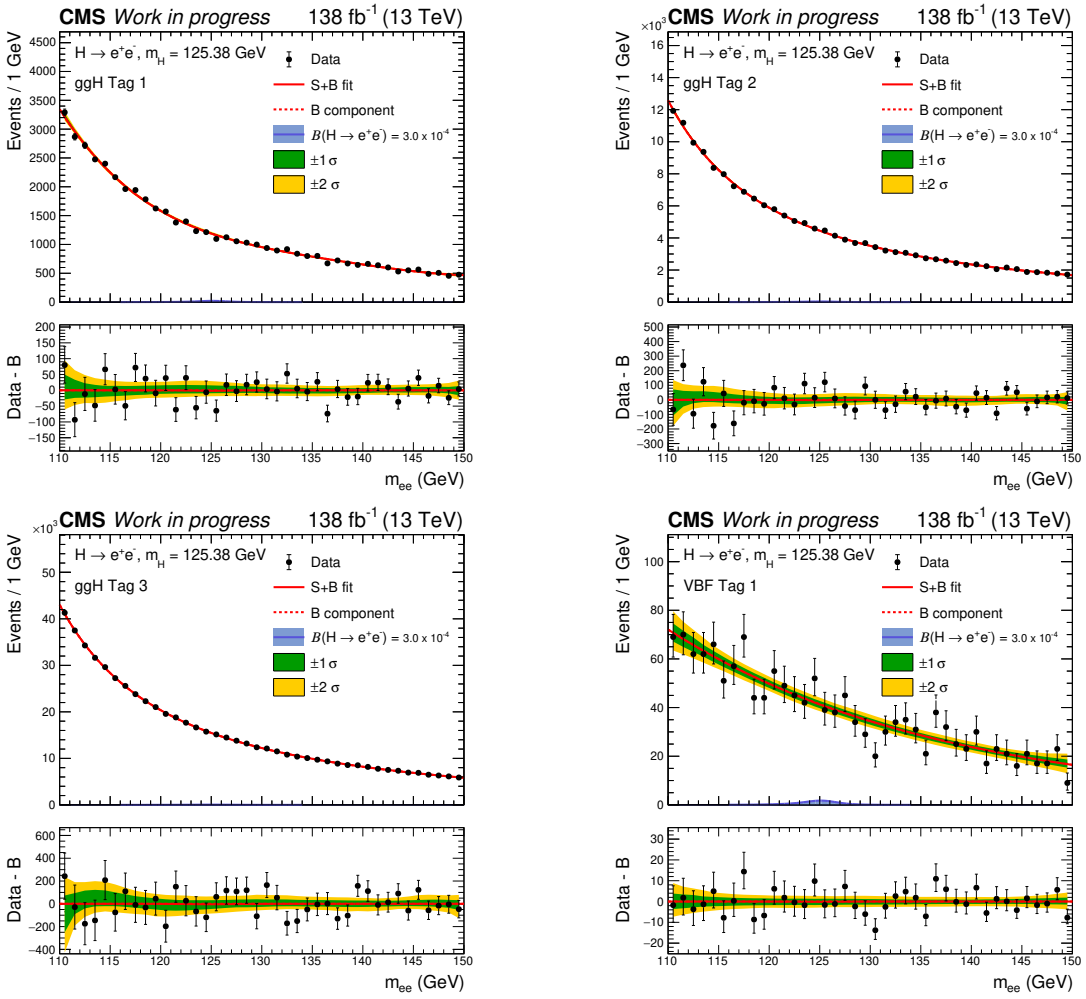

Figure E.1: The dielectron mass distribution for analysis categories targeting ggH production (Tags 1-3), and categories targeting VBF production (Tag 1). The signal-plus-background model fit to the distributions is shown by the solid red line, while the background only component is shown by the dashed red line. The signal model in each category, shown in blue, is scaled to the expected limit at $m_H = 125.38$ GeV. The one (green) and two (yellow) standard deviation bands show the uncertainties in the background component of the fit. The lower panel shows the residuals after subtraction of this background component.

Bibliography

[1] CMS Collaboration. "Search for the Higgs boson decay to a pair of electrons in proton-proton collisions at $\sqrt{s} = 13$ TeV". Tech. rep. (2022).

