

**From Collisions to Predictions:
Search for Non-Resonant Higgs Boson
Pair Production in the Fully Leptonic
 $b\bar{b}W^+W^-$ Final State using
CMS Run 3 Data**

Doctoral dissertation presented by

Ahmet Oğuz Güzel

in fulfilment of the requirements for the degree of Doctor in Sciences

Thesis support committee

Prof. Christophe Delaere (Supervisor)

UCLouvain, Belgium

Dr. Andrea Giammanco (Supervisor)

UCLouvain, Belgium

Prof. Eduardo Cortina Gil (chair)

UCLouvain, Belgium

Prof. Agni Bethani

UCLouvain, Belgium

Dr. Maria Cepeda

CIEMAT, Spain

Dr. Matthias Schroeder

University of Hamburg, Germany

August, 2026

Acknowledgement

“For everything in the world—for life, for civilization, for success—the truest guide is science.”

— Mustafa Kemal Atatürk

This thesis could not have come into existence without the support of certain people, and I would like to express my sincere gratitude to them.

First of all, my heartfelt thanks go to my supervisors. Christophe has always listened to me carefully and appreciated my work; his guidance during times of uncertainty is something every PhD student needs. Andrea has been very supportive and encouraging, and I have learned so much from his experience. I owe my independent research spirit to both of them for their continuous encouragement to pursue what interested me. I am especially grateful to them for giving me the opportunity to pursue this PhD.

Secondly, thank you to the members of my jury: Eduardo, Agni, Maria, and Matthias. All of your input was invaluable, and the private defense went far smoother and better than I had expected.

Special thanks go to Agni, who helped me get onboarded to the analysis group, guided me through every aspect of working in an international collaboration, and provided crucial help whenever I needed it. This thesis would not be in such great shape without her.

I also extend my gratitude to my colleagues at the University of Hamburg—Matthias, Mathis, and the other members of their group—for their seamless collaboration and relentless work.

Warm thanks to my friend-colleagues: Altan, Kerem, Sezen, and Michele for their continuous support, Yıldırım for his companionship, and many others I met within the academic environments at CP3 and CERN.

Last but not least, I want to thank my friends and family, who have supported my passion for science from the very beginning. And to my anchor and light, beloved İlayda—I don’t have enough words to express my gratitude.

Funding Acknowledgments and Remarks

This PhD was funded by the Fonds de la Recherche Scientifique (FNRS) under the IISN convention No. 4.4503.12, and I am grateful to UCLouvain for providing me with the opportunity to pursue my research in such a stimulating environment. Computational resources were provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

I also acknowledge the use of AI tools during the preparation of this thesis — they were used to debug code, refine grammar and flow, and assist with LaTeX formatting; all outputs were thoroughly reviewed, edited, and verified by the author, and full responsibility for the content and integrity of this work rests with the author.

Final words

Looking back over the past four years, I am grateful to have spent my time on something truly purposeful. In that spirit, I leave a quote from Jack London that serves as a timeless strategy for life:

“The function of man is to live, not to exist. I shall not waste my days trying to prolong them. I shall use my time.”

— Jack London

Contents

Acknowledgement	iii
Preface	1
1 The Standard Model and Higgs boson pair production	3
1.1 Standard Model of particle physics	4
1.1.1 Quantum Field Theory for collisions	4
1.1.2 Standard Model - particles and symmetries	7
1.1.3 Gauge Invariance	11
1.1.4 Standard Model Lagrangian	14
1.1.5 Electroweak interactions and BEH mechanism	16
1.1.6 Simulation and reconstruction of proton–proton collisions	21
1.2 Higgs boson pair production	26
1.2.1 HH production in the Standard Model	26
1.2.2 Introduction to Effective Field Theory	30
1.2.3 Double Higgs Production in Linear EFT (SMEFT)	33
1.2.4 Double Higgs Production in Non-linear EFT (HEFT)	34
1.2.5 Modeling of non-resonant BSM effects on HH production	35
1.2.6 Review of current experimental results on HH production	35
1.2.7 Overview and Outlook	40
2 Experimental Setup: CERN and The Compact Muon Solenoid Experiment	43
2.1 The European Organization for Nuclear Research (CERN)	43
2.2 The Large Hadron Collider: Engineering Marvel and Scientific Instrument	44
2.2.1 The Accelerator Complex	46
2.3 The LHC and the CMS Experiment	47
2.3.1 CMS Subsystems and Layer Structure	47
2.3.2 Data Collection and Processing	52
2.3.3 Object reconstruction	54

2.4	Personal contribution	58
3	Machine Learning Implementation	59
3.1	Machine learning in high-energy physics	59
3.1.1	Supervised learning	60
3.1.2	Common supervised ML models in HEP	60
3.1.3	General training procedure and evaluation metrics	62
3.1.4	Practical considerations in HEP analyses	67
3.2	Neural networks	69
3.2.1	Deep neural networks	69
3.2.2	DNNs in HEP	70
3.2.3	DNN Implementation for $HH \rightarrow bbWW$	72
3.3	Transformers and attention mechanisms	73
3.3.1	Transformers in HEP	76
3.3.2	Transformer-based model for $HH \rightarrow bbWW$	78
3.4	Boosted decision trees	79
3.4.1	BDTs in HEP	80
3.4.2	BDT implementation for the $HH \rightarrow bbWW$ (DL) analysis	82
3.5	Comparison of models	83
3.5.1	Interpretability and physics features	83
3.5.2	Robustness to data calibration	87
3.5.3	Evaluation	90
4	Search for HH to $bbWW$ using Run-3 Data from CMS	93
4.1	Analysis setup	93
4.1.1	Samples	94
4.1.2	Object definitions	97
4.1.3	Event selection	99
4.1.4	Signal extraction	100
4.1.5	Control Regions	101
4.1.6	Categorization	103
4.1.7	Binning strategy	103
4.2	Corrections to the data and simulation	105
4.2.1	Issues in data taking in Run-3	107
4.2.2	Pile-up re-weighting	107
4.2.3	Trigger selection and efficiency	108
4.2.4	Noise filters in Run-3	108
4.2.5	Top quark p_T reweighting	109
4.2.6	b-tagging corrections	109
4.2.7	Boosted bb-jet tagging	111
4.2.8	Corrections to leptons	112

4.2.9	Z boson p_T reweighting	113
4.2.10	Jet energy corrections and resolution	113
4.2.11	Jet veto in Run-3	114
4.3	Systematic uncertainties	114
4.3.1	Experimental uncertainties	115
4.3.2	Theoretical uncertainties	116
4.4	Results	117
4.4.1	Impact of systematic uncertainties	119
4.4.2	Constraints on the Higgs boson self-coupling (κ_λ)	121
4.4.3	Constraints on the VVHH coupling (κ_{2V})	123
4.4.4	Two-dimensional coupling constraint	124
4.4.5	Software tools	126
4.5	Comparison with Run-2 and Advanced Run-3 Analysis Regimes	126
4.6	HH production in future searches	128
5	Conclusion	131
A	Additional Material	133
A.1	ML input features	133
A.2	ML performance plots	145
A.2.1	Feature importances and attention weights	145
A.2.2	ROC plots	148
A.2.3	Loss plots	149
A.3	Event yields	150
A.4	Dilepton trigger scalefactors	151
A.5	Boosted bb-jet tagging scalefactors	153
A.6	Inclusive cross-section uncertainties	155
A.6.1	Goodness-of-Fit tests	157
A.7	Likelihood scans	158
	Acronyms	159
	References	163

Preface

Humankind has always been curious trying to understand why things behave in a certain way. For instance, we have observed the nature around us and predicted the seasonal changes or developed tools. For this matter, theories have been developed and observations have been made to link to them. Around 4000 BC, the Mesopotamians postulated ideas to describe their observations of the heavenly bodies, suggesting that the Earth was at the center of the universe and the others bodies orbited it. But the astronomy wasn't the only scope of their endeavor. They extracted iron which gave rise to the Iron Age, experimented with plants to cure where some of those still are in use to produce medicines, Pythagoras, Aristotle, Plato and many others developed mathematics, geometry and suggested the existence of an indivisible building block of matter.

However, it wasn't until the XIII. century that these works are brought together which form the science as we know it today. But, the progress was still very slow. It took another 4 centuries to revolutionize the way we looked at the Universe by the work of Copernicus. The birth of the modern science dates back to the XVII. century where the nature was observed with advanced tool such as telescopes, microscopes and measuring gadgets. The law of gravity was shown, the relations between the volume and the temperature of a gas were shown. Later on in the following century during the Age of Enlightenment, groundwork of biology and chemistry were established.

XIX. century saw many great scientists; Dalton by the theory of atomic matter, Faraday and Maxwell by their theories on the electricity and magnetism, and others, including Curie, Gauss, Mendeleev and many more. Each of the discoveries made by these people helped us to understand better how the stuff around us worked.

It wasn't until the last century that the quantum mechanics and relativity formulated. These two concepts only revolutionized completely the way we perceived world around us.

The Particle Physics, one of the torch carriers of this endeavor, building on top of the modern physics as it is now, studies the building blocks of matter that are invisible, sometimes indivisible. It comes here, the famous Standard Model (SM) of Particle Physics; the most successful theory that paints the picture

depicting how these elementary particles interact with each other. Initially being the collection of numerous individual scientific results in the second half of the 20th century, it has now become mainly the product of collaborative efforts, given the mathematical maturity it had gained throughout the years.

The SM's predictions have been confirmed by many experimental results; the discovery of the $W^\pm$ and Z^0 bosons [1,2] in 1983, the observation of the top quark [3,4] in 1995 and the Higgs boson [5,6] in 2012, along with many others. It's a very successful theory describing three of the four known interactions of the nature; electromagnetic, weak and strong interactions. The fourth one, the gravitational interaction, is too weaker than the others on many orders of magnitude, which is a small enough contribution to be safely ignored [7]. And it's not covered in this thesis, too.

In the following chapters, you will see a handful of fronts of conducting a typical research; theory, experiment, interpretation of results, respectively. Let's break these down into finer pieces; in Chapter 1, we will develop the theory of the SM and focus on the Brout-Englert-Higgs Mechanism. This is where you should pay attention if you would like to understand the underlying mechanism using mathematical formalism.

The Chapter 2 will present a great experimental apparatus in which the most powerful electromagnet in the world is hosted. We will see what it is composed of, how it is commissioned and how the data it produces is analyzed. This is the part that you will like if you are more of an experimentalist.

Chapter 3 is my personal favorite. We will elaborate here on the state-of-art machine learning algorithms. Tune in here if you would like to dive into a realm where mathematics and programming swirls around each other.

A combination utilizing all these efforts comes in Chapter 4. A delicate way of searching for rare signals is in this part. Check out this chapter if you would like to find out how a data analysis at the CMS is conducted. You will not only see how detector data is handled, but also learn how the experimental difficulties are tackled.

I hope you will enjoy the journey as much as I did.

The Standard Model and Higgs boson pair production

In this introductory chapter, we set the theoretical foundation for the thesis by reviewing the Standard Model (SM) and the description of double Higgs (HH) production. The material introduced here is used throughout the following chapters, and it is complemented when needed by Beyond Standard Model (BSM) ideas.

The objectives of this thesis are:

- to review the theoretical tools needed to describe Higgs boson pair production;
- to study the experimental strategy used to search for HH events in the $b\bar{b}W\bar{W}$ final state;
- to develop and validate the analysis methods used to improve the sensitivity of the measurement;
- to compare the resulting limits with the available Run-2 and Run-3 references;
- to assess where machine learning methods can provide additional gains.

This chapter is divided into two parts. The first part focuses on Quantum Field Theory (QFT), which combines quantum mechanics and special relativity to describe fundamental particles and their interactions, and on the structure of the SM, which provides the current description of the electromagnetic, weak, and strong interactions. The derivations in this part follow the standard references by Peskin and Schroeder [8], Srednicki [9], and Griffiths [7]. The second part is devoted to Higgs boson pair production at proton-proton (pp) colliders, covering its production mechanism, main properties, and BSM interpretations, with emphasis on the experimental results that motivate this work.

The structure of the thesis is as follows. Chapter 2 presents the CMS detector, the CERN accelerator complex, and the experimental conditions relevant for the

analysis. Chapter 3 introduces the machine learning methods used in this work and explains how they improve the separation between signal and background events. Chapter 4 details the search for HH production in the $b\bar{b}WW$ channel using Run-3 data, including the event selection, background estimation, and statistical interpretation. Chapter 5 concludes the thesis by summarizing the main results and outlining possible directions for future improvements and searches.

1.1

Standard Model of particle physics

The mathematical representation that we will follow is based on the following conventions adopted by Peskin and Schroeder. The space-time coordinates are denoted by $x = x^\mu = (t, \mathbf{x})$, and the momentum by $p = p^\mu = (E, \mathbf{p})$, and the corresponding derivatives are $\partial^\mu = \partial/\partial x_\mu = (\partial/\partial t, \nabla)$. The metric tensor ($g_{\mu\nu}$) is $(+, -, -, -)$, so that a free massive particle satisfies $p^2 = p_\mu p^\mu = E^2 - |\mathbf{p}|^2$. Scalar fields and Dirac spinors are designated by ϕ and ψ , respectively. The Weyl representation for Dirac matrices i.e. $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, and it is used to define the fermion conjugate $\bar{\psi} \equiv \psi^\dagger\gamma^0$. The symbols used for particles designate either the particle itself or its associated field, where the meaning is inferred from the text. The Feynman slash is given by

$$\not{A} \stackrel{\text{def}}{=} \gamma^0 A_0 + \gamma^1 A_1 + \gamma^2 A_2 + \gamma^3 A_3, \quad (1.1)$$

where A is a covariant vector. Using the Einstein summation convention, the expression simply becomes $\not{A} \stackrel{\text{def}}{=} \gamma^\mu A_\mu$.

Two different unit systems are used depending on the quantities. The first one, International system of units (SI), is the scheme used to quantify macroscopic phenomena, e.g. kg, s, m, K. The second system of units is called *natural units*, which imposes $\hbar = c = 1$ and is forming a more suitable scheme for particle physics usage. This system of units results in

$$[\text{energy}] = [\text{mass}] = [\text{length}]^{-1} = [\text{time}]^{-1} = \text{eV} = 1.602 \times 10^{-19} \text{ J}, \quad (1.2)$$

where electron Volts (eV) is chosen as the reference. As an example, the mass of the electron can be expressed as $m_e = 9.109 \times 10^{-31} \text{ kg} = 0.511 \text{ MeV}$. A proton's mass in this sense is 0.928 GeV, which makes GeV a very useful unit in particle physics given hadrons generally have masses on the order of 1 GeV.

1.1.1

Quantum Field Theory for collisions

Quantum Field Theory is a theoretical framework that combines quantum mechanics and special relativity to describe the behavior of subatomic particles

and fields. In this section, we'll take a quick look at its general principles and observable quantities before moving on to the SM.

The principle of least action is a fundamental principle in classical mechanics and states that the path taken by a system between two states is the one for which the action is minimized. It is used to derive the equations of motion for a system. In QFT, we can derive the equations of motion using a free field $\phi(x)$ where $x = (t, \mathbf{x})$ denotes space-time coordinates. The action S , which is calculated as the integral of the Lagrangian density $\mathcal{L}$, is the mathematical function of the state of the physical system at hand, and can then be written as

$$S = \int \mathcal{L}(\phi, \partial_\mu \phi) d^4x. \quad (1.3)$$

Using the principle of least action by varying the path and requiring $\delta S = 0$

$$\delta S = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi} \delta \phi + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta (\partial_\mu \phi) \right] \quad (1.4)$$

integrating by parts yields

$$= \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) \right] \delta \phi + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta \phi \right), \quad (1.5)$$

where the last term is a total derivative and vanishes obeying $\delta \phi(\vec{x}, t_1) = \delta \phi(\vec{x}, t_2) = 0$. Finally, requiring $\delta S = 0$ for all paths yields the Euler-Lagrange equations of motion for the field ϕ :

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \quad (1.6)$$

Using this equation, an operator representing the total energy of a quantum system, the so called *Hamiltonian density* can be written as follows

$$H = \int \mathcal{H} d^3x \Rightarrow \mathcal{H} = \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \partial_\mu \phi - \mathcal{L}. \quad (1.7)$$

If we let the fields interact with each other, we will need to divide the Hamiltonian into free and interacting parts i.e. $H = H_0 + H_I$. We are now working in the *interaction picture*. The time dependence in the states is controlled in the interacting part $\mathcal{H}_I$ where the free part $\mathcal{H}_0$ controls the time dependence in operators, such as

$$\frac{d|\Psi(t)\rangle}{dt} = iH_I(t) |\Psi(t)\rangle, \quad (1.8a)$$

$$H_I(t) = e^{iH_0(t-t_0)} (H_{\text{int}}) e^{-iH_0(t-t_0)}. \quad (1.8b)$$

The Equations (1.8a) and (1.8b) show the state operator and the interaction Hamiltonian in the interaction picture, respectively.

Some of our main interests in the collider experiments are cross-sections and decay rates. In order to access these variables, we need to describe the transition probability between initial and final states denoted by $|i\rangle$ and $|f\rangle$ corresponding to non-interacting collections of particles with momenta p_i^{in} and p_i^{out} , respectively. These probabilities can be written in terms of a unitary S-matrix, such as $|\Psi(t \rightarrow +\infty)\rangle = S |\Psi(t \rightarrow -\infty)\rangle = S |i\rangle$ which describes the evolution of the initial state in time. Thus, the projection $\langle f|S|i\rangle$ gives the transition amplitude to the final state. Considering the interaction part $\mathcal{H}_I$ small compared to the free Lagrangian, the S-matrix can be derived as a perturbative series, such as

$$S = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int \cdots \int d^4x_1 \cdots d^4x_n T\{\mathcal{H}_I(x_1) \cdots \mathcal{H}_I(x_n)\} \quad (1.9)$$

where T denotes the time-ordered multiplication. Every term in the series in Equation (1.9) can be written as a finite sum of products, which allows the amplitudes to be calculated. These amplitudes can then be represented by Feynman diagrams implementing an easy scheme to interpret terms contributing to a given process at a given order in the perturbative expansion.

We can derive this probability for our collision experiment. The main kind of collision experiments is either a head-on collision or a fixed-target experiment. *Cross-section* becomes a very useful observable in this endeavor. It is defined as the rate of the scattering of a collision per incident flux with area dimensions and it is directly related to the probability of a given interaction to occur. The *differential cross-section* $d\sigma(X)/dX$ is defined as a function of certain quantity X , and is generally more useful in collision systems. The quantity X is a variable inheriting the final momenta p_f^{out} , hence deriving S

$$S = 1 + (2\pi)^4 \delta^4(p_1^{in} + p_2^{in} - \sum p_f^{out}) \cdot i\mathcal{M}(p_{1,2}^{in} \rightarrow \{p_f^{out}\}) \quad (1.10)$$

to incorporate the part of the S-matrix regarding the interactions between the particles into the *matrix element* $\mathcal{M}$, the differential cross-section can be written as follows

$$\frac{d\sigma}{d\Phi} = \frac{|\mathcal{M}(p_{1,2}^{in} \rightarrow \{p_f^{out}\})|^2}{F} \quad (1.11)$$

where the flux F and the phase space differential are as follows

$$F = 4\sqrt{(p_1^{in} \cdot p_2^{in})^2 - (p_1^{in})^2 (p_2^{in})^2} = 4E_i^{in} E_2^{in} \Delta v \quad (1.12)$$

$$d\Phi = (2\pi)^4 \delta^4 \left(p_1^{in} + p_2^{in} - \sum_f p_f^{out} \right) \prod_f \frac{d^3 \mathbf{p}_f^{out}}{(2\pi)^3} \frac{1}{2E_f^{out}} \quad (1.13)$$

where Δv is the relative velocity between the incident particles in Equation (1.12). It's important to note that the cross-section is invariant under Lorentz boost along the beam direction. By integrating the Equation (1.11) over the desired phase space, the total or differential cross-section can be obtained, and decay rates can be calculated similarly by taking a single initial particle of M into consideration and the flux factor F as $2M$.

Cross-section can be computed as a perturbative expansion just as the S-matrix is. A *coupling constant* g denoting the interaction strength and a parameter such as $\alpha = g^2 / (4\pi)$ allow the expansion to be written as

$$\sigma = \sigma_{LO} \cdot \left(1 + \frac{\alpha}{2\pi} \sigma_1 + \left(\frac{\alpha}{2\pi} \right)^2 \sigma_2 + \dots \right) \quad (1.14)$$

where the interaction parameter α should be weak, meaning that it's a small factor i.e. $\alpha \ll 1$, to be computable in perturbative theory.

Another important notion arises here. *Leading order* (LO) calculation, which means taking the lowest-order variable in the Equation (1.14) for the numerical computation is relatively easy. *Next-to-leading order* (NLO) or *next-to-next-to leading order* (NNLO) and so on calculations i.e. taking higher-order terms in the equation proves being a challenging task but provides better theoretical prediction. The reason to the complexity in the NLO and so forth calculations relies on the fact that they involve factorially larger number of Feynman diagrams with the order of the calculation, which makes the evaluation hugely complicated and computationally expensive.

Loop integrals, which are integrals over the momenta of virtual particles in the diagrams, often contain ultra-violet (UV) divergences, which arise from the contributions of very high-momentum virtual particles. These divergences need to be carefully regularized and renormalized to obtain physically meaningful results. To do so, a *renormalisation* procedure, which can absorb infinities into the parameters of the theory, and define renormalized parameters, is applied. This procedure requires introducing a *renormalisation scale* μ_R , which should not affect the physical observables. In this way, bare coupling constants found in the Equation (1.14) are replaced by constants which depend on μ_R .

1.1.2

Standard Model - particles and symmetries

The SM is the ruling theory that describes the fundamental forces and elementary particles that make up all matter in the universe (except gravity). It's a highly successful and well-tested theory, although it's not considered to be

the final theory of everything. The SM systematizes all known elementary particles, which are considered to be the fundamental building blocks of matter. These particles are point-like and indivisible (as far as we know). They are divided into two main categories:

- **Fermions:** These particles have half-integer spin and obey the Pauli exclusion principle, which states that no two identical particles with half-integer spin can occupy the same quantum state simultaneously. This principle has significant implications for the structure of atoms, molecules, and even the stability of stars. They are the constituents of ordinary matter that we see around us. Fermions are further divided into:
 - **Quarks:** These are the building blocks of protons and neutrons, which make up the nuclei of atoms. There are six types of quarks: up, down, charm, strange, top, and bottom. They have fractional electric charges, e.g. $2/3$ or $1/3$ in units of e (see Table 1.1). They are also charged under the strong interaction, which is called the *color*.
 - **Leptons:** These are particles that don't interact with the strong nuclear force. They include electrons, muons, taus, which have a charge of -1 and their associated neutrinos which are electrically neutral and have extremely small mass.
- **Bosons:** These particles have integer spin and act as mediating particles of the fundamental forces. The bosons in the SM are:
 - **Gluons:** Mediate the strong nuclear force, which binds quarks together inside hadrons.
 - **Photons:** Mediate the electromagnetic force, responsible for interactions between charged particles.
 - **W and Z bosons:** Mediate the weak nuclear force. The weak force, mediated by $W^\pm$ and Z bosons, interacts with lepton and quark pairs, causing flavor changes. The Z boson mediates neutral currents, analogous to electromagnetic interactions but with flavor instead of electric charge, enabling it to interact with neutrinos. Importantly, these weak interactions are observed to affect only left-handed particles.
 - **Higgs boson:** Explains how particles acquire mass without breaking gauge invariance (more about that in Section 1.1.5).

Each charged particle in the SM has a corresponding *antiparticle* with the same mass and spin, but with opposite internal quantum numbers, including electric charge, color charge, weak isospin, and all associated flavor numbers (baryon number and lepton number). For example, the antiparticle of the electron is the positron (e^+), which carries a charge of $+1$, while quarks have corresponding

antiquarks (e.g., $\bar{u}$, $\bar{d}$) that bear opposite electric and color charges. It is important to note that the W^+ and W^- bosons are mutually antiparticles, while neutral bosons such as the photon, the Z boson, and the Higgs boson are their own antiparticles.

As experimentally shown, the fermions have three *generations*. Each generation contains heavier versions of the particles in the previous generation, called *flavor*. Neutrinos, in this regard, are extremely light, and their exact masses are still being studied. The generations differ in stability, too. The first-generation particles (up and down quarks, electron, and electron neutrino) are stable. Whereas the heavier quarks and leptons are unstable and decay rapidly into lighter particles. This is a consequence of lepton flavor conservation; for instance, the electron is stable simply because it is the lightest charged lepton and has no lighter particle to decay into. The existence of three generations of fermions is a puzzle that physicists are still trying to solve. While the cancellation of quantum anomalies requires a specific balance of charges within each generation, it does not predict the total number of generations.

Every charged particle of matter is a Dirac fermion, meaning it has a corresponding antiparticle with the opposite charge but otherwise identical properties. Neutral, massive neutrinos, however, present a different situation. They could be either Dirac or Majorana fermions, and research is ongoing to determine which is the case. Given all the information so far, we can list the particles as in Table 1.1.

In order to understand the force-carrier particles, we should take a look at the symmetry concept. Symmetry is fundamental to the SM's structure. Specifically, the SM is built upon the principle of gauge invariance. This means that the theory remains unchanged under certain transformations. These are abstract transformations acting on the internal properties of the particles, and they can be different at every point in spacetime (local gauge transformations). These transformations can be continuous (like rotations and boosts) or discrete (like parity and charge conjugation). They can act on spacetime coordinates (Lorentz transformations) or internal degrees of freedom (gauge transformations).

It is to maintain this gauge invariance that the SM introduces new force-carrying fields, called gauge bosons. These bosons are the mediators of the fundamental forces. Each type of symmetry corresponds to a specific type of force and its associated gauge boson. For example, the symmetry related to electric charge gives rise to electromagnetism, and the corresponding gauge boson is the photon.

The mass of the gauge boson is directly related to the range of the force it carries. Massless bosons, like photons and gluons, mediate forces with infinite range. Massive bosons, like the W and Z bosons, mediate forces with a very

TABLE 1.1 | This table summarizes the properties of leptons and quarks within the SM. Mass values are sourced from reference [10]. While charged lepton masses are presented with four significant digits, they are actually known with much greater precision. Electric charges are expressed as multiples of the electron's absolute charge. For neutrinos, only upper bounds on their masses are available, and the values shown here are derived from cosmological measurements of the total mass of stable neutrinos [11].

Generation	1	2	3
Leptons			
Name, symbol	electron neutrino, ν_e	muon neutrino, ν_μ	tau neutrino, ν_τ
Mass	$\longleftarrow \sum \nu_i < 0.0926 \text{ eV @ 90\% C.L.} \longrightarrow$		
Electric charge	0	0	0
Name, symbol	electron, e	muon, μ	tau, τ
Mass	511.0 keV	105.7 MeV	1.777 GeV
Electric charge	-1	-1	-1
Quarks			
Name, symbol	up, u	charm, c	top, t
Mass	$2.2^{+0.6}_{-0.4} \text{ MeV}$	1.28(3) GeV	173.5(6) GeV
Electric charge	2/3	2/3	2/3
Name, symbol	down, d	strange, s	bottom, b
Mass	$4.7^{+0.5}_{-0.4} \text{ MeV}$	96^{+8}_{-4} MeV	$4.18^{+0.04}_{-0.03} \text{ GeV}$
Electric charge	-1/3	-1/3	-1/3

short range; a consequence of the uncertainty principle is that the heavier the virtual particle, the shorter the distance it can travel. The properties of these gauge bosons, including their masses and spins, are determined by the underlying symmetries of the SM. Table 1.2 lists the SM's gauge bosons. A boson's mass directly affects the interaction's range. The massless photon, for example, results in electromagnetism's infinite range.

It's important to mention the Noether's theorem which highlights a crucial link between symmetries and conservation laws: every continuous symmetry corresponds to a conserved quantity. For the Poincaré group, these are energy, momentum, and angular momentum. Gauge symmetries, on the other hand, lead to conserved charges like electric charge.

Importantly, gauge symmetry-based interactions guarantee renormalizability, a desirable theoretical property. However, this framework initially predicts massless gauge bosons, contradicting experimental evidence that the weak interaction's bosons are massive. This discrepancy is resolved by the mechanism of spontaneous symmetry breaking, which is explained in Section 1.1.5.

1.1.3 Gauge Invariance

Gauge transformations are specific mathematical operations that can be performed on the quantum fields representing particles. Importantly, these transformations can be different at different points in spacetime – this is what makes them "local" gauge transformations. The principle of gauge invariance on the other hand, demands that the physical predictions of a theory must not change even if we perform these local gauge transformations. In order to maintain this gauge invariance, we need to introduce force-carrying particles called gauge bosons.

Consider the Lagrangian density for a free Dirac field ψ :

TABLE 1.2 | This table describes the gauge bosons within the SM and the fundamental forces that they mediate. Mass values are from reference [10], with values in parentheses reporting the experimental upper limits.

Name	Mass	Interaction
Photon γ	0 ($< 1 \times 10^{-18}$ eV)	Electromagnetism
Z boson	91.1876 GeV	Weak interaction
$W^\pm$ bosons	80.385 GeV	
Gluons, g	0 (< 1.3 MeV)	Strong interaction

$$\mathcal{L} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi. \quad (1.15)$$

The Lagrangian density is invariant under a phase transformation of the fermion field, such as:

$$\psi \rightarrow e^{iQ\omega} \psi, \quad \bar{\psi} \rightarrow e^{iQ\omega} \bar{\psi} \quad (1.16)$$

where Q is the charge operator, i.e. $Q\psi = +\psi$, $Q\bar{\psi} = -\bar{\psi}$, ω is a real constant independent of x and $\bar{\psi}$ is the conjugate field. The set of all number $e^{-i\omega}$ form an *abelian* group, i.e. any two elements of the group commute such that $e^{-i\omega_1} e^{i\omega_2} = e^{-i\omega_2} e^{i\omega_1}$. This particular group is called $U(1)$ meaning the group of all unitary 1×1 matrices.

For simplicity purposes, we will take infinitesimal group transformations where the parameter ω is small enough that we can perform an expansion over it and take only the linear term into account, i.e.

$$e^{-i\omega} = 1 - i\omega + Q(\omega^2) \quad (1.17)$$

and the field ψ changes as

$$\psi \rightarrow \psi + \delta\psi = \psi + iQ\omega\psi \quad (1.18)$$

$$\bar{\psi} \rightarrow \bar{\psi} + \delta\bar{\psi} = \bar{\psi} + iQ\omega\bar{\psi} = \bar{\psi} - i\omega\bar{\psi}. \quad (1.19)$$

such that the Lagrangian density remains unchanged. But at this point, in order to perform same transformation at every space-time point, all of these points should 'be aware' of the transformation, which doesn't seem like a good option. Instead, we assume that the invariance under gauge transformations is a fundamental property of nature. This assumption imposes the invariance under local transformations, meaning that the parameter ω depends on space-time point x . In that case, the fields transform as

$$\delta\psi(x) = i\omega(x)\psi(x); \quad \delta\bar{\psi}(x) = -i\omega(x)\bar{\psi}(x). \quad (1.20)$$

But now the Lagrangian density is no longer invariant under these transformations because of the partial derivative between $\bar{\psi}$ and ψ . We can now assume that the Lagrangian in Equation (1.15) is invariant under $U(1)$ transformations. This derivative acts on the space-time parameter $\omega(x)$ which imposes that the Lagrangian density is altered by an amount $\delta\mathcal{L}$, where

$$\delta\mathcal{L} = -\bar{\psi}\gamma^\mu [\partial_\mu Q\omega(x)] \psi(x), \quad (1.21)$$

where the square brackets indicate that the derivative is to be acted only inside them.

In order to restore the gauge invariance, we assume that the fermion field interacts with a vector field A_μ , called a *gauge field*, with an interaction term i.e.

$$-g\bar{\psi}\gamma^\mu A_\mu Q\psi, \quad (1.22)$$

where g is the *gauge coupling constant* and replaced often by the electromagnetic charge e in this context. Adding this term to the Lagrangian density, which now becomes

$$\mathcal{L} = \bar{\psi} \left(i\gamma^\mu (\partial_\mu + ieQA_\mu) - m \right) \psi. \quad (1.23)$$

In order to maintain invariance property, the gauge field should also change as in the following

$$-eQA_\mu \rightarrow -eQ(A_\mu + \delta A_\mu(x)) = -eQA_\mu + Q\partial_\mu\omega(x) \quad (1.24)$$

hence

$$\delta A_\mu(x) = -Q\partial_\mu\omega(x)/e \quad (1.25)$$

which cancels each other along with Equation (1.21), hence the gauge invariance is restored. The Equation (1.23) is known as the fermionic part of the Lagrangian density for Quantum Electrodynamics (QED), where e denotes the electric charge of the electron and A_μ denotes the photon field. A proper quantum field theory, one where the photon field A_μ can be expressed using creation and annihilation operators, requires a kinetic term for the photon. This term must be quadratic in the derivative of A_μ . Without it, the Euler-Lagrange equation for the gauge field would be algebraic, allowing us to eliminate the gauge field entirely from the Lagrangian. However, this kinetic term must be introduced without breaking gauge invariance. This is accomplished by defining the field strength tensor, $F_{\mu\nu}$, as

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu, \quad (1.26)$$

which under gauge transformation in Equation (1.24) remains invariant meaning that any term which depends on $F_{\mu\nu}$ may be added to the Lagrangian density. If we choose this term to be $-\frac{1}{4}F_{\mu\nu}^2$, then the Lagrange equations of motion match Maxwell's equations exactly. We now have achieved the derivation of the Lagrangian density for QED from the framework of $U(1)$ gauge invariance, which is

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi} \left(i\gamma^\mu (\partial_\mu + igQA_\mu) - m \right) \psi. \quad (1.27)$$

Before moving on to the SM Lagrangian, it is useful to introduce the concept of *covariant derivative* D_μ , which is defined as

$$D_\mu \equiv \partial_\mu + igA_\mu. \quad (1.28)$$

It has the same property as the transformations of the fermion field in Equation (1.20) and the gauge field in Equation (1.24) such that the quantity $D_\mu\psi$ transforms in the same manner under the gauge transformations as ψ . Given that, we can rewrite the QED Lagrangian density as

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma D_\mu - m)\psi. \quad (1.29)$$

What's more, the field strength $F_{\mu\nu}$ can now be expressed in terms of the commutator of two covariant derivatives as in the following equation:

$$\begin{aligned} F_{\mu\nu} &= -\frac{i}{g}[D_\mu, D_\nu] \\ &= -\frac{i}{g}[\partial_\mu, \partial_\nu] + [\partial_\mu, A_\nu] + [A_\mu, \partial_\nu] + ig[A_\mu, A_\nu]. \end{aligned} \quad (1.30)$$

Additionally, a local gauge invariant Lagrangian under a certain group requires a covariant derivative $D_\mu = \partial_\mu - igA_\mu^a T^a$, with the T^a as the group generator. Hence, a gauge field tensor can be written as

$$igF_{\mu\nu}^a T^a = [D_\mu, D_\nu]. \quad (1.31)$$

1.1.4 Standard Model Lagrangian

The SM Lagrangian includes terms for fermions, gauge bosons and Higgs boson, and is built on the gauge group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where C refers to the color, L to the left-handed fermions and Y to the weak hypercharge representing the symmetries of the strong, weak and electromagnetic forces, respectively.

The first group, $SU(3)_C$, has 8 generators - abstract mathematical objects that describe infinitesimal transformations in Lie groups like $SU(3)$ - whose representation is provided by the Gell-Mann matrices λ_a . Each of these generators corresponds to a gauge boson, i.e. G_μ^a . The covariant derivative is given by

$$D_\mu^{ij} = \partial_\mu - ig_s G_\mu^a \frac{\lambda_a^{ij}}{2}, \quad (1.32)$$

and the field tensor can be derived using the gauge field tensor from Equation (1.31), as

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c \quad (1.33)$$

where f^{abc} is the structure constant of the group and g_s is the coupling constant of the strong interaction. The Quantum Chromodynamics (QCD) Lagrangian density can then be expressed as

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} + \bar{\psi} \left(i \not{D}^{ij} \right) \psi. \quad (1.34)$$

The second and third groups, $SU(2)_L \otimes U(1)_Y$, describe the weak and the electromagnetic interactions. The covariant derivatives for this interaction acts differently on the left- and right-handed fermions based on the chirality, and can be expressed as

$$D_\mu = \partial_\mu - ig \frac{\sigma_i}{2} W^{\mu i} - ig' \frac{Y}{2} B_\mu \quad (1.35)$$

$$D'_\mu = \partial_\mu - ig' \frac{Y}{2} B_\mu \quad (1.36)$$

where g and g' are the two coupling constants, σ_i are the Pauli matrices, $W^{\mu i}$ and B_μ are the three and one spin-1 boson associated to the $SU(2)_L$ and $U(1)_Y$ groups, respectively. The corresponding field tensors can be built using Equation (1.31) as

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk} W_\mu^j W_\nu^k \quad (1.37)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad (1.38)$$

where ϵ is the fully antisymmetric structure constant of the $SU(2)$ group.

The gauge and fermionic part of the SM Lagrangian can therefore be given as

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} - \frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu} \quad (1.39)$$

$$\mathcal{L}_{\text{fermion}} = \sum_{\psi, I} \bar{\psi}^I (i \not{D}) \psi^I, \quad (1.40)$$

where I represents the three generations of particles and with the various fields $\psi = L, Q, e, u, d$, which are

$$\mathbf{L} = \left[\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \right], \quad \mathbf{Q} = \left[\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \right] \quad (1.41)$$

$$\mathbf{e} = [e_R, \mu_R, \tau_R], \quad \mathbf{u} = [u_R, c_R, t_R], \quad \mathbf{d} = [d_R, s_R, b_R], \quad (1.42)$$

where, $\mathbf{L}$ and $\mathbf{Q}$ are the left-handed lepton and quark doublets of the $SU(2)$ group, while the right-handed fermions are kept in singlets as $\mathbf{e}$, $\mathbf{u}$ and $\mathbf{d}$ where the relation is given by

$$P_{L,R} = \frac{1}{2}(1 \pm \gamma_5), \quad \psi_{L,R} = P_{L,R} \psi. \quad (1.43)$$

The reason to this composition of only left-handed fermions is discussed towards the end of the Section 1.1.5.

So far, all the particles are massless since a mass term of type $m\bar{\psi}\psi$ would break the gauge invariance under $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry. They have to acquire their mass via another mechanism, which is the topic of the next section.

1.1.5 Electroweak interactions and BEH mechanism

Experiments have shown that many fundamental particles, including leptons, quarks and the W and Z bosons, possess mass. However, the SM's Lagrangian, as constructed based on gauge invariance, only permits massless particles. To resolve this contradiction between theory and observation, physicists proposed a complex scalar field, denoted as ϕ , which is a doublet under the $SU(2)$ weak interaction symmetry, has a hypercharge of $Y=1$, and carries no color charge. This field was independently postulated in several publications [12–14]. The process by which leptons, quarks, and bosons acquire mass is known as the *Brout-Englert-Higgs* (BEH) mechanism. The Lagrangian describing this scalar field is given by

$$\mathcal{L}_{\text{scalar}} = D_\mu \phi^\dagger D^\mu \phi - V(\phi^\dagger \phi). \quad (1.44)$$

The potential energy associated with this field, denoted as $V(\phi^\dagger \phi)$, was defined using the most general form that could be consistently calculated within the theory while also adhering to the required gauge symmetries i.e.

$$V(\phi^\dagger \phi) = -\mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2 \quad (1.45)$$

with $\mu^2 > 0$. The lowest energy state of this potential (its minimum) occurs when the quantity $\langle 0 | \psi^\dagger \psi | 0 \rangle$ equals $\mu^2 / \lambda \equiv v^2$. Here, v represents the vacuum expectation value (VEV) of the scalar field. This non-zero VEV breaks the electroweak symmetry group $SU(2)_L \otimes U(1)_Y$. By expressing the field's behavior around this VEV using polar coordinates, we obtain

$$\phi(x) = \frac{1}{\sqrt{2}} \exp(i\sigma^i \xi_i(x)) \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \xrightarrow{\text{EWSB}} \phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (1.46)$$

where $\xi_i(x)$ represent Goldstone bosons, and the Electroweak symmetry breaking mechanism (EWSB) has been applied. The covariant derivative of this field is

$$D_\mu = \partial_\mu - ig \frac{\sigma_i}{2} W^{i\mu} - i \frac{g'}{2} B_\mu. \quad (1.47)$$

Substituting the covariant derivative in the Lagrangian and developing further yields

$$\begin{aligned}
 \mathcal{L}_{\text{scalar}} &= D_\mu \phi^\dagger D^\mu \phi + \mu^2 (\phi^\dagger \phi) - \lambda (\phi^\dagger \phi)^2 \\
 &= \frac{v^2}{8} \left(2W_\mu^i W^{i\mu} + g'^2 B_\mu B^\mu - 2g' g B_\mu W^{3\mu} \right) \left(1 + \frac{h}{v} \right)^2 \\
 &\quad + \frac{1}{2} (\partial_\mu h \partial^\mu h) - \lambda v^2 h^2 - \lambda v h^3 - \frac{\lambda}{4} h^4 - \frac{\lambda v^4}{4}.
 \end{aligned} \tag{1.48}$$

Equation (1.48) reveals how the $W^{1,2,3}$ and B gauge bosons acquire mass, as indicated by its second line. Furthermore, the expression's final term demonstrates a mixing phenomenon between the W^3 and B fields. To understand the actual particles we observe, we must redefine these gauge bosons into the physical bosons:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \pm iW_\mu^2) \tag{1.49}$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \tag{1.50}$$

Here, θ_W represents the Weinberg angle, defined by the relationship $\tan \theta_W = g'/g$. Ultimately, the four degrees of freedom initially present in the scalar field ϕ are reinterpreted as contributing to the mass of the Higgs boson and the masses of the weak bosons. Crucially, the photon remains massless throughout this process:

$$\begin{aligned}
 m_w &= \frac{gv}{2} \\
 m_z &= \sqrt{g^2 + g'^2} \frac{v}{2} \\
 m_a &= 0
 \end{aligned} \tag{1.51}$$

By utilizing the relationship $m_Z = m_w / \cos \theta_W$, which connects the masses of the weak bosons, the initial line of Equation (1.48) can be expressed in a different, but equivalent, form as

$$\mathcal{L}_{\text{scalar}} \supset \left(m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} m_Z^2 Z_\mu Z^\mu \right) \left(1 + \frac{h}{v} \right)^2 \tag{1.52}$$

Equation (1.52), which represents a portion of the scalar Lagrangian, serves a dual purpose: it not only predicts the masses of the weak bosons but also details how these bosons interact with the scalar field.

The second line of Equation (1.48) suggests the existence of an additional scalar particle, the Higgs boson. This particle acquires mass, given by $m_H = \sqrt{\lambda} v$,

utilizing the remaining degree of freedom. The experimentally determined mass of the Higgs boson is 125.09 ± 0.24 GeV. The VEV, v , which is approximately 246 GeV, can be derived from measurements of the weak boson masses in Equation (1.51). Combining these values allows us to predict the value of λ , which in turn predicts the strength of trilinear ($\lambda v h^3$) and quartic ($\frac{\lambda}{4} h^4$) Higgs self-interactions.

Directly adding mass terms for fermions violates gauge invariance, a crucial requirement for the SM. However, experiments clearly indicate that fermions do possess mass. To resolve this conflict, we introduce Yukawa terms, which couple the scalar field ϕ to the fermions:

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}} &= - \sum_i Y_l^{ij} \left(\bar{L}^i \cdot \phi \right) l^j + h.c. && \text{(leptons)} \\ &= - \sum_{ij} Y_u^{ij} \left(\bar{Q}^i \cdot \phi^C \right) u^j - \sum_{ij} Y_d^{ij} \left(\bar{Q}^i \cdot \phi^C \right) d^j + h.c. && \text{(quarks)} \end{aligned} \quad (1.53)$$

Here, ϕ^C is defined as $i\sigma_2 \phi^*$. Y_l , Y_u , and Y_d represent the Yukawa coupling matrices, which govern the interactions of the scalar field with leptons (electron, muon, tau), up-type quarks (up, charm, top), and down-type quarks (down, strange, bottom), respectively. Importantly, these Yukawa couplings are matrices within the flavor space, meaning that they can mix different generations of fermions.

After electroweak symmetry breaking, as described in Equation (1.46), the portion of the Yukawa interaction involving leptons transforms into

$$\mathcal{L}_{\text{Yukawa}} \subset - \frac{Y_l v}{\sqrt{2}} (\bar{l}_L l_R + \bar{l}_R l_L) \left(1 + \frac{h}{v} \right). \quad (1.54)$$

This portion of the Lagrangian reveals a new mass term for charged leptons, expressed as $m_l = Y_l v / \sqrt{2}$. Intriguingly, the Yukawa coupling for the top quark, Y_t , is approximately 1. The top quark's Yukawa coupling is notable due to its magnitude, indicating a strong Higgs interaction. This proximity to unity, unlike other fermion Yukawa couplings, suggests a more "natural" coupling strength and potential relevance to electroweak symmetry breaking or beyond-Standard-Model physics. It highlights the flavor hierarchy problem and possible top-Higgs sector connections. Additionally, it describes an interaction term between lepton-antilepton pairs and the Higgs boson. Figure 1.1 provides experimental evidence supporting these scalar field couplings.

For leptons, a single Yukawa matrix is used to relate left-handed flavor eigenstates to mass eigenstates. This matrix can be diagonalized in flavor space, resulting in a unique Yukawa coupling value for each lepton flavor.

While a similar expression applies to quarks, they involve two Yukawa matrices

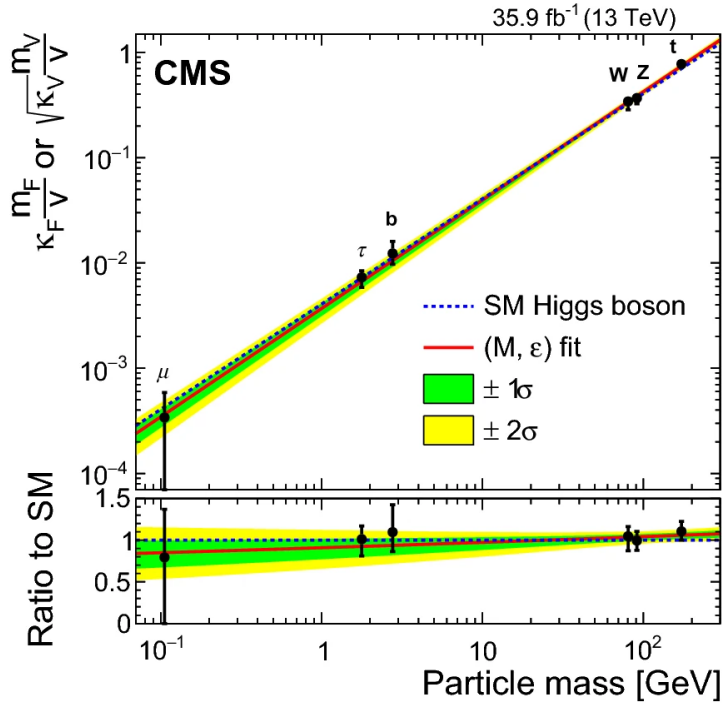


FIGURE 1.1 | The CMS collaboration's [15] measurements of the Higgs boson's interactions with various fermions (top and bottom quarks, tau and muon leptons) and weak bosons (W and Z) through Yukawa couplings show excellent agreement with the SM predictions outlined in Equations (1.52) and (1.54). This agreement strongly supports the inclusion of the scalar field in the Lagrangian as the mechanism responsible for giving mass to elementary particles.

(Y_u and Y_d). Consequently, both matrices cannot be simultaneously diagonalized. Unlike leptons, a direct one-to-one correspondence between mass and flavor eigenstates isn't possible for all quarks. By convention, the up-type quark matrix is diagonalized, leading to a non-diagonal matrix, known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which relates the mass and flavor eigenstates of down-type quarks. The CKM matrix is characterized by four parameters, one of which introduces charge-parity (CP) violation.

In the lepton sector with massive neutrinos, the mismatch between flavor and mass eigenstates is encoded in the unitary Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, which relates the charged-current weak eigenstates (e, μ, τ) to the neutrino mass eigenstates (ν_1, ν_2, ν_3) [16]. Unlike the nearly diagonal CKM matrix, the PMNS matrix exhibits large mixing angles and is typically parameterized by three angles ($\theta_{12}, \theta_{23}, \theta_{13}$) and one Dirac CP-violating phase δ if neutrinos are Dirac particles; if neutrinos are Majorana, two additional Majorana phases become physical but do not affect oscillations. This structure reflects that the charged-lepton Yukawa matrix can be chosen diagonal by convention, while the neutrino mass matrix remains non-diagonal, leading to observable neutrino oscillations and potential leptonic CP violation.

Beyond the Higgs interactions described in Equation (1.54), EWSB leads to leptons interacting with W and Z bosons, as well as photons, through terms that preserve lepton flavor. While this principle also applies to quarks, the inability to simultaneously diagonalize both Yukawa matrices necessitates a modification of the charged currents mediated by the W boson. This modification introduces transitions between up-type and down-type quarks, meaning quarks of different flavors can change into one another.

The SM includes 26 parameters that are not predicted by the theory itself and must be determined experimentally. In addition to the local gauge symmetries already discussed, experiments have revealed global symmetries, such as lepton number conservation (one for each generation) and baryon number conservation. However, these global symmetries were not explicitly imposed during the model's construction, leading them to be classified as "accidental symmetries".

Finally, the lack of experimental detection of right-handed neutrinos suggests that a Yukawa term cannot be included in the SM Lagrangian to give neutrinos mass. Consequently, within the SM framework, neutrinos remain massless after electroweak symmetry breaking, which contradicts experimental observations of neutrino oscillations which require neutrinos to have mass.

1.1.6

Simulation and reconstruction of proton–proton collisions

Theorists face a major challenge in collider experiments: accurately calculating observable quantities like total cross-sections and decay rates. This requires controlling both the perturbative description of the hard scattering and the non-perturbative dynamics of QCD. At hadron colliders such as the Large Hadron Collider (LHC), this is particularly difficult because quarks and gluons are never observed as free particles, but instead undergo parton showering, hadronisation, and the decay of short-lived hadrons before leaving signals in the detector.

When quarks or gluons, collectively referred as *partons*, are created in a high-energy collision, a phenomenon known as the *hard process*, they subsequently emit additional partons. This emission phenomenon is called Final State Radiation (FSR). The cascade of these emitted particles is referred to as a *parton shower*, and importantly, this shower’s development is independent of the specific details of the initial hard process or how the initial partons were generated. This universality, coupled with the significant difference in energy scales between the high-momentum transfers of the hard process and the lower-energy processes occurring within hadrons, allows for the factorisation of these processes in cross-section calculations.

As the parton shower evolves, the relevant momentum scale decreases, and the strong coupling $\alpha_s(Q)$ increases. This growth of α_s reduces the probability for further emissions, so the shower gradually dies out until a non-perturbative scale is reached. At that point, the quarks become confined within hadrons, which are composite particles with neutral color charge. These hadrons, along with their potential decay products, are then detected as collimated clusters of particles known as *jets*.

To effectively interpret jet data and other objects (leptons, photons etc.), simulations are crucial, allowing for comparisons with experimental results. To model the proton-proton collisions, Parton Distribution Functions (PDFs), denoted as $f_a(x)$, are employed. These functions represent the probability of finding a parton of flavor ‘a’ carrying a fraction ‘x’ (where x is between 0 and 1) of the proton’s total energy. The flavor ‘a’ can be gluons, valence quarks, or virtual (sea) quarks, though sea quarks are less probable at high momentum fractions due to their virtual nature.

The total cross-section for a collision between two hadrons (hh) producing a final state X is calculated by summing over all possible pairs of partons (a and b). The calculation involves integrating over the momentum fractions of the partons (x_a and x_b) and the phase space of the final state (Φ). The probability of finding a parton with a specific momentum fraction is given by

$f_a(x_a, \mu_f^2)$ and $f_b(x_b, \mu_f^2)$, which depend on the factorization scale μ_f . The hard scattering cross-section for the interaction of the partons a and b is represented by $d\sigma_{a,b \rightarrow X}(x_a, x_b, \mu_R^2)$, which depends on the renormalisation scale μ_R . This can be expressed as:

$$d\sigma_{hh \rightarrow X} = \sum_{a,b} \int dx_a dx_b d\Phi f_a(x_a, \mu_f^2) f_b(x_b, \mu_f^2) d\sigma_{a,b \rightarrow X}(x_a, x_b, \mu_R^2), \quad (1.55)$$

where X is an arbitrary final state and $d\sigma_{a,b \rightarrow X}(x_a, x_b, \mu_R^2)$ is the scattering cross-section of two partons a and b coming from the two hadrons, whose fraction of momentum and flavor are integrated and summed over, respectively. When hadrons (like protons) collide in a particle accelerator, the resulting event is incredibly complex, producing a large number of particles. To understand the overall properties of these collisions and the particles they create, physicists need to simulate the entire process from the initial interaction to the final detection of the particles. This simulation is done using computer programs called Monte Carlo (MC) generators, which build up a detailed, step-by-step history of the event, as depicted in Figure 1.2. These simulations are crucial for accurately measuring the overall characteristics of the collision.

The simulation of a collision event begins with computing the parton-level differential cross-section at a chosen perturbative order and particle multiplicity. Due to the factorial growth of Feynman diagrams with final-state particles, generators are typically limited to around 10 particles. The sampled points either estimate the total cross-section or represent simulated hard scatterings, with parton distribution functions used to assign parton probabilities. Since events are not sampled with exact probabilities, each is assigned a weight.

Once the hard process is simulated, the parton shower is applied under the assumption that it factorizes from the hard process and evolves independently. Each parton, whether from the initial or final state, undergoes radiative corrections, emitting additional partons that form jets. At the end of the parton shower, radiation stops and partons no longer have enough energy to remain free. The simulation then transitions to color-neutral hadrons using non-perturbative hadronisation models.

Alongside the hard scattering, leftover partons from the colliding hadrons can undergo multiple parton interactions. Their products also pass through parton showering, hadronisation, and interact via color reconnection with the initial and final states. The combination of these parallel processes is known as the Underlying Event (UE).

In many cases, LO (see Section 1.1.1 for explanations of LO and NLO) predictions lack precision due to large uncertainties related to the choice of the renormalisation scale or limited kinematic coverage. To improve accuracy,

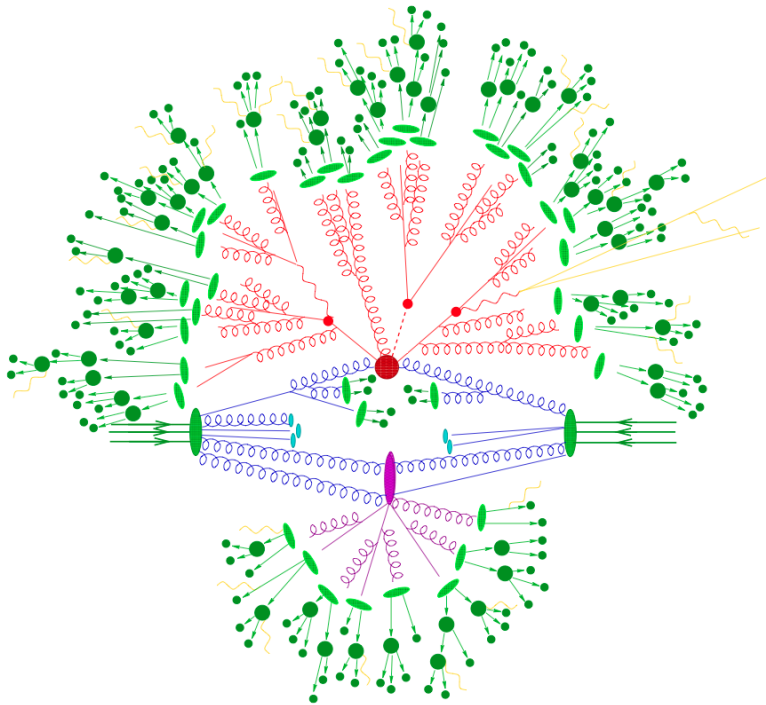


FIGURE 1.2 | Imagine a particle collision in an accelerator, like smashing two protons together. A computer program (a Monte Carlo event generator) simulates this. The core of the collision, where the particles directly interact, is shown as a red blob in the centre. As particles fly out, they can emit radiation, called Bremsstrahlung, which is shown as a branching, tree-like structure surrounding the red blob. Sometimes, there might even be a second, separate hard collision, shown as a purple blob. The quarks and gluons that are created eventually turn into the stable particles we see in detectors; this process is shown as light green blobs. Some of these stable particles then decay into other particles, shown as dark green blobs. Finally, sometimes photons (light particles) are emitted, shown as yellow lines. Essentially, this diagram shows a step-by-step breakdown of all the complex processes that happen in a particle collision, as simulated by a computer [17].

NLO calculations are used. Since parton showers already include higher-order effects beyond LO, matching procedures are needed to avoid double counting. In MG5_aMC@NLO [18], this introduces negative-weight events, reducing statistical efficiency, as more events must be simulated to balance the cancellations. Powheg [19], in contrast, avoids negative weights by specialized methods. Other generators like Whizard [20] and Sherpa [21] also support LO or NLO simulations with parton showers and additional features.

These simulations provide a detailed understanding of the underlying physics processes, which is essential for accurately interpreting the signals observed in the detector and reconstructing the properties of the particles produced in the collisions. To reconstruct jets in a consistent and infrared-safe way (i.e., stable under soft or collinear particle emissions), experimental analyses use clustering algorithms that group the measured particles into jets according to their proximity in the (y, ϕ) plane. The most widely used class at hadron colliders is the sequential recombination family, where a distance measure is defined between pairs of particles and between each particle and the beam. Popular choices include the k_T , Cambridge–Aachen, and, in particular, the anti- k_T algorithm, which is the LHC standard. The anti- k_T algorithm tends to produce approximately conical jets that are stable under soft and collinear radiation, making them well suited for comparisons between theory and experiment.

In addition to jets, the reconstruction of other objects such as leptons, photons, and missing transverse energy (E_T^{miss}) plays a crucial role in collider experiments. Leptons, including electrons and muons, are identified using their distinct signatures in the detector, such as tracks in the inner tracker and energy deposits in the electromagnetic calorimeter. Photons are reconstructed based on their energy deposits in the electromagnetic calorimeter without associated tracks. E_T^{miss} , which indicates the presence of neutrinos or other undetected particles, is calculated as the negative vector sum of the transverse momenta of all reconstructed objects in the event. These objects, along with jets, are essential for accurately characterizing the final states of collisions and for disentangling the complex structure of events.

A significant challenge in these experiments is to accurately infer the properties of the original quark or gluon that initiated the jet, as well as to identify and measure the properties of leptons, photons, and E_T^{miss} from the energy deposits and tracks measured by the detector. This requires sophisticated reconstruction techniques, including machine learning (ML), to disentangle the complex structure of jets and other objects, improve the reconstruction of the initial parton’s characteristics, and ensure precise identification of leptons and photons. These techniques are further complemented by comparisons with theoretical simulations, which are crucial for validating the methods and

ensuring consistency with the underlying physics models.

Accurate measurements in particle physics experiments rely heavily on precise corrections, calibrations, and the careful treatment of uncertainties. Detector responses must be calibrated to ensure that the measured signals, such as energy deposits in calorimeters or tracks, accurately reflect the true properties of the particles. This involves correcting for detector effects like non-uniformities, energy scale shifts, and resolution smearing. Additionally, systematic uncertainties arise from various sources, including the modeling of the detector response, theoretical predictions (e.g., parton distribution functions and renormalization scales), and the simulation of background processes. These uncertainties are quantified and propagated through the analysis to assess their impact on the final results. Robust calibration procedures and uncertainty evaluations are critical for ensuring the reliability of the reconstructed quantities, such as jets, leptons, photons, and missing transverse energy, and for enabling meaningful comparisons between experimental data and theoretical predictions.

1.2

Higgs boson pair production

The 2012 discovery of the Higgs boson by the ATLAS and CMS experiments yielded the first measurement of its mass [5,6]. These measurement, combined with the VEV, allowed physicists to determine the value of the λ parameter in Equation (1.48). This, therefore, led to predictions for how Higgs bosons interact with each other via trilinear and quartic couplings and the rates at which associated processes occur at the leading order, specifically the production of HH and three Higgs bosons (HHH). While observing three Higgs bosons is unlikely with current and near-future detectors due to its extremely low production rate, the study of HH production is ongoing.

Understanding HH production could provide valuable insights into the limitations of the SM as an effective theory at high energies or hint at a more complex Higgs sector. The strength of the trilinear Higgs interaction is particularly important because measuring it would test the SM's prediction of λ . This could significantly improve our understanding of the Higgs potential and the mechanism by which the Higgs boson obtained its VEV and gives mass to all fundamental particles.

1.2.1

HH production in the Standard Model

At a proton-proton collider, the primary way to produce pairs of Higgs bosons is through gluon-gluon fusion (ggF). This process, initially explored in references [22,23], is a loop-induced interaction. Even at the leading order, it involves Feynman diagrams containing loops of top and bottom quarks, as illustrated in Figure 1.3. One type of diagram features a triangular loop that produces a virtual (off-shell) Higgs boson, which then decays into two real (on-shell) Higgs bosons via the triple-Higgs interaction. However, two Higgs bosons can also be produced through box loop diagrams that do not involve the direct interaction of three Higgs bosons. These two sets of diagrams contribute to the probability of the process with opposite signs, resulting in a destructive interference. The reason that the rate at which pairs of Higgs bosons are produced is significantly low is primarily due to the contributions from the box and triangle Feynman diagrams largely cancelling each other out, as shown in Figure 1.4

The total rate for producing Higgs boson pairs in pp collisions is known with high precision (NNLO) in QCD). Reference [24] suggests applying a roughly 10% correction for the top quark mass, observed at NLO, to the total NNLO rate. Using a centre-of-mass energy of $\sqrt{s} = 13$ TeV and a Higgs boson mass of $m_H = 125$ GeV, this yields a cross-section (σ_{HH}) of 33.45 fb with uncertainties of $+4.3\% / -6\%$ (due to the energy scale of the process), $\pm 2.1\%$

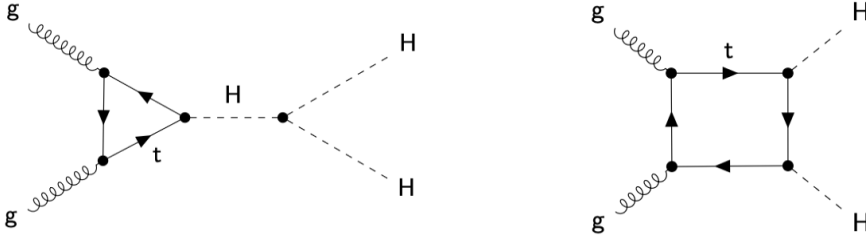


FIGURE 1.3 | Examples of the Feynman diagrams that contribute to the production of two Higgs bosons at the leading order. Notably, only triangle diagrams, like the one on the left, involve the Higgs boson's self-interaction strength, denoted by λ . Both top and bottom quarks can exist within the loop of these diagrams, but the top quark's contributions are much more significant to the overall probability due to its strong interaction with the Higgs field (large Yukawa coupling), appearing once in triangle diagrams and twice in box diagrams.

(due to uncertainties in the PDFs), and $\pm 2.3\%$ (due to the uncertainty in the strong coupling constant, α_s), as stated in Equation (1.56).

$$\sigma_{HH} = 34.45 fb_{-6\%}^{+4.3\%} (\text{scale}) \pm 2.1\%(\text{PDF}) \pm 2.3\%(\alpha_s) \quad (1.56)$$

Current research on non-resonant double Higgs production primarily aims to measure the trilinear Higgs coupling and compare it to the value predicted by the SM based on the known Higgs mass and VEV. Contrary to what might initially be expected, the HH production rate doesn't necessarily increase as the strength of this coupling increase due to interference effects previously discussed. The relationship between the cross-section and the coupling modifier ($\kappa_\lambda = \lambda/\lambda_{SM}$) is shown in Figure 1.5.

The second most significant way to produce pairs of Higgs bosons is via Vector Boson Fusion (VBF) production mode. In this process, the Higgs bosons originate from the exchange of virtual W or Z bosons in the t-channel, or through a similar mechanism where a single, virtual Higgs boson is produced and then decays into a pair. Other production methods, such as double Higgs-strahlung (where a Z or W boson is also produced) and production alongside a top quark-antiquark pair, are analogous to their single Higgs boson production counterparts but are not expected to be observed until the next generation of more powerful experiments. In general, any Feynman diagram that can produce a single Higgs boson can also lead to double Higgs production by adding a trilinear Higgs coupling vertex to the diagram, which significantly reduces the probability of the process. The cross-sections for double Higgs production, including QCD corrections and assuming a Higgs boson mass of 125 GeV and a collision energy of $\sqrt{s} = 13$ TeV, are given by [25]:

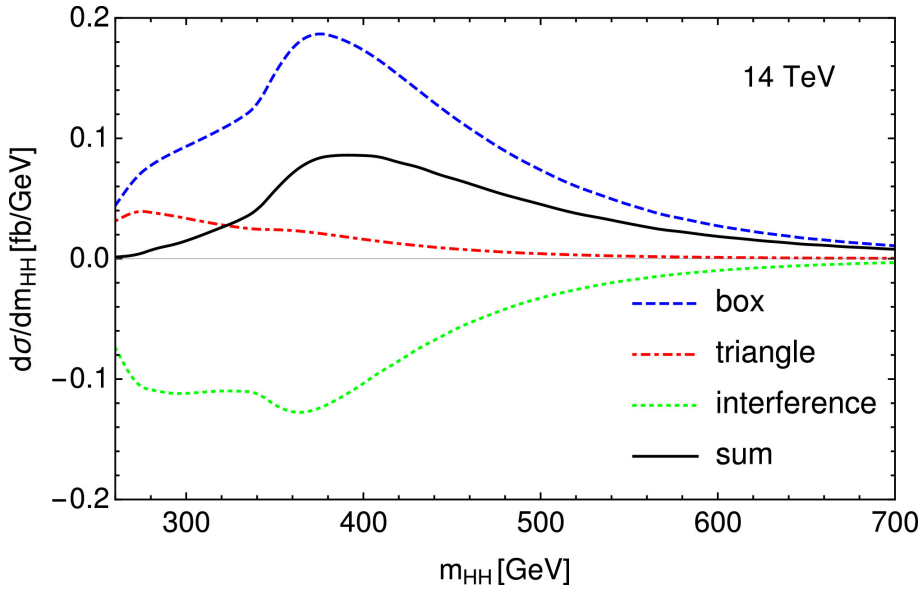


FIGURE 1.4 | Differential cross-section of producing a pair of Higgs bosons at the LO as function of the invariant mass of the Higgs pair. The figure breaks down the total probability into contributions from different underlying Feynman diagrams involved in gluon-gluon fusion, including how these contributions interfere with each other. The final, observable quantity is the sum of all these individual contributions and their interferences [25].

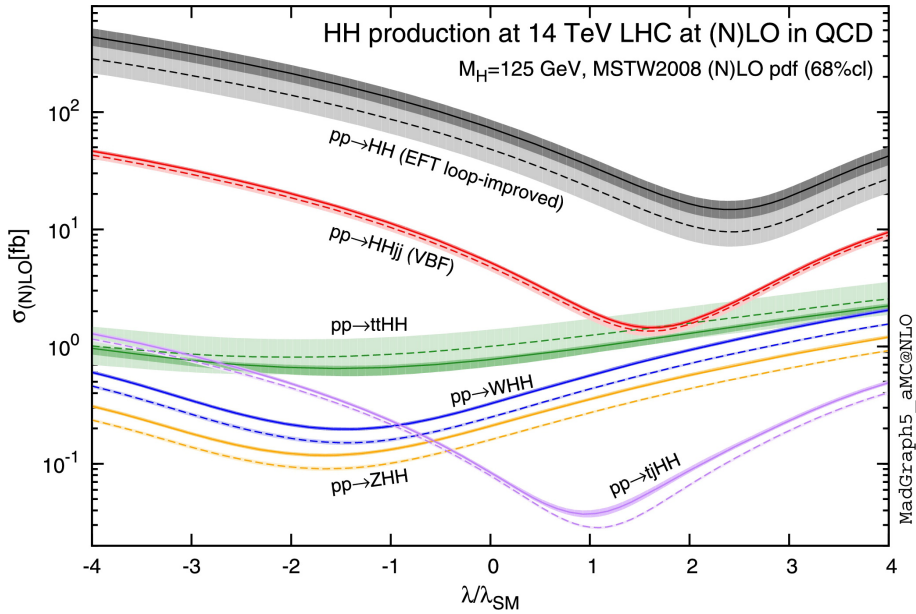


FIGURE 1.5 | Cross-sections for different production modes of pairs of Higgs bosons as function of the strength of the trilinear Higgs coupling (λ) is normalized over its SM value ($\kappa_\lambda = \lambda/\lambda_{SM}$). The dashed lines and lighter-colored bands represent calculations at the LO, while the solid lines and darker-colored bands show the results at the NLO, which are more accurate calculations. The colored bands around the lines indicate the theoretical uncertainties arising from the choice of energy scale in the calculations and from the uncertainties in the PDFs, with these two types of uncertainties added together linearly to show the total uncertainty. [26]

$$\begin{aligned}
\sigma_{\text{GGF}} &= 31.05 \text{ fb}_{-5.0\%}^{+2.2\%} (\text{scale}) \pm 3.0\%(\alpha_s + \text{PDF}) \\
\sigma_{\text{VBF}} &= 1.73 \text{ fb}_{-0.04\%}^{+0.03\%} (\text{scale}) \pm 2.1\%(\alpha_s + \text{PDF}) \\
\sigma_{\text{ZHH}} &= 0.363 \text{ fb}_{-2.7\%}^{+3.4\%} (\text{scale}) \pm 1.9\%(\alpha_s + \text{PDF}) \\
\sigma_{\text{W}^+\text{HH}} &= 0.329 \text{ fb}_{-0.41\%}^{+0.32\%} (\text{scale}) \pm 2.2\%(\alpha_s + \text{PDF}) \\
\sigma_{\text{W}^-\text{HH}} &= 0.173 \text{ fb}_{-1.3\%}^{+1.2\%} (\text{scale}) \pm 2.8\%(\alpha_s + \text{PDF}) \\
\sigma_{t\bar{t}\text{HH}} &= 0.775 \text{ fb}_{-4.3\%}^{+1.5\%} (\text{scale}) \pm 3.2\%(\alpha_s + \text{PDF}),
\end{aligned} \tag{1.57}$$

and the corresponding Feynman diagrams are shown in Figure 1.6.

The κ -framework is a widely used approach in experimental particle physics to parameterize potential deviations from the Standard Model (SM) predictions. In this framework, the couplings of the Higgs boson to other particles and to itself are modified by scaling factors, denoted as κ parameters. For example, κ_λ modifies the trilinear Higgs self-coupling, while κ_t scales the top quark Yukawa coupling. These modifiers are defined such that $\kappa = 1$ corresponds to the SM prediction. Any measured deviation of κ from 1 would indicate new physics beyond the Standard Model (BSM).

The κ -framework is particularly valuable because it provides a model-independent way to test the consistency of experimental results with the SM. By introducing these scaling factors, experimentalists can quantify how well the data align with the SM or identify potential hints of BSM physics. However, while the κ -framework is effective for detecting deviations, it does not provide direct insight into the underlying theoretical model responsible for the deviations. For this reason, it is often complemented by more comprehensive approaches, such as Effective Field Theories (EFTs), which offer a deeper connection to possible BSM scenarios.

1.2.2 Introduction to Effective Field Theory

If the energy scale at which new physics emerges is significantly higher than the electroweak scale, then one can describe its indirect effects on low-energy observables in a way that doesn't depend on the specific details of the underlying model. These effects manifest as modifications to the SM Lagrangian, including both shifts in SM parameters such as couplings and masses, and the introduction of new, non-renormalisable local interactions among SM particles [27].

In this framework, the SM Lagrangian is extended into what's called an effective field theory [28–30], expressed as:

$$\mathcal{L}_{\text{SM}} \rightarrow \mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_{d>4} \sum_i \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)} \tag{1.58}$$

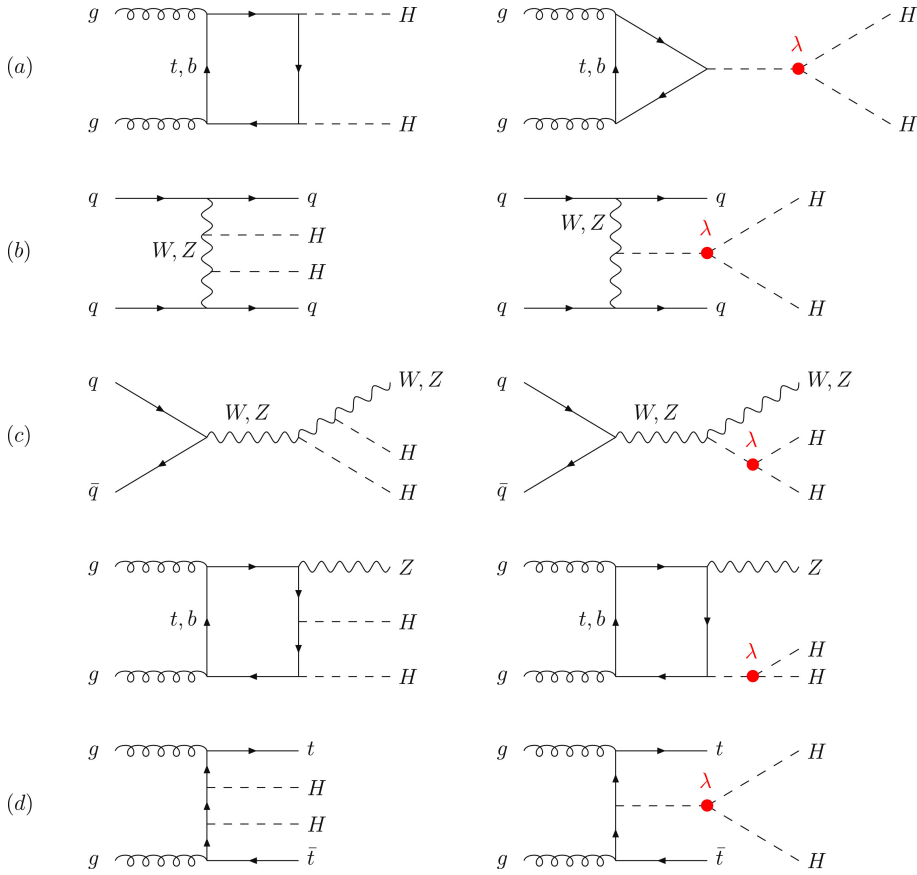


FIGURE 1.6 | Representative Leading Order Feynman diagrams for the dominant double Higgs production mechanisms: (a) gluon-gluon fusion, (b) vector-boson fusion, (c) double Higgs-strahlung, and (d) $t\bar{t}HH$ associated production. The trilinear Higgs self-coupling (λ_{HHH}) vertex is highlighted in red [25] .

Here, $\mathcal{O}_i^{(d)}$ are higher-dimensional operators (of mass dimension d) that obey Poincaré and SM gauge symmetries, and $c_i^{(d)}$ are known as Wilson coefficients, which determine the strength of these new interactions. The SM is recovered either if all $c_i^{(d)}$ vanish or in the limit where $\Lambda \rightarrow \infty$.

Though the expansion includes an infinite number of terms, only a finite number of them contribute at any given order in $1/\Lambda$. Moreover, the effect of these higher-dimensional operators on physical processes with energy E becomes increasingly negligible as powers of E/Λ , allowing the series to be truncated and only the most significant terms considered.

The following two conditions must be met for this approach to yield well-defined predictions:

1. The typical energy scale E of the processes under consideration must be smaller than the new physics scale Λ . When $E \approx \Lambda$, all operators in the effective expansion contribute comparably, the expansion becomes invalid, and the theory must be replaced by a UV-complete description of the new degrees of freedom.
2. The underlying UV theory must allow for a perturbative expansion in its couplings [31].

Although effective interactions result in a non-renormalisable theory, this does not prevent the calculation of higher-order corrections, since there is now a well-defined cut-off scale Λ [32]. For a discussion on the validity of EFT methods in Higgs sector measurements, see ref. [31].

In the EFT framework, the experimental objective is to constrain or detect non-zero values of the operator coefficients $c_i^{(d)}/\Lambda^{d-4}$. Assuming reasonable values for the Wilson coefficients, these constraints can be translated into limits at which new physics might appear. EFTs also help identify observables that are especially sensitive to deviations from SM predictions.

A minimal set of independent operators is known as a *basis*. At dimension six, two commonly used bases are the **Warsaw** basis [33] and the **SILH** basis [34,35]. There are 2,499 baryon-number-preserving dimension-six operators that also respect the required symmetries. However, under the widely used assumption of *Minimal flavor Violation* — which postulates that new physics respects the same flavor structure dictated by the SM Yukawa couplings — this number reduces significantly to just 76 parameters. If one also neglects CP-odd operators, this set further reduces to 53 independent operators, of which only a subset will typically contribute to any given physical process.

1.2.3 Double Higgs Production in Linear EFT (SMEFT)

There are various approaches to determining the EFT modifications to HH production, depending on the chosen operator basis. One common framework is the SM Effective Field Theory (SMEFT), which assumes the Higgs boson is part of an electroweak doublet, and new physics is significantly heavier than the electroweak scale. In this approach, higher-dimensional operators (such as those in the Warsaw or SILH bases) modify the SM Lagrangian.

For instance, specific dimension-six operators modify the Higgs self-coupling and the top quark Yukawa interaction, which are already present in the leading-order Feynman diagrams for double Higgs production. Furthermore, these operators can generate entirely new contact interactions not present in the SM, such as $t\bar{t}HH$, HGG , and $HHGG$ vertices. These new interactions lead to additional tree-level Feynman diagrams, as illustrated in Figure 1.7.

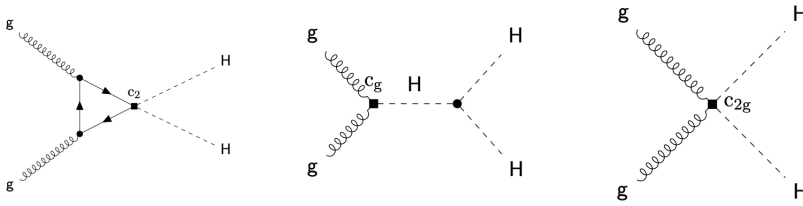


FIGURE 1.7 | Additional Feynman diagrams potentially contributing to Higgs pair production. The new vertices arising from BSM physics are depicted as black squares. The vertex on the left is induced by effective operators generating a $t\bar{t}HH$ contact term. The vertices in the centre and right are induced by operators generating HGG and $HHGG$ contact terms. Note that the labels c_2 , c_g and c_{2g} within the squares are common shorthand notations for these effective interactions, often used in the context of non-linear EFT (discussed in Section 1.2.4).

While SMEFT provides a rigorous, explicitly gauge-invariant formulation, the experimental results presented in this thesis primarily rely on the κ -framework and its natural extension, the non-linear EFT (HEFT). Note that the labels c_2 , c_g , and c_{2g} shown in Figure 1.7 are common shorthand notations for these effective interactions, bridging the gap to the non-linear framework discussed in the following section. This non-linear approach allows for a more direct translation of the κ_λ and κ_{2V} coupling modifiers that are explicitly targeted and constrained in the final experimental analyses.

1.2.4

Double Higgs Production in Non-linear EFT (HEFT)

The use of non-linear EFT to describe BSM effects in Higgs pair production was discussed in refs. [24, 36, 37]. The relevant part of the non-linear Lagrangian is given by:

$$\begin{aligned} \mathcal{L}_{\text{nonlin}} \supset & -\kappa_\lambda \frac{m_H^2}{2v} H^3 - m_t \bar{t}t \left(\kappa_t \frac{H}{v} + c_2 \frac{H^2}{v^2} \right) \\ & + \frac{\alpha_s}{12\pi} \left(c_g \frac{H}{v} + c_{2g} \frac{H}{2v^2} \right) G_{\mu\nu}^a G^{a,\mu\nu}, \end{aligned} \quad (1.59)$$

In this expression, five couplings need to be determined, but the relationship between the contact terms involving one or two Higgs bosons is lost, and thus the connection to single-Higgs processes is obscured. The SM is recovered when $\kappa_\lambda = \kappa_t = 1$ and $c_2 = c_g = c_{2g} = 0$. Here, κ_λ and κ_t are anomalous couplings that modify the cubic Higgs self-coupling and the top quark Yukawa coupling, respectively. The other couplings lead to diagrams identical to those in the linear EFT, as shown in Figure 1.7.

In scenarios where $c_2 = c_g = c_{2g} = 0$, and denoting by $\mathcal{A}_\Delta$ and $\mathcal{A}_\square$ the LO amplitudes corresponding to the triangle and box diagrams in Figure 1.3, the matrix element squared for HH production can be written as:

$$|\mathcal{M}|^2 = \left| \kappa_\lambda \kappa_t \mathcal{A}_\Delta + \kappa_t^2 \mathcal{A}_\square \right|^2 = \kappa_t^4 \left| \frac{\kappa_\lambda}{\kappa_t} \mathcal{A}_\Delta + \mathcal{A}_\square \right|^2. \quad (1.60)$$

This expression highlights the quadratic dependence of the cross-section on the Higgs self-coupling κ_λ . Furthermore, since the matrix element's kinematic features are determined only through the amplitudes $\mathcal{A}_\Delta$ and $\mathcal{A}_\square$, it follows that any observable sensitive to HH kinematics will depend only on the ratio κ_λ/κ_t . Importantly, this observation remains valid even beyond LO, despite the absence of clearly distinguishable box and triangle amplitudes at higher orders.

This framework (Equation (1.60)) also allows us to connect the HEFT parameters (κ_λ, κ_t) to the SMEFT deviations. By comparing the amplitude expressions from both EFTs, we can infer how these parameters relate, providing a bridge between the two approaches.

In practical terms, the most relevant HEFT parameters for Higgs pair production are κ_λ , which rescales the Higgs self-coupling and therefore the triangle diagram; κ_t , which rescales the top Yukawa coupling and affects both the box and triangle diagrams; and the non-linear contact couplings c_2 , c_g , and c_{2g} , which generate the $t\bar{t}HH$, HGG , and $HHGG$ vertices, respectively. For

vector-boson-fusion topologies, the analogous modifiers are κ_V and κ_{2V} , which govern the single- and double-Higgs couplings to electroweak vector bosons. These are the standard coupling labels that appear in the corresponding HEFT Feynman-diagram representations.

1.2.5 Modeling of non-resonant BSM effects on HH production

While the sensitivity of HH production to new physics makes it an excellent probe for deviations from SM predictions, it presents significant challenges for experimental analyses. Efficiently modeling these deviations, especially those arising from EFT effects, is crucial for optimizing analyses, setting limits, or characterizing new physics. The straightforward approach of generating Monte Carlo samples across the multi-dimensional EFT parameter space is computationally impractical.

To overcome this, experimental collaborations often rely on the matrix-element reweighting technique. By defining a limited set of points in the parameter space as kinematic benchmarks, physicists can represent the possible types of kinematic behavior of HH within the non-linear EFT.

This procedure yields a small, manageable set of representative points in the EFT parameter space. This allows experimental searches to be optimized for broad kinematic variations, increasing the sensitivity of the analyses to a wider range of potential signal kinematics without the need for continuous Monte Carlo generation. Despite the fact that the chosen benchmarks represent discrete points, it is possible to model the kinematics of HH production for any values of the anomalous couplings by reweighting the events of the benchmark samples. These weights, derived from the matrix element of the process, are functions of the EFT couplings and allow for smooth interpolations across the parameter space during statistical analysis.

1.2.6 Review of current experimental results on HH production

The current experimental results on HH production from pp collisions at the LHC are reviewed in this section, focusing exclusively on the reported upper limits from the ATLAS and CMS collaborations. The discussion includes integrated luminosities, centre-of-mass energies, observed and expected upper limits on production cross-sections or signal strengths, reported significances, and derived constraints on the Higgs boson trilinear self-coupling modifier and the quartic HHVV coupling modifier (k_{2V}) across all investigated decay channels.

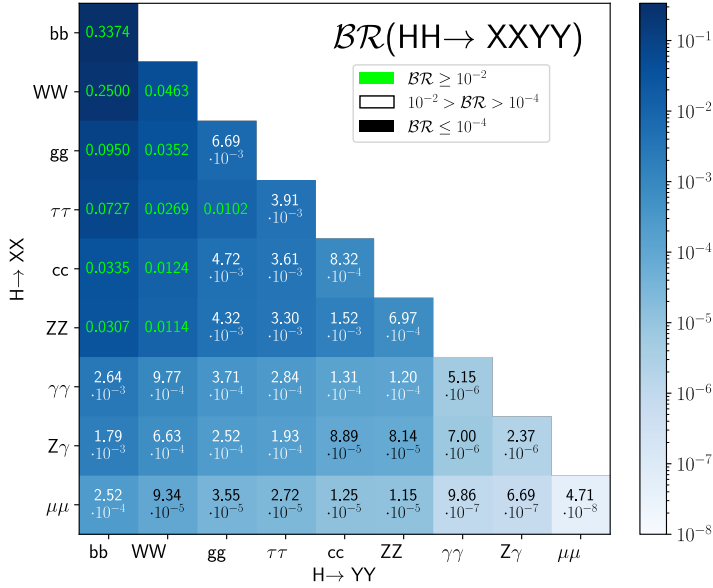


FIGURE 1.8 | Branching ratios of the HH decay modes. The values are calculated using the single Higgs boson branching ratios taken from ref. [24].

HH $\rightarrow b\bar{b}b\bar{b}$ Channel This decay channel represents the most probable final state for Higgs boson pairs, with an SM branching fraction of approximately 34% (see Figure 1.8). Despite this high branching ratio, experimental searches in this channel face considerable challenges due to overwhelming hadronic backgrounds, predominantly originating from multi-jet QCD processes. The difficulty in accurately modeling these large backgrounds, which simulations often struggle to predict, significantly impacts the sensitivity of this channel, illustrating that a high branching ratio does not automatically guarantee the most stringent experimental limits. Another issue is the ambiguity in determining which Higgs boson each b quark originated from. These necessitate the development and application of complex data-driven background estimation methods.

The latest searches for non-resonant $HH \rightarrow b\bar{b}b\bar{b}$ production by both the ATLAS and CMS collaborations, utilizing integrated luminosities of 126 fb^{-1} and 138 fb^{-1} respectively, of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ from LHC Run 2, found no evidence of a signal [38,39]. Consequently, no observed or expected significance in terms of standard deviations was reported by either experiment. ATLAS determined an observed (expected) upper limit on the cross-section of 5.4 (8.1) times the SM prediction at 95% Confidence Level (CL), while CMS reported a tighter observed (expected) limit of 3.9 (7.8) times the

SM prediction. For individual production modes, ATLAS reported observed (expected) 95% CL upper limits on the signal strength for ggF of 5.5 (8.2) and for VBF production of 130 (130). Both collaborations derived constraints on κ_λ and κ_{2V} . ATLAS observed (expected) 2σ constraints for κ_λ were $[-3.5, 11.3]$ $([-5.4, 11.4])$, and for κ_{2V} were $[-0.0, 2.1]$ $([-0.1, 2.1])$. CMS observed (expected) 95% CL allowed intervals for κ_λ were $[-2.3, 9.4]$ $([-5.0, 12.0])$, and for κ_{2V} were $[-0.1, 2.2]$ $([-0.4, 2.5])$, with CMS noting their κ_{2V} constraints were the most stringent at the time of publication.

Methodologically, the analyses diverged significantly. For background estimation, ATLAS utilized a neural network with deep ensembles to reweight events with exactly two b-tagged jets to model the signal region, extracting results from a fit to the reconstructed m_{HH} spectrum. Conversely, CMS employed a Boosted Decision Tree (BDT) to reweight events with three b-tagged jets, extracting signals using the output of a multivariate BDT discriminant. Jet pairing strategies also differed, with ATLAS prioritizing angular proximity and CMS prioritizing mass consistency.

HH $\rightarrow b\bar{b}\tau\bar{\tau}$ Channel This channel presents a more balanced experimental view compared to the $b\bar{b}b\bar{b}$ channel. With a moderate SM branching fraction of 7.3%, it benefits from relatively lower background contamination. This makes it an attractive channel for HH searches, offering a cleaner signature than fully hadronic final states.

The ATLAS and CMS collaborations have both conducted searches for non-resonant $HH \rightarrow b\bar{b}\tau\bar{\tau}$ production using integrated luminosities of 139 fb^{-1} and 138 fb^{-1} respectively, of proton-proton collision data collected at $\sqrt{s} = 13 \text{ TeV}$ during LHC Run 2 [40,41]. Neither search observed a significant excess of events above the expected background. While the specific ATLAS publication did not explicitly provide derived constraints on κ_λ and κ_{2V} , such contributions are incorporated into broader ATLAS combined analyses. CMS set an observed (expected) upper limit on the inclusive HH production cross-section of 3.3 (5.2) times the SM prediction at 95% CL, and for VBF HH production, an observed (expected) limit of 124 (154) times the SM prediction. ATLAS reports an observed (expected) κ_λ constraint of $[-2.7, 9.5]$ $([-3.1, 10.2])$ at 95% CL. CMS, however, derived constraints for κ_λ as $[-1.7, 8.7]$ at 95% CL and for κ_{2V} as $[-0.4, 2.6]$ at 95% CL. Both analyses employed multivariate discriminants, such as BDT and Deep Neural Networks (DNN), for signal identification and background rejection, with improvements in b-tagging, tau lepton identification, and trigger strategies being crucial for enhanced sensitivity.

The analyses employed distinct multivariate strategies. CMS utilized an ensemble of DNN to perform multi-class classification, separating the signal from specific backgrounds across three distinct final states ($\tau_e\tau_h$, $\tau_\mu\tau_h$, $\tau_h\tau_h$). In

contrast, ATLAS employed a BDT to discriminate signal from background in combined $\tau_{lep}\tau_{had}$ and $\tau_{had}\tau_{had}$ channels.

HH $\rightarrow b\bar{b}\gamma\gamma$ Channel This channel is characterized by a very small SM branching fraction of around 0.26%. However, its distinct advantage lies in the clean experimental signature provided by the precisely reconstructible diphoton invariant mass, which leads to relatively low background contamination. This cleanliness of the final state allows for effective signal extraction even from rare events, positioning it as a "golden channel" for precision measurements due to its excellent mass resolution.

The ATLAS and CMS collaborations have conducted searches for HH $\rightarrow b\bar{b}\gamma\gamma$ production [42, 43]. The latest ATLAS result combined full LHC Run 2 and early Run 3 datasets, accumulating over 300 fb^{-1} of proton-proton collision data at 13 TeV and 13.6 TeV. This led to a measured signal strength of $\mu_{HH} = 0.9^{+1.4}_{-1.1}$ (expected $1.0^{+1.3}_{-1.0}$), and an observed (expected) upper limit on the signal strength of 3.8 (2.6) times the SM prediction at 95% CL. The observed statistical significance of the SM signal was 0.84σ , consistent with expectations. ATLAS derived 95% CL constraints for κ_λ of $[-1.7, 6.6]$ and for κ_{2V} of $[-0.5, 2.6]$. The CMS search utilized 137 fb^{-1} of Run 2 data at $\sqrt{s} = 13 \text{ TeV}$, yielding an observed (expected) upper limit on the inclusive HH production cross-section of 7.7 (5.2) times the SM prediction at 95% CL, and for VBF HH production, 225 (208) times the SM prediction. CMS provided 95% CL constraints on κ_λ of $[-3.3, 8.5]$ (expected $[-2.5, 8.2]$), further refining these to exclude values outside $[-2.7, 8.6]$ when combined with $t\bar{t}H(\rightarrow\gamma\gamma)$ and fixing $\kappa_t=1$. Both collaborations relied on advanced multivariate techniques, including machine learning classifiers and precise object identification, to mitigate backgrounds, with CMS also using data-driven methods for non-resonant $\gamma(\gamma)$ +jets backgrounds and fitting two-dimensional $(m_{\gamma\gamma}, m_{jj})$ distributions.

The analyses diverged in their signal extraction and resolution strategies. CMS performed a maximum likelihood fit to the two-dimensional $(m_{\gamma\gamma}, m_{jj})$ distribution, utilizing a discrete profiling method to account for the bias associated with the choice of the background function. Conversely, ATLAS extracted results from a one-dimensional fit to the $m_{\gamma\gamma}$ spectrum. To enhance sensitivity, ATLAS implemented a two-step kinematic fit that leverages the high-precision photon measurements to improve the b -jet momentum resolution, alongside the usage of a Transformer-based b -tagging algorithm.

HH $\rightarrow b\bar{b}WW/b\bar{b}ZZ$ (Multilepton) Channels Searches for HH production in channels where one Higgs decays to $b\bar{b}$ and the other to vector bosons (W or Z)—collectively HH $\rightarrow b\bar{b}VV$ —are motivated by sizable branching ratios and distinctive experimental signatures involving leptons and E_T^{miss} . Both ATLAS

and CMS have analyzed the full Run 2 dataset at $\sqrt{s} = 13$ TeV (about 140 fb^{-1}), focusing on final states such as $2b + 2\ell + \text{missing } E_T$ (which can be originated from either $bbWW$ or $bbZZ$), single- and di-lepton channels, and multilepton ($2b4\ell$) signatures [44–47]. Advanced multivariate techniques were used to suppress dominant backgrounds, particularly from top-quark production. No significant excess above SM expectations was observed. ATLAS set a 95% confidence level upper limit of 9.7 (expected 16.2) times the SM prediction for non-resonant HH production in the $bbWW$ channel, with constraints on the Higgs self-coupling modifier κ_λ of $[-6.2, 13.3]$ and on κ_{2V} of $[-0.17, 2.4]$ [44]. CMS reported a corresponding limit of 14 (expected 18) times the SM value, with κ_λ constrained to $[-7.2, 13.8]$ (expected $[-8.7, 15.2]$) and κ_{2V} to $[-1.1, 3.2]$ (expected $[-1.4, 3.5]$), and a much weaker limit for VBF production [46].

In multilepton channels, CMS set an observed (expected) upper limit of 30 (37) times the SM cross-section for $HH \rightarrow bb4\ell$, with κ_λ constrained to $[-9, 14]$ (expected $[-11, 16]$) [48]. Although a dedicated ATLAS result for $bbZZ$ is not available, such final states are included in their broader multilepton searches [45].

Methodologically, the analyses employed distinct classification strategies. CMS implemented a DNN augmented by a Lorentz Boost Network (LBN) to process four-vector inputs for the $bbWW$ channel, while utilizing a BDT for the statistically limited $bbZZ$ channel. In contrast, the ATLAS multilepton analysis employed BDTs across all categories, classifying events strictly by the multiplicity of light leptons and hadronic taus (e.g., 2ℓ same-charge, 3ℓ , 4ℓ) rather than by specific intermediate vector boson decays.

Combined Results from ATLAS and CMS Both the ATLAS and CMS collaborations have performed comprehensive statistical combinations of searches for Higgs boson pair production across multiple decay channels, substantially improving sensitivity to this rare process and yielding the most stringent constraints to date on the Higgs self-coupling.