[2] CMS Collaboration. "Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2021), p. 027. DOI: 10.1007/JHEP07(2021)027. arXiv: 2103.06956 [hep-ex].

[3] S. L. Glashow. "Partial Symmetries of Weak Interactions". *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

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

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

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

[7] P. Higgs. "Broken symmetries, massless particles and gauge fields". *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

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

[9] P. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[10] T. Kibble. "Symmetry breaking in non-Abelian gauge theories". *Phys. Rev.* 155 (1967), pp. 1554–1561. DOI: 10.1103/PhysRev.155.1554.

[11] CMS Collaboration. "The CMS experiment at the CERN LHC". *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[12] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[13] L. Evans and P. Bryant. "LHC Machine". *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[14] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys. Lett. B* 716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[15] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Phys. Lett. B* 716 (2012), p. 30. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[16] CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". *JHEP* 06 (2013), p. 081. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571 [hep-ex].

[17] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003.

[18] CMS Collaboration. "Observation of $t\bar{t}H$ Production". *Phys. Rev. Lett.* 120 (23 2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801.

[19] ATLAS Collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". *Phys. Lett. B* 784 (2018), p. 173. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425 [hep-ex].

[20] CMS Collaboration. "Observation of Higgs boson decay to bottom quarks". *Phys. Rev. Lett.* 121 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[21] ATLAS Collaboration. "Observation of $H \to bb$ decays and VH production with the ATLAS detector". *Phys. Lett. B* 786 (2018), p. 59. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].

[22] CMS Collaboration. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector". *Phys. Lett. B* 779 (2018), p. 283. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex].

[23] ATLAS Collaboration. "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Phys. Rev. D* 99 (2019), p. 072001. DOI: 10.1103/PhysRevD.99.072001. arXiv: 1811.08856 [hep-ex].

[24] CMS Collaboration. "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV". *Phys. Lett. B* 791 (2019), p. 96. DOI: 10.1016/j.physletb.2018.12.073. arXiv: 1806.05246 [hep-ex].

[25] CMS Collaboration. "Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Submitted to EPJC* (2021). arXiv: 2103.04956 [hep-ex].

[26] CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". *Nature* (2022), 9 p. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043.

[27] CMS Collaboration. "Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2021), p. 027. DOI: 10.1007/JHEP07(2021)027. arXiv: 2103.06956 [hep-ex].

[28] ATLAS Collaboration. "Measurements of gluon–gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\mu e\mu$ decay channel in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector". *Physics Letters B* 789 (2019), pp. 508–529. DOI: 10.1016/j.physletb.2018.11.064.

[29] ATLAS Collaboration. "Measurement of the Higgs boson coupling properties in the $H \to ZZ^* \to 4\ell$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Journal of High Energy Physics* 2018.3 (2018). DOI: 10.1007/jhep03(2018)095.

[30] ATLAS Collaboration. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". *Nature* 607.7917 (2022), pp. 52–59. DOI: 10.1038/s41586-022-04893-w. arXiv: 2207.00092 [hep-ex].

[31] CMS Collaboration. "Evidence for Higgs boson decay to a pair of muons". *JHEP* 01 (2021), p. 148. DOI: 10.1007/JHEP01(2021)148. arXiv: 2009.04363 [hep-ex].

[32] LHC Higgs Cross Section Working Group. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". *CERN Yellow Rep. Monogr.* 4 (2013). DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].

[33] ACME Collaboration. "Improved limit on the electric dipole moment of the electron". *Nature* 562.7727 (2018), pp. 355–360. DOI: 10.1038/s41586-018-0599-8.

[34] CMS Collaboration. "Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC". *Phys. Lett. B* 744 (2015), p. 184. DOI: 10.1016/j.physletb.2015.03.048. arXiv: 1410.6679 [hep-ex].

[35] ATLAS Collaboration. "Search for the Higgs boson decays H $\rightarrow$ ee and H $\rightarrow$ eμ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Phys. Lett. B* 801 (2020), p. 135148. DOI: 10.1016/j.physletb.2019.135148. arXiv: 1909.10235 [hep-ex].

[36] S. D. Breeze. "Precision measurement of the Z invisible width with the CMS experiment". PhD thesis. Imperial College London, (2019).

[37] E. Noether. "Invariant variation problems". *Transport Theory and Statistical Physics* 1.3 (1971), pp. 186–207. DOI: 10.1080/00411457108231446.

[38] M. Thomson. "Modern Particle Physics". Cambridge University Press, (2013).

[39] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". *Phys. Rev.* 105.4 (1957), 1413–1415. DOI: {10.1103/PhysRev.105.1413}.

[40] LHC Higgs Cross Section Working Group. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the nature of the Higgs sector". *CERN Yellow Rep. Monogr.* 2 (2017). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[41] W. Altmannshofer, J. Brod, and M. Schmaltz. "Experimental constraints on the coupling of the Higgs boson to electrons". *Journal of High Energy Physics* 2015.5 (2015). DOI: 10.1007/jhep05(2015)125.

[42] A. Dery, C. Frugiuele, and Y. Nir. "Large Higgs-electron Yukawa coupling in 2HDM". *Journal of High Energy Physics* 2018.4 (2018). DOI: 10.1007/jhep04(2018)044.

[43] G. F. Giudice and O. Lebedev. "Higgs-dependent Yukawa couplings". *Physics Letters B* 665.2-3 (2008), pp. 79–85. DOI: 10.1016/j.physletb.2008.05.062.

[44] M. Feickert and B. Nachman. "A Living Review of Machine Learning for Particle Physics". (2021). arXiv: 2102.02770 [hep-ph].

[45] T. Hastie, R. Tibshirani, and J. Friedman. "The Elements of Statistical Learning". Springer New York Inc., (2001).

[46] Y. LeCun et al. "Efficient BackProp". *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*. Springer-Verlag, (1998), pp. 9–50.

[47] I. Goodfellow, Y. Bengio, and A. Courville. "Deep Learning". http://www.deeplearningbook.org. MIT Press, (2016).

[48] A. P. Bradley. "The use of the area under the ROC curve in the evaluation of machine learning algorithms". *Pattern Recognition* 30.7 (1997), pp. 1145–1159. DOI: https://doi.org/10.1016/S0031-3203(96)00142-2.

[49] J. R. Quinlan. “Induction of Decision Trees”. *Mach. Learn.* 1.1 (1986), pp. 81–106. DOI: 10.1023/A:1022643204877.

[50] C. M. Bishop. “Pattern Recognition and Machine Learning (Information Science and Statistics)”. Springer-Verlag, (2006).

[51] R. E. Schapire. “The Strength of Weak Learnability”. English (US). *Machine Learning* 5.2 (1990), pp. 197–227. DOI: 10.1023/A:1022648800760.

[52] N. Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. *J. Mach. Learn. Res.* 15.1 (2014), pp. 1929–1958.

[53] P. W. Battaglia et al. “Relational inductive biases, deep learning, and graph networks”. (2018). arXiv: 1806.01261.

[54] Y. Bengio, P. Simard, and P. Frasconi. “Learning long-term dependencies with gradient descent is difficult”. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* 5 (1994), pp. 157–66. DOI: 10.1109/72.279181.

[55] S. Hochreiter and J. Schmidhuber. “Long Short-term Memory”. *Neural computation* 9 (1997), pp. 1735–80. DOI: 10.1162/neco.1997.9.8.1735.

[56] K. Cho et al. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”. (2014). arXiv: 1406.1078 [cs.CL].

[57] K. Greff et al. “LSTM: A Search Space Odyssey”. *IEEE Transactions on Neural Networks and Learning Systems* 28.10 (2017), pp. 2222–2232. DOI: 10.1109/tnnls.2016.2582924.

[58] LHCb Collaboration. “The LHCb Detector at the LHC”. *Journal of Instrumentation* 3.08 (2008), S08005.

[59] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. *Journal of Instrumentation* 3.08 (2008), S08002.

[60] D Contardo et al. “Technical Proposal for the Phase-II Upgrade of the CMS Detector”. CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02 (2015).

[61] CERN. “LEP design report”. CERN-LEP-84-01, (1984).

[62] CMS Collaboration. “CMS Luminosity - Public Results”. http://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. [Online; accessed 10-Jan-2022].

[63] T. Sakuma and T. McCauley. “Detector and Event Visualization with SketchUp at the CMS Experiment”. (2014).