The latest ATLAS combination, using $126\text{--}140 \text{ fb}^{-1}$ of Run 2 data at $\sqrt{s} = 13$ TeV, sets an observed (expected) upper limit of 2.9 (2.4) times the SM prediction at 95% confidence level, with no significant excess above background expectations. The analysis constrains the Higgs boson self-coupling modifier κ_λ to the range $[-1.2, 7.2]$ and the quartic $HHVV$ coupling modifier κ_{2V} to $[0.6, 1.5]$, both consistent with the SM [49]. These results are achieved by combining a broad set of decay channels, including $bbbb$, $bb\tau\tau$, $bb\gamma\gamma$, multilepton, and $bb\ell\ell + E_T^{\text{miss}}$, with improved sensitivity arising from advances in individual analyses and refined statistical treatments to ensure compatibility and avoid event overlap. The ongoing accumulation of Run 3 data is expected to further enhance sensitivity. CMS has similarly combined results from a wide range of channels, utilizing

138 fb^{-1} of Run 2 data at $\sqrt{s} = 13 \text{ TeV}$. The observed (expected) upper limit on the inclusive HH production cross-section is 3.5 (2.5) times the SM prediction, with no significant signal observed [50]. The combined analysis constrains κ_λ to $[-1.4, 6.8]$ (expected $[-1.0, 7.0]$) and κ_{2V} to $[0.6, 1.4]$ (expected $[0.7, 1.4]$), consistent with SM expectations. These combinations include channels such as $bbbb$, $bb\tau\tau$, $bb\gamma\gamma$, $bbWW$, multilepton, and $bbZZ$.

Looking ahead, both collaborations anticipate further progress with Run 3 and the HL-LHC, which are expected to yield even tighter constraints and potentially the first evidence for Higgs boson pair production.

1.2.7 Overview and Outlook

Given the extreme rarity of HH production, a multi-channel approach is essential, leveraging the complementary advantages and challenges of different final states. Progress in experimental sensitivity is driven by two main factors: the increasing integrated luminosity delivered by the LHC and the continuous refinement of analysis methodologies. The transition from Run 2 to early Run 3 data, combined with the widespread adoption of advanced machine learning algorithms, such as BDTs and DNNs, has remarkably improved the precision and reach of these searches. As a result, physicists are able to extract more information from the available data, continually pushing the boundaries of experimental capability.

Current combined analyses from ATLAS and CMS experiments constrain the Higgs boson self-coupling modifier and the quartic HHVV coupling modifier to ranges consistent with SM predictions as shown in Figure 1.9. However, the allowed parameter space remains broad, and a definitive measurement of the Higgs self-interaction is still beyond the present statistical reach of the LHC. Looking ahead, the HL-LHC is expected to transform the search for HH production. With significantly increased integrated luminosity, the HL-LHC is anticipated to provide the first evidence—and potentially the observation—of HH production, enabling much tighter constraints on the Higgs self-coupling. This will offer unprecedented insights into the Higgs potential and the mechanism of electroweak symmetry breaking. The continued synergy between growing data samples and innovative analysis techniques will be crucial in unraveling the mysteries of the Higgs boson self-interaction and exploring new frontiers in particle physics.

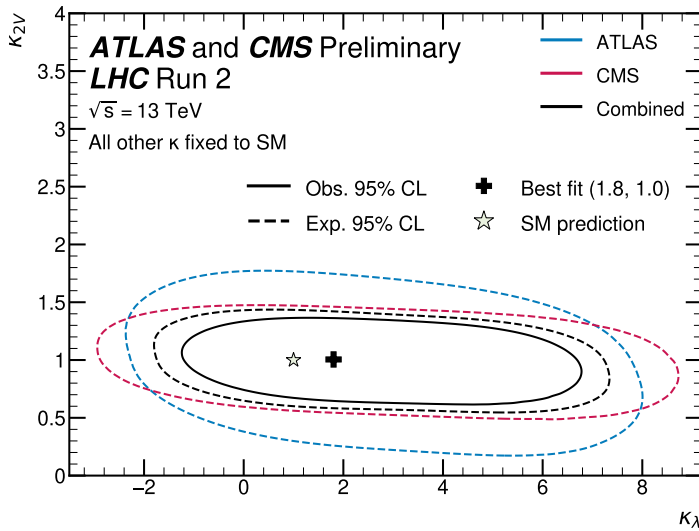


FIGURE 1.9 | Observed and expected 95% confidence level contours from a simultaneous profile likelihood scan in the $(\kappa_\lambda, \kappa_{2V})$ parameter space. The plot also displays the expected constraints from each experiment individually. All other Higgs boson couplings are set to their SM values [51].

Chapter

2

Experimental Setup: CERN and The Compact Muon Solenoid Experiment

This chapter focuses on the experimental equipment deployed at CERN, with particular emphasis on the Compact Muon Solenoid experiment (CMS), one of the Large Hadron Collider's flagship detectors. The CMS is an immense and technically complex device designed to detect a wide variety of particles and phenomena arising from high-energy proton-proton collisions at unprecedented energies, creating conditions similar to those just moments after the Big Bang. Through its layered structure and specialized subsystems, the detector captures detailed information about particle trajectories, energies, and interactions, enabling physicists to probe fundamental questions about the composition of matter and the forces governing the universe. In the following sections, we describe briefly the CERN's LHC and CMS detector's design, operational principles, and the technologies that support its broad scientific program.

2.1

The European Organization for Nuclear Research (CERN)

The European Organization for Nuclear Research, commonly known by its French acronym CERN (Conseil Européen pour la Recherche Nucléaire), is the world's largest particle physics laboratory and one of the most significant scientific collaborations in history. Founded in 1954 by 12 European countries, CERN emerged from post-World War II efforts to unite European scientists and restore Europe's position in fundamental physics research. The organization was strategically located near Geneva, Switzerland, at the Franco-Swiss border, after Geneva was selected as the site in 1952 and approved by referendum in

1953.



FIGURE 2.1 | An aerial view of the LHC and the Super Proton Synchrotron (SPS) ring showing the locations of the main experiments and sites along with the French-Swiss border [52].

CERN's fundamental mission focuses on providing physicists with the necessary tools and infrastructure to conduct unprecedented experiments in particle physics, advancing our understanding of the building blocks of the universe. Today, CERN has expanded to include 25 member states along with numerous worldwide affiliations, making it one of the best international scientific collaborations that brings together people from diverse backgrounds to push the frontiers of science and technology.

2.2

The Large Hadron Collider: Engineering Marvel and Scientific Instrument

The Large Hadron Collider represents the pinnacle of particle accelerator technology and stands as the world's largest and most powerful particle accelerator. Constructed between 1998 and 2008 through collaboration with over 10,000 scientists and hundreds of universities across more than 100 countries, the LHC occupies a circular tunnel 27 kilometers (17 miles) in circumference, reaching depths of up to 175 meters beneath the France-Switzerland border (see Figure 2.1).

The name "Large Hadron Collider" reflects its key characteristics: "Large" refers to its massive 27-kilometer circumference, "Hadron" indicates it accelerates protons or ions (particles belonging to the hadron family), and "Collider" describes its function of creating head-on collisions between counter-rotating particle beams. The LHC reuses the tunnel originally built for the Large

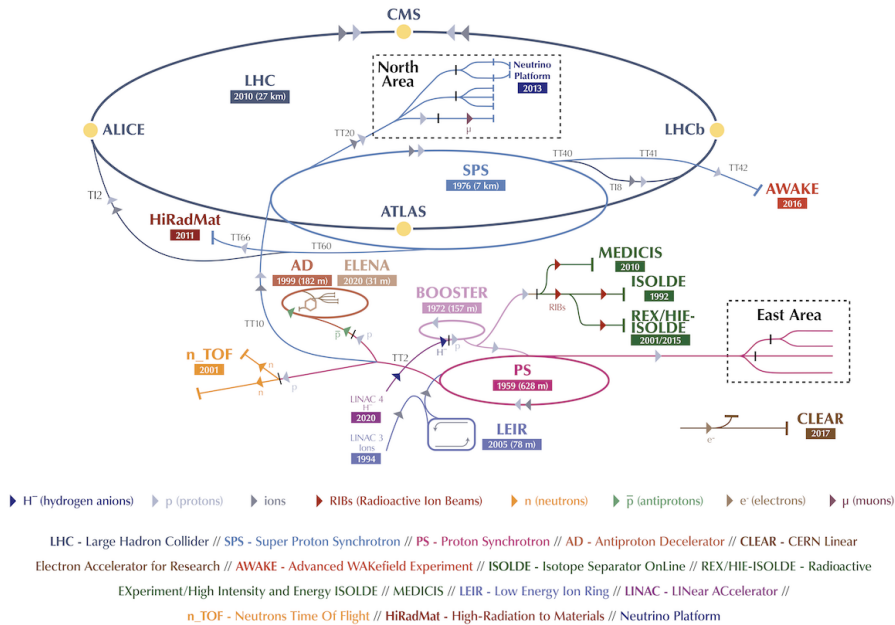


FIGURE 2.2 | Schematic diagram of CERN's accelerator complex showing the LHC and its injection [53].

Electron-Positron Collider (LEP), as an example of the evolutionary nature of particle physics infrastructure (see Figure 2.2).

The LHC operates by accelerating protons to extraordinary energies of 6.8 TeV per beam, creating collision energies of 13.6 TeV when the beams meet. The accelerator achieves these extreme energies using superconducting dipole magnets that generate magnetic fields of 8.33 tesla (T) (see Figure 2.3).

The engineering challenges to create world record collision energy represent one of the most advanced technological achievements in modern physics, requiring unprecedented advances across multiple engineering disciplines to achieve its revolutionary capabilities. The LHC's superconducting magnet system constitutes perhaps the most technically demanding component, with 1,232 main dipole magnets each generating 8.3 T magnetic fields while operating at the extreme temperature of 1.9 Kelvin (-271°C). These 15-meter-long, 35-tonne magnets must maintain field uniformity to within micrometers across their aperture, requiring precision manufacturing tolerances that push industrial capabilities to their absolute limits. The superconducting coils utilize 11,080 amperes of current flowing through niobium-titanium cables cooled by superfluid helium, demanding revolutionary cryogenic systems that consume 40 MW of electricity and 120 tonnes of helium. The LHC's vacuum system presents

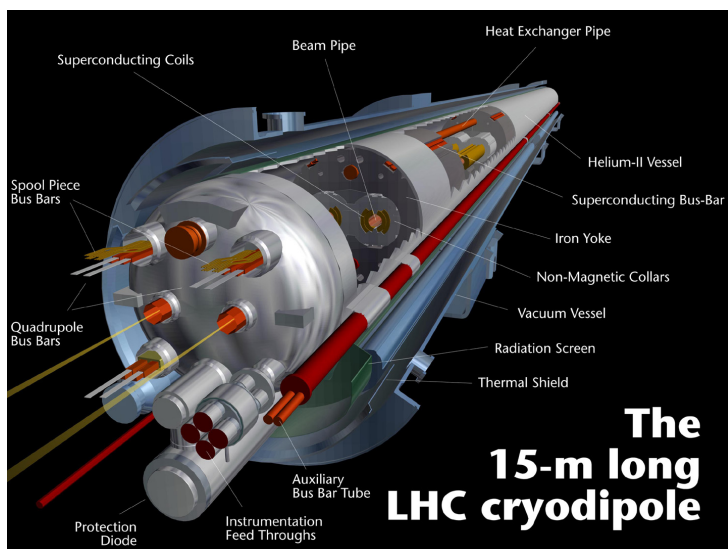


FIGURE 2.3 | Computer-generated cross-sectional diagram of the Large Hadron Collider tunnel showing beam pipes and magnetic systems [54].

equally extraordinary challenges, encompassing 104 kilometers of ultra-high vacuum piping maintained at pressures as low as 10^{-11} mbar achieved through 250,000 welded joints and 18,000 vacuum seals. Beam stability requirements demand micron-level precision in magnet alignment and orbit control, with feedback systems capable of correcting positional drifts as small as 10-25 micrometers. The extreme scale of coordinating these systems represents an engineering achievement that required pushing numerous technologies "well beyond today's frontiers" [55,56].

The LHC's beam structure can consist of up to 2,808 bunches of protons in each beam, with each bunch containing approximately 115 billion protons. These bunches are separated by 25 nanoseconds, creating a collision rate of 40 MHz at the four intersection points where experiments are located. The design luminosity—a measure of collision rate—reaches $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, first achieved in June 2016 and subsequently exceeded by a factor of two.

2.2.1 The Accelerator Complex

The LHC operates as the final stage in CERN's sophisticated accelerator complex, which progressively increases particle energies through a series of interconnected machines (see Figure 2.2). The acceleration chain begins with Linac4, which generates 160 MeV negative hydrogen ions, progresses through the Proton Synchrotron Booster, the 28 GeV Proton Synchrotron, and the SPS, which

accelerates particles to 450 GeV before injection into the LHC.

This complex system allows particles to be gradually accelerated over a period of approximately 20 minutes before reaching their final energy and circulating for 5 to 24 hours during physics runs. The LHC can also accelerate heavy ions, particularly lead ions, which reach energies of 5.3 TeV per nucleon pair for studies of quark-gluon plasma—the state of matter believed to have existed in the early universe.

2.3 The LHC and the CMS Experiment

The LHC hosts four major experiments at its collision points, each designed to investigate different aspects of particle physics. The two largest general-purpose detectors, ATLAS and CMS, jointly announced the discovery of the Higgs boson in 2012 and continue investigating its properties along with searches for new physics beyond the Standard Model.

ATLAS (A Toroidal LHC ApparatuS) represents the largest-volume particle detector ever constructed, measuring 46 meters in length, 25 meters in diameter, and weighing 7,000 tonnes (see Figure 2.4). Located in a cavern about 100 meters underground near Meyrin, Switzerland, ATLAS employs a distinctive magnet system built around a large central solenoid for inner tracking and two air-core toroidal systems (barrel and end-cap) for muon bending, yielding excellent standalone muon momentum resolution across a wide pseudorapidity range (η). The collaboration comprises more than 5,500 scientists from 245 institutes across 42 countries. For detailed information about ATLAS, see Ref. [57].

The Compact Muon Solenoid takes a different technical approach while pursuing similar scientific goals to ATLAS. Built around a massive 4-T solenoid magnet, the CMS detector measures 21 meters long, 15 meters wide, and 15 meters high, with a total weight of 14,000 tonnes (see Figure 2.5). Uniquely among LHC detectors, CMS was constructed in 15 sections at ground level before being lowered into an underground cavern near Cessy, France, and re-assembled. The collaboration involves approximately 5,500 particle physicists from 241 institutes in 54 countries. Since the work of this thesis depends on the data collected at the CMS, we will elaborate on it a little more.

2.3.1 CMS Subsystems and Layer Structure

The CMS experiment is designed to explore a broad range of studies, from the Standard Model (including the Higgs boson) to searching for signals of physics beyond the Standard Model such as extra dimensions and dark matter. The full detector measures 21 meters long, 15 meters wide, and 15 meters high, and

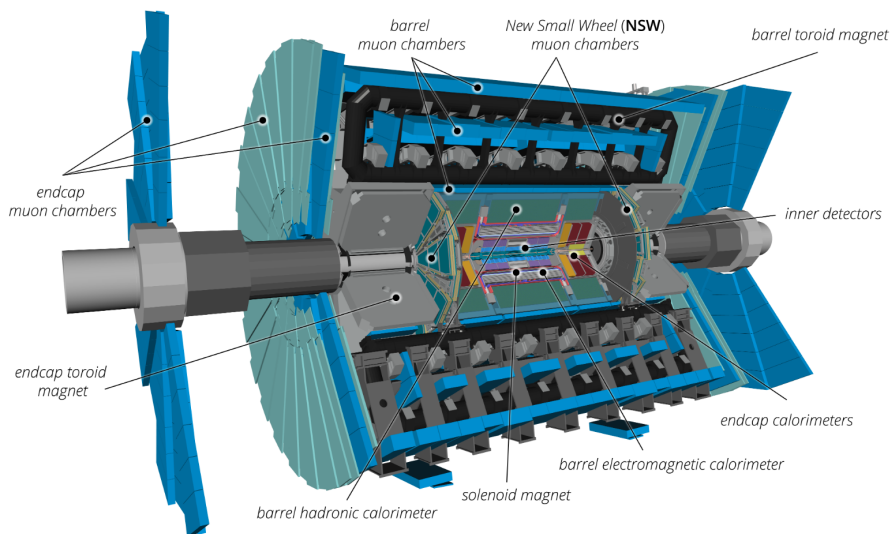


FIGURE 2.4 | Cut-away view of the Run 3 configuration of the ATLAS detector indicating notable sub-systems. [57].

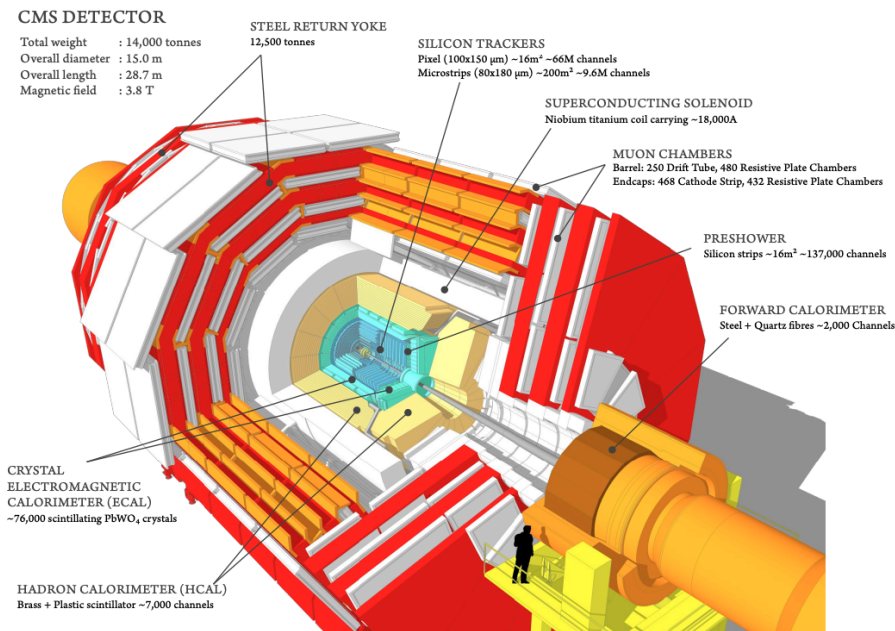


FIGURE 2.5 | Schematic drawing of the CMS detector, taken from Ref. [58].

the design emphasizes compactness, high magnetic field, and precision muon detection, inspiring the name "Compact Muon Solenoid". Its structure features concentric cylindrical layers around the beam pipe, each designed for precise detection of different particles (see Figure 2.6).

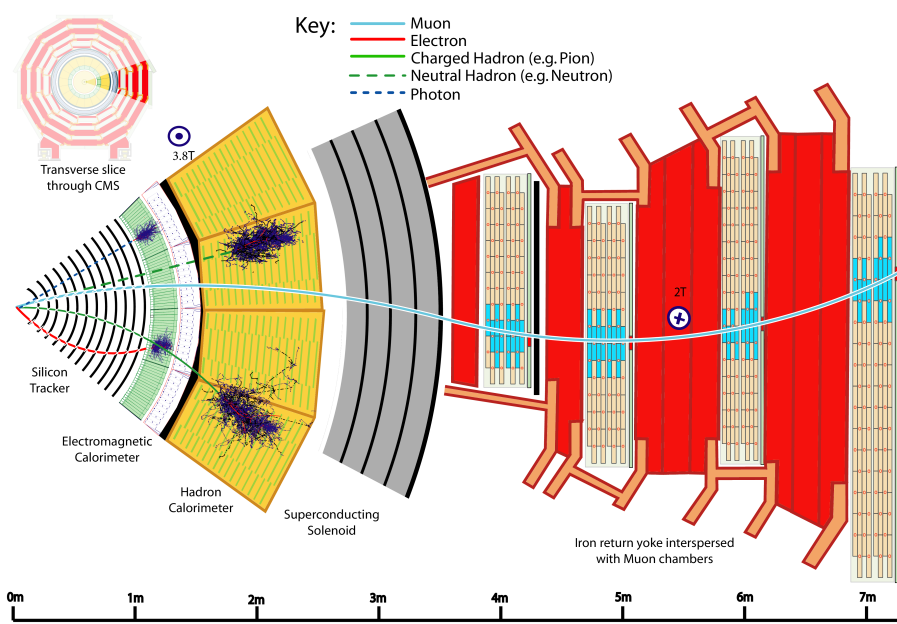


FIGURE 2.6 | Cutaway diagram of the CMS detector at CERN showing key components involved in the detection and data collection from particle collisions. [59].

Inner Tracking System

The core consists of four cylindrical layers of **silicon pixel detectors** for high-precision vertex and track reconstruction, at radii ranging from 2.9 cm to 16 cm, plus three forward endcap disks on each side located at $|z| = 29.1, 39.6,$ and 51.6 cm from the interaction point [60]. This innermost detector contains 124 million pixels, each with a size of $100\mu\text{m} \times 150\mu\text{m}$, enabling a spatial resolution of about $10\mu\text{m}$. This seemingly sub-pixel precision is achieved because the hit position is reconstructed from clusters spanning multiple pixels: the deposited charge diffuses during drift and is displaced by the magnetic field (Lorentz drift), leading to charge sharing between neighbouring pixels. Using the measured charge distribution within a cluster (rather than only the index of the hit pixel), the position can be interpolated with template- or charge-weighted methods, yielding a resolution substantially better than the binary limit of

pitch/ $\sqrt{12}$ (the standard deviation for a uniform hit position within one pixel when only the hit pixel is known).

Surrounding the pixel layers is a **silicon strip detector** comprising four inner barrel layers, six outer barrel layers, and two endcap sections (one at each side of the interaction point along $\pm z$) composed of multiple strip disks (the Tracker Inner Disks and Tracker Endcaps). With over 15,000 modules, it provides a total of about 10 million strips. This lightweight design minimizes material to limit multiple scattering and preserves tracking accuracy in the high radiation environment.

The intrinsic hit resolution of the strip tracker is set primarily by the strip pitch typically $\mathcal{O}(80\text{--}180)$ μm depending on the layer and radius and the track incidence angle. The strip-hit resolution is evaluated with the overlap “pair method” (see Figure 2.7), where two hits from overlapping modules in the same layer are crossed by the same track to suppress multiple-scattering effects. The study uses tracks with $p_T > 3$ GeV/ c , at least 8 valid hits, and track-fit probability $\geq 10^{-3}$, and retains hit pairs with cluster width ≤ 4 strips, identical cluster width in both modules, predicted propagation path < 7 cm, and predicted-distance uncertainty in the bending coordinate below 25 μm . The single-sensor resolution is extracted from

$$\sigma_{\text{hit}} = \sqrt{\frac{\sigma_{\text{pred-meas}}^2 - \sigma_{\text{meas}}^2}{2}}, \quad (2.1)$$

where $\sigma_{\text{pred-meas}}$ is the double difference between measured and predicted positions across the two sensors and σ_{meas} is the difference between measured positions. As illustrated in Figure 2.7, smaller strip pitch gives better resolution, while cluster sizes above one strip improve performance via charge sharing and analogue readout.

Electromagnetic Calorimeter (ECAL)

The ECAL uses nearly 76,000 lead tungstate (PbWO_4) crystals in a barrel and endcap geometry. These crystals are highly dense and radiation-hard, scintillate rapidly, and feature high granularity for accurate energy measurement of electrons and photons. Photo-detectors attached to the back of each crystal convert scintillation light into electrical signals, amplified and sent for analysis.

This sub-detector’s relative energy resolution improves with energy and is commonly parameterized as $\sigma_E/E = S/\sqrt{E} \oplus N/E \oplus C$, where E is the particle energy (in GeV), S is the stochastic term, N the noise term, C the constant term, and $\oplus$ denotes addition in quadrature. For the barrel, a representative test-beam result implies a relative resolution at the percent level at $E \sim \mathcal{O}(10)$ GeV and below 0.5% at $E \sim 100$ GeV [63].

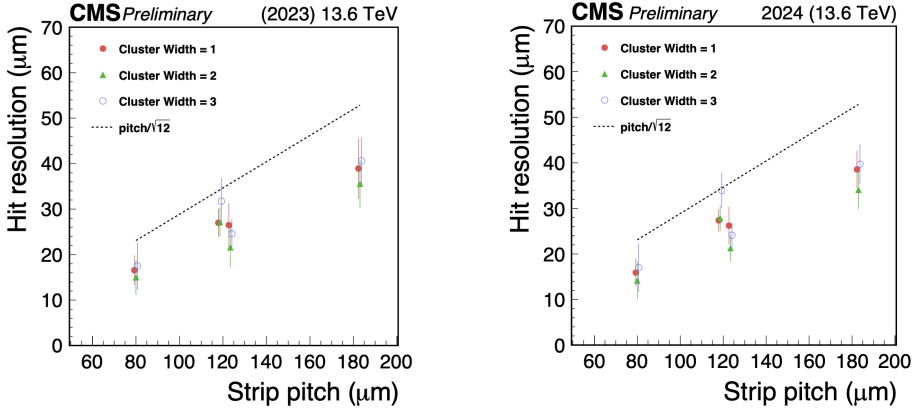


FIGURE 2.7 | Silicon strip hit resolution for 13.6 TeV pp collision runs in 2023 (left) and 2024 (right). Strip hit resolution is derived by selecting pairs of hits in different types of overlapping sensors and for different cluster widths expressed in units of number of strips. The theoretically expected resolution from a binary readout tracking detector (i.e., hit or no hit) is also displayed. By measuring the fraction of charge collected by adjacent strips, the hit position can be interpolated more precisely, leading to a resolution significantly better than the binary limit. [61, 62].

Hadronic Calorimeter (HCAL)

The hadronic calorimeter surrounds the ECAL and measures energies of hadrons. It comprises a barrel, two endcaps, an outer section beyond the solenoid, and forward calorimetry. While the central and endcap regions use layers of brass absorber and plastic scintillator tiles, the forward calorimeter (HF) employs a different technology to withstand the unprecedented particle fluxes and extreme radiation doses near the beamline. The HF consists of a steel absorber structure embedded with radiation-hard quartz fibers. Instead of scintillation, the active medium in the HF generates signals via Cherenkov light produced by charged shower particles above the Cherenkov threshold. Across the entire sub-detector, these optical signals are summed over "towers" to determine energy. The overall design ensures hermetic coverage with no gaps and strong resistance to radiation.

Following the same resolution parameterization described for the ECAL, a representative HCAL energy resolution curve demonstrates a relative resolution of approximately 100% at 1 GeV, improving to around 10% at 100 GeV, and approaching a constant term limit of about 5% at energies above 1000 GeV [63].

Superconducting Solenoid Magnet

The solenoid magnet is 12.5 meters long, 6 meters in diameter, and provides a 3.8 T field, the largest ever built for a particle detector. It is constructed from aluminum-stabilized Nb-Ti/Cu cable, cooled to 4 K (-269°C). The magnet bends the paths of charged particles, allowing accurate momentum measurement in conjunction with the tracker and muon system. The 12,500-ton steel yoke returns the field and supports the massive detector sections.

Muon Detection System

Outside the calorimeters and solenoid sits the CMS muon system, consisting of drift tubes, cathode strip chambers, resistive plate chambers and gas electron multipliers. Muons are tracked over large distances with high accuracy with chambers embedded at strategic locations enabling redundancy and trigger capabilities essential for physics analyses.

Unlike the calorimeters, the muon system measures momentum (p) rather than energy, determining it from the track curvature in the magnetic field. The offline momentum resolution of the standalone muon system is approximately 9% for small pseudorapidities (η) and momenta up to 200 GeV/ c , varying between 15% and 40% at $p = 1$ TeV/ c depending on η . However, when these measurements are combined with data from the inner tracking system to perform a “global” muon fit, the resolution improves dramatically. At lower momenta, the tracker dominates and improves the global resolution by an order of magnitude (to roughly 1–2%). At high momenta ($p \sim 1$ TeV/ c), the muon system’s long lever arm complements the inner tracker, yielding an overall global momentum resolution of approximately 5%.

See Figure 2.5 for a wide view of all sub-detectors of CMS.

2.3.2 Data Collection and Processing

The CMS experiment represents one of the most sophisticated data collection and processing systems among modern physics experiments. With proton bunches colliding at a rate of 40 MHz and producing terabytes of raw data per second, the experiment requires an intricate multi-stage data flow system to capture, process, and analyze collision events.

The CMS experiment employs a sophisticated two-level trigger system designed to reduce the incoming data rate from 40 MHz to approximately 1 kHz for permanent storage. This reduction is essential as recording every collision would produce unmanageable data volumes.

Level-1 Trigger (L1) operates using custom-designed electronics located in the underground service cavern. It processes data from the calorimeter and muon systems exclusively, with a maximum decision latency of $3.2\ \mu\text{s}$. During this brief window, event data is held in front-end pipeline buffers awaiting the trigger decision. The L1 trigger reduces the event rate from 40 MHz to approximately 100 kHz by identifying events containing high transverse energy electrons, photons, jets, muons, or significant missing transverse energy.

High-Level Trigger (HLT) represents the second stage, implemented in software running on a farm of approximately 20,000 CPU cores housed on the surface above the CMS cavern. The HLT performs more sophisticated reconstruction and selection algorithms, using faster versions of offline reconstruction software to further reduce the event rate to about 1 kHz for prompt reconstruction and permanent storage.

Data Acquisition and Trigger System

The Data Acquisition (DAQ) system serves as the critical bridge between the detector hardware and the computing infrastructure. It is responsible for collecting, inspecting, and assembling data fragments from the myriad of electronic channels following a Level-1 trigger acceptance. By aggregating these individual fragments into complete collision events, the DAQ system manages massive, high-bandwidth data flows and routes them to the High-Level Trigger processors for final online filtering. The Data Quality Monitoring (DQM) system operates online during data collection to provide rapid feedback to experiment operators, enabling prompt identification of detector issues and ensuring data integrity.

Computing Model and Data Processing Workflow

Once an event passes all trigger selections, CMS relies on a comprehensive, multi-stage computing workflow to transform raw detector signals into refined physics objects. The initial stage, known as prompt reconstruction, employs sophisticated algorithms—including track fitting, vertexing, and particle flow reconstruction—to build a complete picture of the collision. This computationally intensive task is performed rapidly after data taking, applying the latest detector calibrations and alignment conditions to ensure high-quality, physics-ready output.

To accommodate the diverse analytical needs of the collaboration while managing storage limits, CMS employs a tiered hierarchy of data formats. These formats progressively reduce event sizes by stripping away lower-level detector details in favor of high-level physics kinematics:

- **RAW & RECO:** The foundational formats containing direct digitized signals and the full output of the reconstruction algorithms, respectively.
- **Analysis Object Data (AOD) & MiniAOD:** Reduced formats containing the essential physics objects required for most analyses, offering a significant reduction in file size.
- **NanoAOD:** An ultra-compact format that reduces the per-event data footprint to a fraction of a percent of the original RECO size, serving as the primary format for modern high-level statistical analysis.

The sheer volume of CMS data necessitates a distributed global computing infrastructure, historically organized into a hierarchical Tier structure (Tier-0, Tier-1, and Tier-2). The central facility handles initial prompt reconstruction and data archiving, while global centers provide the processing power and distributed storage required for large-scale simulated data production, reprocessing campaigns, and end-user physics analysis. Throughout this pipeline, a comprehensive Data Quality Monitoring (DQM) system provides continuous feedback—operating both online during active collisions and offline during reconstruction—to guarantee data integrity and detector health.

The scale and efficiency of CMS data operations are immense, processing tens of petabytes of data annually. Demonstrating the robust capabilities of this entire workflow, during the 2025 proton run of the LHC, the CMS experiment successfully recorded a record-breaking integrated luminosity of approximately 125 fb^{-1} . This remarkable milestone underscores the seamless integration of hardware triggers, high-throughput software processing, and distributed grid computing required to drive discoveries in modern high-energy physics.

2.3.3 Object reconstruction

Reconstructed physics objects in CMS and used in this analysis are built with the particle-flow (PF) algorithm, which combines information from the tracking detectors and the calorimeters into a common set of particle candidates [64]. The PF reconstruction operates on PF blocks, i.e. groups of low-level detector elements that are likely to originate from the same particle, and iteratively identifies muons, electrons, photons, charged and neutral hadrons, and finally the missing transverse energy, removing used elements at each step. After PF reconstruction, objects are further classified into working points (WPs) of differing efficiency and purity. The high-level trigger follows the same principles, with simplified and more regional implementations.

Muons

Muon PF candidates are required to be isolated, using the scalar sum of transverse momenta of nearby PF candidates within a cone, and to pass identification working points based on track quality and compatibility with the primary vertex. Loose muons include candidates from both prompt production and heavy-flavour decays, while tight muons must be reconstructed in both the inner tracker and muon system with stringent requirements on hit multiplicity, impact parameters, and fit quality. Several reconstruction fits (tracker-only, combined, and specialised high- p_T fits) are considered, and the TuneP algorithm selects the optimal momentum estimate. The CMS muon reconstruction is highly performant; the tight selection criteria typically yield an identification efficiency between 95% and 99% across the detector acceptance, while maintaining a hadron-to-muon misidentification probability well below 0.1% [65,66].

Electrons and photons

Electrons and photons are reconstructed using a combined treatment, since bremsstrahlung and conversions couple their signatures. Electron candidates are built by associating Gaussian-sum filter tracks to ECAL clusters, with the momentum estimated from both the track curvature and the ECAL energy, while photons are reconstructed from ECAL clusters with no matching tracks. Isolation is defined from the transverse energy in a cone around the candidate. Prompt electrons and photons are selected using either cut-based criteria or multivariate discriminants that exploit shower-shape variables, track-cluster matching, and isolation. Depending on the chosen multivariate working point, typical electron and photon identification efficiencies range from 80% to 90%. At these efficiencies, the probability for a hadronic jet to be misidentified as an electron or isolated photon is strictly controlled, typically falling between 1% and 5% [67].

Jets

After leptons and isolated photons are removed from consideration, the remaining PF candidates are clustered into jets. Charged hadrons, neutral hadrons, and non-isolated photons are combined using the anti- k_T algorithm [68] with distance parameters $R = 0.4$ (AK4) and $R = 0.8$ (AK8) to reconstruct jets from single-parton and boosted topologies, respectively. The PF information from the tracker, ECAL, and HCAL is combined to avoid double counting and to improve the jet energy resolution. This global event description achieves a typical jet energy resolution of 15% to 20% at 30 GeV, about 8% at 100 GeV, and

4% at 1 TeV [69]. Following rigorous calibration, the residual uncertainty on the jet energy scale is tightly constrained, remaining below 1% for central jets with $p_T > 30$ GeV. To further suppress background, a multivariate pileup jet identification algorithm is applied to low- p_T AK4 jets, which retains 90% to 95% of genuine hard-scatter jets while rejecting over 90% of pileup-induced jets. For boosted jets, substructure and grooming techniques such as soft-drop [70] are used to mitigate soft and pileup contamination and to resolve two-prong decays of heavy resonances inside AK8 jets.

Hadronic taus

Hadronically decaying τ leptons (τ_h) are reconstructed starting from jets clustered with the anti- k_T algorithm. The hadron-plus-strips (HPS) algorithm [71] identifies τ_h candidates by combining charged hadrons and photons into decay-mode hypotheses, constrained by invariant mass, charge, and collimation requirements. A convolutional neural network-based multiclassifier, known as DeepTau [72], then discriminates genuine τ_h from quark- or gluon-initiated jets and misidentified electrons or muons, using both high-level variables and low-level calorimeter and tracking information in (η, ϕ) grids. The latest iteration of this algorithm incorporates domain adaptation by backpropagation to significantly reduce performance discrepancies between collision data and simulation. This refined algorithm provides exceptional separation power; at a genuine τ_h identification efficiency of 60% (the Tight working point), it restricts the jet-to- τ_h misidentification probability to approximately 0.25%, representing a 50% reduction in background compared to previous iterations. Furthermore, dedicated discriminators against light leptons heavily suppress electroweak backgrounds. The algorithm limits the electron misidentification probability to roughly 0.15% at an 80% τ_h signal efficiency (a factor of two improvement), while the muon misidentification probability remains tightly controlled at approximately 0.01% for a 99.5% τ_h signal efficiency [73].

Missing transverse energy

The missing transverse momentum $\vec{p}_T^{\text{miss}}$ is defined as the negative vector sum of the transverse momenta of all PF candidates, and its magnitude is referred to as E_T^{miss} . It provides sensitivity to undetected particles such as neutrinos, but is also affected by detector acceptance, misreconstruction, and pileup. After incorporating jet energy corrections (Type-1 corrections), the E_T^{miss} response is restored to near unity. The corresponding E_T^{miss} resolution depends heavily on the overall event topology and pileup conditions, but typically ranges from 15 to 30 GeV for standard collision environments [74].

To mitigate the severe effects of high pileup in Run 3, the pileup per particle identification (PUPPI) algorithm is now recommended as the default over the standard PF-based method, providing reduced pileup dependence and improved resolution. While advanced machine-learning algorithms such as DeepMET [75] have recently been developed—offering an additional 10% to 30% improvement in E_T^{miss} resolution by learning the optimal weight for each PF candidate—they are not employed in this work. Instead, this analysis relies entirely on the robust, default PUPPI E_T^{miss} reconstruction.

b-tagging

Jets originating from b quarks are identified using dedicated b -tagging algorithms that exploit the presence of long-lived B hadrons and their displaced decay vertices [76]. Tracks associated with jets are required to satisfy quality and impact-parameter criteria, and secondary vertices are reconstructed when possible. A variety of multivariate discriminants, such as DeepJet [77] and ParticleNetAK4 [78], combine information on track impact parameters, secondary-vertex properties, and possible leptons inside jets to distinguish b jets from light-flavor and charm jets.

Starting from Run 3, ParticleNet is the recommended algorithm and used in this analysis. The ParticleNetAK4 architecture treats jets as an unordered "particle cloud" to efficiently incorporate low-level jet information. For a standard medium working point, it provides state-of-the-art separation power, achieving a b -jet identification efficiency of approximately 75% to 80%, while limiting the misidentification probability to roughly 1% for light-flavor jets and 10% for charm jets. Furthermore, recent CMS commissioning studies demonstrate that the ParticleNetAK4 discriminants exhibit excellent data-to-simulation agreement across various Run 3 phase spaces, ensuring robust performance in the active collision environment [79].

Pileup mitigation

Multiple proton–proton interactions per bunch crossing introduce additional soft particles that deteriorate the reconstruction of jets, E_T^{miss} , and lepton isolation. For AK4 jets, charged-hadron subtraction is used to remove charged PF candidates associated with pileup vertices before clustering, while AK8 jets are reconstructed using the PUPPI algorithm [80], which assigns per-particle weights based on local event shape and vertex compatibility, down-weighting particles likely originating from pileup. Similar corrections are applied to lepton isolation and E_T^{miss} to reduce their dependence on the number of pileup interactions.

2.4 Personal contribution

This work benefited from extensive operational responsibilities within the CMS collaboration and contributions to the CMS Tracker performance. Detector operations included regular shifts as Detector Control System/Detector Safety System shifter ensuring safe and stable running conditions of CMS subdetectors, real-time response to alarms, and coordination with sub-system experts during LHC fills. Additional operational service comprised trigger shifter duties within the CMS online trigger team, monitoring High-Level Trigger menus and rates, validating changes during commissioning, and interfacing with DAQ to maintain data-taking efficiency in evolving beam conditions. Complementing these, service as Tracker detector-on-call provided rapid expert support for silicon tracker issues during data taking, including diagnosis of module trips, DAQ problems related to tracker and recovery procedures, and feedback to operations for sustained data quality. Periods as shift leader involved end-to-end coordination of CMS data-taking, leading the control room crew, communicating with LHC, and executing run plans while enforcing safety and data quality protocols.

In parallel to operations, a substantive contribution was made to the CMS Tracker project within the Detector Performance Group through a hit resolution study using proton–proton datasets from Run 3 [61, 62]. The core of this contribution was the implementation, validation, and interpretation of the strip overlap pair-method workflow summarized in Figure 2.7 and equation (2.1): defining the track and hit-pair selection, evaluating the measured and predicted inter-module residuals, extracting single-sensor resolution from their quadratic combination, and producing consistent layer-by-layer comparisons across detector regions and data-taking periods. The analysis measured hit resolution as a function of incident angle, cluster size, local track parameters, and instantaneous luminosity, with dedicated checks of alignment and calibration updates and their impact on apparent resolution shifts. The resulting plots and summaries were used to document strip-tracker performance and to support detector-performance monitoring in Run 3.

Chapter

3

Machine Learning Implementation

Machine learning (ML) has become an indispensable tool in modern particle physics, enabling physicists to extract rare signals from overwhelming backgrounds in complex experimental data. This chapter introduces the fundamental concepts, methodologies, and applications of machine learning relevant to this thesis. The general machine learning concepts presented in the following sections are based primarily on the textbooks by Bishop [81] and Goodfellow, Bengio, and Courville [82].

This chapter first reviews supervised learning paradigms and training practices. It then discusses the three ML models compared for optimizing the signal extraction in the analysis of $HH \rightarrow bbWW$ channel in the di-leptonic (DL) final state introduced in Chapter 4, namely a Transformer, a Deep Neural Network (DNN), and a Boosted Decision Tree (BDT). Finally, it provides a comparison of these models.

3.1

Machine learning in high-energy physics

Particle physics experiments generate petabytes of data annually, with each collision event comprising thousands of reconstructed objects whose properties must be analyzed to identify rare signal processes. Traditional selection strategies, relying on hand-crafted cuts on individual observables, often lack sensitivity to subtle kinematic correlations that distinguish signal from background. Machine learning techniques, by contrast, learn complex relationships between observables, automatically capturing non-linear relationships and high-dimensional feature interactions that human analysts might overlook.

The application of machine learning in high-energy physics (HEP) is particularly valuable for several reasons. First, ML algorithms can exploit correlations among large numbers of input variables without requiring explicit theoretical

guidance. Second, they provide a probabilistic interpretation of events—for example, assigning a “signal probability” to each event—that naturally integrates with the likelihood-based statistical inference frameworks used in HEP. Third, ML methods offer interpretability through feature importance rankings and partial-dependence plots, allowing physicists to validate that learned patterns reflect genuine physics rather than statistical artifacts. Finally, advances in ML architectures have enabled the incorporation of complex detector information (e.g., calorimeter and tracker deposits arranged in spatial grids) directly into models, improving sensitivity [83,84].

3.1.1 Supervised learning

Supervised learning refers to the class of machine learning methods where a model is trained on examples for which the desired output is known. Concretely, the training dataset consists of pairs $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i$ is a vector (or a collection) of input features describing an event, and y_i is a target label such as “signal” or “background”. In the context of this analysis, these perfectly labeled datasets are generated using MC simulations. Because the true physical origin of an event in collision data is fundamentally unknown, MC simulations provide the indispensable ground truth—simulating both the targeted signal and the various SM backgrounds—to construct the exact $(\mathbf{x}_i, y_i)$ pairs required for training. The goal is to learn a function $f(\mathbf{x})$ that generalizes to unseen events (such as the actual collision data) by predicting the label (classification) or a continuous quantity (regression).

In **classification**, y is discrete (binary or multi-class) and the model outputs either a class label or a score that can be interpreted as a class probability. In **regression**, y is continuous and the model predicts a real-valued quantity (e.g. a particle’s energy). In HEP analyses, classification is most commonly used for signal-background discrimination, while regression is frequently used for calibration tasks or improved kinematic reconstruction.

3.1.2 Common supervised ML models in HEP

HEP analyses typically choose ML models by trading off discrimination power, robustness to systematic variations, interpretability for validation, and training/inference cost. Mostly used architectures are:

- **Boosted decision trees:** A decision tree is a non-parametric model that partitions the feature space through a sequence of simple binary splits, leading to a final prediction at the “leaves” (Figure 3.1a). Because individual trees are prone to overfitting and can be weak learners (performs only slightly better than random guessing), boosting is employed: an

ensemble technique that combines multiple trees sequentially, where each new tree is trained to correct the residual errors of the previous ensemble. In HEP, BDTs are a primary tool for event-level classification with tabular, high-level features, offering strong performance, fast training, and easy diagnostics e.g. with intrinsically-offered feature importance [85]. See Section 3.4 for more information.

- **Deep neural networks:** Artificial neural networks consist of interconnected nodes (neurons) organized in layers, where each connection has a learnable weight and neurons apply non-linear activation functions to their inputs. Multi-layer Perceptrons (MLPs) are neural networks that contain at least one hidden layer between the input and output. When scaled with multiple hidden layers, they become a form of DNNs, exploiting this depth to learn highly complex, hierarchical representations. (Figure 3.1b). They are often used when non-linear feature combinations are complex or when learning a shared representation across channels/categories is beneficial [86,87]. See Section 3.2.1 for detailed information on DNNs.
- **Convolutional neural networks (CNNs):** CNNs process data by using small, sliding windows (filters) to scan across the input, automatically learning to detect specific shapes and features. Crucially, they can recognize a pattern no matter where it appears (translational invariance property). This makes them ideal for analyzing grid-like data (Figure 3.1c). In particle physics, CNNs are commonly used to process calorimeter energy deposits by treating the detector cells like pixels in a 2D image, allowing the network to easily identify the distinct “shapes” of different particle showers [88,89].
- **Graph Neural Networks (GNNs):** GNNs are designed to process data structured as a network (or “graph”), where individual items are represented as nodes connected by relational links. Instead of analyzing items in isolation, GNNs learn by passing information back and forth between connected nodes to understand how they interact. This architecture is uniquely suited for input data that varies in size and has no fixed order. In particle physics, this paradigm is extremely powerful for analyzing collections of particles, such as those inside a jet. By treating each reconstructed particle as a node in a graph, the network can dynamically map out the complex spatial and kinematic relationships between them, making GNNs the dominant standard for modern jet tagging [78,90].
- **Recurrent Neural Networks (RNNs):** RNNs process sequential data one element at a time, maintaining an internal memory or “hidden state” that updates at each step based on the current input and the previous state. Because they process variable-length sequences via these hidden states, they were historically dominant for jet constituent analysis and event-level

sequences. However, they are increasingly replaced by transformers due to superior parallel scaling and long-range dependency modeling [91,92].

- **Normalizing Flows / Invertible Neural Networks:** Normalizing flows construct complex probability distributions by applying a sequence of invertible, differentiable transformations (bijections) to a simple base distribution (like a Gaussian), allowing for exact likelihood evaluation. These generative models are often adapted for supervised regression and unfolding. Unlike standard networks that output a single value, conditional flows learn the full conditional probability density $p(y|x)$, making them ideal for tasks requiring precise uncertainty quantification or inverse problem solving (e.g. detector unfolding) [93,94].
- **Transformers (attention models):** The core mechanism of a Transformer is self-attention, which allows the model to dynamically weigh the importance of different parts of the input relative to one another, regardless of their distance in a sequence or physical space. Because they do not impose a fixed neighborhood structure, they are increasingly used to model global correlations among many objects/particles. In collider physics, modern transformer-based jet-tagging and b-tagging models such as Unified Particle Transformer (UParT) and related Particle Transformer variants have become important benchmarks [95]. Transformers are also the primary choice for the foundation models in particle physics [96–98]. See Section 3.3 for more details on transformer architecture.