[64] CMS Collaboration. "CMS Technical Design Report for the Pixel Detector Upgrade". Tech. rep. CERN-LHCC-2012-016, CMS-TDR-011. (2012).

[65] CMS Collaboration. "The Phase-2 Upgrade of the CMS Tracker". Tech. rep. (2017). DOI: 10.17181/CERN.QZ28.FLHW.

[66] A. Benaglia. "The CMS ECAL performance with examples". CMS-CR-2013-430 (2013). DOI: 10.1088/1748-0221/9/02/C02008.

[67] CMS Collaboration. "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, (1997).

[68] CMS Collaboration. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". *JINST* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[69] CMS Collaboration. "The CMS trigger system". *JINST* 12 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[70] G. Apollinari et al. "High Luminosity Large Hadron Collider HL-LHC". *CERN Yellow Rep.* 5 (2015), pp. 1–19. DOI: 10.5170/CERN-2015-005.1. arXiv: 1705.08830 [physics.acc-ph].

[71] CMS Collaboration. "The Phase-2 Upgrade of the CMS Endcap Calorimeter". CERN-LHCC-2017-023. CMS-TDR-019 (2017).

[72] CMS Collaboration. "Sensitivity projections for Higgs boson properties measurements at the HL-LHC" (2018).

[73] CMS Collaboration. "The Phase-2 Upgrade of the CMS Level-1 Trigger". CERN-LHCC-2020-004. CMS-TDR-021 (2020).

[74] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". Proceedings for the Association for Computing Machinery, (2016). DOI: 10.1145/2939672.2939785.

[75] L. S. Shapley. "17. A Value for n-Person Games". *Contributions to the Theory of Games (AM-28), Volume II*. Princeton University Press, (2016), pp. 307–318. DOI: doi:10.1515/9781400881970-018.

[76] S. M. Lundberg and S. Lee. "A Unified Approach to Interpreting Model Predictions". *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., (2017), pp. 4765–4774.

[77] M. Kursa, A. Jankowski, and W. Rudnicki. "Boruta - A System for Feature Selection". *Fundam. Inform.* 101 (2010), pp. 271–285. DOI: 10.3233/FI-2010-288.

[78] CMS Collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". *JINST* 16 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[79] CMS Collaboration. "Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the CMS experiment". *Journal of High Energy Physics* 2011.10 (2011). DOI: 10.1007/jhep10(2011)132.

[80] 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". *Journal of High Energy Physics* 2014.7 (2014). DOI: 10.1007/jhep07(2014)079.

[81] T. Sjöstrand et al. "An introduction to PYTHIA 8.2". *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[82] CMS Collaboration. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". *Eur. Phys. J. C* 80 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[83] P. Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].

[84] S. Frixione, P. Nason, and C. Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[85] S. Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[86] S. Frixione, P. Nason, and G. Ridolfi. "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction". *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[87] S. Agostinelli et al. "GEANT4: a simulation toolkit". *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[88] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[89] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". *JINST* 9 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[90] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. DOI: https://doi.org/10.1016/0168-9002(87)90887-4.

[91] CMS Collaboration. "Technical Proposal for the Phase-II Upgrade of the CMS Detector" (2015). Ed. by D. Contardo et al.

[92] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ex].

[93] W. Adam et al. "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC". *Journal of Physics G: Nuclear and Particle Physics* 31.9 (2005), N9–N20. DOI: 10.1088/0954-3899/31/9/n01.

[94] CMS Collaboration. "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$=8 TeV". *JINST* 10.08 (2015), P08010. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702 [physics.ins-det].

[95] CMS Collaboration. "ECAL 2016 refined calibration and Run 2 summary plots". CMS Detector Performance Note CMS-DP-2020-021. (2020).

[96] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST* 13 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[97] CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". *JINST* 12 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[98] CMS Collaboration. "Pileup mitigation at CMS in 13 TeV data". *JINST* 15 (2020), P09018. DOI: 10.1088/1748-0221/15/09/P09018. arXiv: 2003.00503 [hep-ex].

[99] D. Rainwater, R. Szalapski, and D. Zeppenfeld. "Probing color-singlet exchange in Z+2-jet events at the CERN LHC". *Physical Review D* 54.11 (1996), pp. 6680–6689. DOI: 10.1103/physrevd.54.6680.