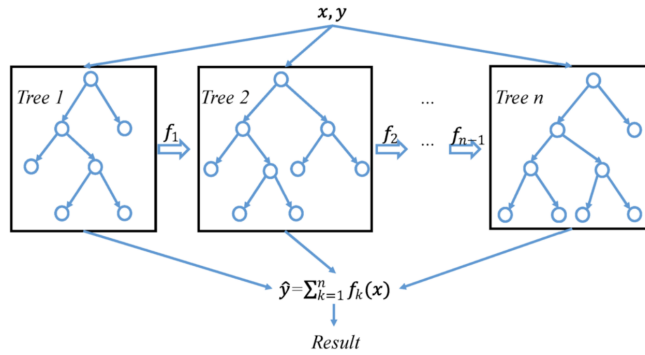
3.1.3

General training procedure and evaluation metrics

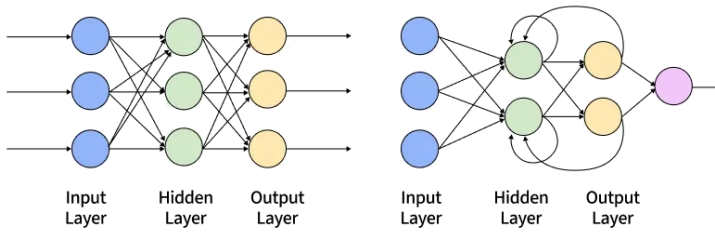
A typical training process starts by defining a **loss function** $\mathcal{L}(f(\mathbf{x}), y)$ that penalizes incorrect predictions and then processed by optimizing model parameters to minimize the average loss over the training set. For binary classification, a standard choice is the binary cross-entropy loss, which encourages the model score to approach 1 for signal and 0 for background.

Optimization is typically performed iteratively using gradient-based methods. The process relies on backpropagation to compute the gradients of the loss function with respect to the network parameters, which indicate the direction in the parameter space that minimizes the loss. While standard Stochastic Gradient Descent (SGD) updates weights using a fixed learning rate, modern deep learning tasks typically employ adaptive methods that adjust the learning rate for each parameter based on historical gradients such as Adagrad, RMSprop, and Adam [102].

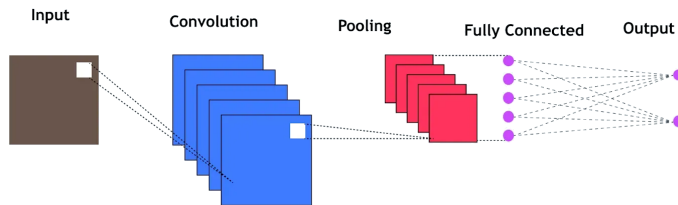
While the loss function guides the optimization, the network’s ability to model



(A) General architecture of a BDT model. The input features x are evaluated sequentially by an ensemble of decision trees, where each tree contributes a prediction $f_k(x)$ and the final output is obtained by combining the predictions of all trees. BDT architecture. Figure taken from ref. [99]



(B) Comparison between a DNN and an RNN. In a DNN, information flows forward through successive fully connected layers, whereas in an RNN, recurrent connections allow information from previous steps to influence the current output. Figure taken from ref. [100]



(C) General architecture of a CNN. The input is processed through convolutional layers that extract local features, followed by pooling layers that reduce the spatial dimensionality, and finally by fully connected layers that map the extracted features to the output prediction.. Figure taken from ref. [101]

FIGURE 3.1 | Schematic representations of four of the supervised machine learning architectures. These diagrams illustrate some of the architectures discussed in Section 3.1.2.

complex data depends fundamentally on the **activation function** applied to the output of each artificial neuron. These functions determine whether a neuron "fires" and passes information to the next layer, transforming the weighted sum of inputs into a meaningful output signal. Mathematically, they serve the critical role of "breaking linearity." A neural network composed exclusively of linear layers (weighted sums) would effectively collapse into a single linear regression model, regardless of how many layers are stacked, limiting it to learning only simple, flat decision boundaries. By introducing non-linear activation functions, the network gains the ability to approximate arbitrarily complex functions, allowing it to model the intricate, non-linear kinematic correlations of particle physics collisions. Common choices for activation functions in high-energy physics include:

- **Sigmoid:** Defined as $\sigma(x) = 1/(1 + e^{-x})$ (Figure 3.2 left), this function maps inputs to the range $(0, 1)$. It is standard for the final output layer in binary classification tasks, where the output is interpreted as a probability.
- **ReLU (Rectified Linear Unit):** Defined as $f(x) = \max(0, x)$ (Figure 3.2 right), ReLU is the default choice for hidden layers in modern architectures. It mitigates the vanishing gradient problem associated with saturating functions, enabling the efficient training of deep networks.
- **Softmax:** A generalization of the sigmoid function for multi-class classification, defined for a vector of raw scores $\mathbf{z} = (z_1, \dots, z_K)$ as:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

This transformation ensures that the outputs lie in the range $(0, 1)$ and sum to unity, representing a categorical probability distribution.

A fundamental challenge in machine learning is achieving the optimal balance

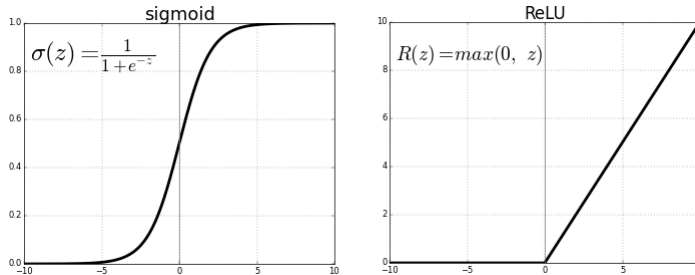


FIGURE 3.2 | Sigmoid (left) and ReLU (right) activation functions. The sigmoid function maps real-valued inputs smoothly to the interval $[0,1]$, while ReLU outputs zero for negative inputs and increases linearly for positive inputs.

between underfitting and overfitting. As illustrated in Fig. 3.3, **underfitting** occurs when a model is too simple to capture the underlying physics encoded in the data, resulting in poor performance on both the training and validation sets. Conversely, **overfitting** arises when a model is excessively complex, leading it to memorize statistical fluctuations and noise in the training data rather than learning generalizable physical patterns. This typically manifests as high performance on the training set but poor performance on a held-out test set.

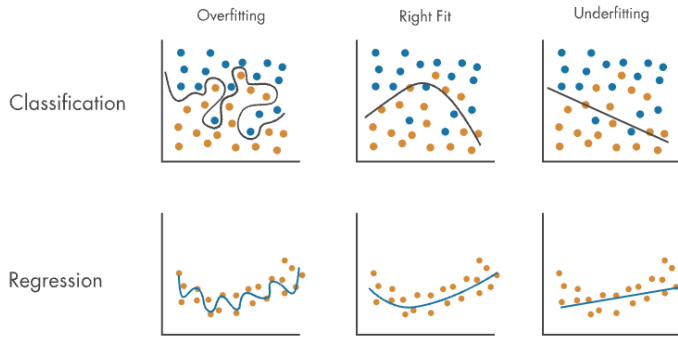


FIGURE 3.3 | Schematic comparison of overfitting, appropriate fitting, and underfitting in classification and regression problems. Complex models may fit noise in the data, while overly simple models fail to capture the true underlying structure [103].

To mitigate the risk of overfitting, several **regularization** techniques are employed. Regularization terms, such as L1 or L2 norms, can be added to the loss function to penalize excessively large weights and discourage complexity. **Dropout** is another effective method that randomly deactivates a fraction of neurons during training, forcing the network to learn robust, redundant representations. Additionally, **early stopping** is used to halt training once the performance on a separate validation set ceases to improve, preventing the model from learning the noise of the training data.

The effectiveness of these techniques in ensuring generalization is measured by splitting the dataset into statistically independent subsets—typically training, validation, and testing samples. A more robust approach, often used in physics analyses to avoid bias and maximize the available statistics, is **stratified k -fold cross-validation (CV)**. In this approach, the data is partitioned into k folds (partitions) while preserving the signal-to-background proportions in each split. Each event is then evaluated by a model version that was trained without using that specific event, ensuring an unbiased estimate of the model's performance.

Beyond these architectural and procedural choices, the performance and generalization capability of a model are heavily influenced by its **hyperparameters**.

Key configuration settings include:

- **Learning Rate:** Controls the step size during the gradient descent update. A value that is too large causes the loss to oscillate or diverge, while a value that is too small results in slow convergence and a risk of getting stuck in local minima. The values are typically in the range $\mathcal{O}(10^{-5})$ to $\mathcal{O}(10^{-2})$, depending on the optimizer and the batch size.
- **Batch Size:** The number of training samples processed before the internal model parameters are updated. Larger batches provide more accurate gradient estimates but require significant memory; smaller batches introduce noise that can act as a regularizer. The batch size is commonly between 256 and a few thousand events, limited by available memory and I/O throughput.
- **Epochs:** One epoch is defined as a complete pass through the entire training dataset. The total training duration is a function of the number of epochs and the dataset size.
- **Architecture Complexity:** Refers to the structural complexity of the model, such as the number of hidden layers and neurons in a DNN, the number of attention heads in a Transformer, or the maximum depth of trees in a BDT. This must be tuned to ensure the model is expressive enough to learn the physics without becoming prone to overfitting.
- **Regularization Strengths:** Hyperparameters such as the *dropout rate* or the coefficients of *L1* and *L2* regularization, which control the strength of the penalty applied to large model weights and are tuned to reduce overfitting and the performance gap between training and validation samples.

The **performance** of a classifier is assessed using several metrics:

- **Accuracy:** The fraction of correctly classified events. However, in presence of severe class imbalance (many background events, few signal events), a naive classifier might achieve high accuracy simply by predicting the majority class.
- **Precision and Recall:** Precision is the fraction of predicted signal events that are truly signal; recall is the fraction of true signal events correctly identified. A high-precision classifier rarely mislabels background as signal, while a high-recall classifier misses few signal events.
- **ROC Curve and AUC:** The Receiver Operating Characteristic (ROC) curve plots the true positive rate (recall) against the false positive rate as the classification threshold varies. The Area Under the ROC Curve (AUC) summarizes the classifier's ability to rank signal events higher than background, independent of threshold choices.

- **Confusion Matrix:** A matrix comparing true and predicted labels, showing the counts of true positives, false positives, false negatives, and true negatives. Normalizing the confusion matrix by row (true label) or column (predicted label) reveals per-class performance.

For physics analyses, the confusion matrix normalized by true label is particularly informative: it reveals the purity (fraction of events in each predicted class that are truly that class) and the efficiency (fraction of true signal/background events assigned to each predicted class). An ideal classifier would yield a purely diagonal confusion matrix. The purity of the signal categories is particularly important, as these classes provide the primary discrimination power between signal and background in the subsequent likelihood fit, while the purity of the background classes helps constrain the corresponding normalizations and shape uncertainties.

3.1.4 Practical considerations in HEP analyses

Class imbalance and weighting

In supervised HEP analyses, models are typically trained on simulated samples where each event is labeled as signal or background, and each event carries a physics-motivated weight (generator weight $\times$ cross section $\times$ integrated luminosity $\times$ corrections) so that the training emphasizes events according to their expected contribution in data [104]. While these weights ensure the training sample is physically accurate, they often lead to a significant numerical asymmetry between categories, requiring specific strategies to manage the resulting sample composition.

In particle physics, signal events are typically far rarer than background events. Naive training on imbalanced data causes the model to ignore the minority (signal) class. Common solutions include:

- **Class reweighting:** Applying additional training weights to signal and background events so that the loss function is not dominated by the more abundant class, for example by scaling the signal contribution to be comparable to that of the background.
- **Oversampling:** A resampling technique for imbalanced classification where the minority class is increased—either by duplicating existing minority examples or by generating synthetic ones—so the model does not favor by default predicting the majority class.
- **Undersampling:** Complementary resampling approach where the majority class is reduced (typically by randomly removing majority-class events from separate strata) to avoid a classifier being dominated by

background-like examples.

Negative event weights

In Monte Carlo event generators, NLO corrections can produce events with negative weights, representing destructive interference in the underlying amplitudes. These negative weights have no meaningful interpretation for machine learning loss functions and can destabilize training. The standard approach is to exclude events with negative weights from the training set, accepting a potential decrease in the performance of the model.

Feature engineering and input selection

The choice of input features significantly impacts model performance. In physics analyses, features are typically grouped into:

- **Low-level variables:** Four-momenta components, object identification scores, isolation measures.
- **High-level variables:** Hand-crafted observables such as missing transverse energy, invariant masses, angular separations, scalar sums.

Despite the ability of deep networks to learn feature representations automatically, including well-motivated physics-inspired high-level variables often improves both performance and interpretability. Moreover, highly correlated features can be retained because modern algorithms exploit non-linear relationships that linear correlation alone does not capture.

Interpretability and validation

Despite their power, machine learning models are often perceived as “black boxes” obscuring the logic of their decisions. Several techniques improve interpretability [104]:

- **Feature importance:** Estimating how strongly each input variable influences the model’s predictions, for example through model-specific feature importance scores or model-agnostic methods such as permutation importance.
- **Partial dependence plots:** Showing how the model’s output varies with one variable while marginalizing over others.
- **SHAP values (SHapley Additive exPlanations):** A game-theoretic approach to quantify each feature’s contribution to individual predictions.

- **Validation on orthogonal samples:** Training on simulation and validating on control regions in data to ensure learned patterns reflect genuine physics.

3.2 Neural networks

Artificial neural networks are a broad class of models constructed from interconnected computational units (neurons) arranged in layers. Each neuron computes a weighted sum of its inputs, adds a bias term, and applies a non-linear activation function. Formally, the output a of a single neuron can be expressed as:

$$a = \sigma \left(\sum_{i=1}^n w_i x_i + b \right)$$

where w represents the weight vector, x the input vector, b the bias, and $\sigma(\cdot)$ the non-linear activation function. Feed-forward neural networks with non-linear activations are universal function approximators, i.e. they can approximate arbitrary continuous functions on compact domains given sufficient capacity [105].

The **training** process involves minimizing a loss function, $L(\mathbf{x})$, which quantifies the mismatch between predictions and targets. Common loss functions include Mean Squared Error (MSE) for regression tasks and cross-entropy for classification. Training is performed via backpropagation, an algorithm that uses the chain rule of calculus to compute the gradient of the loss with respect to every weight and bias in the network. These gradients are then used to update the parameters using an optimization algorithm, typically Gradient Descent or adaptive variants such as Adam or RMSprop.

3.2.1 Deep neural networks

Deep neural networks (DNNs) are neural networks with multiple hidden layers between the input and output. Increasing the depth enables the model to learn hierarchical representations, where early layers can capture relatively simple patterns and later layers can combine these into more abstract and task-relevant features. In practice, this additional representational hierarchy often allows a deep model to achieve a similar discrimination power with fewer units than a purely shallow architecture, albeit at a higher training and tuning cost.

For a feed-forward DNN, the computation can be written compactly as a sequence of layer transformations. Starting from an input feature vector $\mathbf{h}^{(0)} \equiv \mathbf{x}$, each hidden layer ℓ produces

$$\mathbf{h}^{(\ell)} = \sigma\left(\mathbf{W}^{(\ell)}\mathbf{h}^{(\ell-1)} + \mathbf{b}^{(\ell)}\right),$$

where $\mathbf{W}^{(\ell)}$ and $\mathbf{b}^{(\ell)}$ are trainable parameters and σ is a non-linear activation function (commonly ReLU in hidden layers). For binary classification, the final layer is typically a single node with a sigmoid activation, yielding a score that can be interpreted as an estimate of the signal probability.

Depth, however, also introduces practical challenges. As networks become deeper, optimization can become unstable due to issues such as vanishing/-exploding gradients and sensitivity to the scale of activations. Modern DNN training therefore commonly relies on a small set of architectural components and regularization techniques that improve convergence and generalization:

- **Normalization layers:** Batch normalization or layer normalization stabilizes the distribution of activations across layers, enabling faster and more reliable training.
- **Regularization:** Weight decay (L2 regularization), dropout, and early stopping are widely used to reduce overfitting, especially when training statistics are limited or when inputs exhibit strong correlations.
- **Initialization and optimization choices:** Suitable weight initialization and adaptive optimizers (e.g., Adam) can significantly improve training stability and reduce sensitivity to hyperparameter choices.

3.2.2 DNNs in HEP

The rapid adoption of DNNs in HEP is driven by several distinct **advantages** over traditional multivariate techniques and shallow machine learning models:

- **Universal function approximation:** Unlike linear discriminants, DNNs can theoretically approximate any continuous function to arbitrary precision, allowing them to model highly complex, non-linear decision boundaries in the multidimensional phase space of particle collisions.
- **Automated feature engineering:** Deep Learning models excel at learning optimal representations directly from low-level inputs, such as four-momenta or raw calorimeter energy deposits [106]. This reduces the reliance on hand-crafted features that may overlook subtle kinematic information.
- **Scalability with data:** DNNs generally benefit more from large datasets than traditional algorithms. As the volume of simulated training data increases, deep networks continue to improve in performance without saturating as quickly as shallow models, making them well-suited for the massive datasets produced by the LHC [88].

- **Flexibility of architecture:** The deep learning framework allows for specialized architectures tailored to the specific geometry of HEP data. For instance, CNNs leverage the grid-like structure of calorimeters, while GNNs and Transformers naturally handle the variable-length, unordered sets of particles typical of jet substructure and event reconstruction [106].

While DNNs offer superior expressivity, their successful deployment in HEP analyses requires careful attention to several **practical implementation** details:

- **Input preprocessing:** DNNs are highly sensitive to the scale of input data. Large numerical values (e.g., particle p_T in MeV) can push activation functions into saturation regions (vanishing gradients), halting learning. It is therefore mandatory to standardize inputs—typically to zero mean and unit variance—to ensure efficient convergence [106].
- **Negative event weights:** As discussed in Section 3.1.4, standard loss functions like cross-entropy are not convergent when negative weights are included, leading to training instabilities. Common mitigation techniques include training on absolute weights, ignoring negative weights, or employing specialized loss functions.
- **Computational resources:** The training DNNs involves dense matrix multiplications that are computationally heavy on CPUs. Efficient training requires GPU acceleration, which introduces additional complexity regarding memory management (batch size tuning) and software dependencies (e.g., CUDA, TensorFlow/PyTorch versions).
- **Hyperparameter optimization:** The hyperparameter space for DNNs is significantly large. Finding the optimal configuration often requires automated scanning strategies, such as Bayesian optimization or grid search, as manual tuning is inefficient and prone to bias.

The application of DNNs in HEP is rapidly evolving beyond standard signal-background classification. **Current research** focuses on integrating DNNs into the core computational pipelines of experiments and expanding their capability to discover unknown physics:

- **Generative models for fast simulation:** Detailed detector simulation (e.g., using GEANT4) is the primary computational bottleneck in modern experiments. Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are being deployed to generate synthetic calorimeter showers orders of magnitude faster than traditional methods, while maintaining high physical fidelity [107, 108].
- **Real-time inference on FPGAs:** To deploy deep learning at the trigger level—where decisions must be made in microseconds—models are being compressed and synthesized for Field-Programmable Gate Arrays

(FPGAs). This allows physicists to convert trained DNNs into firmware, enabling complex event tagging at the full collision rate [109].

- **Anomaly detection (unsupervised learning):** Rather than training on specific signal models, autoencoders are used to learn the distribution of Standard Model background events. When applied to data, events with high reconstruction error are flagged as potential anomalies, allowing for model-independent searches for new physics [110].

3.2.3 DNN Implementation for $HH \rightarrow bbWW$

The DNN model for $HH \rightarrow bbWW$ (DL) analysis is trained (the code implementation is archived at ref. [111]) as a supervised binary classifier to separate non-resonant HH signal from the dominant backgrounds after the baseline event selection and categorization described in Chapter 4.

The input feature set include a mixture of low-level kinematic information (four-momenta components of leptons and jets, E_T^{miss} -related variables) and high-level observables constructed from these objects (masses and angular separations), including flags for resolved and boosted topologies and ggF- and VBF-like events. The full list of input features is given in Table 3.1.

A fully connected multilayer perceptron (MLP) architecture with 3 hidden layers comprising 32, 16, and 8 neurons respectively is chosen. Each hidden layer is followed by batch normalization to stabilize learning and a ReLU activation function. The output layer consists of a single node with a sigmoid activation, producing a score in $[0, 1]$. To ensure the sigmoid score can be interpreted strictly as a probability, the model output is calibrated using Platt scaling [112] (logistic regression) on the validation set.

Events are weighted during training using the same nominal analysis weights applied in histogram production (cross section, integrated luminosity, and all relevant correction factors), ensuring that the loss emphasizes phase-space regions according to their expected impact in data. Negative-weight events are excluded from training and an additional class reweighting (calculated to balance the sum of weights for signal and background) is applied, which stabilizes optimization and improves sensitivity in the most discriminating regions. The model is trained with a binary cross-entropy loss, which measures the error between predicted probabilities and true binary labels, and the Adam optimizer, an adaptive gradient-based optimization method that updates parameters efficiently during training. A batch size of 2048 is used, and early stopping with a patience of 5 epochs monitors the validation loss and stops training when no improvement is observed for five consecutive epochs, helping limit overfitting and preserve generalization performance.

Hyperparameters were optimized via a manual heuristic search centered around standard best practices, as a full grid search was computationally prohibitive. The best hyperparameters were selected based on the highest validation AUC, and the final model was retrained using these settings.

The stability of the DNN is assessed using a stratified k -fold splits with $k = 5$, and performance is summarized with weighted ROC curves and AUC values on held-out folds.

The results are shown in Section 3.5 in comparison with the other two models.

3.3

Transformers and attention mechanisms

Transformers represent a paradigm shift in machine learning, introducing the self-attention mechanism as an alternative to recurrent architectures. Originally developed for natural language processing [113], transformers are increasingly adopted in particle physics for tasks ranging from event classification to anomaly detection. This section provides a detailed explanation of transformers, their advantages over recurrent models, and emerging applications to high-energy physics.

The core innovation of transformers is the **self-attention mechanism**, which computes a weighted sum of all input elements, with weights determined dynamically based on element similarities. Unlike recurrent networks, which process inputs sequentially and thus struggle with long-range dependencies due to vanishing gradients, attention enables direct communication between distant elements in a single operation.

The fundamental building block is **scaled dot-product attention**, defined as

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V},$$

where $\mathbf{Q}$ (query), $\mathbf{K}$ (key), and $\mathbf{V}$ (value) are learned linear projections of the input. More concretely, the **query** represents the specific features an element (e.g., a reconstructed particle) is “looking for” to better contextualize itself within the event. The **key** acts as an identifier broadcasted by every element in the sequence, advertising the nature of the information it holds. The **value** contains the actual content or physical kinematics of that element. Intuitively, each query vector attends to all available items by computing a similarity score against their keys ($\mathbf{Q}\mathbf{K}^T$). These scores are normalized into attention weights via a softmax function, which then dictate how much of each corresponding value ($\mathbf{V}$) should be aggregated to form the final, context-aware representation

of the element. The scaling factor $\sqrt{d_k}$ (where d_k is the dimension of the key) prevents the dot products from becoming too large, which would saturate the softmax and suppress gradients.

Rather than relying on a single global attention function, transformers utilize **multi-head attention** to capture multiple types of correlations simultaneously. This mechanism executes h independent attention operations—or heads—in parallel:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^O,$$

where each head projects the input into a distinct representation subspace of dimension $d_k = d/h$ using its own set of learned weight matrices. This decomposition allows different heads to specialize in extracting different physical features from the same event; for instance, one head might learn to attend to angular relationships while another focuses on kinematics. The outputs of all heads are concatenated and mixed through a final learned linear projection $\mathbf{W}^O$, synthesizing these independent insights into a single, rich, context-aware latent representation.

Transformer encoder

A standard transformer encoder layer combines multi-head self-attention with a feed-forward network (see Figure 3.4):

$$\begin{aligned} \mathbf{h}' &= \text{MultiHead}(\mathbf{h}, \mathbf{h}, \mathbf{h}), \\ \mathbf{h}'' &= \text{FFN}(\mathbf{h}') = \text{ReLU}(\mathbf{h}' \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2, \end{aligned}$$

where residual connections and layer normalization are applied around each sub-layer:

$$\begin{aligned} \mathbf{h}'_{\text{residual}} &= \text{LayerNorm}(\mathbf{h} + \text{MultiHead}(\mathbf{h}, \mathbf{h}, \mathbf{h})), \\ \mathbf{h}_{\text{out}} &= \text{LayerNorm}(\mathbf{h}'_{\text{residual}} + \text{FFN}(\mathbf{h}'_{\text{residual}})). \end{aligned}$$

Residual connections enable stable training of very deep networks by avoiding the vanishing gradient problem and allows the model to simply add the new relational information to the existing features which is like an "update" to the particle's description rather than a total replacement. Layer normalization normalizes activations across the feature dimension, improving training stability and generalization. Finally, stacking multiple encoder layers allows the model to build hierarchical representations, with each layer refining features learned by previous layers.

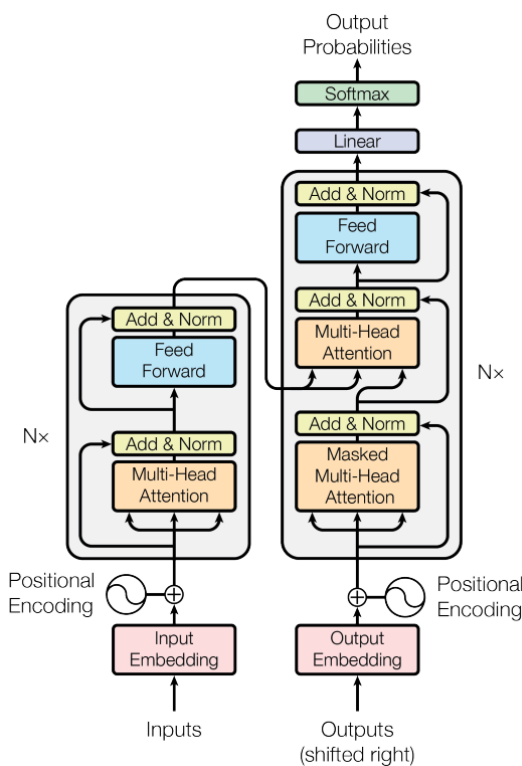


FIGURE 3.4 | Encoder–decoder transformer: the encoder (left) produces contextual representations used for downstream tasks such as classification, while the decoder (right) is primarily used for auto-regressive generation [113].

Transformer decoder

The original transformer architecture is composed of an encoder and a decoder: the encoder produces contextual embeddings for the input sequence, while the decoder generates an output sequence auto-regressively, using masked self-attention over previously generated tokens and cross-attention to the encoder outputs. In particle-physics applications, encoder-only variants are typically sufficient for event or jet classification, since the task is to map a set/sequence of objects to a single score or class label. By contrast, the decoder component is mainly used for generative tasks, where one aims to sequentially generate structured outputs (e.g. particle-level representations). We will not elaborate on the details of the transformer decoder given it's not used in this thesis.

3.3.1 Transformers in HEP

Transformers offer several **advantages** over DNN architectures in the context of particle physics:

- **Global correlations:** Unlike sequential or local architectures, the attention mechanism operates globally over all input elements. This allows the model to capture long-range dependencies and complex symmetries between particles regardless of their position in the input sequence.
- **Parallelizable:** Transformers process all elements in parallel, enabling efficient implementation on GPUs.
- **Variable-length inputs:** Transformers naturally handle variable numbers of particles per event via masking, without requiring padding to fixed sizes.
- **Interpretability:** Attention weights reveal which particles the model focuses on for its predictions, providing physics-interpretable results.
- **Scalability:** Transformer models have demonstrated strong scaling properties—larger models trained on more data consistently improve performance.

Practical considerations for transformer deployment in particle physics

The quadratic attention complexity $\mathcal{O}(n^2)$ in sequence length can become prohibitive for events with hundreds of reconstructed particles. Strategies to reduce **computational cost** include:

- **Sparse attention:** Restricting the attention mechanism to a local neighborhood or a fixed subset of nodes to reduce the number of dot-product

operations.

- **Hierarchical attention:** Implementing a multi-stage approach, such as first processing constituents into jets and subsequently applying attention at the jet or event level.
- **Linear attention:** Utilizing approximations of the softmax function to achieve $\mathcal{O}(n)$ complexity [114].
- **Input Pruning:** Pre-selecting physics objects based on kinematic thresholds (e.g., p_T cuts) or isolation criteria to reduce the initial sequence length.

For typical HEP events with $\mathcal{O}(10)$ to $\mathcal{O}(50)$ reconstructed objects, the quadratic cost is manageable.

Unlike language models, which process discrete tokens (words), particle physics events contain continuous variables (four-momenta, energies, angular positions). The choice of how to encode particles as **embeddings** significantly impacts model performance:

- **Dense embeddings:** Concatenating particle four-momenta and identification scores and projecting through a learned linear layer.
- **Lorentz-invariant embeddings:** Constructing Lorentz-invariant combinations (masses, angles, rapidity differences) and including them in the embedding.
- **Learnable embeddings:** Using neural networks to project particle properties into a learned representation space, allowing the model to discover optimal embeddings.

Standard transformers use **positional encodings** to distinguish between different positions in a sequence. In particle physics, there is no natural “position”—particles are not ordered. To mitigate this, either positional encoding is omitted to rely on the particle properties themselves to distinguish particles, or a learned or fixed positional encoding can be introduced by associating each particle with an embedding corresponding to its rank e.g. by p_T providing information about the event structure.

Current research in HEP using Transformers

Several **recent applications** demonstrate transformers’ potential in HEP:

- While not strictly transformers, ParticleNet and similar architectures [78, 115, 116] use graph neural networks and set-based operations (invariant to permutations of jet constituents) that share the spirit of attention. These models process jet constituents as a set, learning permutation-invariant

representations that naturally encode the irregular structure of hadronic decays.

- Transformers have been explored for event-level classification tasks, processing the full set of reconstructed objects (leptons, jets, missing energy) as a sequence [117, 118]. A simple application encodes each particle as a vector of its properties (four-momenta, identification scores, etc.), feeds the sequence through transformer layers, and applies a pooling operation (e.g., averaging all output embeddings or using a dedicated classification token) to produce a single event-level prediction.
- Transformers also excel in unsupervised settings. VAEs and diffusion models (i.e. models that transform random noise back into a structured representation) based on transformers can learn the distribution of background events in a control region, then flag deviations in data as potential anomalies [119, 120]. This complements targeted signal searches by identifying unexpected structures in data without presupposing the signal's properties.

As transformer models scale to larger event datasets and computational infrastructure matures, several **trends** emerge:

- **Foundation models for HEP:** Pre-training transformers on large unlabeled datasets of collision events (via self-supervised learning such as masked particle prediction) could enable efficient fine-tuning for downstream physics tasks [98].
- **End-to-end learning from raw detector data:** Transformers that operate directly on detector readouts, bypassing hand-crafted reconstruction algorithms [121, 122].
- **Generative models for simulation:** Transformers can learn to generate synthetic collision events, potentially reducing the computational cost of Monte Carlo simulation [123, 124].
- **Online triggering:** Deploying lightweight transformer variants on trigger systems with FPGA to make real-time physics decisions at the full event rate [125, 126].

3.3.2 Transformer-based model for $HH \rightarrow bbWW$

An architecture based on transformer encoder is implemented for the $HH \rightarrow bbWW$ (DL) analysis (The code implementation is archived at ref. [127]). The model processes the event as a sequence of reconstructed objects, learning to distinguish the signal from background processes.

The model receives a sequence of 8 tokens representing the event: 2 leptons, 4 AK4 jets, 1 AK8 jet, and the E_T^{miss} . Each object type is described by a

distinct feature vector containing four-momentum components (p_x, p_y, p_z, E) and object-specific properties. For leptons, this includes the charge and the PDG ID—a unique integer assigned by the Particle Data Group to identify particle species (e.g., ± 11 for electrons/positrons and ± 13 for muons). For jets, the features include b -tagging scores and other substructure variables. Full list of input features is same as the ones for the DNN implementation given in Table 3.1. These raw feature vectors are projected into a shared 32-dimensional latent space using separate MLPs for leptons, jets, and E_T^{miss} . Each embedding network consists of three fully connected layers with ReLU activations and layer normalization.

The core architecture consists of 6 stacked transformer encoder layers. Each layer employs multi-head self-attention with 4 heads and a feed-forward network dimension of 128. After the final encoder layer, the learned object embeddings are aggregated via mean pooling across the sequence dimension to produce a single, global event-level representation vector. The global representation is passed through a final linear layer and Platt scaling to produce probabilities.

The results are shown in Section 3.5.

3.4 Boosted decision trees

Boosted decision trees are ensemble methods that combine many weak learners (shallow decision trees) into a single strong learner. A **decision tree** recursively partitions the feature space using axis-aligned splits on individual variables. At each node, a split is chosen to minimize impurity (e.g., entropy) in the resulting child nodes. Trees are easy to interpret but prone to overfitting.

Boosting is an additive ensemble technique (combining multiple individual models) that improves predictive performance by sequentially training weak learners—typically shallow decision trees. Unlike bagging, which builds trees in parallel, boosting is iterative: each new tree is specifically designed to minimize the errors, or residuals, left by its predecessors.

In the gradient boosting approach, this "correction" is treated as an optimization problem. At each iteration, the algorithm calculates the negative gradient of the loss function with respect to the current model's predictions. This gradient effectively represents the direction and magnitude of the steepest descent required to minimize error. By fitting a new tree to these gradients rather than the original target, the model performs gradient descent in function space. The final prediction is the accumulated result of these updates, often scaled by a learning rate to ensure generalization and prevent overfitting.

The key hyperparameters controlling model complexity include:

- **Number of boosting rounds:** The total number of trees added to the ensemble. Too few rounds lead to underfitting; too many risk overfitting.
- **Learning rate:** A shrinkage factor applied to each tree's contribution. Lower learning rates require more boosting rounds but often yield better generalization.
- **Tree depth:** Controlling the maximum depth of individual trees. Shallow trees are weak learners; excessively deep trees capture noise.
- **Subsampling:** Training each tree on a random subsample of the data and features reduces overfitting and improves stability.

3.4.1 BDTs in HEP

While Deep Learning has gained prominence, BDTs remain a primary choice in particle physics due to their robustness, interpretability, computational efficiency and performance on structured kinematic data. The primary distinctions lie in the boosting strategy and optimization objective:

- **Gradient Boosted Decision Trees (GBDT):** A generalization of boosting where, instead of reweighting events, new trees are trained to predict the residual errors (gradients) of the loss function from the previous step [128].
- **XGBoost (Extreme Gradient Boosting):** The modern standard in collider physics analyses. It is an optimized implementation of GBDT that introduces regularization terms (L1 and L2) into the objective function to prevent overfitting—a critical feature when training on limited Monte Carlo statistics. It also offers significant improvements in computational speed and scalability [129].
- **LightGBM (Light Gradient Boosting Machine):** A high-performance framework that differs from XGBoost primarily by employing a *leaf-wise* (best-first) tree growth strategy rather than a level-wise approach. This allows the algorithm to grow deeper, asymmetrical trees by expanding the specific leaf that results in the largest reduction in the loss function. This strategy often leads to faster convergence and higher accuracy on large-scale datasets, though it requires careful tuning of the maximum depth to prevent overfitting [130].

Despite the growing dominance of Deep Learning, BDTs remain the workhorse of many high-energy physics analyses due to several distinct **advantages**:

- **Performance on tabular data:** For analyses relying on high-level, structured kinematic variables (e.g., invariant masses, angular separations), BDTs often match or outperform DNNs. They partition the phase

space using sequential, multi-dimensional orthogonal cuts (forming hyperrectangles). This aligns perfectly with the physics-driven logic of traditional, cut-based signal region definitions, making them highly interpretable [106,131].

- **Interpretability:** Unlike "black box" neural networks, BDTs offer intrinsic transparency. Metrics such as feature importance allow physicists to rank input variables by their discrimination power, providing immediate feedback on which physical properties drive the classification. This is critical for validating that the model relies on robust physics rather than simulation artifacts [132].
- **Training speed and efficiency:** BDTs typically train orders of magnitude faster than deep networks and require significantly fewer computational resources. This rapid training time facilitates extensive optimization cycles, allowing analysts to test hundreds of feature combinations and hyperparameter settings in the time it takes to train a single large Transformer.
- **Robustness to small datasets:** In analyses with limited Monte Carlo statistics (e.g., rare signal grids), BDTs are generally less prone to overfitting than over-parameterized deep networks, as their model complexity (tree depth) is more easily constrained.

Practical considerations for BDT training

While robust, BDTs present specific challenges that must be addressed to ensure optimal performance [133]:

- **Feature engineering dependency:** Unlike DNNs or Transformers, which learn directly from raw data, BDTs rely on the quality of input features. They cannot automatically extract high-level concepts from low-level four-vectors; physicists must explicitly provide engineered variables to maximize sensitivity.
- **Correlated variables:** Standard decision trees tackle redundant variables by arbitrarily selecting one and ignoring the other. While this does not degrade performance, it can skew feature importance rankings, making a strong physical variable appear irrelevant if a correlated proxy was chosen by the algorithm first.
- **Discrete nature and gradient descent:** Because decision trees make discrete cuts (non-differentiable), they cannot be trained via standard backpropagation. This prevents them from being easily integrated into "end-to-end" differentiable pipelines where the detector simulation or systematic uncertainties are optimized simultaneously with the classifier.

Current trends in BDT research

Research into BDTs continues to evolve, focusing on integrating them into modern differentiable workflows and real-time systems:

- **Differentiable Decision Trees:** To bridge the gap between BDTs and Deep Learning, recent work focuses on "soft" decision trees that replace hard cuts with sigmoid functions. This makes the tree differentiable, allowing it to be inserted as a layer within larger neural networks or optimized via gradient descent for tasks like systematic uncertainty minimization [134].
- **Fast inference on FPGAs:** For trigger systems where decisions must be made in nanoseconds, BDTs are being synthesized directly onto FPGAs. Tools that convert trained BDTs into firmware, enable complex machine learning algorithms to run at the full 40 MHz collision rate of the LHC [135].
- **Gradient boosting for reweighting:** BDTs are increasingly used not just for classification, but for density estimation. By training a BDT to distinguish between data and simulation, the output score can be converted into an event weight that corrects the simulation to match the data, offering a high-dimensional alternative to traditional histogram-based reweighting [136,137].

3.4.2 BDT implementation for the $HH \rightarrow bbWW$ (DL) analysis

In the $HH \rightarrow bbWW$ (DL) analysis, a BDT is trained using the XGBoost [129] library. The code implementation is archived at [138]. To prevent overfitting while maintaining computational efficiency, the model is configured with the following hyperparameters:

- **Architecture:** A maximum of 200 boosting rounds (trees) with a maximum depth of 3.
- **Optimization:** An initial learning rate of 0.1, which decays exponentially by a factor of 0.99 in every iteration to ensure stable convergence.
- **Regularization:** Training is monitored via specific validation sets, employing early stopping with a patience of 30 rounds to determine the optimal number of trees. The model utilizes default L2 regularization.
- **Hardware:** GPU-accelerated histogram construction is used to speed up training.

The dataset was initially partitioned into a 20% independent testing set and an 80% development set. The latter was further subdivided into training (80% of

the development set, or 64% of the total) and validation (20% of the development set, or 16% of the total) sets for model optimization and hyperparameter tuning. Unlike neural networks, the BDT input features—corresponding to the kinematic variables listed in Table 3.1—are not scaled, as decision trees are invariant to monotonic feature transformations. The output is a raw logistic probability score in $[0, 1]$.

The results are shown in Section 3.5.

3.5 Comparison of models

This section details the comparative application of the three machine learning architectures—Transformer, Deep Neural Network, and Boosted Decision Tree—to the $HH \rightarrow bbWW$ (DL) analysis. To ensure a fair comparison, all three models are trained on an identical set of input features, as listed in Table 3.1. The correlation matrix of input features is given in Figure 3.5

3.5.1 Interpretability and physics features

To interpret the decision-making process of the algorithms and verify that they are utilizing physically motivated kinematic features, the feature importance for three distinct architectures are analyzed.

DNN Permutation Importance

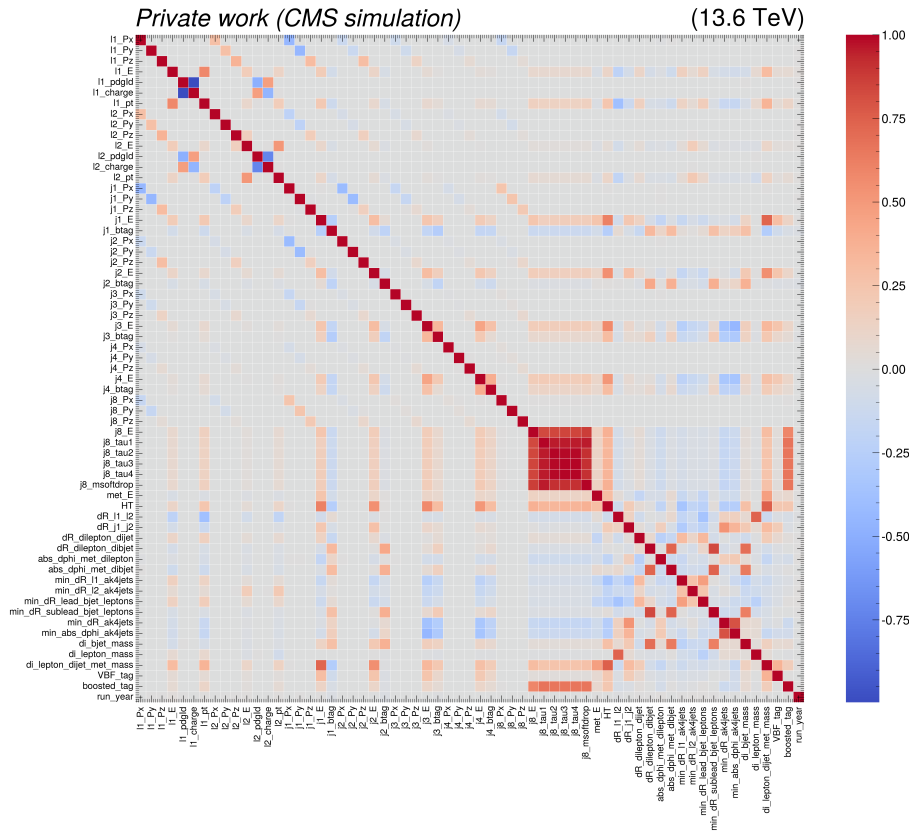
For the DNN model, feature importance is evaluated using permutation importance, as shown in Figure 3.6. This metric measures the decrease in the model's AUC when the values of a single feature are randomly shuffled, breaking the correlation with the target. The results identify the dilepton invariant mass ($m_{\ell\ell}$) as the dominant driver of the DNN's performance; permuting this variable causes a substantial drop in AUC of nearly 0.10. Following the mass term, the DNN relies on other lepton, jet and E_T^{miss} variables. This indicates the DNN has learned a decision boundary that is highly sensitive to the topology of the leptons and jets, the presence of heavy-flavor jets and the E_T^{miss} .

Transformer Attention Mechanism

To interpret the Transformer model, the mean attention weights (aggregated over all heads, layers, and batches) assigned to the input features are analyzed, as shown in Figure 3.7. This metric reflects the intrinsic "focus" of the model's self-attention mechanism during the learning process. Consistent with the DNN,

<i>l1(2)_Px</i>	x component of the (sub)leading lepton's p_T
<i>l1(2)_Py</i>	y component of the (sub)leading lepton's p_T
<i>l1(2)_Pz</i>	z component of the (sub)leading lepton's p_T
<i>l1(2)_E</i>	energy of the (sub)leading lepton
<i>l1(2)_pdgId</i>	PDG ID of the (sub)leading lepton
<i>l1(2)_charge</i>	charge of the (sub)leading lepton
<i>l1(2)_pt</i>	p_T of the (sub)leading lepton
<i>j(i)_Px</i>	x component of the i^{th} leading jet's p_T ($i = 1, 2, 3, 4$)
<i>j(i)_Py</i>	y component of the i^{th} leading jet's p_T ($i = 1, 2, 3, 4$)
<i>j(i)_Pz</i>	z component of the i^{th} leading jet's p_T ($i = 1, 2, 3, 4$)
<i>j(i)_E</i>	energy of the i^{th} leading jet ($i = 1, 2, 3, 4$)
<i>j(i)_btag</i>	b-tagging score of the i^{th} leading jet ($i = 1, 2, 3, 4$)
<i>j8_Px</i>	x component of the leading AK8 jet's p_T
<i>j8_Py</i>	y component of the leading AK8 jet's p_T
<i>j8_Pz</i>	z component of the leading AK8 jet's p_T
<i>j8_E</i>	energy of the leading AK8 jet
<i>j8_tau(N)</i>	N-subjettiness of the leading AK8 jet ($N = 1, 2, 3, 4$)
<i>j8_msoftdrop</i>	soft drop mass of the leading AK8 jet
<i>met_E</i>	magnitude of the missing transverse energy
<i>HT</i>	scalar sum of the p_T of all jets
<i>dR_l1_l2</i>	ΔR between the two leptons
<i>dR_j1_j2</i>	ΔR between the two leading jets
<i>dR_dilepton_dijet</i>	ΔR between the dilepton and dijet systems
<i>dR_dilepton_dibjet</i>	ΔR between the dilepton and dibjet systems
<i>abs_dphi_met_dilepton</i>	$ \Delta\phi $ between the missing transverse energy and the dilepton system
<i>abs_dphi_met_dibjet</i>	$ \Delta\phi $ between the missing transverse energy and the dibjet system
<i>min_dR_l1_ak4jets</i>	Minimum ΔR between the leading lepton and all AK4 jets
<i>min_dR_l2_ak4jets</i>	Minimum ΔR between the subleading lepton and all AK4 jets
<i>min_dR_lead_bjet_leptons</i>	Minimum ΔR between the leading b-tagged jet and both leptons
<i>min_dR_sublead_bjet_leptons</i>	Minimum ΔR between the subleading b-tagged jet and both leptons
<i>min_abs_dphi_ak4jets</i>	Minimum $\Delta\phi$ between all AK4 jets and the missing transverse energy
<i>di_bjet_mass</i>	Mass of the dijet system formed by the two leading b-tagged jets
<i>di_lepton_mass</i>	Mass of the dilepton system
<i>di_lepton_dijet_met_mass</i>	Mass of the dilepton + dijet + missing transverse energy system
<i>VBF_tag</i>	Binary VBF tagging variable
<i>boosted_tag</i>	Binary boosted topology tagging variable
<i>run_year</i>	Run year of the event (2022 or 2023)

TABLE 3.1 | Input features to the three ML models. A value of zero is passed whenever a variable was not available e.g. AK8 jet variables for resolved categories.



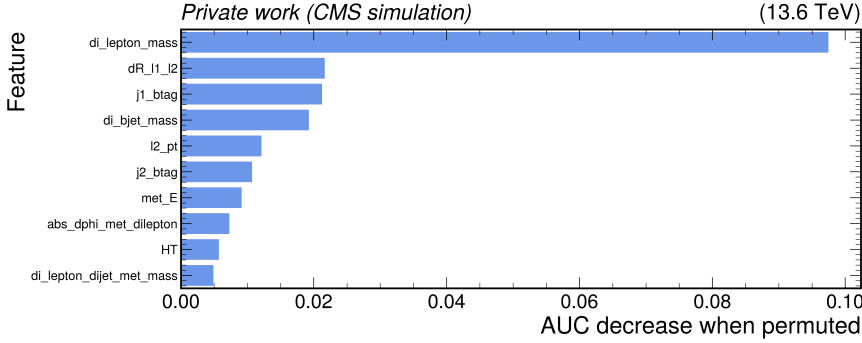


FIGURE 3.6 | The top 10 input features for the DNN classifier ranked by permutation importance. The importance is quantified as the decrease in the Area Under the ROC Curve (AUC) when the values of a specific feature are randomly shuffled while keeping others fixed. The dilepton invariant mass ($m_{\ell\ell}$), the angular separation between leptons ($\Delta R(\ell_1, \ell_2)$), and the leading jet b-tag score are identified as the most critical features for predictive performance. For the full list, see Appendix A.2.1

the Transformer assigns the highest attention score to the dilepton invariant mass ($m_{\ell\ell}$). However, the secondary features reveal a different prioritization compared to the DNN. The Transformer places significant attention on the transverse momentum imbalance, ranking the azimuthal separation between the missing transverse energy and the dilepton system ($|\Delta\phi(p_T^{miss}, \ell\ell)|$) and the magnitude of the missing energy (p_T^{miss}) as the next most important features. This suggests the Transformer architecture builds its internal representation by focusing more on the global event alignment and the recoil of the system against the leptons, rather than the specific jet flavor or lepton-lepton ΔR prioritization seen in the DNN.

BDT Feature Importance

The feature importance for the BDT classifier, derived from the weight of each feature in the decision trees, is presented in Figure 3.8. Here, the "weight" metric counts the number of times a variable is used to split the data across all trees in the ensemble. The BDT model also ranks the dilepton invariant mass ($m_{\ell\ell}$) as the most critical discriminator. However, unlike the neural networks, BDT places a much stronger emphasis on the hadronic resonance structure. The di-b-jet mass (m_{bb}) is ranked as the clear second most important feature. This implies that the BDT structure effectively isolates the signal by explicitly cutting on the mass windows of both the leptonic ($m_{\ell\ell}$) and hadronic (m_{bb}) decay legs, whereas the deep learning models rely more on complex correlations involving

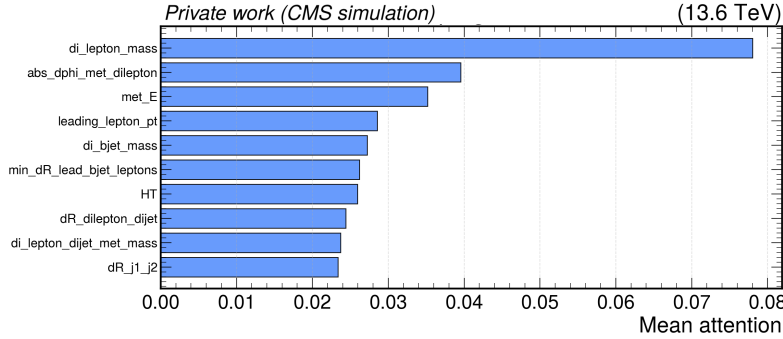


FIGURE 3.7 | Top 10 mean attention scores assigned by the Transformer model to the input features, aggregated over all attention heads, encoder layers, and the evaluation batch. The ranking reveals that the model prioritizes leptonic and missing energy variables. For the full list, see Appendix A.2.1.

the missing energy and angular separations.

3.5.2

Robustness to data calibration

To further evaluate the robustness of the three architectures against potential mismodeling or calibration inaccuracies, a sensitivity study was performed by applying a 10% shift and a 10% smearing to the input features. The impact on model performance, quantified by the change in the AUC (ΔAUC), reveals a stark contrast between the methods. While the DNN and Transformer models exhibit remarkable stability—showing negligible performance degradation ($\Delta\text{AUC} \leq 2 \times 10^{-4}$)—the XGBoost BDT demonstrates significantly higher sensitivity to these perturbations, particularly under the smearing scenario where the AUC drops by 3.4×10^{-3} (Figure 3.9). This discrepancy is likely rooted in the fundamental differences between their decision-making processes: while the BDT relies on discrete, axis-aligned splits that can be easily bypassed by small feature shifts, the neural networks utilize smooth activation functions and high-dimensional representations that inherently offer greater resilience to localized input noise. Despite the BDT's distributed feature importance profile, these results suggest that the "softer" decision boundaries of deep learning models provide superior mathematical robustness to systematic variations in input data.

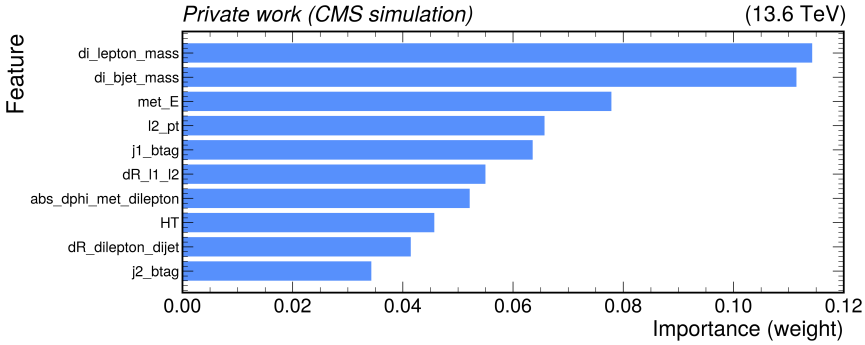


FIGURE 3.8 | The top 10 input features for the BDT classifier model. The ranking is dominated by the high-level engineered variables, specifically the invariant masses of the dilepton and dijet systems. The importance metric reflects the relative contribution of each feature to the construction of the boosted decision trees. Unlike the neural networks, the BDT model places high importance on the di-b-jet mass (m_{bb}) alongside the dilepton mass ($m_{\ell\ell}$), indicating a strong reliance on reconstructing both resonant decay systems. For the full list, see Appendix A.2.1

Summary

All three models consistently identify the dilepton invariant mass as the single most discriminative feature. However, the analysis reveals distinct strategies for background rejection among the architectures.

- The DNN relies on topological closeness (ΔR) and b-tagging.
- The Transformer focuses on global alignment with the missing transverse energy ($\Delta\phi$).
- The BDT model exploits the kinematic resonance of the di-b-jet system (m_{bb}).

A critical distinction emerges when comparing the distribution of importance scores across the three architectures. Both the DNN and the Transformer exhibit a highly concentrated dependency, where the dilepton invariant mass ($m_{\ell\ell}$) disproportionately dominates the decision-making process, effectively treating other variables as minor corrections. This suggests that the deep learning models primarily optimize their loss functions by relying on this single high-level discriminant. In contrast, the BDT model demonstrates a more distributed utilization of the feature space. Its importance profile is significantly flatter, with comparable weights assigned to the top several features (e.g., $m_{\ell\ell}$, m_{bb} , p_T^{miss} , and lepton p_T). This indicates that the gradient-boosted decision trees are successfully capturing a broader range of kinematic correlations, constructing a robust decision boundary that relies on the interplay of multiple physical

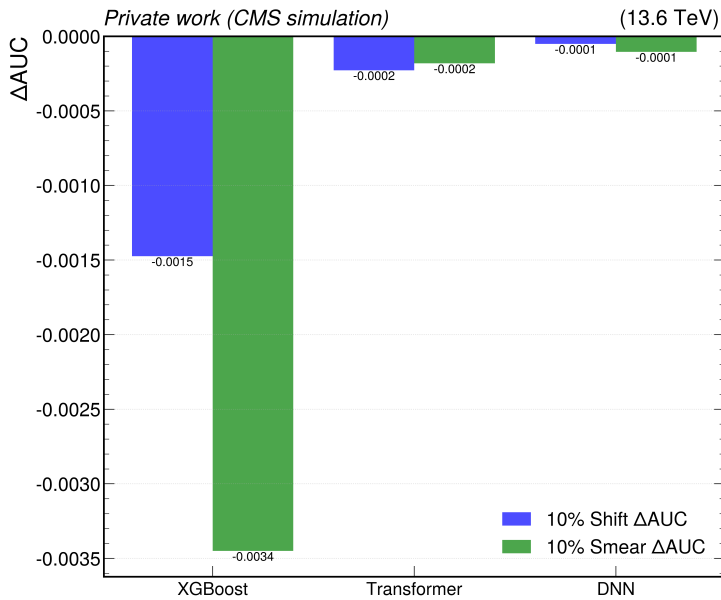


FIGURE 3.9 | Robustness comparison of the XGBoost, Transformer, and DNN models under 10% shift (blue) and 10% smearing (green) of the input features. The y-axis represents the change in AUC relative to the nominal performance evaluated on CMS simulation. The results highlight that deep learning architectures (DNN and Transformer) maintain stable discrimination power under these perturbations, whereas the BDT model exhibits a more pronounced performance loss, indicating a higher sensitivity to the exact calibration of input variables.

observables rather than hinging the classifier’s performance on a single variable. This broader dependency may offer the BDT model greater resilience against systematic uncertainties that might affect the dilepton invariant mass.

3.5.3 Evaluation

The models are evaluated using a robust validation strategy to prevent overfitting and ensure generalization:

- **Stratified k-fold cross-validation:** The dataset is partitioned into statistically independent folds, preserving the signal-to-background ratio in each split. This ensures that every event is assigned a score by a model trained on a disjoint subset of data, preventing bias in the final statistical inference. The results are listed in Table 3.2
- **Performance metrics:** The primary metric for comparison is the AUC, which quantifies the model’s ability to rank signal events higher than background events. The results are listed in Table 3.3 and shown as ROC curve plots in Appendix A.2.2.
- **Purity and efficiency:** Confusion matrices (shown in Figure 3.10) are analyzed to determine the purity of signal categories, which is critical for maximizing sensitivity in the likelihood fit.

Fold	1 st	2 nd	3 rd	4 th	5 th	Average
AUC	0.9670	0.9674	0.9676	0.9666	0.9662	0.96696

TABLE 3.2 | To validate the BDT’s performance stability, AUC values were calculated for each fold of a k -fold cross-validation study. This rigorous assessment was performed exclusively for the BDT, as it was the selected model for the final inference study; other architectures were excluded from this process due to the prohibitive computational time required. The results indicate a robust implementation.

Model	DNN	Transformer	BDT
AUC	0.965	0.969	0.966

TABLE 3.3 | Comparison of the AUC performance for the DNN, Transformer, and BDT models, evaluated on the independent test dataset. While the Transformer achieves the highest discrimination power (0.969), all three architectures demonstrate comparable performance.

The **comparison** aims to identify the optimal balance between discrimination power and computational cost. While the discrimination performance of the three models is comparable as shown in Table 3.3 and figure 3.10, the

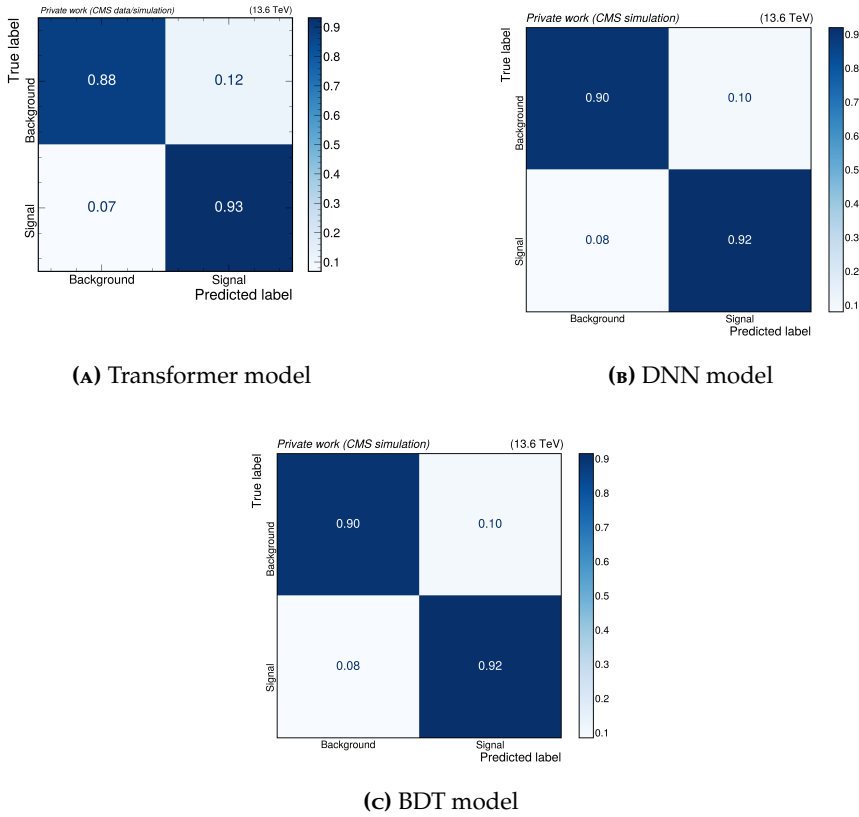


FIGURE 3.10 | Confusion matrices for the Transformer (top left), DNN (top right), and BDT (bottom) models, evaluated on the test set. The matrices are normalized by true class label (rows), such that the diagonal elements represent the recall (efficiency) for background and signal, respectively. While the Transformer yields the highest signal efficiency (0.93), the DNN and BDT provide stronger background rejection (0.90).

final selection of the BDT is driven by practical considerations regarding computational efficiency. As summarized in Table 3.4, the training costs vary by orders of magnitude. The BDT model converges in under one minute, whereas the DNN requires approximately 30 minutes, and the Transformer necessitates 1.5 hours per training cycle on the same hardware.

Model	Total Time	Epochs*	Time / Epoch	Early Stopping
Transformer	90 min	4	22.5 min	No
DNN	28 min	16	106 s	Yes
BDT	45 s	200	0.2 s	Yes

TABLE 3.4 | Comparison of computational costs for the three architectures. The BDT offers a significant speed advantage, enabling rapid iteration.

*For the BDT, “Epochs” refers to boosting rounds.

This efficiency is crucial for the analysis workflow, as it facilitates extensive hyperparameter optimization, feature engineering iterations, and the rigorous k -fold cross-validation required to ensure unbiased inference.

Beyond wall-time, the BDT offers several additional advantages in the context of this analysis. First, it is computationally efficient and straightforward to retrain across multiple folds and configurations, which is particularly useful when k -fold strategies are employed to avoid bias in downstream statistical inference workflows. Second, BDTs provide strong interpretability via established diagnostics (e.g., feature importance and related ranking methods), enabling faster validation that the classifier is learning physically meaningful correlations. Third, the reduced training cost makes it feasible to perform broader robustness studies—such as repeated trainings under alternative weighting schemes or systematic variations—without prohibitive training time, thereby improving confidence in the stability of the final template shapes used in the fit.

Search for HH to bbWW using Run-3 Data from CMS

This chapter focuses on the experimental search for the Higgs boson pair production using Run-3 data of the CMS Experiment. The following sections present the search in the decay mode involving two bottom quarks and two W bosons where both W boson decay leptonically. This analysis utilizes proton-proton collision data collected at a center-of-mass energy of $\sqrt{s} = 13.6$ TeV by the CMS experiment during first two data taking years of Run-3 (2022–2023), amounting to a total integrated luminosity of 62.4 fb^{-1} .

4.1

Analysis setup

Since the SM Higgs boson can decay into several different final states, Higgs boson pair production can be studied in a wide variety of channels. In this analysis, the $HH \rightarrow bb\ell\nu\ell\nu$ final state provides several advantages:

1. The two most frequent Higgs boson decay modes are $H \rightarrow b\bar{b}$ and $H \rightarrow W^{\pm}W^{\mp}$. Through the subsequent decay $W \rightarrow \ell\nu$, this leads to a clean signal provided by the two leptons with a branching ratio of 2.64%
2. The dominant background process in this topology, top quark pair ($t\bar{t}$) production is well modeled in simulation.
3. This channel has been shown to be complementary to other HH decay modes already explored by CMS. Given that the observation of SM HH production is expected to be extremely challenging, combining as many channels as possible will be essential.
4. In the event that a deviation from the SM prediction for HH production is observed in another channel, it is crucial to provide several independent cross-checks in different final states.

While the hadronically decaying Higgs boson can be reconstructed from the jets

originating from the hadronization of the b quarks in $H \rightarrow bb$, the kinematics of the leptonically decaying Higgs boson cannot be fully reconstructed because of the two neutrinos in the final state. As a consequence, the di-Higgs invariant mass m_{HH} cannot be directly accessed, which largely removes the distinction between resonant and non-resonant production in terms of analysis strategy. The most distinctive features of the signal, common to all signal hypotheses, are a pair of b jets with an invariant mass peaking near the Higgs boson mass, on top of a smooth background distribution, and a pair of prompt, isolated charged leptons. The dilepton (DL) system provides a clean experimental signature that can be efficiently selected by DL triggers.

The most frequent SM processes contributing to the selected event topology are $t\bar{t}$ and Drell–Yan (DY) production of $Z/\gamma^* (\rightarrow \ell^+ \ell^-)$ in association with jets [139]. The dileptonic $t\bar{t}$ production cross section is known at NNLO in QCD and is approximately 98 pb [140]. Since $t\bar{t}$ production leads to an identical visible final state, it constitutes an irreducible background. Its rate can only be reduced by exploiting differences in the kinematic properties of the reconstructed objects to enhance the signal sensitivity.

In contrast, the DY process is reducible: it can be significantly suppressed by requiring the presence of b -tagged jets. This requirement does not, however, reduce the contribution from Z/γ^* production in association with genuine b jets, for which additional strategies are needed. Further subdominant backgrounds include single top quark production in association with a W boson (tW) [141], single top production in the t - and s -channels [142], diboson production (ZZ , WW , ZW , collectively denoted as VV) [143], $t\bar{t}$ production in association with vector bosons ($t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}\gamma$, denoted as $t\bar{t}V$) [144], and various single Higgs boson production modes, primarily $t\bar{t}H$ and ZH [145, 146]. Backgrounds arising from jets misidentified as leptons, originating from W +jets or QCD multijet events, are rendered negligible by the tight requirements imposed on the identification and isolation of reconstructed electrons and muons.

4.1.1 Samples

The analysis is based on proton–proton collision data collected with the CMS detector at a center-of-mass energy of $\sqrt{s} = 13.6$ TeV. Only luminosity sections certified as suitable for physics analyses are used. The resulting integrated luminosities are 34.7 fb^{-1} in 2022 and 27.8 fb^{-1} in 2023, for a total of $62, 4 \text{ fb}^{-1}$.

Events are selected using the High Level Trigger (HLT) paths. The sensitivity of the dileptonic final states is enhanced by combining single- and double-lepton triggers. Events with two offline reconstructed electrons (muons) must satisfy at least one of the corresponding single-electron (single-muon) or double-electron (double-muon) triggers, whereas events with one electron and one

muon are required to pass at least one of the single-electron, single-muon, or electron–muon trigger paths.

To ensure a high signal acceptance while maintaining manageable data rates, the analysis relies on a combination of single and double lepton HLT paths. These triggers together require the presence of at least two isolated leptons (electrons or muons) with transverse momentum thresholds optimized for the Run 3 data-taking conditions. Events are selected if they fire at least one of the HLT paths corresponding to their final state. The primary triggers used in this analysis are:

- **Single Muon:**

- HLT_IsoMu24

Requires a single isolated muon with $p_T > 24$ GeV. This path acts as a primary collector and a safety net for dilepton events where the subleading lepton may be too soft to satisfy double-lepton thresholds.

- **Single Electron:**

- HLT_Ele30_WPTight_Gsf
- HLT_Ele50_CaloIdVT_GsfTrkIdT_PFJet165

Provides coverage for events with at least one high- p_T electron. The WPTight requirement ensures a robust, tight selection at 30 GeV, while the 50 GeV path utilizes very tight calorimetric (CaloIdVT) and tight tracking (GsfTrkIdT) identification to maintain trigger stability at higher energy scales.

- **Double Muon:**

- HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8

Requires two muons with $p_T > 17$ and 8 GeV, respectively. It applies track isolation (TrkIsoVVL), a vertex compatibility cut (d_z), and a low invariant mass cut ($m_{\mu\mu} > 3.8$ GeV) to suppress backgrounds from low-energy resonances and pileup.

- **Double Electron:**

- HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
- HLT_DoubleEle33_CaloIdL_MW

The primary path requires two electrons with $p_T > 23$ and 12 GeV, satisfying loose calorimetric (CaloIdL) and tracking (TrackIdL) identification requirements.

- **Muon-Electron:**

- HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
- HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ

– HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL

These cross-triggers allow for asymmetric thresholds, maximizing acceptance for events where one lepton is significantly softer than the other.

Data are organized into several primary datasets (PDs): Muon, MuonEG, and EGamma. Each event is recorded in one or more PDs if the corresponding trigger path fires. To avoid double counting, a prioritization scheme is applied so that each event is selected once among the PDs. The priority order is: Muon→EGamma→MuonEG.

Simulated samples are used to model the signal and the background processes. Background events are generated at leading order or next-to-leading order accuracy using matrix-element generators such as MG5_aMC@NLO and POWHEG. For the DY process, inclusive samples covering the full phase space are combined with additional samples binned in generator-level jet multiplicity and DL mass using a stitching procedure [147] to improve statistical precision in relevant regions of phase space.

Parton showering, hadronization, τ decays, pileup interactions and polarization effects are simulated with PYTHIA. The underlying-event tune CP5 is used for all the samples [148]. The parton distribution functions are primarily NNPDF3.1 [149].

Signal events are generated following the same strategy. Non-resonant ggF production is simulated at NLO, whereas non-resonant vector-boson fusion is generated at LO. Each decay channel is represented by a dedicated signal sample with its corresponding branching fraction except for the $HH \rightarrow bbWW \rightarrow bb\ell^+\nu\ell^-\nu$ and $HH \rightarrow bbZZ \rightarrow bb\ell^+\nu\ell^-\nu$ modes, which are merged because their final states are experimentally indistinguishable.

The analysis employs four GGF samples at NLO with different κ_λ values and nine VBF samples at LO with different values of the κ_λ , c_V and c_{2V} couplings. Their cross sections are summarized in Tables 4.1 and 4.2. Under the SM hypothesis, this limited grid of samples enables a continuous scan of the coupling modifier parameter space.

TABLE 4.1 | The production cross-section of the double Higgs GGF signal samples at NLO precision.

Sample	κ_λ	Cross-section [fb]
SM	1	34.13
ggF 2	0	65.30
ggF 3	2.45	12.85
ggF 4	5	83.49

TABLE 4.2 | List of VBF HH signal samples with their corresponding coupling modifiers used in the analysis. The parameters κ_V , κ_{2V} , and κ_λ correspond to the C_V , C_{2V} , and C_3 parameters in the sample generation, respectively. The values of the HEFT Wilson coefficients in the SM and in eight BSM benchmark models are defined in Ref. [150]

Sample	κ_V	κ_{2V}	κ_λ	Cross-section [fb]
SM	1.00	1.00	1.00	1.9
VBF 2	1.74	1.37	14.40	394.6
VBF 3	-0.96	0.96	-1.43	1.11
VBF 4	-0.76	1.44	-19.30	354.8
VBF 5	-1.60	2.72	-1.36	11.58
VBF 6	-0.01	0.03	10.20	0.01
VBF 7	-1.21	1.94	-0.94	3.76
VBF 8	-1.83	3.57	-3.39	16.63
VBF 9	-2.12	3.87	-5.96	672.1

4.1.2 Object definitions

Leptons Electrons and muons used in the analysis are defined in three WPs, referred to as “loose”, “fakeable”, and “tight”. Loose leptons are employed to veto low-mass DL pairs, fakeable leptons are used for object cleaning, and tight leptons correspond to the final lepton selection in the signal regions.

The selections summarized in Tables 4.3 and 4.4 include requirements on transverse momentum (p_T) and pseudorapidity η , lepton isolation, the transverse ($|d_{xy}|$) and longitudinal ($|d_z|$) impact parameters of the lepton track with respect to the primary vertex, the three-dimensional impact parameter significance d/σ_d , the b -tagging discriminator of the closest jet, and the lepton MVA score [67]. For electrons passing the fakeable selection, additional criteria are imposed on the electromagnetic shower shape ($\sigma_{\eta\eta}$), the hadronic-to-electromagnetic energy ratio (H/E), and the difference between the inverse of the cluster energy and the inverse of the track momentum ($1/E - 1/p$), in order to resemble the identification applied at trigger level.

Hadronically decaying τ leptons (τ_h) are required to have $p_T > 20$ GeV, $|\eta| < 2.3$, and to pass the medium working point of the DeepTau algorithm [72]. They are cleaned against fakeable electrons and muons by vetoing τ_h within $\Delta R < 0.3$ of any such lepton to ensure orthogonality with other HH analyses targeting final states with hadronic τ decays.

Jets AK4 (AK8) jets [68] are required to satisfy tight identification criteria, to have $|\eta| \leq 2.4$, and $p_T \geq 25(200)$ GeV. AK8 jets must additionally contain at

TABLE 4.3 | Loose, fakeable, and tight selection criteria defined for electrons.¹ Barrel / endcaps² Defined as $p_T^{jet}/p_T^e - 1$ if the electron is matched to a jet within $\Delta R < 0.4$ or as the PF-relative isolation computed within a cone of fixed size $\Delta R = 0$ otherwise[†] Fails (passes) the requirement prompt-electron MVA ≥ 0.30

Observable [unit]	Loose	Fakeable	Tight
p_T [GeV]	> 5	> 10	> 10
$ \eta $	< 2.5	< 2.5	< 2.5
$ d_{xy} $ [cm]	< 0.05	< 0.05	< 0.05
$ d_z $ [cm]	< 0.1	< 0.1	< 0.1
d/σ_d	< 8	< 8	< 8
I_e	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
$\sigma_{i\eta i\eta}$	–	$< 0.011/0.030^1$	$< 0.011/0.030^1$
H/E	–	< 0.1	< 0.1
1/E - 1/p	–	> -0.04	> -0.04
Conversion rejection	no	yes	yes
Missing hits	≤ 1	$= 0$	$= 0$
EGamma POG MVA	$> \text{WP-80}$	$> \text{WP-90 no-iso}$	$> \text{WP-90 iso}$
PNet of the nearby jet	–	$< \text{WP-tight} (< \text{WP-medium})^\dagger$	$< \text{WP-medium}$
Jet relative isolation ²	–	$< 0.7(-)^\dagger$	–
Electron MVA	–	–	> 0.3

TABLE 4.4 | Loose, fakeable, and tight selection criteria defined for muons.¹ Barrel (endcaps)² Defined as $p_T^{jet}/p_T^\mu - 1$ if the electron is matched to a jet within $\Delta R < 0.4$ or as the PF-relative isolation computed within a cone of fixed size $\Delta R = 0$ otherwise[†] Fails (passes) the requirement prompt- μ MVA ≥ 0.5

Observable [unit]	Loose	Fakeable	Tight
p_T [GeV]	> 5	> 10	> 10
$ \eta $	< 2.4	< 2.4	< 2.4
$ d_{xy} $ [cm]	< 0.05	< 0.05	< 0.05
$ d_z $ [cm]	< 0.1	< 0.1	< 0.1
d/σ_d	< 8	< 8	< 8
I_μ	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
PF muon	$> \text{WP-loose}$	$> \text{WP-loose}$	$> \text{WP-medium}$
PNet of the nearby jet	–	$< \text{WP-interp.} (< \text{WP-medium})^1$	$< \text{WP-medium}$
Jet relative isolation ²	–	$< 0.8(-)^\dagger$	–

least two subjets with $p_T \geq 20$ GeV and $|\eta| \leq 2.4$, a soft-drop mass¹ in the range $30 \leq m_{\text{SD}} \leq 210$ GeV, and an N-subjettiness² ratio $\tau_2/\tau_1 \leq 0.75$, ensuring a two-prong topology compatible with a boosted $H \rightarrow b\bar{b}$ decay.

A jet cleaning procedure is applied to avoid double counting of energy deposits. AK4 (AK8) jets overlapping with fakeable electrons or muons within $\Delta R < 0.4$ ($\Delta R < 0.8$) are removed from the jet collection. To enhance the sensitivity to VBF HH production, an additional category of jets, referred to as VBF jets, is defined. VBF jets share the same reconstruction as AK4 jets, but with a relaxed pseudorapidity requirement $|\eta| \leq 4.7$ and a higher p_T threshold $p_T \geq 30$ GeV, increased further to $p_T \geq 60$ GeV in the region $2.7 < |\eta| < 3.0$ to suppress noise-induced jets in the ECAL endcaps. VBF jets are cleaned with respect to the $H \rightarrow b\bar{b}$ candidates by discarding any jet within $\Delta R < 0.8$ of the two AK4 jets, or within $\Delta R < 1.2$ of a merged AK8 jet.

Identification of b-jet candidates among AK4 jets and of subjets of AK8 jets are performed using the ParticleNet (PNet) algorithm [78]. Medium btagging working point is used for AK4 jets whereas an AK8 jet is considered btagged if it has PNet vs QCD (particleNet_XbbVsQCD, PNet score to discriminate AK8 jets from QCD background processes) score higher than 0.92 which corresponds to an efficiency of approximately 80% and a 1% misidentification probability for light-flavor quark and gluon induced jets, and at least one of its subjets' p_T is greater than 30 GeV.

Missing transverse momentum The missing transverse momentum ($\vec{p}_T^{\text{miss}}$) is determined using the “Type-1” correction algorithm [151]. This method propagates the jet energy corrections applied to reconstructed jets into the raw missing energy calculation. By adjusting the raw $\vec{p}_T^{\text{miss}}$ vector by the difference between the corrected and uncorrected jet momenta, the variable ensures that the momentum imbalance is consistent with the calibrated hadronic activity in the event.

4.1.3 Event selection

The event selection consists of a set of simple requirements applied to reconstructed events in data and simulation, designed to retain as many signal events as possible while suppressing the various background contributions. The selection is applied in successive stages, so that the agreement between data and simulation can be validated at multiple levels.

¹Softdrop mass is a groomed jet mass that uses a recursive algorithm to strip away soft, wide-angle radiation and pileup from large-radius jets, allowing for the precise identification of heavy boosted particles like Higgs or W bosons.

²N-subjettiness (τ_N) quantifies how well the radiation inside a jet aligns with the hypothesis that the jet is composed of N subjets.

Data events are first collected using a set of trigger path listed in Section 4.1.1. As no emulation of the trigger is available in simulation, the offline selection is defined to closely follow the trigger requirements. The selection begins by requiring two oppositely charged leptons (muons or electrons) that satisfy the identification and isolation criteria introduced in Section 4.1.2. Selected events are then categorized according to the lepton flavors into three channels: $\mu^+\mu^-$ and e^+e^- (collectively referred to as “same-flavor”), and $\mu^\pm e^\mp$ (“mixed-flavor”). If more than two leptons are present, the pair with the largest transverse momenta is chosen. Veto on additional leptons is applied. The leading and subleading leptons are required to have p_T above 30 GeV and 15 GeV in ee , 24 GeV and 15 GeV in $\mu\mu$ and 25 GeV and 15 GeV in $e\mu$. These thresholds are aligned with the trigger thresholds used in each category. A DL invariant mass requirement $m_{\ell\ell} > 12$ GeV is applied in all categories to suppress backgrounds from quarkonium resonances and jets misidentified as leptons.

Jets are required to be separated from the selected leptons by $\Delta R > 0.4$. At least two such jets must be present in the event. If more than two jets are available, the pair with the highest combined scores of the PNet b -tagging discriminator is chosen to define the hadronically decaying Higgs boson candidate.

Finally, a requirement, $m_{\ell\ell} < 110(100)\text{GeV}$ in the same-flavor (mixed) channels, efficiently suppresses $t\bar{t}$ events, at the cost of discarding only a tiny fraction of signal events. Another requirement, $m_{\ell\ell} < 70$ GeV in the same-flavor channels is imposed, which removes the bulk of the DY background for which $m_{\ell\ell}$ peaks around the Z -boson mass $m_Z \simeq 91$ GeV. The predicted and observed event yields in this region for the three flavor channels are listed in Figure A.24, showing good agreement between data and the sum of expected backgrounds and signal contributions.

4.1.4 Signal extraction

Signal extraction is based on a BDT performing binary classification of events that pass the full selection, as explained in Section 3.4.2. The discriminant distributions per category enter the statistical inference procedure.

The BDT model is provided with a rich set of input features built from the 2022 and 2023 Run 3 datasets with a total integrated luminosity of $62,4\text{ fb}^{-1}$, combining all final states. Illustrative distributions of some of these input variables are shown in Figure 4.1. The input to the BDT model consists of a mixture of low-level variables, such as the four-momenta of leptons and b -tagged jets, and high-level observables built from these objects. The output layer uses a sigmoid activation function, yielding a single node with values between 0 and 1 that sum to unity for each event and passed through a calibration to be able to

interpret as signal probability. To verify the stability of the training and estimate the model's generalization performance, a stratified k -fold cross-validation is conducted using five partitions.

4.1.5 Control Regions

To robustly constrain the normalization of the dominant background processes, $t\bar{t}$ and DY, the analysis defines dedicated control regions (CRs) enriched in these specific backgrounds. These regions are orthogonal to the SR and are included simultaneously in the final maximum likelihood fit. By leveraging the high statistics available in these regions, the background normalizations are constrained directly by the data, reducing the reliance on theoretical cross-section predictions and their associated large uncertainties. The definition of the control regions is driven primarily by the dilepton invariant mass, $m_{\ell\ell}$, as illustrated in the categorization scheme (Figure 4.2). The event selection is divided into three distinct regions:

- **Signal Region (SR):** Defined by $m_{\ell\ell} < 70$ GeV. This low-mass region retains the majority of the HH signal while suppressing the Z-peak contribution.
- **Drell-Yan Control Region (DY CR):** Defined by $70 \leq m_{\ell\ell} \leq 110$ GeV. This region captures the Z boson mass peak ($m_Z \approx 91$ GeV), providing a pure sample of $Z/\gamma^* \rightarrow \ell\ell$ events to constrain the DY normalization and heavy-flavor modeling.
- **$t\bar{t}$ Control Region ($t\bar{t}$ CR):** Defined by $m_{\ell\ell} > 110$ GeV. This high-mass tail is dominated by $t\bar{t}$ production, allowing for a precise measurement of the top background rate in the phase space relevant to the analysis.

In the statistical inference, the normalization of these backgrounds is not fixed to the SM prediction but is instead treated as a set of free-floating parameters. To account for potential kinematic differences and modeling discrepancies between different topologies, these normalization parameters are split for the $t\bar{t}$ normalization as decoupled between the boosted and resolved kinematic regimes to address potential modeling discrepancies in the high- p_T phase space. Additionally, the DY normalization is decoupled between the ggF and VBF production to compensate for the modeling inaccuracy in the high p_T , forward region.

The inclusion of dedicated control regions in the simultaneous fit is crucial for constraining both background normalizations and systematic shape uncertainties. When a systematic effect is treated as fully correlated between a high-statistics CR and the SR, the fit leverages the large CR event yield to precisely measure the associated nuisance parameter. This process, known

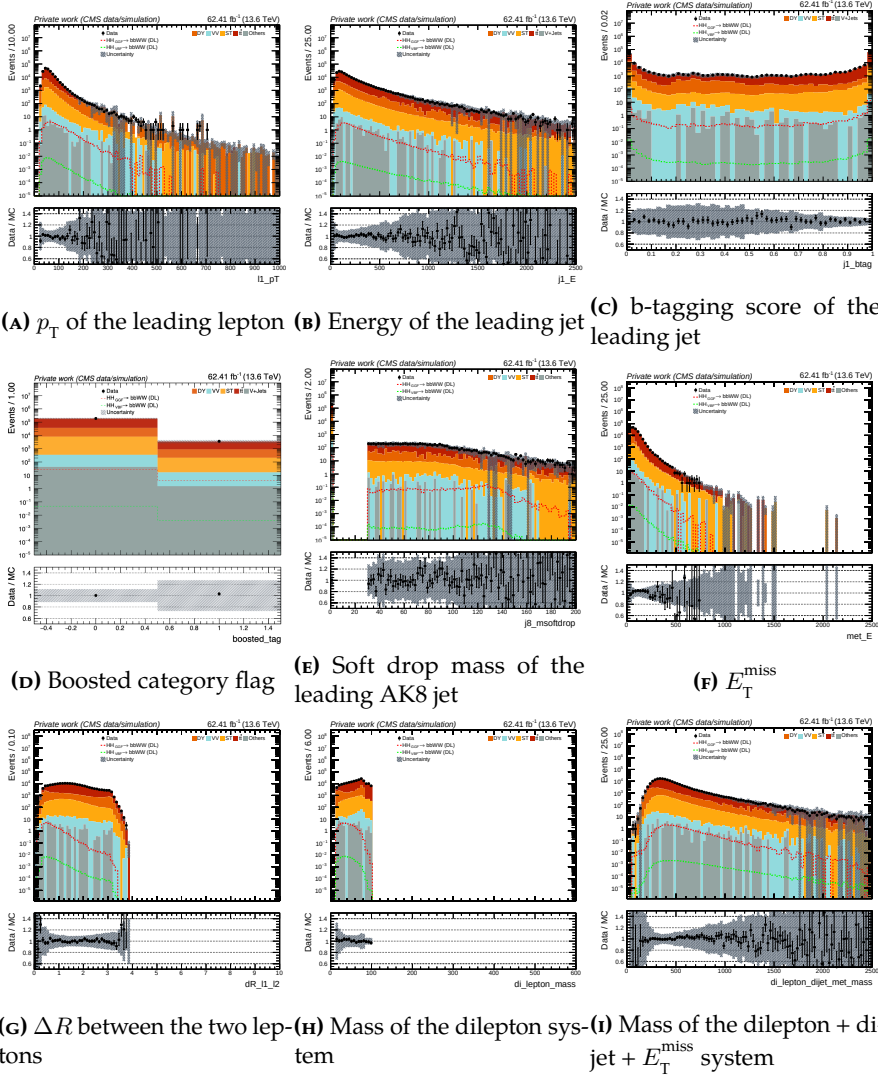


FIGURE 4.1 | Distributions of selected input variables used in the multivariate analysis, utilizing the 2022 and 2023 dataset (62.4 fb^{-1}) and combining the different lepton channels (e^+e^- , $\mu^+\mu^-$ and $e^\pm\mu^\mp$). The variables include kinematic and topological properties of the leptons and missing energy (leading lepton p_T , p_T^{miss} , $\Delta R(\ell_1, \ell_2)$, $M_{\ell\ell}$), features specific to resolved jets (leading jet's energy and b -tagging score), boosted AK8 jet properties (boosted tag, M_{softdrop}), and the invariant mass of the total system ($M_{2\ell 2j + \text{MET}}$).

as profiling, tightly constrains the parameter's variance. Because the regions are mathematically linked, this constraint propagates directly to the SR, significantly shrinking the background uncertainty band. Conversely, if these shape uncertainties were modeled as uncorrelated, the CR would provide no constraining power, leaving the SR degraded by large prior uncertainties.

4.1.6 Categorization

To maximize sensitivity to different signal kinematics and production modes, the analysis divides the selected events into orthogonal categories. This categorization is primarily driven by the hadronic topology of the event—specifically the reconstruction of large-radius jets ("Boosted") versus standard jets ("Resolved")—and the multiplicity of b-tagged jets. The categorization strategy proceeds as follows:

1. **Boosted Topology:** The boosted category targets the regime where the decay products of a heavy resonance (e.g., a Higgs boson) are collimated into a single large-radius jet. Events fall into this category if they contain at least one AK8 jet ($p_T > 200$ GeV) identified as a b-jet.
2. **Resolved Topology:** Events that do not satisfy the boosted criteria (i.e., contain zero b-tagged AK8 jets) are considered for the resolved categories. These rely on standard anti- k_T $R = 0.4$ jets (AK4). The resolved events are further split based on b-jet multiplicity to improve the signal-to-background discrimination:
 - **Resolved 1b:** Events containing exactly one b-tagged AK4 jet.
 - **Resolved 2b:** Events containing at least two b-tagged AK4 jets.
3. **VBF Topology:** Dedicated categories are defined to target VBF production signatures. These events are characterized by the presence of two additional "tagging" jets with a large invariant mass ($m_{jj} > 500$ GeV) and a large pseudo-rapidity separation ($|\Delta\eta_{jj}| > 3.0$). VBF candidates are identified in both boosted and resolved event topologies, using jet collections cleaned against the signal jets to avoid overlap.

Figure 4.2 summarizes the categories set up in the analysis.

4.1.7 Binning strategy

The output of the neural network classifier serves as the final discriminating variable used in the statistical inference. To maximize the sensitivity to the HH signal while ensuring the stability of the maximum-likelihood fit, the binning of these distributions is optimized to balance resolution in the signal-rich region with sufficient statistical precision in background-dominated regions.

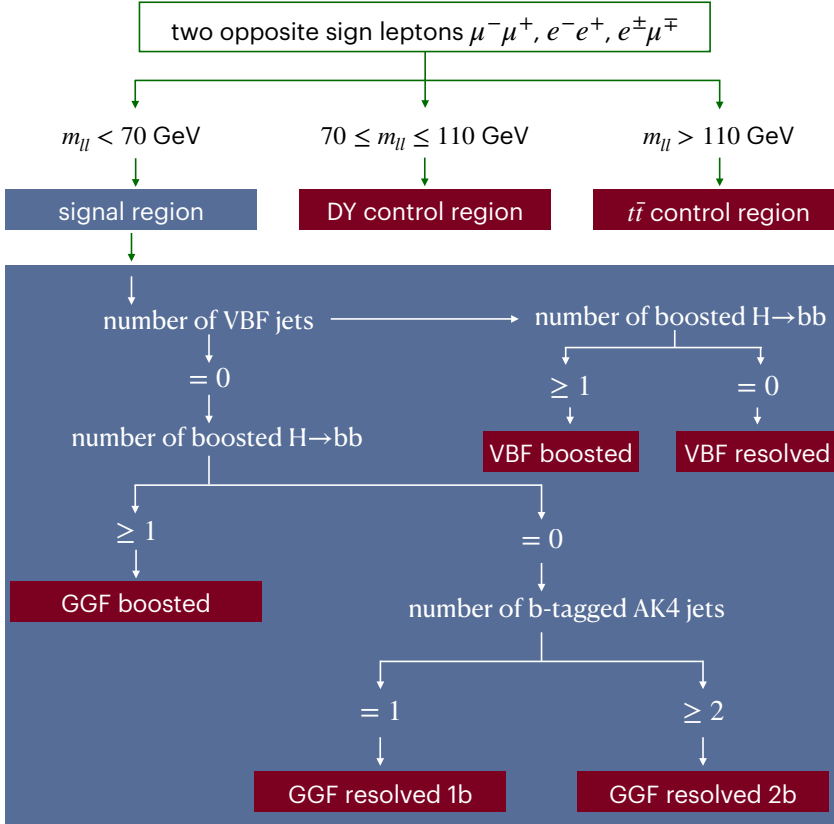


FIGURE 4.2 | Schematic overview of the event categorization strategy. Events containing two opposite-sign leptons are first classified based on the dilepton invariant mass ($m_{\ell\ell}$) into a Signal Region ($m_{\ell\ell} < 70$ GeV), a Drell-Yan Control Region ($70 \leq m_{\ell\ell} \leq 110$ GeV), and a $t\bar{t}$ Control Region ($m_{\ell\ell} > 110$ GeV). Events falling into the Signal Region are further subdivided based on hadronic activity: VBF categories are given the priority and those containing at least one boosted $H \rightarrow b\bar{b}$ candidate are classified as "Boosted," while the remaining events are classified as "Resolved 1b" or "Resolved 2b" based on the multiplicity of b-tagged AK4 jets (1 or ≥ 2). VBF resolved categories do not have the 1b and 2b separation.

A simple uniform binning is generally suboptimal for machine learning discriminants, which typically exhibit strongly peaked shapes near 0 (background-like) and near 1 (signal-like). Uniform binning can therefore dilute sensitivity by providing insufficient granularity where the signal is concentrated, while simultaneously creating sparsely populated bins where backgrounds dominate, increasing statistical fluctuations and potentially destabilizing the fit.

To address this, a custom variable-width binning strategy is applied to the classifier output histograms. The bin boundaries are optimized iteratively, starting from the highest classifier scores (the signal-enriched region) and moving toward the lowest scores (the background-dominated region). In the sensitive high-score tail, bins are defined to satisfy strict statistical safety thresholds—specifically, maintaining a minimum of 10 expected background Monte Carlo events and bounding the maximum fractional statistical uncertainty per bin between 20% and 30%, depending on the specific phase-space category. Moving toward the low-score region, where the signal is increasingly depleted, progressively coarser bins are constructed. This ensures robust constraints on the background normalizations while preventing the introduction of sparsely populated bins that could destabilize the maximum-likelihood fit.

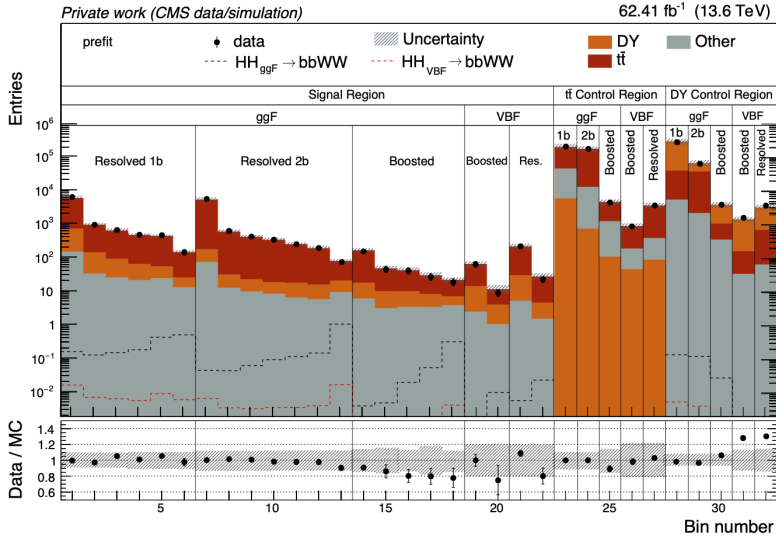
In background control regions, where the discriminant shape is not required, the workflow is simplified to use inclusive yields (a single bin) per category rather than a multi-bin shape. This avoids introducing low-signal bins while still providing strong constraints on background normalizations in the global fit. Figure 4.3 shows the resulting rebinned ML score distributions in the signal region.

4.2

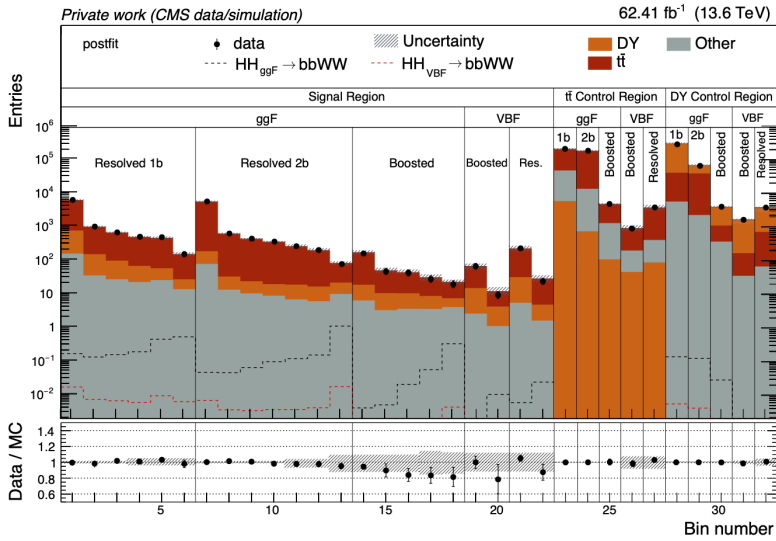
Corrections to the data and simulation

MC simulations are indispensable for estimating background processes and modeling signal characteristics in high-energy physics analyses. However, despite the high precision of modern event generators and the detailed detector modeling provided by software such as `GEANT4` [152], simulated samples cannot perfectly reproduce the complex experimental conditions of the CMS detector. Discrepancies between data and simulation arise from various sources, including imperfect modeling of the detector, approximations in the description of underlying event activity, and fluctuations in instantaneous luminosity. Furthermore, localized hardware malfunctions and time-dependent shifts in detector acceptance encountered during the Run-3 data-taking periods introduce additional non-uniformities that require specialized mitigation, as detailed in Section 4.2.1.

To ensure that the simulation accurately represents the data collected, a series



(A) Prefit.



(B) Postfit.

FIGURE 4.3 | ML score distributions before (A) and after (B) the fit.

of corrections must be applied. These corrections are typically derived by comparing the performance of physics objects (such as leptons and jets) in data against those in simulation using standard reference processes (e.g., $Z \rightarrow \ell\ell$).