[100] CMS Collaboration. "Jet algorithms performance in 13 TeV data". Tech. rep. CMS-PAS-JME-16-003. (2017).

[101] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an].

[102] The CMS Collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". *Physics Letters B* 805 (2020), p. 135425. DOI: https://doi.org/10.1016/j.physletb.2020.135425.

[103] E. Bols et al. "Jet flavour classification using DeepJet". *Journal of Instrumentation* 15.12 (2020), P12012–P12012. DOI: 10.1088/1748-0221/15/12/p12012.

[104] F. Chollet et al. "Keras". https://keras.io. (2015).

[105] A. Martin et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems". Software available from tensorflow.org. (2015).

[106] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". (2014). eprint: 1412.6980.

[107] S. L. Smith et al. "Don't Decay the Learning Rate, Increase the Batch Size". (2017). eprint: arXiv:1711.00489.

[108] CMS Collaboration. "Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel". *Physical Review Letters* 125.6 (2020). DOI: 10.1103/physrevlett.125.061801.

[109] J. Wright. "Study of Higgs boson production through vector boson fusion at the CMS experiment using a dense convolutional neural network". PhD thesis. Imperial College London, (2018).

[110] P. Dauncey et al. "Handling uncertainties in background shapes: the discrete profiling method". *JINST* 10.04 (2015), P04015. DOI: 10.1088/1748-0221/10/04/P04015. arXiv: 1408.6865 [physics.data-an].

[111] CMS Collaboration. "Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 11 (2018), p. 185. DOI: 10.1007/JHEP11(2018)185. arXiv: 1804.02716 [hep-ex].

[112] CMS Collaboration. "Measurements of Higgs boson production via gluon fusion and vector boson fusion in the diphoton decay channel at $\sqrt{s} = 13$ TeV". CMS-PAS-HIG-18-029 (2019).

[113] M. Oreglia. "A study of the reactions $\psi' \to \gamma\gamma\psi$". SLAC Report SLAC-R-236. PhD thesis. Stanford University, (1980).

[114] R. A. Fisher. "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P". *Journal of the Royal Statistical Society* 85.1 (1922), pp. 87–94.

[115] J. Butterworth et al. "PDF4LHC recommendations for LHC Run II". *J. Phys.* G43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[116] CMS Collaboration. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". *Eur. Phys. J. C* 81 (2021), p. 800. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[117] CMS Collaboration. "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV". CMS Physics Analysis Summary CMS-PAS-LUM-17-004. (2018).

[118] CMS Collaboration. "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV". CMS Physics Analysis Summary CMS-PAS-LUM-18-002. (2019).

[119] W. Verkerke and D. Kirkby. "The RooFit toolkit for data modeling". *eConf* C0303241 (2003). arXiv: physics/0306116 [physics].

[120] W. Eadie et al. "Statistical Methods in Experimental Physics". *Journal of the American Statistical Association* (2013). DOI: 10.2307/2284113.

[121] A. L. Read. "Presentation of search results: The CL_s technique". *J. Phys. G* 28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.

[122] T. Junk. "Confidence level computation for combining searches with small statistics". *Nucl. Instrum. Meth. A* 434 (1999), p. 435. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex].

[123] ATLAS and CMS Collaborations, and the LHC Higgs Combination Group. "Procedure for the LHC Higgs boson search combination in Summer 2011". Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. (2011).

[124] M. Bicer et al. "First look at the physics case of TLEP". *Journal of High Energy Physics* 2014.1 (2014). DOI: 10.1007/jhep01(2014)164.

[125] The FCC Collaboration. "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1". *Eur. Phys. J. C* 79.6 (2019), p. 474. DOI: 10.1140/epjc/s10052-019-6904-3.

[126] D. d'Enterria, A. Poldaru, and G. Wojcik. "Measuring the electron Yukawa coupling via resonant s-channel Higgs production at FCC-ee". *The European Physical Journal Plus* 137.2 (2022). DOI: 10.1140/epjp/s13360-021-02204-2.

[127] CMS Collaboration. "2015 ECAL detector performance plots". CMS-DP-2015-057 (2015).

[128] CMS Collaboration. "Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 12 (2019), p. 061. DOI: 10.1007/JHEP12(2019)061. arXiv: 1909.04133 [hep-ex].