The primary method for implementing these efficiency corrections is through the application of scale factors (SFs). An SF is defined as the ratio of the efficiency measured in data (ϵ_{data}) to the efficiency measured in simulation (ϵ_{MC}):

$$\text{SF} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}.$$

These factors are applied as event weights, re-normalizing the simulated events to match the selection probabilities observed in reality.

4.2.1 Issues in data taking in Run-3

The Run-3 data taking period encountered a few detector-specific issues that required dedicated mitigation strategies in both data processing and simulation [153,154]. These effects are taken into account in this analysis through dedicated event cleaning, corrections, and period-specific simulations.

During the 2022 data taking, a leak in the ECAL endcap (EE) water cooling circuit led to abnormally high energy deposits in several ECAL towers. The leak affected the cooling of electronics serving 20 communication and control units, corresponding to 491 ECAL crystals, and the impacted region was masked from 17 September until the end of the 2022 run, effectively rendering about 7% of the ECAL endcap unusable. Jets falling in the corresponding event-cleaning region are vetoed, and simulations used for 2022 are split into pre-EE and post-EE periods to correctly mimic the impact of the masked crystals; the leak was repaired during the year-end technical stop, so 2023 data are unaffected.

In 2023, after Technical Stop 1 (19–24 June), 27 modules in Barrel Pixel (BPix) layers 3 and 4 became inoperable because of a problem in the distribution of the LHC clock signal. To account for the resulting loss of jet energy and tracking performance, a dedicated jet energy scale and resolution correction is applied, and 2023 simulations are divided into pre-BPix and post-BPix periods to reproduce the changed detector conditions.

4.2.2 Pile-up re-weighting

The high instantaneous luminosity of the LHC results in multiple simultaneous proton-proton interactions per bunch crossing, referred to as pileup. To ensure the simulation accurately reflects experimental conditions, MC events are reweighted to match the pileup distribution observed in the data. The data

distribution is derived from the per-bunch instantaneous luminosity, assuming a total inelastic proton-proton cross section of 69.2 mb with a precision of $\pm 4.6\%$ [155]. Along with this uncertainty, the resulting weights are propagated to the final results to cover uncertainties in the luminosity measurement and the underlying physics model.

4.2.3 Trigger selection and efficiency

To correct for differences in trigger response between data and simulation, the trigger efficiency is measured as a function of the leading and sub-leading lepton p_T . The efficiency is defined as:

$$\epsilon(p_{T_1}, p_{T_2}) = \frac{N_{\text{pass trigger}}}{N_{\text{total}}}$$

where N_{total} is the number of events passing the full offline selection and $N_{\text{pass trigger}}$ is the subset of those events that also fired the relevant HLT path.

This efficiency is computed separately for Data and MC simulation. The final correction applied to MC events is the trigger scale factor, defined as the ratio of the observed efficiencies:

$$SF(p_{T,1}, p_{T,2}) = \frac{\epsilon_{\text{data}}(p_{T,1}, p_{T,2})}{\epsilon_{\text{MC}}(p_{T,1}, p_{T,2})}$$

The scale factors are derived in 2D histograms with variable binning up to 300 GeV in lepton p_T . The nominal weight along with "up" and "down" systematic variations are derived from the statistical uncertainty of the efficiency measurement. The calculated scale factors are shown in Appendix A.4

4.2.4 Noise filters in Run-3

For Run-3 analyses utilizing NanoAOD, a standard set of noise filters is recommended by CMS to reject anomalous events [156]. The baseline recommendation for 2022–2023 Data and MC includes the following flags:

Flag_goodVertices: Ensures the event contains a valid, well-reconstructed primary vertex.

Flag_globalSuperTightHalo2016Filter: Vetoes events containing beam halo particles (muons originating from upstream interactions).

Flag_EcalDeadCellTriggerPrimitiveFilter: Rejects fictitious MET generated by energy deposited in dead cells of the ECAL.

Flag_BadPFMuonFilter & Flag_BadPFMuonDzFilter: Reject events with erroneous muons based on track quality and vertex compatibility.

Flag_hfNoisyHitsFilter: Removes events with anomalous noise in the Hadronic Forward calorimeter.

Flag_eeBadScFilter: Vetoes noise spikes arising from specific bad supercrystals in the ECAL Endcap.

Flag_ecalBadCalibFilter: Rejects events with energy deposits in ECAL crystals known to have incorrect calibration factors, which would otherwise create artificial MET tails.

4.2.5 Top quark p_T reweighting

Measurements of differential $t\bar{t}$ cross-sections during Run 1 and Run 2 have revealed a discrepancy between experimental data and standard NLO MC predictions [157,158]. Specifically, the p_T spectrum of top quarks observed in data is significantly softer than that predicted by generators such as POWHEG+PYTHIA8. This mismodeling is generally attributed to missing higher-order QCD and electroweak corrections in the simulation. To mitigate this effect and improve the modeling of the detector response (e.g., trigger and b-tagging efficiencies), a reweighting procedure is applied to simulated $t\bar{t}$ events.

The correction is applied at the parton level using SFs derived from the ratio of the unfolded differential cross-section in data to the MC prediction. The total event weight, w_{top} , is defined as the geometric mean of the scale factors for the top quark (t) and top antiquark ($\bar{t}$):

$$w_{\text{top}} = \sqrt{\text{SF}(p_T^t) \cdot \text{SF}(p_T^{\bar{t}})}.$$

For the standard POWHEG+PYTHIA8 samples, the data-driven scale factor is given by the TOP PAG as:

$$\text{SF}(p_T) = 0.103 \cdot e^{-0.0118 \cdot p_T} - 1.34 \cdot 10^{-4} \cdot p_T + 0.973,$$

where p_T is measured in GeV. To avoid extrapolating into unmeasured phase space, the weight for quarks with $p_T > 500$ GeV is capped at the value calculated for 500 GeV.

4.2.6 b-tagging corrections

To accurately model the performance of the b-tagging algorithm in simulation, corrections are applied to the b-tagging discriminant shape. Unlike standard methods that correct efficiencies at specific working points, this “shape correction” method reweights simulated events to match the full discriminant distribution observed in data. This approach is essential for analyses that

rely on the discriminant shape for signal separation or background estimation. These are provided by the b -tagging and Vertexing (BTV) POG [159].

The correction is implemented as an event weight derived from the properties of all jets in the event. For a given jet i , the scale factor SF_i depends on the jet's transverse momentum ($p_{T,i}$), pseudorapidity (η_i), and its specific b -tagging discriminant value (D_i). The total event weight, ω_{event} , is the product of the scale factors for all selected jets:

$$\omega_{\text{event}} = \prod_{i=1}^{N_{\text{jets}}} SF(D_i, p_{T,i}, \eta_i).$$

This method ensures that events are reweighted without migrating them between jet multiplicity bins.

The application of these scale factors can inadvertently alter the total predicted event yield before any b -tagging selection is applied. To preserve the inclusive normalization of the sample—ensuring that the tagging calibration does not affect the total theoretical cross-section—a renormalization factor r is required. This factor is defined as the ratio of the sum of weights before and after the b -tagging correction:

$$r = \frac{\sum \omega_{\text{nominal}}}{\sum \omega_{\text{btag}}}. \quad (4.1)$$

Since the impact of the reweighting varies with the phase space, this ratio r must be derived and applied separately for each process, jet flavor composition, and jet multiplicity bin.

The uncertainty on the shape correction is evaluated through several independent variations, which account for detector effects and statistical limitations in the derivation of the scale factors.

- **Jet energy scale (JES):** The scale factors are re-derived using jets with JES-shifted energies to propagate the impact of JES uncertainties to the b -tagging performance.
- **Purity (lf, hf):** This accounts for the contamination of light-flavor jets in heavy-flavor control regions (hf) and vice-versa (lf) by varying the contamination rates.
- **Statistics (hfstats, lfstats):** These variations account for statistical fluctuations in the Data and MC samples used to derive the scale factors. They are split into linear (1) and quadratic (2) terms for both heavy-flavor (hfstats1/2) and light-flavor (lfstats1/2) jets.
- **Charm-jet mis-tagging (cferr1/2):** Accounts for the probability of charm quarks being mis-identified as b -jets. Due to the lack of pure

calibration samples, c -jets are assigned a nominal scale factor of 1.0, with uncertainties extrapolated from the b -jet envelope. These are split into linear ($cferr1$) and quadratic ($cferr2$) terms to allow the fit to profile potential mismodeling in charm-enriched backgrounds like $t\bar{t}$ and W +jets.

4.2.7 Boosted $b\bar{b}$ -jet tagging

To identify the boosted decay of the Higgs boson into a pair of bottom quarks ($H \rightarrow b\bar{b}$), this analysis utilizes AK8 PUPPI¹ jets tagged with the ParticleNet algorithm with the mass-correlated discriminant (`particleNetWithMass - _HbbVsQCD`).

To account for differences between data and simulation in tagging efficiency, SFs are derived using the μ -tagged method [160], as recommended by the BTV POG. The μ -tagged method relies on selecting a high-purity sample of heavy-flavor jets from data to serve as "proxies" for the signal jets. Since the hadronization of b quarks has a significant probability of producing a muon via semileptonic decay, the presence of a soft, non-isolated muon within a jet cone is a strong indicator of heavy-flavor content.

The calibration workflow utilizes AK8 jets containing a reconstructed soft muon to model inclusive heavy-flavor jets. To ensure accurate modeling of the background, a multi-dimensional reweighting procedure is applied. A 3D reweighting map based on jet p_T , $|\eta|$, and the substructure variable τ_{21} is computed to correct the simulation to match the jet kinematics observed in data. This step is critical for minimizing kinematic differences between the proxy jets and the target signal jets.

The SFs are extracted as the ratio of tag efficiencies in data and simulation for the chosen working point for the discriminant (> 0.92), in bins of jet p_T (300-350, 350-425 and 425-infinity).

This analysis specifically utilizes the mass-correlated version of the ParticleNet discriminant. However, because the standard μ -tagged calibration procedure was originally designed for the mass-decorrelated tagger, a noticeable shape discrepancy was observed in the proxy jet soft-drop mass (m_{SD}) distribution after applying the calibration reweighting. To quantify the systematic uncertainty associated with this effect, the scale factors were re-derived by shifting the lower m_{SD} threshold applied to the proxy jets from 40 GeV up to 100 GeV (shown

¹A PUPPI jet is reconstructed by applying a clustering algorithm to particle-flow candidates that have been individually weighted by the Pileup Per Particle Identification (PUPPI) algorithm. Because each constituent particle's four-momentum is scaled by its probability of originating from the primary hard-scatter vertex rather than from pileup interactions, the resulting jet exhibits improved energy and mass resolution, offering greater stability against pileup contamination.

in Figure A.29). While the scale factors remain stable for moderate threshold changes, extreme variations (specifically 90 and 100 GeV) act as outliers and yield noticeably different values. Since the baseline analysis requires a loose, inclusive threshold of $m_{SD} \geq 30$ GeV, the maximum deviation caused by these high-mass cuts is adopted as a conservative systematic envelope to robustly cover the observed m_{SD} shape discrepancy.

Flavor dependence is treated explicitly: AK8 jets matched to B hadrons are corrected with bb scale factors, while jets matched to C hadrons are corrected with cc scale factors. Light-flavor jets remain unweighted as their contribution in the tagged region is negligible. The resulting scale factors are consistent with unity within their uncertainties, as shown in Figure A.28.

In this analysis, I applied the μ -tagged calibration procedure myself, and the resulting scale factors and associated uncertainties are included in the public Run-3 analysis [161].

4.2.8 Corrections to leptons

To ensure accurate modeling of lepton kinematics and selection efficiencies, specific corrections are derived by the Muon and EGamma POGs. These corrections address discrepancies between Data and MC simulation arising from detector effects, material modeling, and reconstruction performance.

Muons

For analyses utilizing muons in the medium transverse momentum range ($15 < p_T < 200$ GeV), corrections are derived by Muon POG using the Tag-and-Probe (TnP) method on Z boson resonances [162]. SFs on identification (ID) efficiency are provided for standard working points (Loose, Medium, Medium Prompt, Tight, Soft, and High- p_T). These efficiencies are parametrized as a function of the muon p_T and η . Isolation SFs are computed using TnP on DY events. Available working points include PFIso (Loose, Tight), MiniIso (Loose, Medium, Tight), and TrackerIso (Loose, Tight). The corrections corresponding to the working points used for the objects are applied.

Electrons

Corrections for electrons are applied for reconstruction, identification, and triggering [67]. SFs for electron reconstruction and ID are derived for standard cut-based IDs (Veto, Loose, Medium, Tight) and MVA-based IDs by the EGamma POG. Trigger SFs on the other hand are dependent on the offline selection. The provided corrections assume standard offline ID requirements;

analyses applying stricter isolation or impact parameter cuts must validate these SFs separately.

To calibrate the electron energy measurement, a "Scale and Smearing" procedure is applied, matching the $Z \rightarrow ee$ peak in Data to MC. For data, energies are scaled (typically shifted) to correct for underestimated energy response and for simulation energies are smeared to widen the resolution, matching the worse resolution observed in data.

4.2.9 Z boson p_T reweighting

The DY process simulation, typically generated at LO or NLO precision, often exhibits discrepancies in the modeling of the vector boson transverse momentum spectrum, $p_T(Z)$, when compared to data [163]. These mismodelings arise primarily from missing higher-order QCD corrections.

To mitigate this, a reweighting procedure is applied to the DY MC samples. In this analysis, the SFs are provided by CMS and parameterized as a function of the generator-level dilepton transverse momentum, $p_T^{\text{gen}}(\ell\ell)$. The generator-level leptons are constructed by selecting particles from the GenPar t (generator level events) collection that satisfy the following criteria:

- **Status:** Stable particles (status code 1).
- **Flavor:** Correct PDG ID (± 11 for ee , ± 13 for $\mu\mu$).
- **Origin:** A check on the status flags to ensure the leptons originate from the hard process.

The reweighting factor is applied event-by-event to the simulation. Furthermore, systematic uncertainties associated with this reweighting are propagated via dedicated variations to cover the uncertainty in the derivation of the weights.

4.2.10 Jet energy corrections and resolution

Accurate reconstruction of jet kinematics is critical for precision measurements and searches. To ensure the simulated jet response matches the data, a series of corrections are applied to the jet energy scale (JES) and jet energy resolution (JER), collectively called Jet Energy Corrections (JECs) [164].

Jet energy scale

The measured energy of reconstructed jets differs from the true particle-level energy due to detector non-uniformities, pileup, and non-linear calorimeter response. A factorized set of corrections is applied to mitigate these effects. For the 2022 and 2023 datasets, corrections are provided for specific eras to account

for changing detector conditions (e.g., the EE leak in 2022 and the BPix issue in 2023). Recommended SFs by the Jets and Missing Transverse Energy POG (JME) are applied corresponding to each data-taking period.

Jet energy resolution

Simulated jets typically exhibit better energy resolution than observed in data. To correct for this, the energy of simulated jets is “smeared” to match the resolution measured in data. The resolution scale factors are provided for each era separately and are applied after the JEC corrections.

4.2.11 Jet veto in Run-3

For analyses utilizing data from Run-3, the application of jet veto maps has become mandatory, as stipulated by the Jet Energy Resolution and Corrections (JERC) group under JME POG in CMS [165]. The primary motivation for this requirement is that JECs are derived excluding specific zones of the detector that exhibit anomalous behavior. Consequently, jets falling within these regions do not possess reliable corrections, requiring their exclusion or the rejection of the entire event (safest as recommended by the POG) to preserve the integrity of the analysis phase space.

To apply the event veto, a “minimal selection” is defined to identify the jets that trigger the veto. A jet is considered for control against the maps if it satisfies the following criteria:

- Transverse momentum: $p_T > 15$ GeV.
- Identification: tightLepVeto Jet ID (strict criteria excluding jets originating from isolated leptons).
- Electromagnetic Fraction: The sum of the charged and neutral electromagnetic energy fractions must be less than 0.9. This selection is motivated by Type-1 MET corrections.

4.3 Systematic uncertainties

Several sources of systematic uncertainties are considered in this analysis, each represented as a nuisance parameter (NP) in the maximum likelihood fit. These uncertainties affect both the normalization and the shape of the distributions used for signal extraction. The different types of uncertainties are grouped into experimental and theory-based. Additionally, a statistical uncertainty to account for the limited size of the simulated samples is employed using the Barlow-Beeston-lite method.

4.3.1 Experimental uncertainties

- **Luminosity:** Annual luminosity uncertainties are determined according to Luminosity Physics Object Group (LUM POG) recommendations. These values are set at 1.4% for 2022 and 1.3% for 2023, affecting the normalization of all simulated processes. These uncertainties are treated as uncorrelated between the two data-taking years [156].
- **Pileup:** Shape uncertainty from mismodeling of extra pp interactions. This uncertainty is considered correlated across all data-taking campaigns.
- **Trigger:** Shape uncertainty from data/MC differences in trigger efficiency which are uncorrelated across campaigns.
- **Lepton identification and isolation:** The analysis applies SFs provided by the respective POGs to align simulated lepton identification and isolation efficiencies with data. These sources are treated as uncorrelated across the different data-taking campaigns.
- **Top-quark p_T reweighting:** Shape uncertainty from the reweighting procedure, assessed by comparing results with and without the reweighting applied as per TOP POG recommendations.
- **JES and JER:** Uncertainties from JEC and JER are estimated by shifting energy scales or smearing factors by $\pm 1\sigma$. This uncertainty is treated as uncorrelated across data-taking campaigns.
- **b-jet tagging:** Uncertainties are applied using the "shape-based" method following BTV POG recommendations.
 - **JEC/JER correlation:** Variations in jet energy scale and resolution are propagated to the b-tag SFs by shifting both simultaneously.
 - **Flavor purity (hf/lf):** SF uncertainties related to heavy-flavor and light-flavor control sample purity are applied to b-quark and light-flavor jets, respectively.
 - **Statistical and c-quark jet uncertainties:** Statistical fluctuations in SF derivation (lfstats1/2, hfstats1/2) and c-quark jet uncertainties (cferr1/2) are parameterized as separate nuisance parameters.
 - **Correlation:** To prevent artificial fit constraints, all b-tagging uncertainty sources are treated as uncorrelated across data-taking campaigns.
- **Boosted bb-jet tagging:** The μ -tagged method introduces specific intrinsic uncertainties related to the definition of the proxy jets. The dominant contribution arises from the choice of the τ_{21} threshold used to select the proxies. By varying this threshold (from 0.2 to 0.4), the sensitivity

of the scale factors to the kinematic similarity between proxy and signal jets is estimated. Other intrinsic uncertainties include the modeling of the QCD background composition and the simulation-to-data reweighting applied to the secondary vertex mass fit variable, though these are sub-dominant compared to the τ_{21} variation. Additionally, a shape uncertainty is assigned based on the residual data-to-MC mismodeling in the m_{SD} distribution. Each uncertainty is treated as uncorrelated between the data-taking campaigns and between the boosted ggF SR, VBF SR, and CR.

4.3.2 Theoretical uncertainties

- **Inclusive Cross Sections and PDF:** These uncertainties apply to the inclusive signal and background cross sections used for MC normalization [166]. They are divided into two categories: those arising from renormalization and factorization scales (μ_R/μ_F) in the matrix-element calculations, and those stemming from the PDF set and α_S value. These are incorporated as rate uncertainties in the final fit for all processes excluding $t\bar{t}$ and Drell-Yan, which are left as free-floating parameters. Common systematic sources are fully correlated between processes. Detailed values are provided in Tables A.1 and A.2.
- **Unconstrained Normalizations:** As noted above, certain background rates are treated as free parameters in the final fit. Additionally, two distinct normalization factors are assigned for $t\bar{t}$ to address potential modeling discrepancies between the resolved and boosted p_T regions. Drell-Yan backgrounds are also split into ggF and VBF normalization parameters, allowing the fit to independently correct for the discrepancy seen in the VBF production mode.
- **Matrix-Element Scale Variations:** The analysis accounts for μ_R and μ_F scale uncertainties by evaluating the envelope of six scale combinations (varying each by factors of 0.5 and 2). These variations, derived from internal generator weights, are applied to the signal and all background processes as uncorrelated parameters. Because the total rate is already constrained by inclusive cross-section uncertainties, these variations are scaled to be "shape-only," preserving the nominal normalization in the pre-selection phase space. This ensures the fit only accounts for changes in the kinematic distributions and relative acceptance.
- **PDF-Induced Shape Variations:** Uncertainties arising from the PDF set choice are evaluated by reweighting event distributions according to NNPDF variation sets and taking their envelope. To avoid overlap with the inclusive cross-section constraints, these variations are treated

as "shape-only," meaning they are normalized to match the nominal yields before selection cuts are applied. This procedure accounts for how PDF uncertainties affect the distribution shapes and relative selection efficiencies [166]. These effects are treated as uncorrelated across all signal and background processes.

- **Parton Shower Scales (ISR/FSR):** Parton shower uncertainties are evaluated by scaling α_S up and down by a factor of 2 within the "default" scheme for ISR and FSR independently [167]. While ISR effects are assumed to be fully correlated across all processes, FSR variations are treated as independent (uncorrelated) for each specific process. These variations are scaled to preserve the nominal normalization at the inclusive level, focusing the fit on shape-driven migrations and relative acceptance.
- **MC statistics:** The statistical uncertainty arising from the finite size of the simulated samples is accounted for in the likelihood fit using the Barlow-Beeston lite approach [168]. This method assigns a single nuisance parameter per bin to account for the total statistical fluctuation of the sum of all MC processes. To ensure computational efficiency, these parameters are only activated for bins where the effective number of unweighted events falls below a specified threshold, allowing the fit to focus on statistically significant fluctuations while maintaining stability in the global likelihood.

4.4 Results

The search for non-resonant Higgs boson pair production in the dileptonic $b\bar{b}WW$ decay channel is performed using the Run-3 proton-proton collision dataset collected during 2022 and 2023. This dataset corresponds to an integrated luminosity of 62.4 fb^{-1} at a center-of-mass energy of $\sqrt{s} = 13.6 \text{ TeV}$.

To quantify the compatibility of the data with the background-only and signal-plus-background hypotheses, a binned maximum likelihood fit is performed simultaneously across all signal categories and background control regions. The global likelihood function $\mathcal{L}(\mu, \theta)$ is constructed as a product of Poisson probabilities over all bins, penalized by constraint terms for the systematic uncertainties:

$$\mathcal{L}(\mu, \theta) = \prod_i \frac{(\mu s_i(\theta) + b_i(\theta))^{n_i} e^{-(\mu s_i(\theta) + b_i(\theta))}}{n_i!} \prod_j \pi(\theta_j)$$

Here, n_i is the observed event yield in bin i , while $s_i(\theta)$ and $b_i(\theta)$ are the expected signal and background yields, which depend on the nuisance parameters θ .

ters θ encompassing all systematic and statistical uncertainties. The parameter $\mu = \sigma_{\text{HH}}/\sigma_{\text{HH}}^{\text{SM}}$ is the signal strength modifier, and $\pi(\theta_j)$ represent the prior probability density functions constraining the systematic variations.

Upper limits on the signal strength μ are established at the 95% CL by employing the modified frequentist CL_s method. This procedure utilizes the profile likelihood ratio, q_μ , as the primary test statistic, with its distributions evaluated under the asymptotic approximation.

No significant excess of events above the SM background expectation is observed. Consequently, 95% CL upper limits are set on the inclusive HH production cross-section. The observed (expected) upper limit is determined to be:

$$\sigma_{\text{HH}}^{\text{obs}} < 35(47) \times \sigma_{\text{HH}}^{\text{SM}} \quad \text{at 95\% CL}$$

as shown in Figures 4.4 and 4.5. The compatibility between the observed data and the multi-category statistical model was further validated using a Goodness-of-Fit test based on the saturated model algorithm. The result indicates that the data are well-described by the model across all categories; detailed results of this test are provided in Figure A.30.

The individual contributions of each analysis category to the final result are summarized in Figure 4.5, revealing a clear hierarchy in sensitivity across the different event topologies. The analysis is primarily driven by the ggF Boosted and ggF 2b (resolved) categories, which provide the most stringent constraints on the signal strength. The Boosted category exhibits the highest individual expected sensitivity ($68 \times \sigma_{\text{SM}}$), as the collimated decay products at high- p_T allow for a cleaner separation of signal from the dominant background processes. The resolved **ggF 2b** category follows closely ($80 \times \sigma_{\text{SM}}$), offering substantial statistical power in the moderate p_T regime. While the ggF 1b category is significantly less sensitive ($299 \times \sigma_{\text{SM}}$) due to its higher background contamination, its inclusion in the simultaneous fit is vital. It ensures maximum signal acceptance across the phase space and provides additional data to constrain background nuisance parameters that are correlated with the more sensitive regions. The power of the multi-category approach is most evident in the combined expected limit of $47 \times \sigma_{\text{SM}}$, which represents an improvement of approximately 30% over the single most sensitive channel. Furthermore, the combination serves as a statistical stabilizer; for instance, the slight under-fluctuation observed in the 2b category and the minor over-fluctuation in the Boosted category are reconciled in the global fit. This leads to an observed limit of $38 \times \sigma_{\text{SM}}$ that is more robust and less susceptible to the statistical noise of any individual topology.

It should be noted that the VBF categories are excluded from the per-channel

summary figure. This exclusion ensures that the visualization of the highly sensitive channels is not constrained to a narrow area of the plot. The upper limit values for these categories are as follows: the VBF resolved category yields an expected limit of 1295 and an observed limit of 808; the VBF boosted category yields an expected limit of 2937 and an observed limit of 5779. Both results are consistent with the background-only hypothesis within the 2σ uncertainty band.

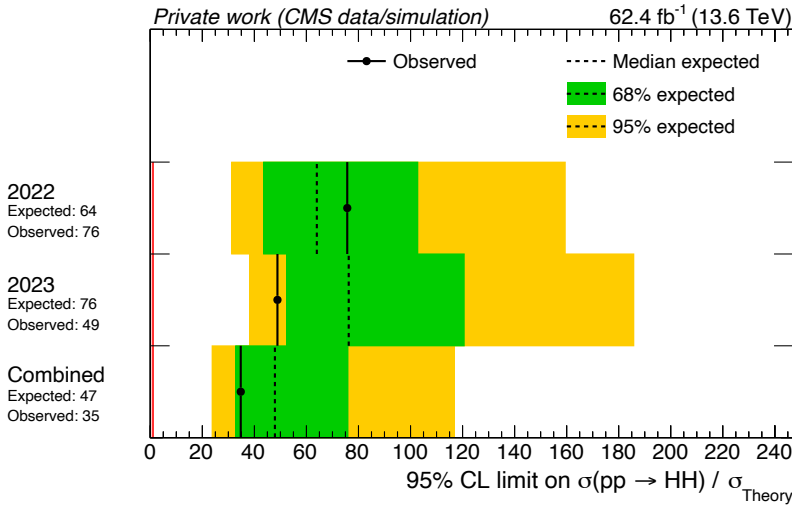


FIGURE 4.4 | Observed and expected 95% CL upper limits on the non-resonant Higgs boson pair production cross-section normalized to the SM prediction ($\sigma/\sigma_{\text{SM}}$). The limits are presented for the 2022 and 2023 datasets individually, as well as for their combination. The solid black line with a marker indicates the observed limit, while the dashed black line represents the median expected limit. The red line indicates the SM prediction value of 1. The green and yellow bands correspond to the 68% and 95% confidence intervals, respectively. The combination of different data-taking periods is essential for a healthier and more robust result, as the increased integrated luminosity significantly enhances the statistical sensitivity and helps mitigate the impact of independent year-to-year fluctuations.

4.4.1 Impact of systematic uncertainties

To assess the influence of systematic uncertainties on the final result, a study of the nuisance parameter impacts is performed. Figure 4.6 shows the "pulls and impacts" summary for the 20 nuisance parameters with the largest effect on the signal strength modifier μ .

The "pull" (black markers) indicates the deviation of the post-fit value of the

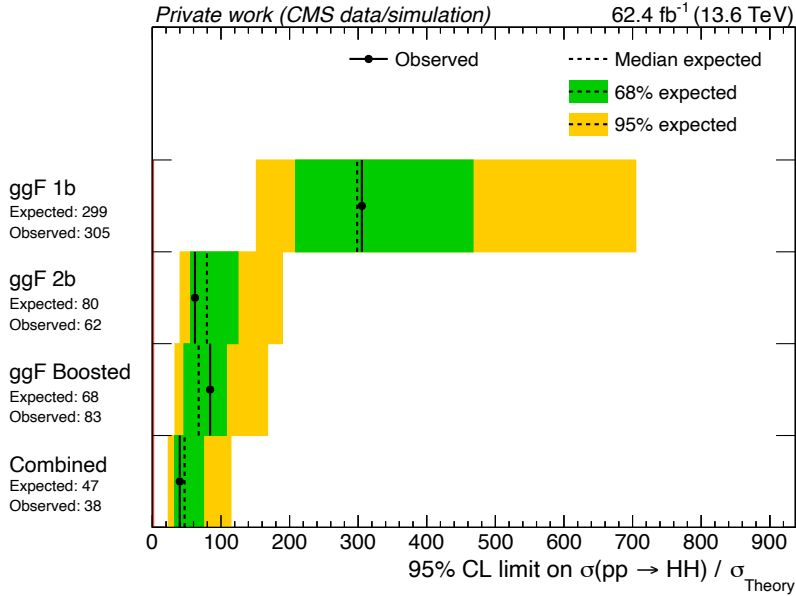


FIGURE 4.5 | Observed and expected 95% CL upper limits on the non-resonant HH production cross-section, displayed for the individual analysis categories (ggF 1b, ggF 2b, and ggF Boosted) and their statistical combination. The solid black line with a marker indicates the observed limit, the dashed line represents the median expected limit, and the green (yellow) bands denote the 68% (95%) confidence intervals. This comparison highlights the complementary nature of the different event topologies; while the ggF 2b and Boosted categories provide the primary sensitivity, the simultaneous fit across all categories—including the 1b region—maximizes the use of the available signal phase space. The combined limit of 47 (expected) represents a significant improvement over any single category, demonstrating that a multi-category approach is vital for achieving the most stringent constraints on the HH cross-section.

nuisance parameter (θ_{post}) from its pre-fit value (θ_{pre}), normalized by the pre-fit uncertainty ($\Delta\theta$). The "impact" (colored bars) represents the shift in the signal strength observed when the specific nuisance parameter is fixed to its best-fit value plus or minus its post-fit uncertainty, while profiling all other parameters. As shown in Figure 4.6, the pulls for all major nuisance parameters are well within one standard deviation of their prior uncertainties, indicating that the background model and systematic uncertainties are well-constrained by the data without exhibiting significant tensions. While the analysis sensitivity is primarily limited by the statistical uncertainty of the data due to the small signal cross-section, the systematic budget is dominated by theoretical modeling and high-level object reconstruction.

Among the systematic sources, the highest impacts arise from theoretical uncertainties—specifically QCD scale variations and parton shower modeling—as well as jet energy scale corrections. Pileup, free-floating background normalization rates and object-related uncertainties, particularly those associated with trigger SFs and b -tagging efficiencies, represent the leading experimental contributors. Significant future reductions in the total uncertainty will rely on a combination of increased data-taking, improved jet energy scale corrections, and advancements in theoretical calculations to constrain modeling discrepancies.

4.4.2 Constraints on the Higgs boson self-coupling (κ_λ)

The result is further interpreted in the context of the trilinear Higgs boson self-coupling, κ_λ ($= \lambda_{HHH}/\lambda_{HHH}^{\text{SM}}$). Upper limits on the HH production cross-section are computed as a function of κ_λ , assuming all other couplings (including κ_{2V}) remain at their SM values.

Changing κ_λ modifies not only the total rate but also the HH kinematics, because it changes the interference pattern between the box and triangle diagrams in gluon-gluon fusion. Near the destructive-interference region, the m_{HH} spectrum is shifted toward lower values and the Higgs bosons tend to be less boosted, while larger positive or negative values of κ_λ populate harder kinematic configurations and increase the available phase space.

Figure 4.7 shows the 95% CL upper limits on the cross-section relative to the SM prediction. The sensitivity varies with κ_λ due to the interference effects between the box and triangle diagrams in the gluon-gluon fusion loop, which alter both the total cross-section and the kinematic distributions of the Higgs bosons. The resulting limit curve is therefore non-monotonic: it is weakest where the signal is most suppressed and kinematically least distinctive, and it tightens again when the cross-section and event hardness increase away

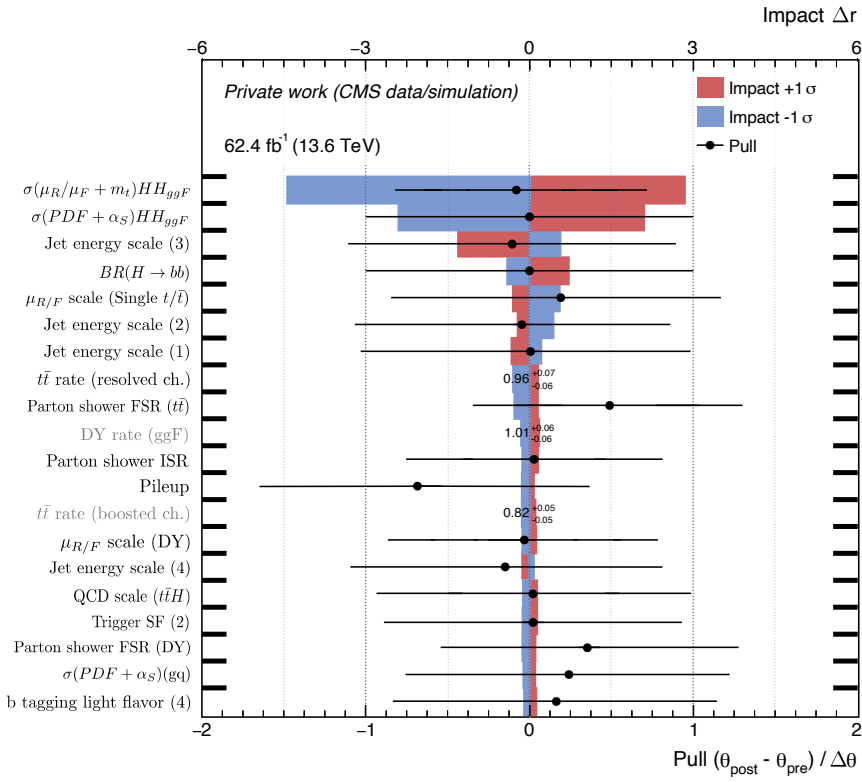


FIGURE 4.6 | The impacts of the 20 most significant nuisance parameters on the signal strength parameter (colored bars) and their post-fit pulls relative to the pre-fit uncertainties (black points).

from that region. The corresponding negative log-likelihood scan is shown in Figure A.31. Values of κ_λ outside the interval $[-13.3, 19.0]$ are excluded at 95% CL, compared to an expected exclusion interval of $[-20.0, 23.7]$.

4.4.3 Constraints on the VVHH coupling (κ_{2V})

Constraints are also placed on the interaction strength between two vector bosons and two Higgs bosons, κ_{2V} ($= c_{2V}/c_{2V}^{\text{SM}}$). This coupling is directly accessible via the VBF production mode. Similar to the κ_λ interpretation, limits are derived by scanning κ_{2V} while fixing other couplings to their SM predictions.

Figure 4.8 displays the cross-section limits, and Figure A.31 presents the likelihood scan. The data exclude values of κ_{2V} outside the interval $[-2.9, 5.0]$ at 95% CL, with an expected exclusion range of $[-2.9, 5.0]$. This constraint demonstrates the sensitivity of the analysis to the VBF production topology, which is explicitly targeted by the categorization strategy.

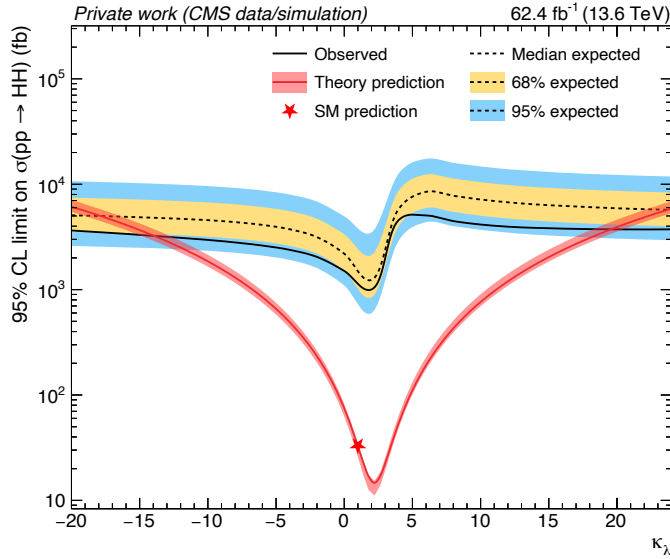


FIGURE 4.7 | Observed and expected 95% CL upper limits on the inclusive HH production cross-section as a function of κ_λ . The red curve represents the theoretical prediction for the cross-section.

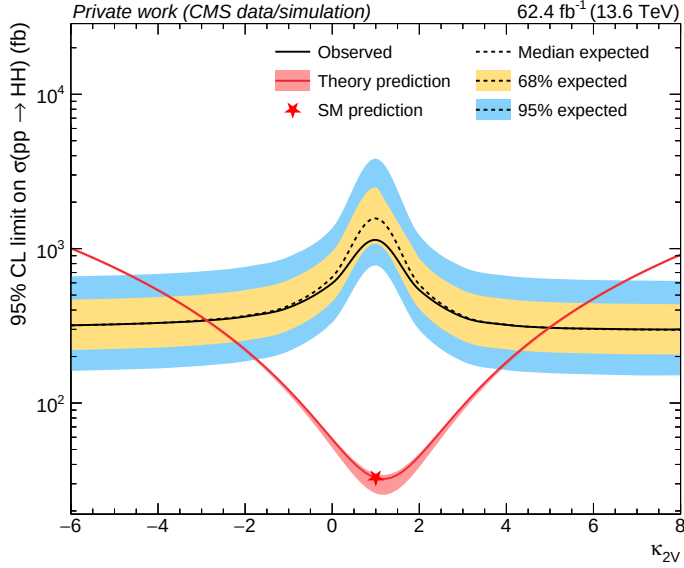


FIGURE 4.8 | Observed and expected 95% CL upper limits on the inclusive HH production cross-section as a function of κ_{2V} . The red curve represents the theoretical prediction for the cross-section.

4.4.4

Two-dimensional coupling constraint

While single-parameter scans provide insight into the sensitivity to individual couplings, the physical cross-sections depend on the simultaneous values of multiple coupling modifiers due to interference effects between diagrams. To capture these correlations, two-dimensional likelihood scan is performed.

The interplay between the ggF and VBF production modes is explored by scanning the interaction strengths κ_λ and κ_{2V} simultaneously. While κ_λ affects both production modes (appearing in the trilinear coupling vertex), κ_{2V} is specific to the VBF process. Figure 4.9 shows the 95% CL exclusion contours in the $(\kappa_\lambda, \kappa_{2V})$ plane. The expected and observed exclusion regions highlight the analysis sensitivity to deviations from the SM prediction ($\kappa_\lambda = 1, \kappa_{2V} = 1$). The shape of the contours reflects the complementary constraints provided by the categorization strategy: the "Boosted" and "Resolved" categories primarily constrain the GGF-driven κ_λ direction, while the dedicated "VBF" categories provide the necessary sensitivity to constrain the κ_{2V} axis.

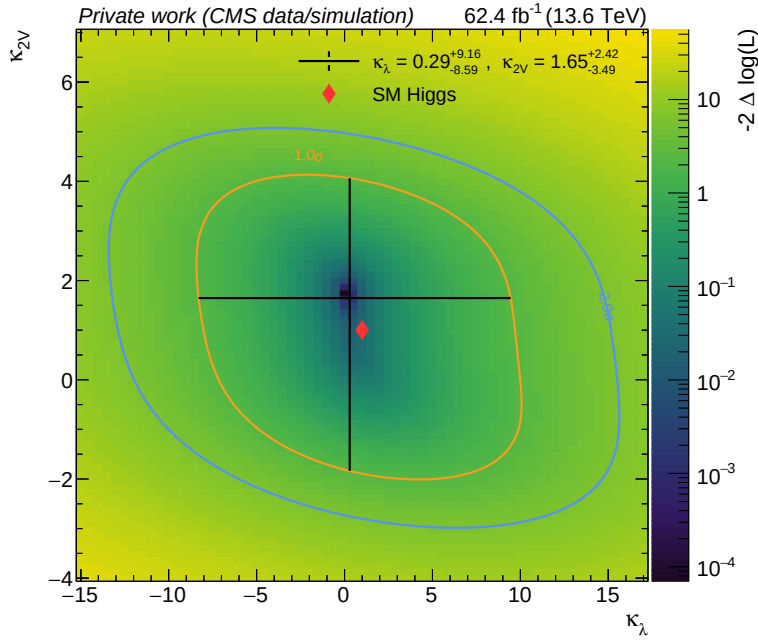


FIGURE 4.9 | Two-dimensional likelihood scan of the Higgs self-coupling modifier κ_λ versus the VVHH coupling modifier κ_{2V} . The solid contours represent the observed 1σ and 2σ exclusion regions. The SM prediction corresponds to the point $(\kappa_{2V}, \kappa_\lambda) = (1, 1)$, marked with a red diamond.

4.4.5 Software tools

The full analysis chain is developed using the BAMBOO [169] analysis framework to construct a comprehensive workflow that handles event selection, object definition, categorization, and histogram production, which are archived at https://github.com/Oguz-Guzel/HHtoWWbb_Run3. This included implementing the complex categorization logic illustrated in Figure 4.2 and ensuring the correct application of all object corrections and systematic uncertainties described in Section 4.2 and Section 4.3.

For the multivariate analysis strategy described in Chapter 3, the deep learning models used for signal discrimination are developed using the PyTorch library to develop the three different architectures explained in the same chapter. This work included the optimization of hyperparameters and the integration of the trained models into the BAMBOO analysis pipeline for inference.

Finally, HH INFERENCE TOOLS [170] is used to perform the simultaneous binned maximum likelihood fit across all analysis categories. This involved setting up the datacards, handling the correlations of systematic uncertainties, and computing the 95% CL upper limits on the inclusive HH production cross-section and coupling modifier scans.

4.5 Comparison with Run-2 and Advanced Run-3 Analysis Regimes

The baseline results obtained in this thesis can be contextualized by comparing them to the full Run-2 analysis of the same final state performed by CMS, which utilized a dataset of 138 fb^{-1} collected at $\sqrt{s} = 13 \text{ TeV}$ [171].

The Run-2 legacy analysis reported an observed (expected) upper limit of 27 (19) times the SM prediction. The baseline limit observed (expected) established in this thesis is 35 ($47 \times \text{SM}$). While this standalone thesis result is less stringent, the difference is largely driven by the available integrated luminosity; the Run-3 dataset used here (62.4 fb^{-1}) represents approximately 45% of the integrated luminosity available in Run-2. Assuming the sensitivity is statistically dominated, a naive scaling with the square root of the luminosity ratio yields:

$$\sqrt{\frac{62.4}{138}} \times 35 \approx 23.5. \quad (4.2)$$

This benchmark demonstrates that the standalone analysis provides a solid foundation for the bbWW channel in the new 13.6 TeV energy regime.

However, the ultimate sensitivity of this early Run-3 dataset extends significantly further when leveraging advanced statistical and machine learning methodologies. In a recent collaborative study using this exact same 62.4 fb^{-1} dataset—to which I actively contributed—the observed (expected) upper limit was significantly improved to 12.7 (18.6) times the SM prediction [161].

This substantial gain in sensitivity is driven by two main methodological advancements over the baseline framework of this thesis:

- **Two-Stage Machine Learning Architecture:** Rather than relying on a single classifier, the advanced study implements a sophisticated, two-staged machine learning model optimized for signal extraction alongside an ML-based multi-classification network.
- **Multi-Category Simultaneous Fit:** Instead of isolating a single signal-dominated region, the advanced framework performs a simultaneous profile likelihood fit across multiple orthogonal categories. Yields from gluon-gluon fusion ($HH \text{ ggF}$), vector boson fusion ($HH \text{ VBF}$), $t\bar{t}$, single top ($t/\bar{t}$), Drell-Yan (DY), and single Higgs production categories all enter the fit concurrently. This allows the background-dominated control regions to constrain systematic nuisances directly, reducing the impact of background uncertainties on the signal extraction.

Remarkably, by deploying these sophisticated techniques, the expected limit of $18.6 \times \text{SM}$ surpasses not only the naive luminosity scaling prediction of 25.6 but also exceeds the sensitivity of the legacy Run-2 analysis ($19 \times \text{SM}$), despite utilizing less than half the integrated luminosity. This comparison underscores a critical thesis conclusion: advanced multi-classification and global fitting strategies can effectively compensate for luminosity limitations in early data-taking phases.

Additionally, the work in this thesis was done during the same timeline of the Run 3 publication in which I personally also participated. A central part of my contribution was the development of a transformer-based ML model as an alternative of the main DNN approach which was very similar to the transformer-based model explained in Section 3.3.2. A second major task was the derivation of scale factors for the boosted $H \rightarrow b\bar{b}$ tagging algorithm (ParticleNet). I worked on extracting the SFs using the μ -tagged method described in Section 4.2.7. My work involved computing the multi-dimensional reweighting maps to correct the simulation kinematics and extracting the final scale factors in bins of jet p_T for both b and c flavors, as shown in Figure A.28.

4.6

HH production in future searches

The search for Higgs boson pair production is entering a critical phase, and the HH prospects studies form one of the central outcomes of this section. I worked extensively on these projection studies, which connect the Run-3 analysis to the expected HL-LHC sensitivity and quantify the physics reach of the $b\bar{b}WW$ channel. While the standalone results presented in this thesis are statistically limited, the success of the multi-category ML framework lays a clear foundation for precision measurements in the coming years. The Run-3 data-taking campaign has already delivered an integrated luminosity exceeding 300 fb^{-1} , providing a statistical power surpassing that of the entire Run-2 era. Beyond the simple accumulation of data, several strategic pathways exist to further enhance the sensitivity of the $b\bar{b}WW$ channel.

First, the analysis sensitivity will benefit significantly from improved object reconstruction and calibration. As shown in the previous chapter, the uncertainty associated with the jet energy scale is currently among the leading systematic limitations. As the understanding of the 13.6 TeV detector conditions matures, the significant uncertainties related to jet energy resolution and $t\bar{t}$ modeling are expected to decrease.

Second, the evolution of multivariate analysis techniques promises substantial gains. Moving beyond the standard BDT architectures and two-stage models discussed here, future iterations can exploit lower-level features, such as constituent-level information via Particle Transformers, to extract maximal kinematic information from the event topology. Such advanced architectures could also improve the resolution of the reconstructed Higgs boson masses, which is a key discriminator in non-resonant searches.

Finally, the $b\bar{b}WW$ channel will play an indispensable role in the global combination of HH searches. While the $b\bar{b}\gamma\gamma$ and $b\bar{b}b\bar{b}$ channels currently drive the sensitivity to the SM signal, the $b\bar{b}WW$ decay mode offers unique complementarity. Its distinct background composition and dependence on the W boson branching ratios make it particularly robust against QCD multijet backgrounds that dominate purely hadronic channels. Moreover, this channel is highly sensitive to the VBF production mode, making it a crucial component for constraining the $VVHH$ coupling. As the LHC progresses toward the High-Luminosity era, where dataset sizes are expected to reach 3000 fb^{-1} , the techniques developed during Run-3 will be useful in the transition from exclusion limits to the first observation of Higgs boson pair production.

To quantify this ultimate physics reach and inform the choice of future colliders at CERN, I contributed extensively to the comprehensive HL-LHC projection studies for the European Strategy for Particle Physics [172]. This collaborative

work combines statistical models across available HH final states to extrapolate sensitivities assuming 3 ab^{-1} of integrated luminosity. These studies demonstrate that the HL-LHC can achieve an observation of Standard Model di-Higgs production with a significance exceeding 7σ (Figure 4.10), and measure the Higgs boson trilinear self-coupling with a precision better than 30%. Furthermore, they provide foundational phenomenological constraints on generic Beyond the Standard Model scalar potentials, highlighting the power of future measurements in establishing the shape of the electroweak vacuum.

My specific contributions centered on compiling and validating the statistical datacards along with executing the combinations for the HL-LHC scenarios for all channels, and optimizing the systematic uncertainty treatments for the $b\bar{b}WW$ channel. The challenges addressed in the current Run-3 $b\bar{b}WW$ inference—such as controlling background shapes in the machine learning score tails—directly informed the systematic uncertainty floors built for these projections.

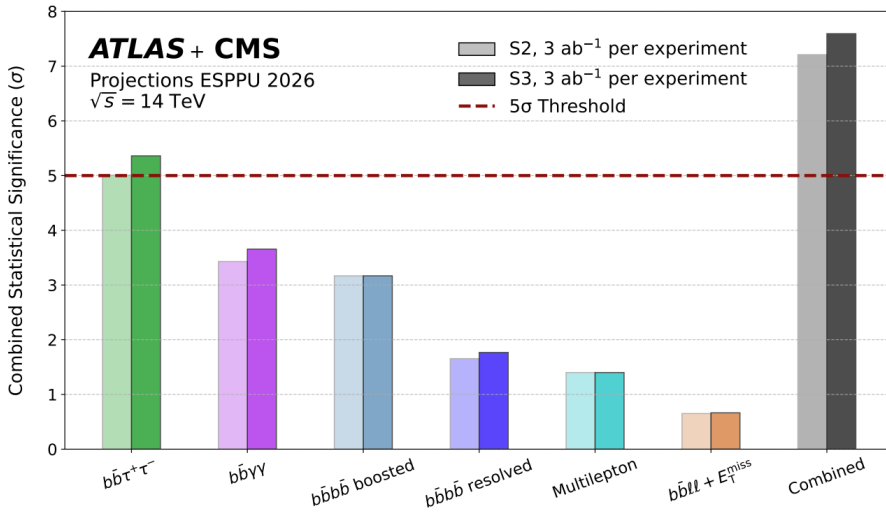


FIGURE 4.10 | Projected combined statistical significance for SM Higgs boson pair production in the ATLAS+CMS combination at the HL-LHC, shown separately for each analysis channel and for the full combination, assuming 3 ab^{-1} per experiment at $\sqrt{s} = 14$ TeV. The light and dark bars correspond to the S2 and S3 systematic uncertainty scenarios, respectively, while the dashed horizontal line indicates the 5 σ observation threshold [172]. Scenario S2 assumes reduced systematic uncertainties with halved theory uncertainties where feasible, negligible Monte Carlo statistical limitations, and object-dependent scaling of experimental uncertainties. Scenario S3 evaluates further analysis-specific improvements beyond S2, motivated by reconstruction or performance gains that have already been demonstrated. Both scenarios are based on the best currently available knowledge and should be considered conservative projections.

Conclusion

This thesis has presented a complete chain for the search for non-resonant Higgs boson pair production in the $b\bar{b}\ell\ell\nu\nu$ final state with early Run 3 CMS data. Starting from the theoretical motivation for Higgs self-interactions in Chapter 1, the work then connected these physics targets to the detector and reconstruction performance of CMS in Chapter 2, with a particular focus on tracking precision and its role in robust object reconstruction. A central pillar of this thesis is the machine-learning study in Chapter 3, where multiple architectures (Transformer, DNN, and BDT) were developed and compared not only in classification power but also in training stability, robustness, and computational cost. This ML work was directly integrated into the full statistical strategy in Chapter 4, from event categorization to profile-likelihood inference. Beyond the final limits themselves, a central outcome of this work is methodological: the analysis demonstrates how sensitivity in a low-signal, high-background channel can be improved through careful feature design, rigorous model comparison, and stable multivariate training coupled to a robust treatment of systematic uncertainties in a simultaneous fit. In particular, the ML comparison established a practical analysis strategy that balances discrimination and speed, enabling efficient iteration, cross-validation, and systematic robustness checks that are essential for reliable final templates. In parallel, detector-performance contributions—notably the strip hit-resolution study based on the pair method—supported the broader CMS tracker performance program and strengthened the detector-level understanding that underpins precision physics analyses.

Placed in a wider perspective, this thesis reflects the multidisciplinary nature of collider physics: progress depends equally on physics interpretation, advanced analysis tools, and sustained detector operations. The work reported here contributed to all three aspects, with a particularly strong emphasis on machine-learning-driven analysis development, alongside operational CMS responsibilities and statistical interpretation for Higgs pair searches.

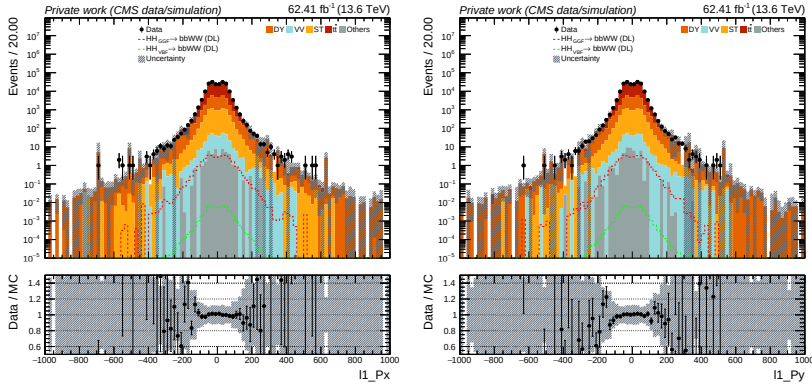
Looking forward, the outlook is strong. With the growing Run 3 dataset and the transition to HL-LHC conditions, the $b\bar{b}WW$ channel is expected to gain

substantially in sensitivity through larger data samples, improved calibrations, and next-generation ML strategies. In combination with other HH final states, these advances will be essential for moving from exclusion-level measurements toward precise constraints on the Higgs potential, and ultimately toward a direct characterization of the Higgs self-coupling.

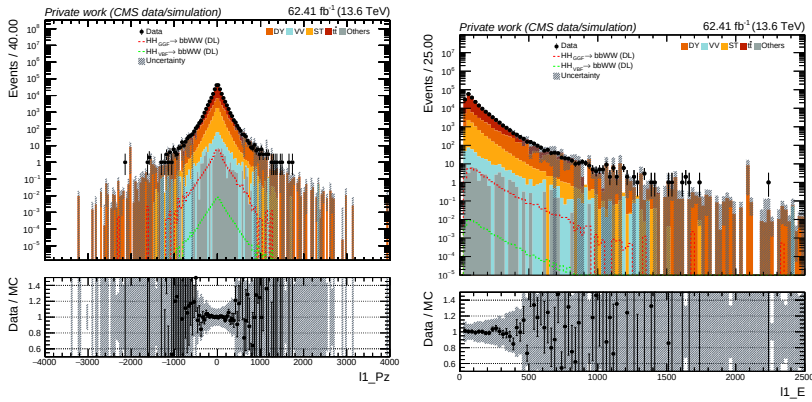
Additional Material

A.1

ML input features



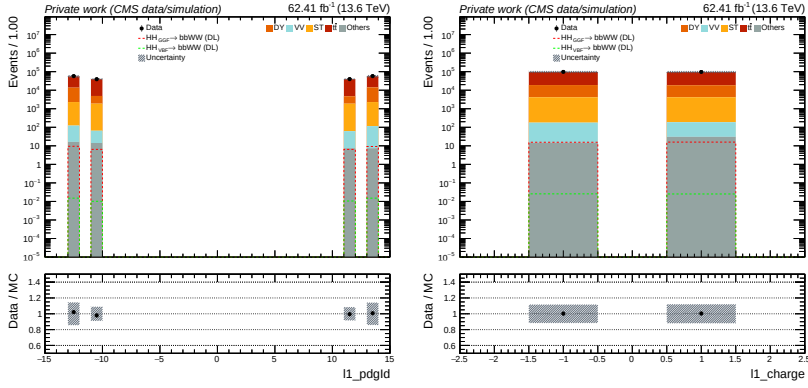
(A) X component of the leading lepton's (B) Y component of the leading lepton's

 p_T p_T 

(C) Z component of the leading lepton's

(D) Energy of the leading lepton

 p_T



(E) PDG ID of the leading lepton

(F) Charge of the leading lepton

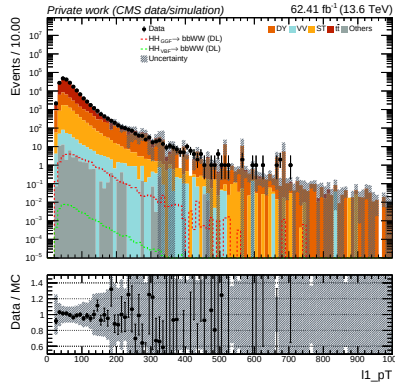
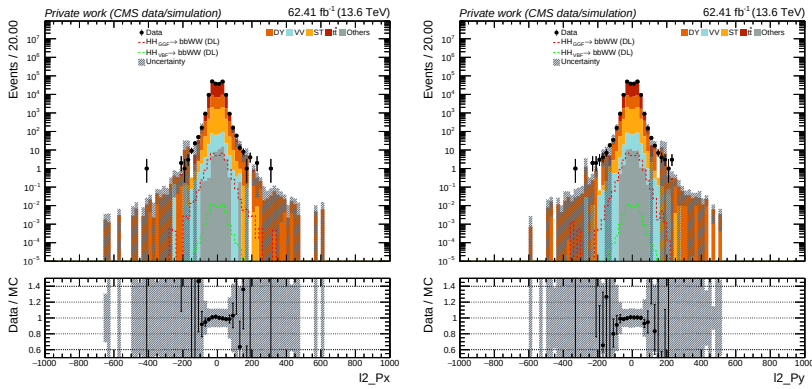
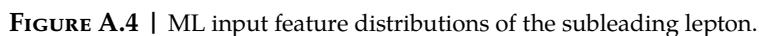
(G) p_T of the leading lepton

FIGURE A.2 | ML input feature distributions of the leading lepton.

(A) X component of the subleading lepton's p_T (B) Y component of the subleading lepton's p_T



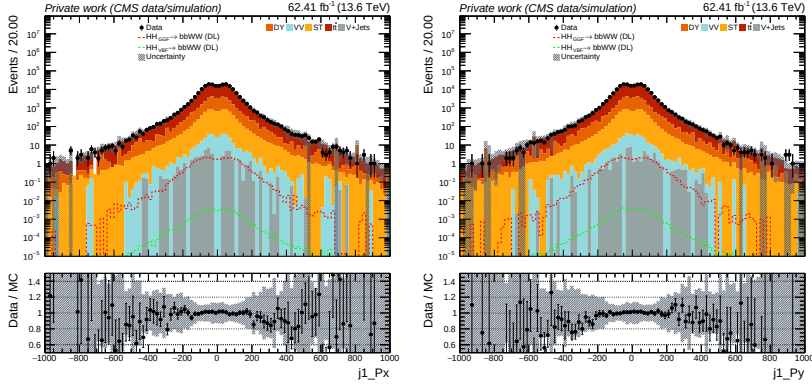
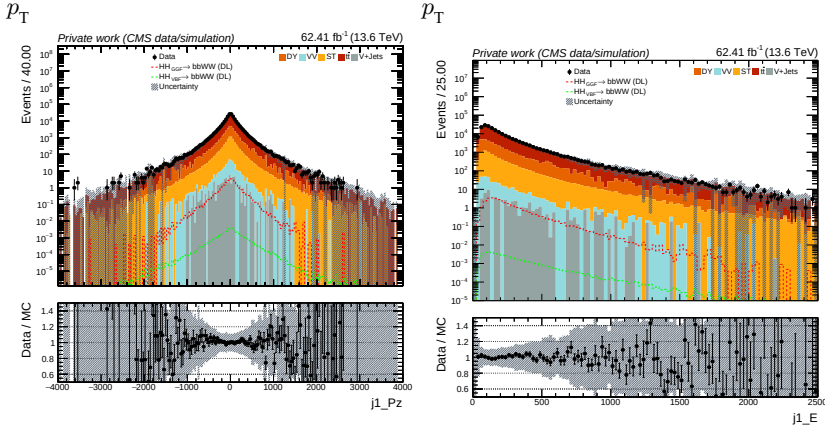
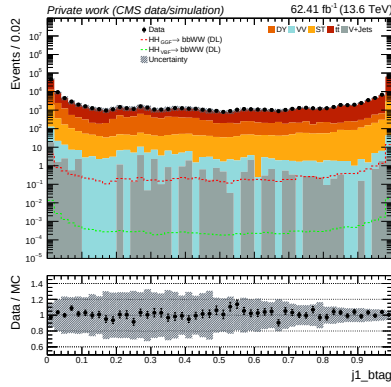
(A) X component of the 1st leading jet's (B) Y component of the 1st leading jet's(C) Z component of the 1st leading jet's (D) Energy of the 1st leading jet(E) b-tagging score of the 1st leading jet

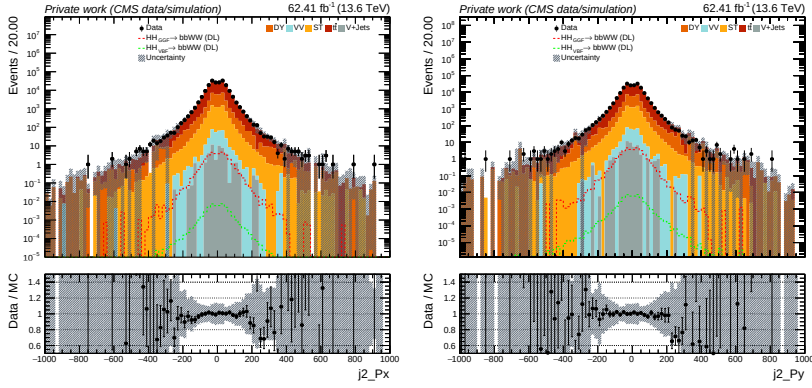
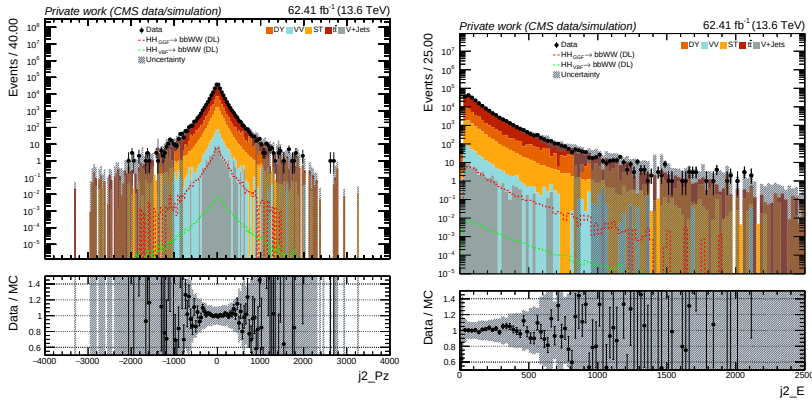
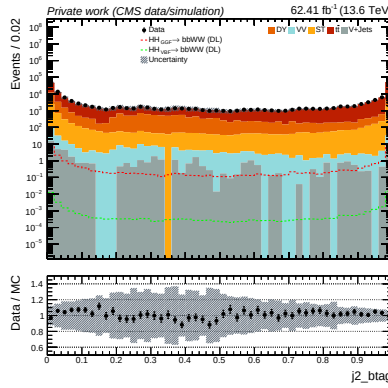
FIGURE A.6 | ML input feature distributions of the 1st leading jet.(A) X component of the 2nd leading jet's p_T (B) Y component of the 2nd leading jet's p_T (C) Z component of the 2nd leading jet's p_T (D) Energy of the 2nd leading jet(E) b-tagging score of the 2nd leading jet

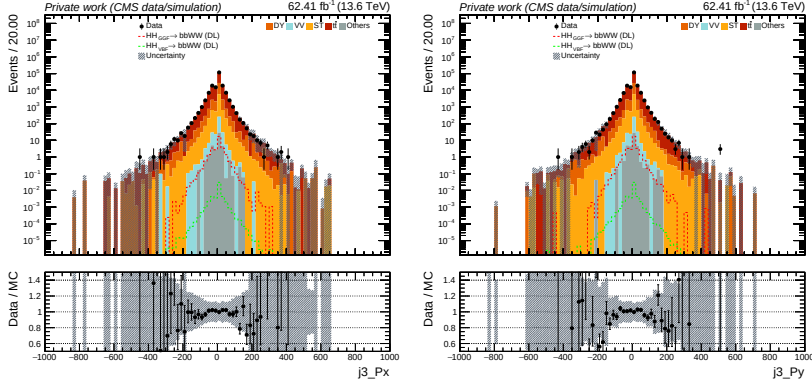
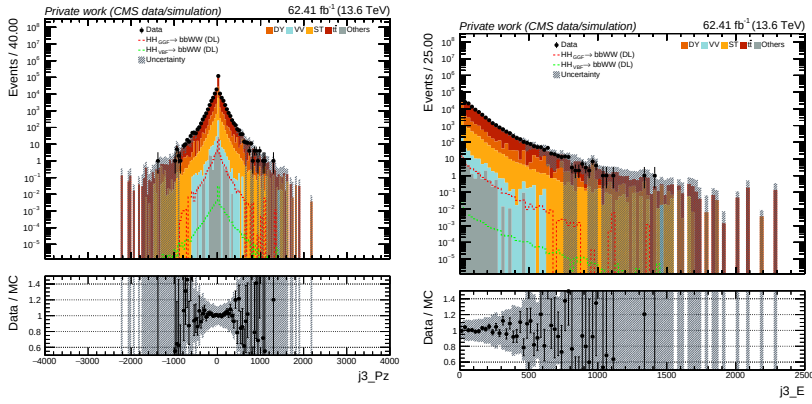
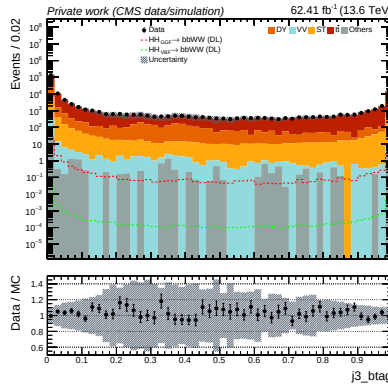
FIGURE A.8 | ML input feature distributions of the 2nd leading jet.**(A)** X component of the 3rd leading jet's **(B)** Y component of the 3rd leading jet's p_T p_T **(C)** Z component of the 3rd leading jet's**(D)** Energy of the 3rd leading jet p_T **(E)** b-tagging score of the 3rd leading jet

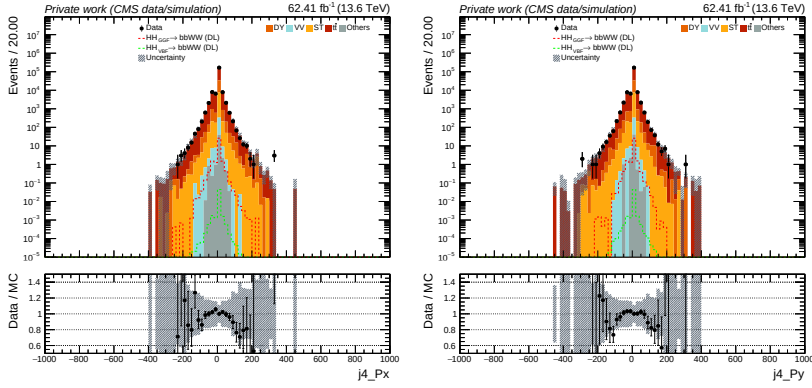
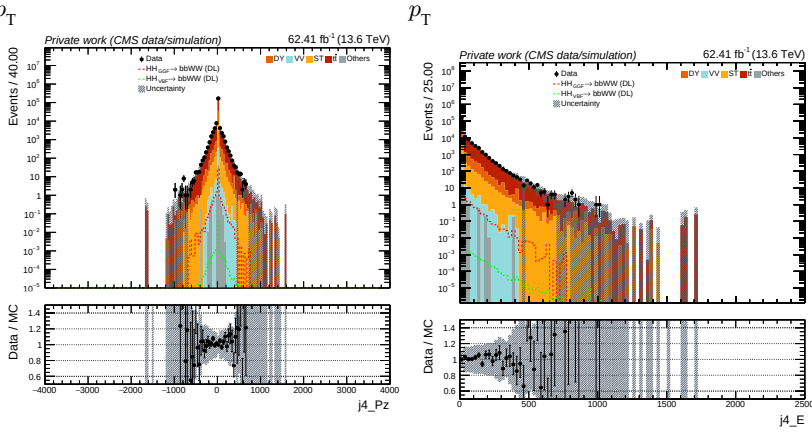
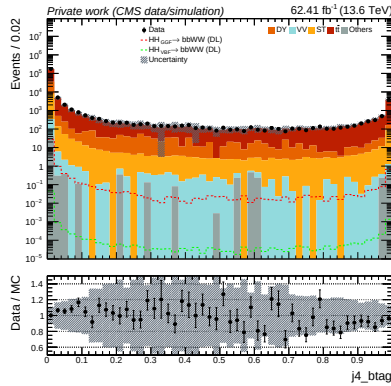
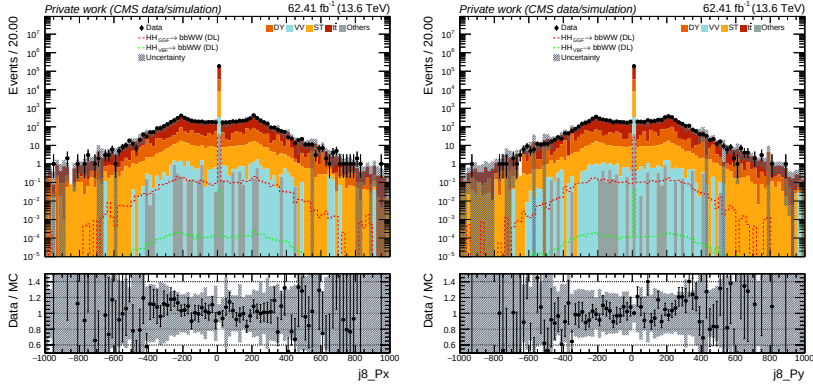
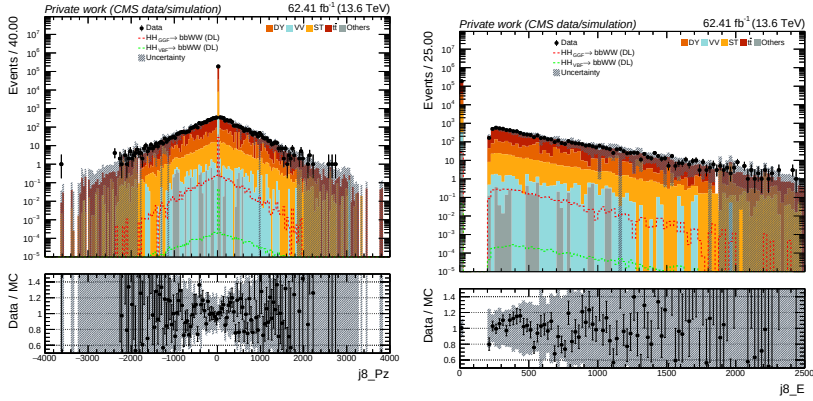
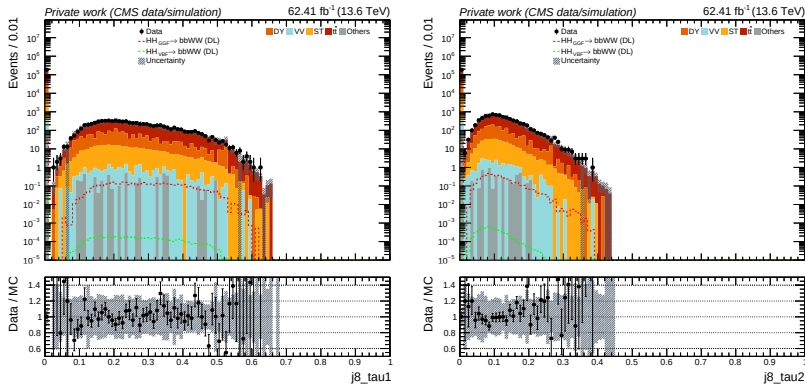
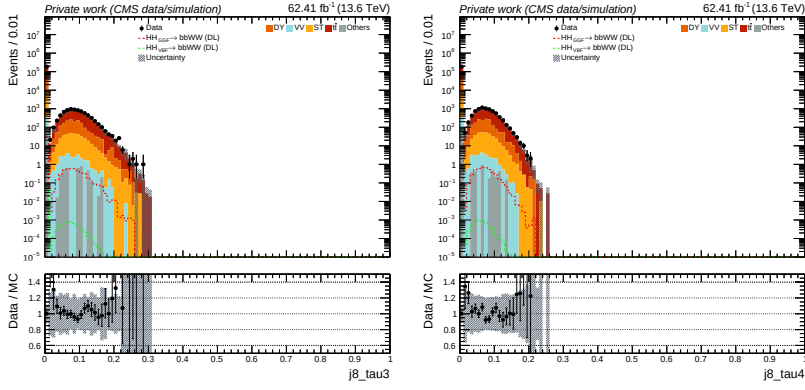
FIGURE A.10 | ML input feature distributions of the 3rd leading jet.(A) X component of the 4th leading jet's p_T (B) Y component of the 4th leading jet's p_T (C) Z component of the 4th leading jet's p_T (D) Energy of the 4th leading jet(E) b-tagging score of the 4th leading jet

FIGURE A.12 | ML input feature distributions of the 4th leading jet.(a) X component of the leading AK8 jet's p_T (b) Y component of the leading AK8 jet's p_T (c) Z component of the leading AK8 jet's p_T (d) Energy of the the leading AK8 jet

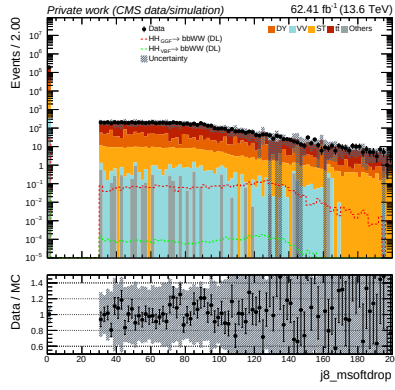
(e) tau 1 of the leading AK8 jet

(f) tau 2 of the leading AK8 jet



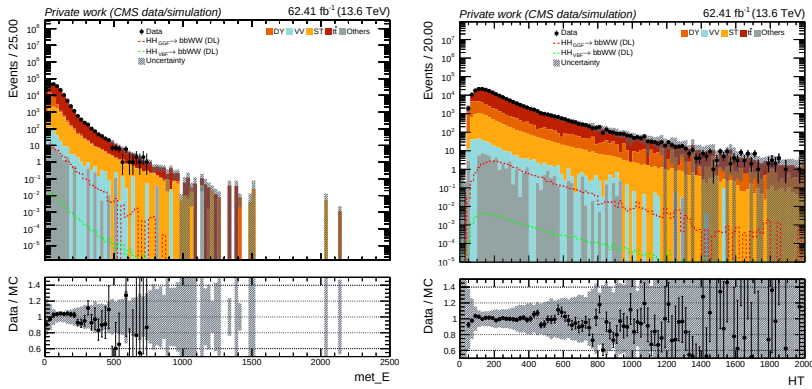
(G) tau 3 of the leading AK8 jet

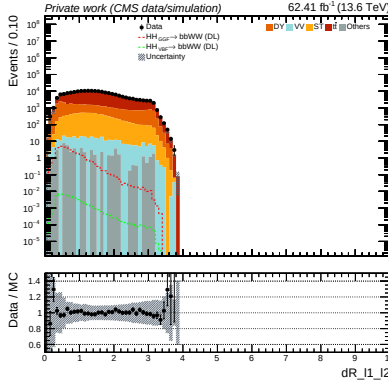
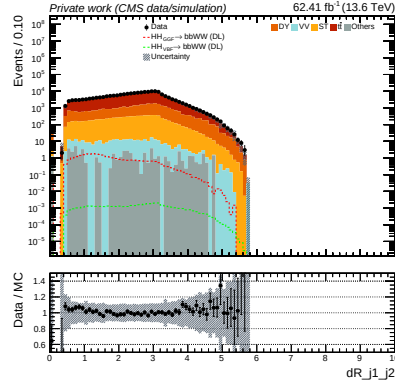
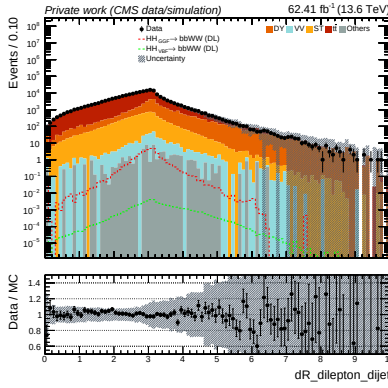
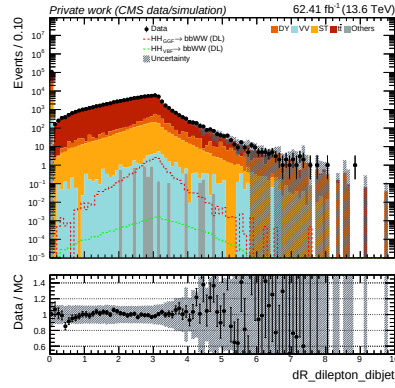
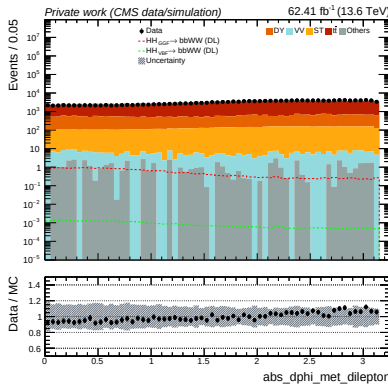
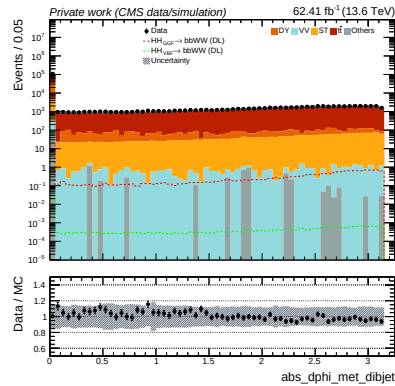
(H) tau 5 of the leading AK8 jet

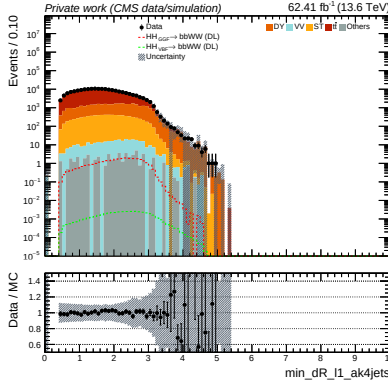
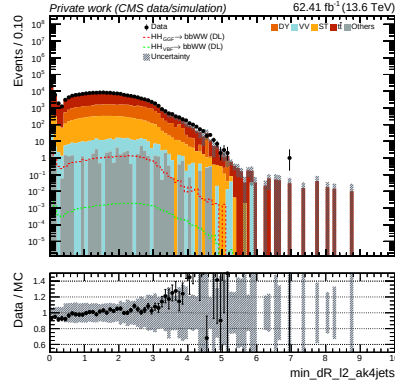
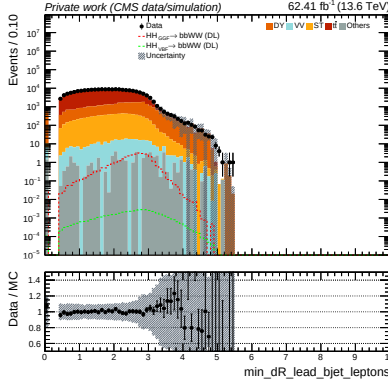
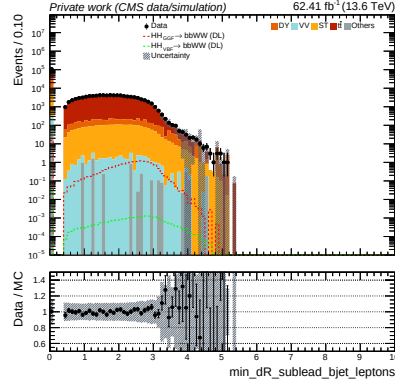
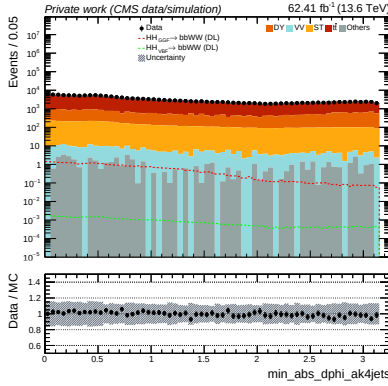
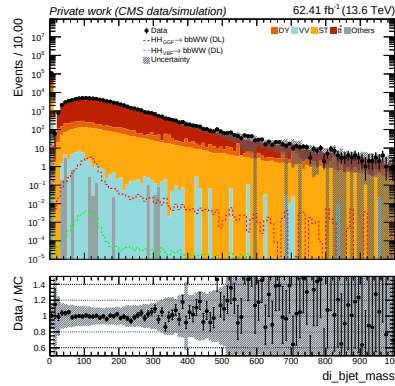


(I) Soft drop mass of the leading AK8 jet

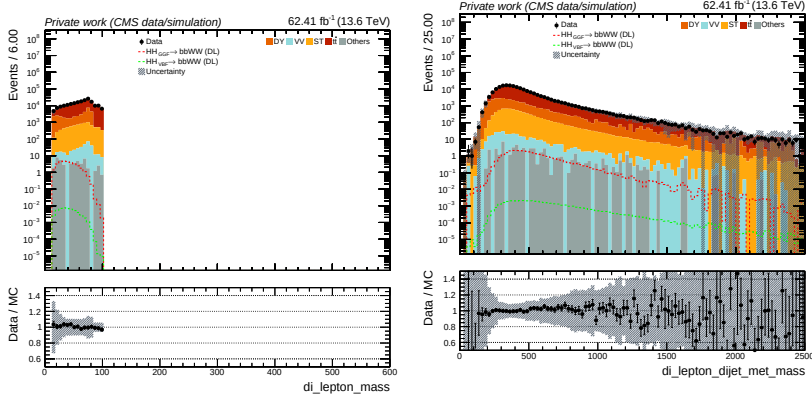
FIGURE A.14 | ML input feature distributions of the leading AK8 jet.

(A) E_T^{miss} (B) Scalar sum of the p_T of all jets

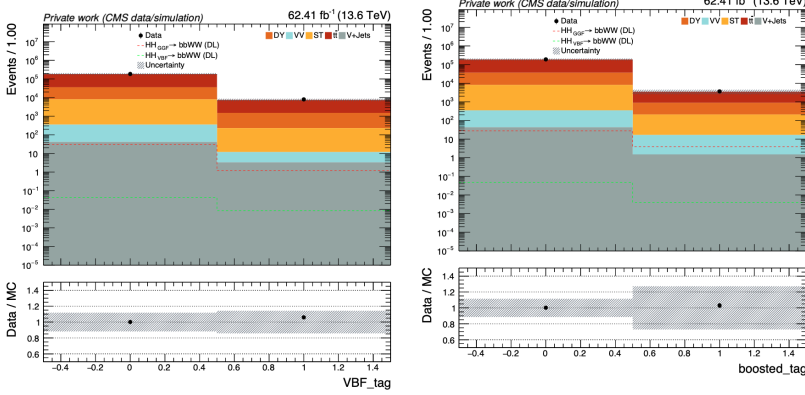
(c) ΔR between the two leptons(d) ΔR between the two leading jets(e) ΔR between the dilepton and dijet systems(f) ΔR between the dilepton and dibjet systems(g) $\Delta\phi$ between the E_T^{miss} and the dilepton system(h) $\Delta\phi$ between the E_T^{miss} and the dibjet system

(i) Minimum ΔR between the leading lepton and all AK4 jets(j) Minimum ΔR between the subleading lepton and all AK4 jets(k) Minimum ΔR between the leading b-tagged jet and both leptons(l) Minimum ΔR between the subleading b-tagged jet and both leptons(m) Minimum $\Delta\phi$ between all AK4 jets and the E_T^{miss} 

(n) Mass of the two leading b-tagged jet system

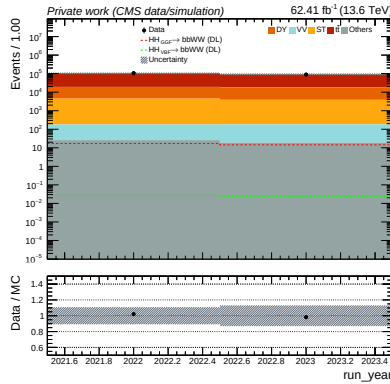


(o) Mass of the dilepton system

(p) Mass of the dilepton + dijet + E_T^{miss} system

(q) Binary VBF tagging variable

(r) Binary boosted topology tagging variable



(s) Run year of the event (2022 or 2023)

FIGURE A.16 | Various other ML input features.

A.2

ML performance plots

A.2.1

Feature importances and attention weights

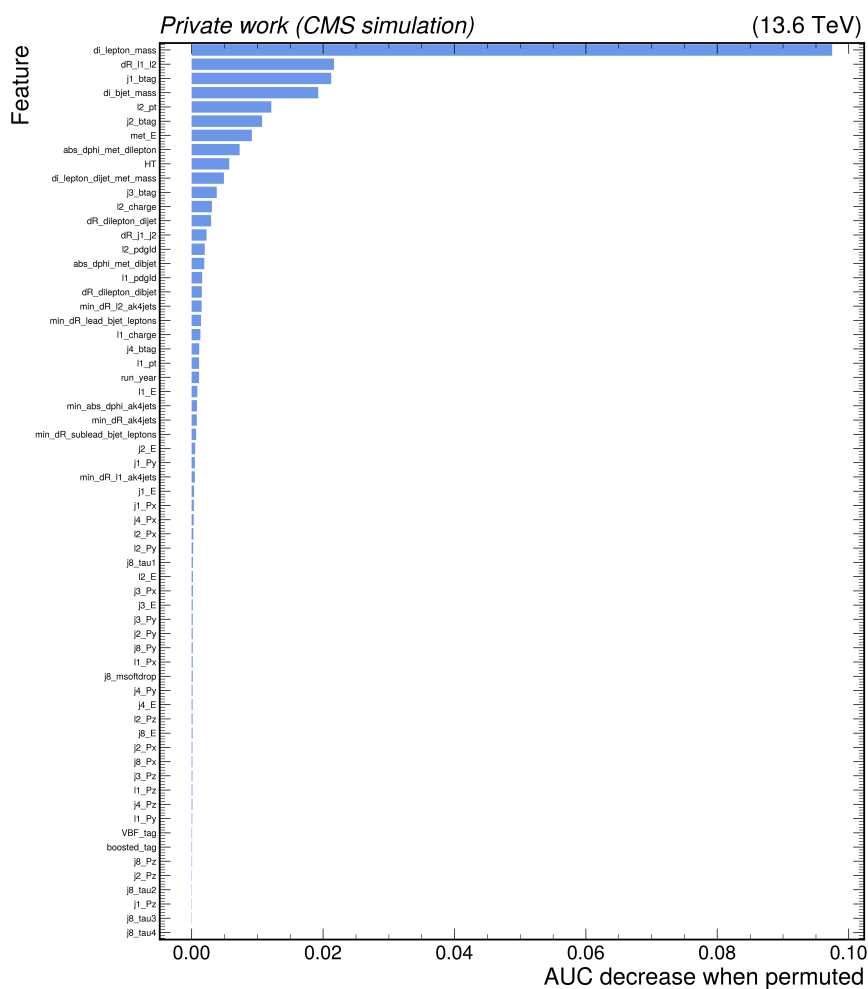


FIGURE A.17 | The full list of input features for the DNN classifier ranked by permutation importance.

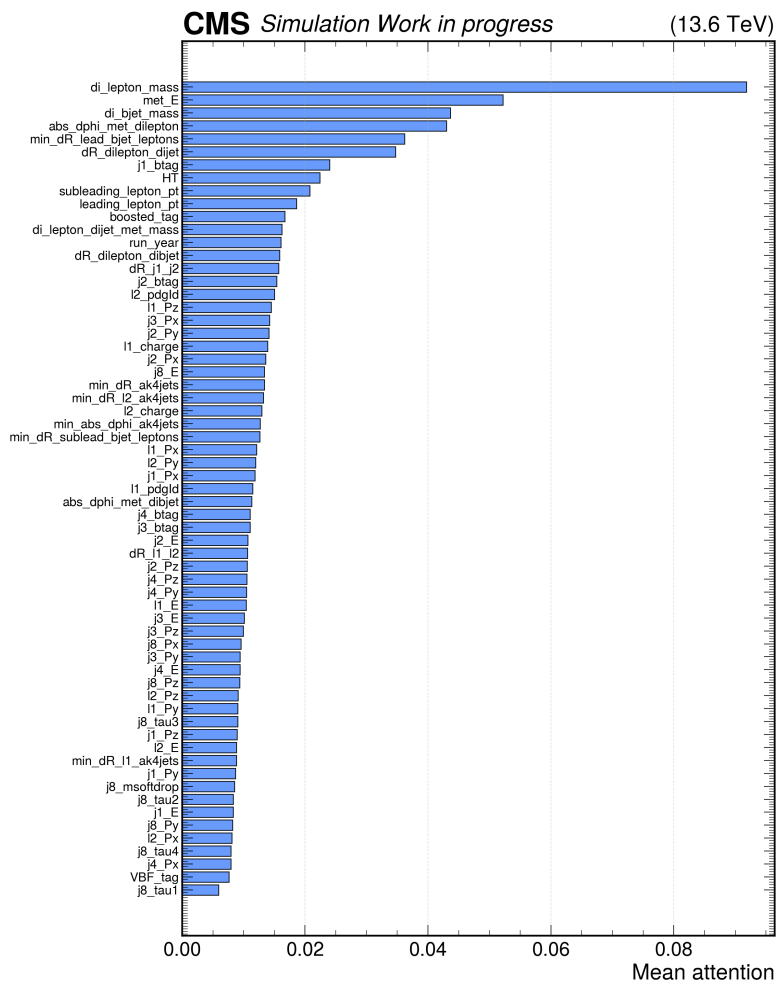


FIGURE A.18 | Mean attention scores assigned by the Transformer model to all input features.

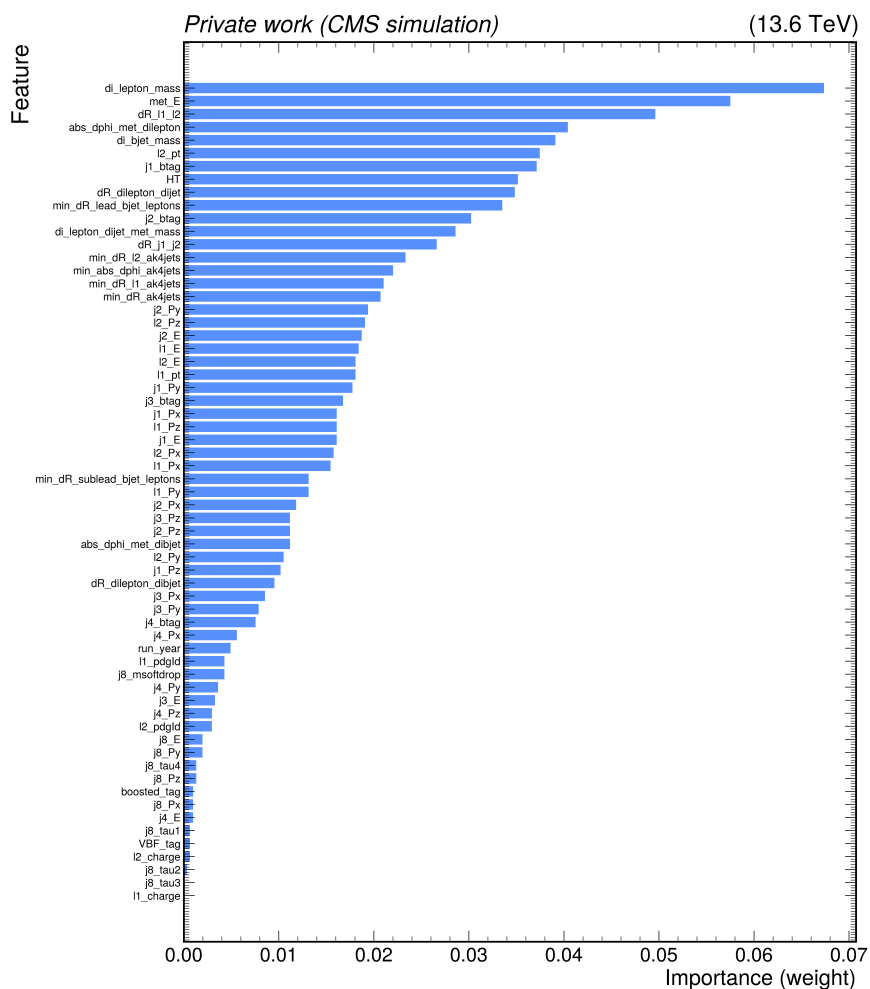
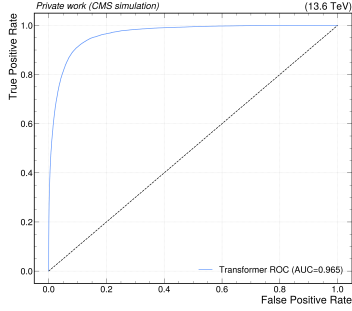


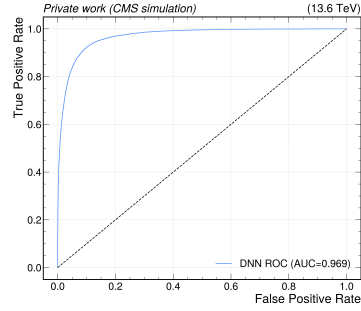
FIGURE A.19 | The full list of input features for the BDT model ranked by feature importance based on the number of times each feature is used to split the data across all trees in the ensemble.

A.2.2

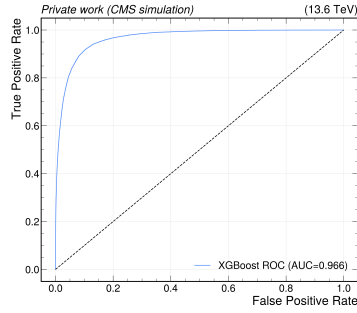
ROC plots



(A) Transformer model.



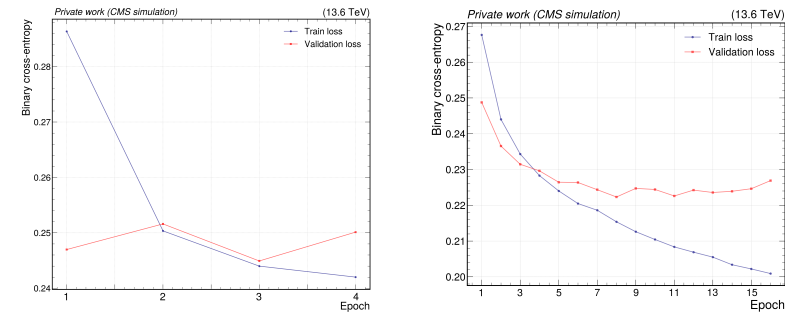
(B) DNN model.



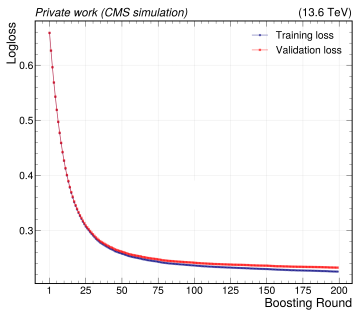
(C) BDT model.

FIGURE A.21 | ROC curve distributions of the models.

A.2.3 **Loss plots**



(A) Loss plot of the transformer model. **(B)** Loss plot of the DNN model.



(c) Loss plot of the BDT model.

FIGURE A.23 | ROC curve distributions of the models.

A.3

Event yields

Table 1: Event yields for various SR categories.

Category	Tot. Signal	TT	DY	ST	Others	Tot. MC	Data	Data/MC
ggF boos. ee	0.7 ± 0.0	33.7 ± 1.2	27.6 ± 4.8	3.3 ± 0.2	5.0 ± 0.5	69.7 ± 5.1	63	0.90 ± 0.13
ggF boos. $\mu\mu$	2.5 ± 0.0	122.9 ± 2.7	95.4 ± 21.6	11.0 ± 0.5	15.3 ± 0.9	244.5 ± 22.0	271	1.11 ± 0.12
ggF boos. $e\mu$	2.0 ± 0.0	189.1 ± 3.7	3.5 ± 0.6	17.8 ± 0.6	12.4 ± 1.0	222.8 ± 4.2	241	1.08 ± 0.07
ggF res. $1b\ ee$	3.9 ± 0.1	3903.9 ± 56.0	2499.8 ± 105.3	301.3 ± 4.8	105.5 ± 3.2	6810.6 ± 138.3	6232	0.92 ± 0.02
ggF res. $2b\ ee$	3.1 ± 0.1	3945.4 ± 56.5	321.4 ± 30.6	124.4 ± 2.3	24.8 ± 1.0	4415.9 ± 69.9	3980	0.90 ± 0.02
ggF res. $1b\ \mu\mu$	12.5 ± 0.2	14220.7 ± 200.4	10180.2 ± 251.1	1036.1 ± 15.1	330.8 ± 6.2	25768.2 ± 416.5	26479	1.03 ± 0.02
ggF res. $2b\ \mu\mu$	10.0 ± 0.1	14889.0 ± 209.7	1114.1 ± 67.1	415.5 ± 6.4	81.7 ± 2.1	16501.2 ± 241.1	17014	1.03 ± 0.02
ggF res. $1b\ e\mu$	12.9 ± 0.2	22592.4 ± 317.6	153.7 ± 6.3	1680.4 ± 24.1	301.0 ± 6.4	24737.9 ± 347.6	25022	1.01 ± 0.02
ggF res. $2b\ e\mu$	10.3 ± 0.2	23257.8 ± 326.8	20.0 ± 2.0	686.8 ± 12.0	85.5 ± 2.3	24050.2 ± 338.0	24096	1.00 ± 0.02
VBF res. ee	0.2 ± 0.0	74.0 ± 1.9	40.2 ± 7.4	1.6 ± 0.2	2.8 ± 0.4	118.6 ± 7.8	124	1.05 ± 0.12
VBF boos. ee	0.0 ± 0.0	14.4 ± 0.7	28.9 ± 10.2	0.6 ± 0.1	1.4 ± 0.3	45.3 ± 10.3	42	0.93 ± 0.25
VBF res. $\mu\mu$	0.6 ± 0.0	244.6 ± 4.5	165.8 ± 25.5	4.7 ± 0.3	8.9 ± 0.8	423.9 ± 56.2	491	1.16 ± 0.19
VBF boos. $\mu\mu$	0.1 ± 0.0	46.6 ± 1.4	82.4 ± 18.7	2.1 ± 0.2	3.4 ± 0.5	134.6 ± 18.8	153	1.14 ± 0.18
VBF res. $e\mu$	0.6 ± 0.0	393.5 ± 6.6	2.5 ± 0.6	8.1 ± 0.6	11.7 ± 1.0	415.5 ± 7.0	434	1.04 ± 0.05
VBF boos. $e\mu$	0.1 ± 0.0	78.6 ± 1.9	2.6 ± 0.4	2.9 ± 0.2	4.4 ± 0.6	88.5 ± 2.2	82	0.93 ± 0.10

FIGURE A.24 | Event yields for various SR categories.

A.4

Dilepton trigger scalefactors

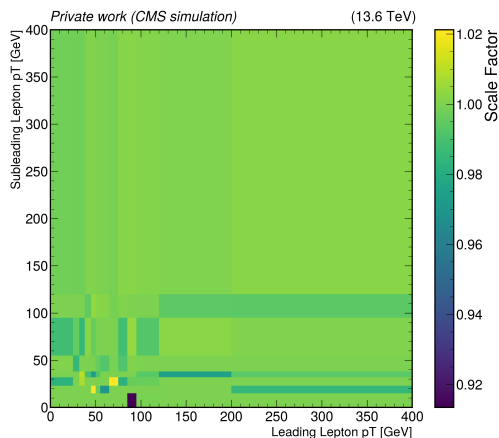


FIGURE A.25 | Di-lepton trigger scalefactors in di-electron channel.

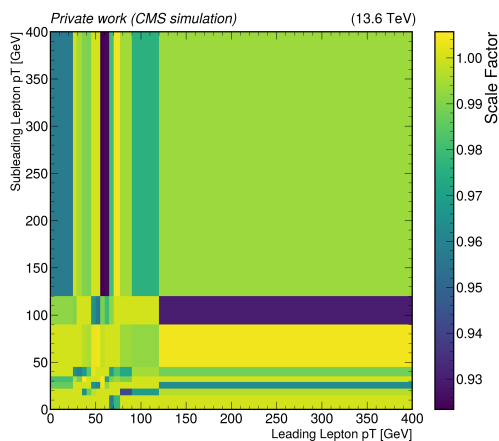


FIGURE A.26 | Di-lepton trigger scalefactors in di-muon channel.

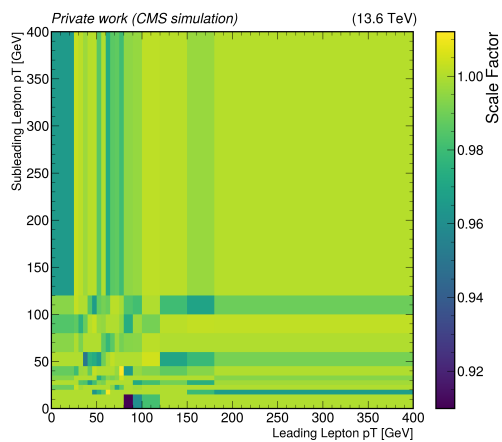


FIGURE A.27 | Di-lepton trigger scalefactors in mixed (electron-muon) channel.

A.5

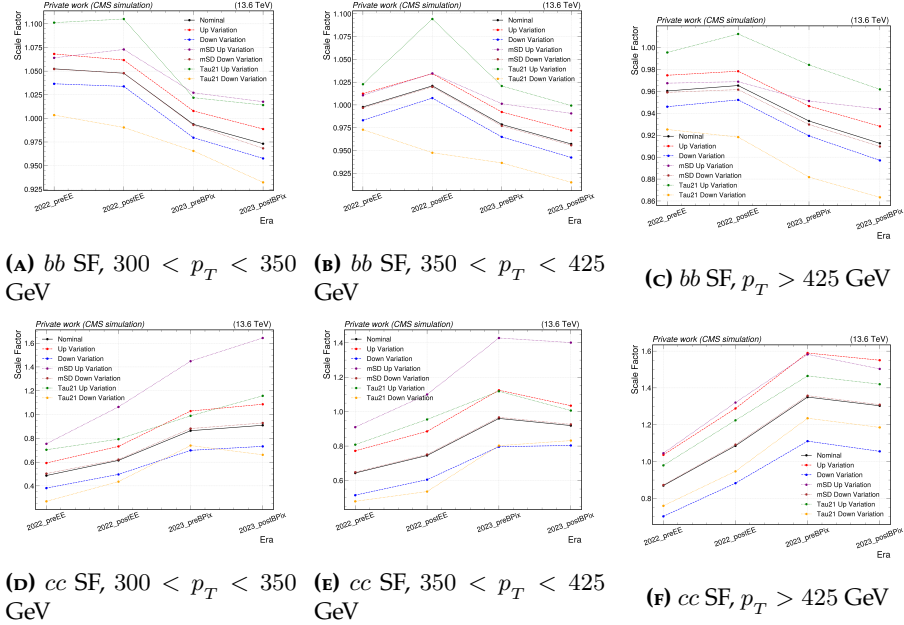
Boosted bb -jet tagging scalefactors

FIGURE A.28 | SFs for the ParticleNet Hbb discriminant derived using the μ -tagged method for bb -jets (top row) and cc -jets (bottom row). The SFs are measured in three different ranges of jet p_T : 300 – 350 GeV (left), 350 – 425 GeV (middle), and > 425 GeV (right). The measurements include robustness checks where the soft-drop mass requirement is varied to estimate systematic uncertainties.

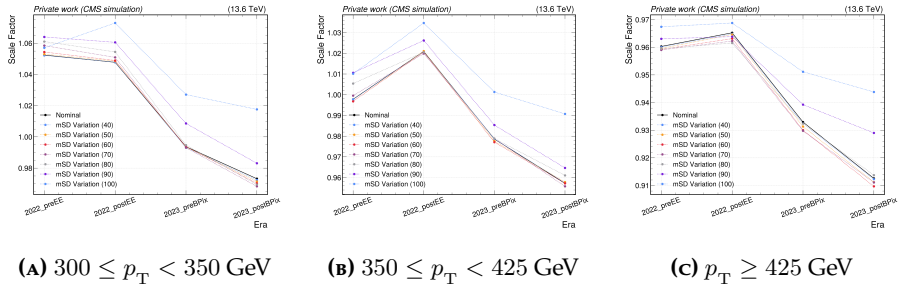


FIGURE A.29 | Boosted b -jet tagging scale factors derived using the μ -tagged method across the Run-3 data-taking eras. The plots show the nominal scale factor alongside variations derived by shifting the proxy jet soft-drop mass (m_{SD}) threshold from 40 to 100 GeV, evaluated in three separate jet p_T bins.

A.6

Inclusive cross-section uncertainties

TABLE A.1 | Inclusive cross-section uncertainties $\sigma(\mu_R/\mu_F)$ due to the renormalisation and factorization scales on the background processes. Uncertainties of processes separated by the horizontal lines are treated as correlated.

QCDscale_ $t\bar{t}$	
single-t (t-channel)	+1.1% / -0.8% [173]
single-t (s-channel)	+0.5% / -0.4% [173]
single-t (tW-channel)	+2.3% / -2.2% [173]
$t\bar{t}W$	+26.1% / -16.2% [174]
$t\bar{t}Z$	+8.6% / -9.5% [174]
$t\bar{t}t\bar{t}$	+8.2% / -17.5% [175]
QCDscale_V	
W + jets	+1.2% / -1.3% [176]
QCDscale_VV	
WW	+2.5% / -2.2% [177]
WZ	+4.1% / -3.2% [177]
ZZ	+2.9% / -2.7% [177]
QCDscale_ggH	
H_{ggF}	+4.6% / -6.7% [178]
QCDscale_qqH	
H_{VBF}	+0.5% / -0.3% [178]
QCDscale_VH	
WH	+0.4% / -0.7% [178]
ZH	+3.7% / -3.2% [178]
QCDscale_ttH	
$t\bar{t}H$	+6.0% / -9.3% [178]
tHq	+6.5% / -14.8% [178]
tHW	+5.0% / -6.8% [178]

TABLE A.2 | Inclusive cross-section uncertainties $\sigma(\text{PDF} + \alpha_s)$ due to the $\text{PDF} + \alpha_s$ on the background processes for different initial states. Uncertainties of processes separated by the horizontal lines are treated as correlated, motivated by the same dominant initial state.

<u>gq initiated</u>	
single-t (tW-channel)	$\pm 2.7\%$ [173]
<u>gg initiated</u>	
$t\bar{t}Z$	$\pm 3.5\%$ [179]
$t\bar{t}t\bar{t}$	$\pm 6.7\%$ [175]
<u>qq initiated</u>	
single-t (t-channel)	$+1.5\% / -0.9\%$ [173]
$W + \text{jets}$	$\pm 0.7\%$ [180]
$t\bar{t}W$	$\pm 2.1\%$ [174]
$t\bar{t}Z$	$\pm 2.3\%$ [174]
<u>single-H processes</u>	
H_{ggF}	$\pm 3.2\%$ [178]
H_{VBF}	$\pm 2.1\%$ [178]
WH	$\pm 1.8\%$ [178]
ZH	$\pm 1.6\%$ [178]
$t\bar{t}H$	$\pm 3.5\%$ [178]
tHq	$\pm 3.7\%$ [178]
tHW	$\pm 6.3\%$ [178]

A.6.1 Goodness-of-Fit tests

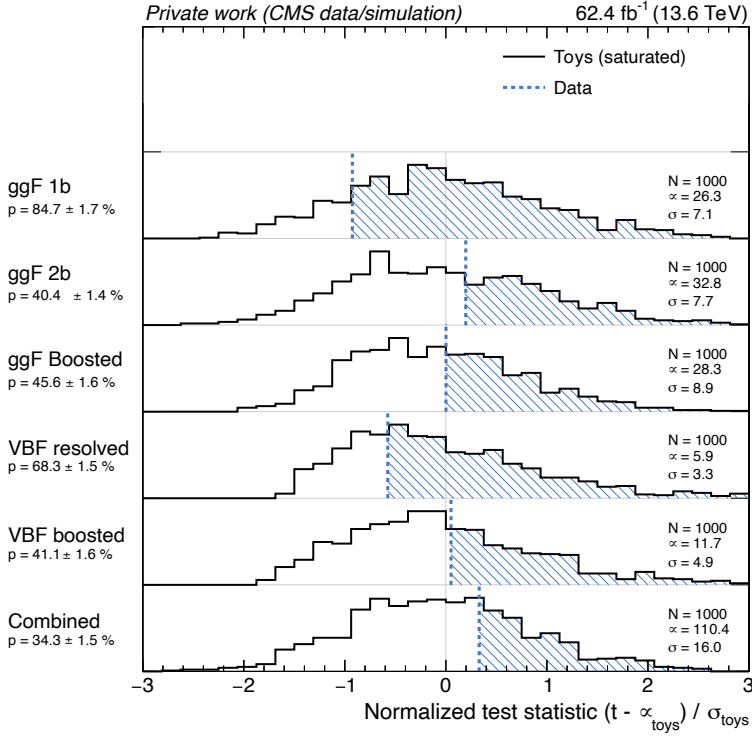


FIGURE A.30 | Goodness-of-Fit test results for individual topologies and the combined fit. The distributions represent the saturated test statistic calculated from $N = 1000$ pseudo-experiments (toys) for each category: ggF (1b, 2b, and Boosted) and VBF (resolved and boosted). The observed test statistic in data is indicated by the dashed blue line, with the shaded area representing the fraction of toys used to calculate the p -value. The normalized test statistic on the x-axis is defined as $(t - \langle t_{\text{toys}} \rangle) / \sigma_{\text{toys}}$. The global combined p -value of $34.3 \pm 1.5\%$ confirms that the simultaneous maximum likelihood fit provides a statistically sound description of the data across all analysis channels.

A.7 Likelihood scans

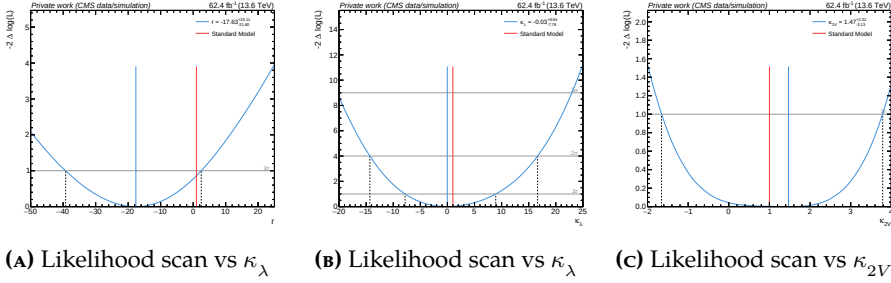


FIGURE A.31 | Observed (blue) and expected (orange) negative log-likelihood scans as a function of signal strength modifier (left), κ_λ (middle) and κ_{2V} (right). The dashed lines indicate the contribution from statistical uncertainties only, highlighting that the measurement is statistically limited.

Acronyms

AOD	Analysis object data 54	DNN	Deep neural network 37, 39, 40, 59, 61, 63, 70, 80
ATLAS	A Toroidal LHC ApparatuS 26, 35	DQM	Data quality monitoring 53, 54
AUC	Area under the ROC curve 66, 83, 87, 89, 90	DY	Drell-Yan 94, 96, 100, 112, 113
BDT	Boosted decision tree 37–40, 59, 63, 80, 100, 128	ECAL	Electromagnetic calorimeter 50, 99, 107–109
BEH	Brout-Englert-Higgs 16	EE	ECAL endcap 107, 114
BPix	Barrel Pixel 107, 114	EFT	Effective field theory 30, 32–35
BSM	Beyond the Standard Model 3, 30, 33, 34	eV	electron volt 4
BTX	b-tagging & Vertexing 110, 111, 115	EWSB	Electroweak symmetry breaking mechanism 16, 20
CERN	The European Organization for Nuclear Research 43	FPGA	Field Programmable Gate Array 72, 82
CKM	Cabibbo-Kobayashi-Maskawa (matrix) 20	FSR	Final state radiation 21, 117
CL	Confidence level 36	GAN	Generative adversarial network 71
CMS	Compact Muon Solenoid 19, 26, 35, 43, 93, 94, 105, 114	GeV	Giga electron volt 4
CNN	Convolutional neural network 61, 63, 71	ggF	Gluon-gluon fusion 26, 37, 72, 96, 124
CP	Charge-Parity 20, 32	GNN	Graph neural network 61, 71
CPU	Central processing unit 71	GPU	Graphics processing unit 71
CR	Control region 101, 103	HCAL	Hadronic calorimeter 51
CV	Cross-Validation 65	HEFT	Higgs effective field theory v, 33, 34
DAQ	Data acquisition system 53	HEP	High-energy physics 59, 60
DL	Di-leptonic vi, 59, 72, 78, 82, 83, 94	HF	HCAL Forward detector 51
		HH	Double Higgs boson 3, 4, 26, 27, 33–38, 40
		HHH	Triple Higgs boson 26

- HL-LHC** High-Luminosity LHC 40
- HLT** High-level trigger 53, 94
- ID** Identification 112
- ISR** Initial state radiation 117
- JEC** Jet energy corrections 113–115
- JER** Jet energy resolution 113, 115
- JERC** Jet Energy Resolution and Corrections POG 114
- JES** Jet energy scale 110, 113, 115
- JME** Jets and Missing Transverse Energy POG 114
- L1** Level-1 (trigger) 53
- LBN** Lorentz Boost Network 39
- LEP** Large Electron-Positron Collider 45
- LHC** Large Hadron Collider 21, 44, 107
- LO** Leading order 7, 22, 24, 28, 29, 34, 113
- LUM** Luminosity Physics Object Group 115
- mb** milibarn 108
- MC** Monte Carlo 22, 60, 105, 107–110, 112, 113, 115–117
- MeV** Mega electron volt 4
- MG5_aMC@NLO** MG5_aMC@NLO framework 24
- ML** Machine learning 24, 59
- MLP** Multi-layer perceptrons 61, 72, 79
- MSE** Mean Squared Error 69
- NLO** Next-to leading order 7, 24, 26, 29, 68, 113
- NNLO** Next-to-next-to leading order 7, 26
- NP** Nuisance parameter 114
- PAG** Physics Analysis Group 109
- PD** Primary dataset 96
- PDF** Parton distribution function 21, 27, 29, 116, 117
- PF** Particle flow 54
- PMNS** Pontecorvo-Maki-Nakagawa-Sakata (matrix) 20
- PNet** Particle Net 99, 100
- POG** Physics object groups 98, 110–112, 114, 115
- pp** Proton-proton 3, 26, 115
- PUPPI** Pileup Per Particle Identification 111
- QCD** Quantum Chromodynamics 15, 26, 36, 94, 99, 109, 113, 116, 121
- QED** Quantum Electrodynamics 13, 14
- QFT** Quantum Field Theory 3, 5
- ReLU** Rectified Linear Unit 64, 79
- RNN** Recurrent Neural Network 61, 63
- ROC** Receiver operating characteristic 66, 73, 90
- SF** Scalefactors 107, 109, 112, 113, 115
- SGD** Stochastic Gradient Descent 62
- SI** International System of Units 4
- SM** Standard Model 1, 3, 5, 7–11, 13–16, 18–20, 26, 27, 29, 30, 32–40, 60, 121, 123
- SMEFT** Standard Model effective field theory v, 33
- SPS** Super Proton Synchrotron 44, 46
- SR** Signal region 101, 103, 150
- T** tesla 45
- TnP** Tag-and-Probe 112

TOP	Top quark Physics Object Group 115	72, 116, 118, 123, 124
UE	Underlying event 22	
UV	Ultra-violet 7, 32	
VAE	Variational Autoencoder 71, 78	
VBF	Vector boson fusion 27, 37–39,	
VEV	Vacuum expectation value 16,	18, 26, 27
WP	Working point 54, 97	
XGBoost	Extreme Gradient Boosting 80	

References

- [1] G. Arnison and others (UA1 Collaboration), *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Physics Letters B* **122** (1983) 103.
- [2] G. Arnison and others (UA1 Collaboration), *Experimental Observation of Lepton Pairs of Invariant Mass around $95 \text{ GeV}/c^2$ at the CERN SPS Collider*, *Physics Letters B* **126** (1983) 398.
- [3] F. Abe and others (CDF Collaboration), *Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab*, *Physical Review Letters* **74** (1995) 2626.
- [4] S. Abachi and others (DØ Collaboration), *Observation of the Top Quark*, *Physical Review Letters* **74** (1995) 2632.
- [5] ATLAS Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, *Physics Letters B* **716** (2012) 1.
- [6] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Physics Letters B* **716** (2012) 30.
- [7] D.J. Griffiths, *Introduction to Elementary Particles*, Wiley-VCH, 2 ed. (2008).
- [8] M.E. Peskin and D.V. Schroeder, *An Introduction to Quantum Field Theory*, Addison-Wesley, Reading, Massachusetts (1995).
- [9] M.A. Srednicki, *Quantum field theory*, Cambridge University Press (2007).
- [10] K. Olive, *Review of Particle Physics*, *Chinese Physics C* **40** (2016) 100001.
- [11] E. Di Valentino, E. Giusarma, O. Mena, A. Melchiorri and J. Silk, *Cosmological limits on neutrino unknowns versus low redshift priors*, *Phys. Rev. D* **93** (2016) 083527.
- [12] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321.
- [13] P. Higgs, *Broken symmetries, massless particles and gauge fields*, *Physics Letters* **12** (1964) 132.
- [14] P.W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508.

- [15] C. Collaboration, *Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C* **79** (2019) .
- [16] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, *Progress of Theoretical Physics* **28** (1962) 870.
- [17] S. Höche, *Introduction to parton-shower event generators*, arXiv:1411.4085. <https://arxiv.org/abs/1411.4085>.
- [18] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer 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** (2014) .
- [19] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *Journal of High Energy Physics* **2007** (2007) 070–070.
- [20] W. Kilian, T. Ohl and J. Reuter, *WHIZARD—simulating multi-particle processes at LHC and ILC*, *The European Physical Journal C* **71** (2011) .
- [21] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al., *Event generation with SHERPA 1.1*, *Journal of High Energy Physics* **2009** (2009) 007–007.
- [22] O. Éboli, G. Marques, S. Novaes and A. Natale, *Twin Higgs-boson production*, *Physics Letters B* **197** (1987) 269.
- [23] E. Glover and J. van der Bij, *Higgs boson pair production via gluon fusion*, *Nuclear Physics B* **309** (1988) 282.
- [24] CERN, *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, 2017. doi:10.23731/CYRM-2017-002.
- [25] B.D. Micco, M. Gouzevitch, J. Mazzitelli and C. Vernieri, *Higgs boson potential at colliders: Status and perspectives*, *Reviews in Physics* **5** (2020) 100045.
- [26] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli et al., *Higgs pair production at the LHC with NLO and parton-shower effects*, *Physics Letters B* **732** (2014) 142.
- [27] T. Appelquist and J. Carazzone, *Infrared singularities and massive fields*, *Phys. Rev. D* **11** (1975) 2856.
- [28] S. Weinberg, *Baryon- and Lepton-Nonconserving Processes*, *Phys. Rev. Lett.* **43** (1979) 1566.
- [29] C.N. Leung, S.T. Love and S. Rao, *Low-energy manifestations of a new interactions scale: Operator analysis*, *Zeitschrift für Physik C Particles and*

- Fields* **31** (1986) 433–437.
- [30] W. Buchmüller and D. Wyler, *Effective lagrangian analysis of new interactions and flavour conservation*, *Nuclear Physics B* **268** (1986) 621.
 - [31] R. Contino, A. Falkowski, F. Goertz, C. Grojean and F. Riva, *On the validity of the effective field theory approach to SM precision tests*, *Journal of High Energy Physics* **2016** (2016) .
 - [32] D. Anselmi, *Renormalization of a class of non-renormalizable theories*, *Journal of High Energy Physics* **2005** (2005) 077–077.
 - [33] B. Grzadkowski, M. Iskrzyński, M. Misiak and J. Rosiek, *Dimension-six terms in the Standard Model Lagrangian*, *Journal of High Energy Physics* **2010** (2010) .
 - [34] G.F. Giudice, C. Grojean, A. Pomarol and R. Rattazzi, *The strongly-interacting light Higgs*, *Journal of High Energy Physics* **2007** (2007) 045–045.
 - [35] R. Contino, M. Ghezzi, C. Grojean, M. Mühlleitner and M. Spira, *Effective Lagrangian for a light Higgs-like scalar*, *Journal of High Energy Physics* **2013** (2013) .
 - [36] A. Azatov, R. Contino, G. Panico and M. Son, *Effective field theory analysis of double Higgs boson production via gluon fusion*, *Physical Review D* **92** (2015) .
 - [37] D. de Florian, I. Fabre and J. Mazzitelli, *Higgs boson pair production at NNLO in QCD including dimension 6 operators*, *Journal of High Energy Physics* **2017** (2017) .
 - [38] ATLAS Collaboration, *Search for nonresonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **108** (2023) 052003.
 - [39] CMS Collaboration, *Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **129** (2022) 081802.
 - [40] A. Collaboration, *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector*, *Journal of High Energy Physics* **2023** (2023) .
 - [41] CMS Collaboration, *Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s}=13$ TeV*, *Physics Letters B* **842** (2023) 137531.
 - [42] ATLAS Collaboration, *Study of Higgs boson pair production in the $HH \rightarrow b\bar{b}\gamma\gamma$ final state with 308 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV and 13.6 TeV by the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2025-005, CERN, Geneva (2025).

- [43] CMS Collaboration, *Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Journal of High Energy Physics* **2021** (2021) .
- [44] ATLAS Collaboration, *Combination of searches for Higgs boson pair production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS Conference Note ATLAS-CONF-2024-006, CERN, Geneva (May, 2024).
- [45] ATLAS Collaboration, *Search for non-resonant Higgs boson pair production in final states with leptons, taus, and photons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Journal of High Energy Physics* **2024** (2024) 164 [2405.20040].
- [46] CMS Collaboration, *Search for Higgs boson pair production in the $b\bar{b}W^+W^-$ decay mode in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Journal of High Energy Physics* **2024** (2024) 293 [2403.09430].
- [47] CMS Collaboration, *Search for HH production in the $b\bar{b}WW$ decay mode*, CMS Physics Analysis Summary CMS-PAS-HIG-21-005, CERN, Geneva (Mar, 2023).
- [48] CMS Collaboration, *Search for nonresonant Higgs boson pair production in the four leptons plus two b jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Journal of High Energy Physics* **2023** (2023) 130 [2206.10657].
- [49] ATLAS Collaboration, *Combination of searches for Higgs boson pair production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Physical Review Letters* **133** (2024) 101801 [2406.09971].
- [50] CMS Collaboration, *Combination of searches for nonresonant Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-HIG-20-011, CERN, Geneva (Nov, 2024).
- [51] CMS Collaboration, *Combination of ATLAS and CMS searches for Higgs boson pair production at 13 TeV*, Tech. Rep. CMS-PAS-HIG-25-014, CERN, Geneva (2025).
- [52] M. Brice, *Aerial view of the Large Hadron Collider (LHC) geography*, 2021. CERN Document Server.
<https://www.forbes.com/sites/startswithabang/2021/09/07/the-3-reasons-why-cerns-large-hadron-collider-cant-make-particles-go-faster/>.
- [53] E. Lopienska, "The CERN accelerator complex, layout in 2022."
<https://cds.cern.ch/record/2800984>, 2022. General Photo.
- [54] L. Rossi, *The longest journey: the LHC dipoles arrive on time*, CERN Courier (2006) . <https://cerncourier.com/a/the-longest-journey-the-lhc-dipoles-arrive-on-time/>.

- [55] L. Rossi, *Superconducting magnets for the LHC main lattice*, *IEEE Trans. Appl. Supercond.* **14** (2004) 153.
- [56] P. Collier, *The technical challenges of the Large Hadron Collider*, *Phil. Trans. Roy. Soc. Lond. A* **373** (2014) 20140044.
- [57] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** (2008) S08003–S08003.
- [58] T. Sakuma, *Cutaway diagrams of CMS detector*, .
<https://cds.cern.ch/record/2665537>.
- [59] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *Journal of Instrumentation* **12** (2017) P10003–P10003.
- [60] CMS Collaboration, *CMS Technical Design Report for the Pixel Detector Upgrade*, Tech. Rep. CERN-LHCC-2012-016; CMS-TDR-11, CERN, Geneva (2012).
- [61] CMS Collaboration, *CMS Silicon Strip Tracker Performance in 2023*, Tech. Rep. CMS-DP-2023-040, CERN, Geneva (2023).
- [62] CMS Collaboration, *CMS Silicon Strip Tracker Performance Results with 2024 Data*, Tech. Rep. CMS-DP-2025-020, CERN, Geneva (2025).
- [63] CMS Collaboration, *The CMS experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08004–S08004.
- [64] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *Journal of Instrumentation* **12** (2017) P10003–P10003.
- [65] CMS Collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JINST* **13** (2018) P06015.
- [66] CMS Collaboration, *Muon ID and Isolation efficiencies with muons in proton-proton collisions at $\sqrt{s} = 13.6$ TeV*, .
<https://cds.cern.ch/record/2904703>.
- [67] CMS Collaboration, *Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC*, *Journal of Instrumentation* **16** (2021) P05014.
- [68] M. Cacciari, G.P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *J. High Energy Phys.* **2008** (2008) 063.
- [69] A.a. Benecke, *Jet performance and pileup mitigation in Run3 in CMS*, in *Proceedings of 42nd International Conference on High Energy Physics — PoS(ICHEP2024)*, ICHEP2024, p. 893, Sissa Medialab, Dec., 2024, DOI.
- [70] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *Journal of High Energy Physics* **2014** (2014) .

- [71] C. Collaboration, *Performance of τ -lepton reconstruction and identification in CMS*, *Journal of Instrumentation* **7** (2012) P01001–P01001.
- [72] CMS Collaboration, *Identification of hadronic tau lepton decays using a deep neural network*, *Journal of Instrumentation* **17** (2022) P07023.
- [73] CMS Collaboration, *Identification of tau leptons using a convolutional neural network with domain adaptation*, *Journal of Instrumentation* **20** (2025) P12032.
- [74] CMS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector*, *JINST* **14** (2019) P07004.
- [75] CMS Collaboration, *Improving missing transverse momentum estimation with a deep neural network*, *Physical Review D* **113** (2026) .
- [76] CMS Collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *Journal of Instrumentation* **13** (2018) P05011–P05011.
- [77] E. Bols, J. Kieseler, M. Verzett, M. Stoye and A. Stakia, *Jet flavour classification using DeepJet*, *Journal of Instrumentation* **15** (2020) P12012–P12012.
- [78] H. Qu and L. Gouskos, *Jet tagging via particle clouds*, *Physical Review D* **101** (2020) .
- [79] CMS Collaboration, *Run 3 commissioning results of heavy-flavor jet tagging at $\sqrt{s} = 13.6$ TeV with CMS data using a modern framework for data processing*, Tech. Rep. CMS-DP-2024-024, CERN, Geneva (2024).
- [80] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup per particle identification*, *Journal of High Energy Physics* **2014** (2014) .
- [81] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York (2006).
- [82] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press (2016).
- [83] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *Journal of High Energy Physics* **2016** (2016) .
- [84] M. Andrews, M. Paulini, S. Gleyzer and B. Pocz, *End-to-End Physics Event Classification with CMS Open Data: Applying Image-Based Deep Learning to Detector Data for the Direct Classification of Collision Events at the LHC*, *Computing and Software for Big Science* **4** (2020) .
- [85] A.S. Cornell, W. Doorsamy, B. Fuks, G. Harmsen and L. Mason, *Boosted decision trees in the era of new physics: a smuon analysis case study*, *Journal of High Energy Physics* **2022** (2022) .

- [86] P. Baldi, P. Sadowski and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, *Nature Communications* **5** (2014) .
- [87] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban and D. Whiteson, *Jet flavor classification in high-energy physics with deep neural networks*, *Phys. Rev. D* **94** (2016) 112002.
- [88] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (2015) 436–444.
- [89] J. Cogan, M. Kagan, L. Mackey and B. Nachman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **01** (2015) 059.
- [90] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy flow networks: deep sets for particle jets*, *Journal of High Energy Physics* **2019** (2019) .
- [91] ATLAS Collaboration, *Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, Tech. Rep. ATL-PHYS-PUB-2017-003, CERN, Geneva (2017).
- [92] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-aware recursive neural networks for jet physics*, *Journal of High Energy Physics* **2019** (2019) .
- [93] D. Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, eds., vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 1530–1538, PMLR, 07–09 Jul, 2015, <https://proceedings.mlr.press/v37/rezende15.html>.
- [94] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn, A. Rousselot, R. Winterhalder et al., *Invertible networks or partons to detector and back again*, *SciPost Physics* **9** (2020) .
- [95] CMS Collaboration, *A unified approach for jet tagging in Run 3 at $\sqrt{s}=13.6$ TeV in CMS*, Tech. Rep. CMS-DP-2024-066, CERN, Geneva (2024).
- [96] H. Qu, C. Li and S. Qian, *Particle Transformer for Jet Tagging*, .
- [97] L. Builtjes, S. Caron, P. Moskvitina, C. Nellist, R.R. de Austri, R. Verheyen et al., *Attention to the strengths of physical interactions: Transformer and graph-based event classification for particle physics experiments*, *SciPost Phys.* **19** (2025) 028.
- [98] A. Hallin, *Foundation models for high-energy physics*, in *2nd European AI for Fundamental Physics Conference*, 9, 2025 [2509 . 21434].
- [99] Y. Wang, Z. Pan, J. Zheng, L. Qian and M. Li, *A hybrid ensemble method for pulsar candidate classification*, *Astrophysics and Space Science* **364** (2019) .
- [100] “Difference Between Feed-Forward Neural Networks and Recurrent Neural Networks.” Accessed March 20, 2026.

- <https://www.geeksforgeeks.org/data-analysis/difference-between-feed-forward-neural-networks-and-recurrent-neural-networks/>.
- [101] “Basic CNN architecture.” Accessed March 20, 2026.
<https://www.upgrad.com/blog/basic-cnn-architecture/>.
- [102] S. Ruder, *An overview of gradient descent optimization algorithms*, 2017.
<https://arxiv.org/abs/1609.04747>.
- [103] “Overfitting.” Accessed March 20, 2026.
<https://nl.mathworks.com/discovery/overfitting.html>.
- [104] K. Albertsson, P. Altoe, D. Anderson, M. Andrews, J.P. Araque Espinosa, A. Aurisano et al., *Machine Learning in High Energy Physics Community White Paper, Journal of Physics: Conference Series* **1085** (2018) 022008.
- [105] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are universal approximators, Neural Networks* **2** (1989) 359–366.
- [106] D. Guest, K. Cranmer and D. Whiteson, *Deep Learning and Its Application to LHC Physics, Annual Review of Nuclear and Particle Science* **68** (2018) 161–181.
- [107] M. Paganini, L. de Oliveira and B. Nachman, *CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks, Physical Review D* **97** (2018) .
- [108] A. Butter, T. Plehn and R. Winterhalder, *How to GAN LHC events, SciPost Physics* **7** (2019) .
- [109] J. Duarte, S. Han, P. Harris, S. Jindariani, E. Kreinar, B. Kreis et al., *Fast inference of deep neural networks in FPGAs for particle physics, Journal of Instrumentation* **13** (2018) P07027–P07027.
- [110] G. Kasieczka, B. Nachman, D. Shih, O. Amram, A. Andreassen, K. Benkendorfer et al., *The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics, Reports on Progress in Physics* **84** (2021) 124201.
- [111] O. Güzel, *DNN for HH-> bbWW Run3 Analysis at the CMS Experiment at CERN*, 2026. doi:10.5281/ZENODO.19109383.
- [112] C. Guo, G. Pleiss, Y. Sun and K.Q. Weinberger, *On calibration of modern neural networks*, in *International conference on machine learning*, pp. 1321–1330, PMLR, 2017.
- [113] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez et al., *Attention Is All You Need, CoRR* **abs/1706.03762** (2017) [1706.03762]. <http://arxiv.org/abs/1706.03762>.
- [114] K. Choromanski, V. Likhoshesterov, D. Dohan, X. Song, A. Kane, T. Sarlos et al., *Rethinking Attention with Performers*, 2022.

- <https://arxiv.org/abs/2009.14794>.
- [115] H. Qu, C. Li and S. Qian, *Particle Transformer for Jet Tagging*, in *Proceedings of the 39th International Conference on Machine Learning*, pp. 18281–18292, 2022 [2202.03772].
- [116] J. Birk, A. Hallin and G. Kasieczka, *OmniJet- α : the first cross-task foundation model for particle physics*, *Machine Learning: Science and Technology* **5** (2024) 035031.
- [117] L. Builtjes, S. Caron, P. Moskvitina, C. Nellist, R. Ruiz de Austri, R. Verheyen et al., *Attention to the strengths of physical interactions: Transformer and graph-based event classification for particle physics experiments*, *SciPost Physics* **19** (2025) .
- [118] A. Hammad, S. Moretti and M. Nojiri, *Multi-scale cross-attention transformer encoder for event classification*, *Journal of High Energy Physics* **2024** (2024) .
- [119] B.M. Dillon, R. Mastandrea and B. Nachman, *Self-supervised anomaly detection for new physics*, *Physical Review D* **106** (2022) .
- [120] T. Golling, L. Heinrich, M. Kagan, S. Klein, M. Leigh, M. Osadchy et al., *Masked particle modeling on sets: towards self-supervised high energy physics foundation models*, *Machine Learning: Science and Technology* **5** (2024) 035074.
- [121] S. Van Stroud, P. Duckett, M. Hart, N. Pond, S. Rettie, G. Facini et al., *Transformers for Charged Particle Track Reconstruction in High-Energy Physics*, *Physical Review X* **15** (2025) .
- [122] F. Mokhtar, J. Pata, D. Garcia, E. Wulff, M. Zhang, M. Kagan et al., *Fine-tuning machine-learned particle-flow reconstruction for new detector geometries in future colliders*, *Physical Review D* **111** (2025) .
- [123] Rizzi, Andrea, Vaselli, Francesco, Cattafesta, Filippo and Asenov, Patrick, *CMS FlashSim: End-to-end simulation with Machine Learning*, *EPJ Web Conf.* **337** (2025) 01014.
- [124] A. Butter, N. Huetsch, S. Palacios Schweitzer, T. Plehn, P. Sorrenson and J. Spinner, *Jet diffusion versus JetGPT – Modern networks for the LHC*, *SciPost Physics Core* **8** (2025) .
- [125] Z. Jiang, D. Yin, E.E. Khoda, V. Loncar, E. Govorkova, E. Moreno et al., *Ultra Fast Transformers on FPGAs for Particle Physics Experiments*, 2024. <https://arxiv.org/abs/2402.01047>.
- [126] L. Laatu, C. Sun, A. Cox, A. Gandrakota, B. Maier, J. Ngadiuba et al., *Sub-microsecond Transformers for Jet Tagging on FPGAs*, 2510.24784.
- [127] O. Güzel, *Transformer model for HH-> bbWW Run3 Analysis at the CMS Experiment at CERN*, 2026. doi:10.5281/ZENODO.19109519.

- [128] J.H. Friedman, *Greedy function approximation: A gradient boosting machine.*, *The Annals of Statistics* **29** (2001) .
- [129] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, p. 785–794, ACM, Aug., 2016, DOI.
- [130] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma et al., *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*, in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.
- [131] Y. Coadou, *Boosted Decision Trees*, in *Artificial Intelligence for High Energy Physics*, p. 9–58, WORLD SCIENTIFIC (2022), DOI.
- [132] S.M. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, in *Neural Information Processing Systems*, 2017, <https://api.semanticscholar.org/CorpusID:21889700>.
- [133] J. Albrecht, A.A. Alves, G. Amadio, G. Andronico, N. Anh-Ky, L. Aphetche et al., *A Roadmap for HEP Software and Computing R&D for the 2020s*, *Computing and Software for Big Science* **3** (2019) .
- [134] S. Popov, S. Morozov and A. Babenko, *Neural Oblivious Decision Ensembles for Deep Learning on Tabular Data*, in *International Conference on Learning Representations*, 2020, <https://openreview.net/forum?id=r1eiu2VtwH>.
- [135] S. Summers, G.D. Guglielmo, J. Duarte, P. Harris, D. Hoang, S. Jindariani et al., *Fast inference of Boosted Decision Trees in FPGAs for particle physics*, *Journal of Instrumentation* **15** (2020) P05026–P05026.
- [136] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, *Journal of Physics: Conference Series* **762** (2016) 012036.
- [137] A. Andreassen and B. Nachman, *Neural networks for full phase-space reweighting and parameter tuning*, *Physical Review D* **101** (2020) .
- [138] O. Güzel, *XGBoost for $HH \rightarrow bbWW$ Run3 Analysis in CMS Experiment at CERN*, 2026. doi:10.5281/ZENODO.19109166.
- [139] S.D. Drell and T.-M. Yan, *Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies*, *Phys. Rev. Lett.* **25** (1970) 316.
- [140] M. Czakon, P. Fiedler and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$* , *Phys. Rev. Lett.* **110** (2013) 252004.
- [141] T.M.P. Tait, *tW^- mode of single top quark production*, *Phys. Rev. D* **61** (1999) 034001.

- [142] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t -channel single top quark production*, *Phys. Rev. D* **83** (2011) 091503(R).
- [143] J.M. Campbell, R.K. Ellis and C. Williams, *Vector boson pair production at the LHC*, *Journal of High Energy Physics* **2011** (2011) .
- [144] S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao and M. Zaro, *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, *Journal of High Energy Physics* **2015** (2015) .
- [145] W. Beenakker, S. Dittmaier, M. Krämer, B. Plümper, M. Spira and P. Zerwas, *NLO QCD corrections to $t\bar{t}$ production in hadron collisions*, *Nuclear Physics B* **653** (2003) 151–203.
- [146] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, *Physics Letters B* **579** (2004) 149–156.
- [147] K. Ehataht and C. Veelken, *Stitching Monte Carlo samples*, *The European Physical Journal C* **82** (2022) .
- [148] CMS Collaboration, *Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements*, *The European Physical Journal C* **80** (2020) .
- [149] R.D. Ball, V. Bertone, S. Carrazza, L.D. Debbio, S. Forte, P. Groth-Merrild et al., *Parton distributions from high-precision collider data: NNPDF Collaboration*, *The European Physical Journal C* **77** (2017) .
- [150] M. Capozzi and G. Heinrich, *Exploring anomalous couplings in Higgs boson pair production through shape analysis*, *Journal of High Energy Physics* **2020** (2020) .
- [151] CMS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector*, *Journal of Instrumentation* **14** (2019) P07004–P07004.
- [152] J.A. et al., *Recent developments in Geant4*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835** (2016) 186.
- [153] M.-Y. Lee, *Run 3 commissioning results of heavy-flavor jet tagging at $\sqrt{s} = 13.6$ TeV with CMS data using a modern framework for data processing*, in *Proceedings of 12th Large Hadron Collider Physics Conference — PoS(LHCP2024)*, LHCP2024, p. 272, Sissa Medialab, Jan., 2025, DOI.
- [154] CMS Collaboration, *Development of the CMS detector for the CERN LHC Run 3*, *Journal of Instrumentation* **19** (2024) P05064.
- [155] CMS Collaboration, *Pileup mitigation at CMS in 13 TeV data*, *Journal of Instrumentation* **15** (2020) P09018–P09018.

- [156] M. Seidel, *Physics objects in CMS Run 3*, *Journal of Subatomic Particles and Cosmology* **5** (2026) 100330.
- [157] CMS Collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, *The European Physical Journal C* **75** (2015) .
- [158] CMS Collaboration, *Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons*, *Journal of High Energy Physics* **2019** (2019) .
- [159] CMS Collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *Journal of Instrumentation* **13** (2018) P05011–P05011.
- [160] CMS Collaboration, *Performance of heavy-flavour jet identification in Lorentz-boosted topologies in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Journal of Instrumentation* **20** (2025) P11006.
- [161] CMS Collaboration, *Search for non-resonant Higgs boson pair production in the $bbWW$ decay channel with two leptons in the final state using proton-proton collision data at $\sqrt{s} = 13.6$ TeV*, Tech. Rep. CMS-PAS-HIG-25-018, CERN, Geneva (2026).
- [162] CMS Collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Journal of Instrumentation* **13** (2018) P06015–P06015.
- [163] CMS Collaboration, *Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Journal of High Energy Physics* **2017** (2017) .
- [164] CMS Collaboration, *Jet Energy Scale and Resolution Measurements Using Run3 Data Collected by CMS in 2022 and 2023 at 13.6 TeV*, Tech. Rep. CMS-DP-2024-039, CERN, Geneva (2024).
- [165] CMS Collaboration, “CMS Jet Energy Resolution and Calibration (JERC) Group Web Page.” <https://cms-jerc.web.cern.ch/>, 2024. [Internal CMS documentation; accessed Jan 2026].
- [166] R.D. Ball, J. Butterworth, A.M. Cooper-Sarkar, A. Courtoy, T. Cridge, A. De Roeck et al., *The PDF4LHC21 combination of global PDF fits for the LHC Run III**, *Journal of Physics G: Nuclear and Particle Physics* **49** (2022) 080501.
- [167] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codebases* (2022) 8.
- [168] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, *Computer Physics Communications* **77** (1993) 219–228.

- [169] P. David, *Readable and efficient HEP data analysis with bamboo*, *EPJ Web Conf.* **251** (2021) 03052 [2103.01889].
- [170] “HH Inference Tools.” <https://gitlab.cern.ch/hh/tools/inference>, 2026. Accessed: 2026-02-05.
- [171] CMS Collaboration, *Search for Higgs boson pair production in the $b\bar{b}W^+W^-$ decay mode in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **07** (2024) 293 [2403.09430].
- [172] ATLAS and CMS Collaborations, *Highlights of the HL-LHC physics projections by ATLAS and CMS*, Tech. Rep. ATL-PHYS-PUB-2025-018, CMS-HIG-25-002, CERN, Geneva (2025).
- [173] ATLAS Collaboration, *Reference Single Top-Quark Cross-Sections for ATLAS and CMS Analyses*, Tech. Rep. ATL-PHYS-PUB-2025-035, CERN, Geneva (2025).
- [174] A. Kulesza, L. Motyka, D. Schwartländer, T. Stebel and V. Theeuwes, *Associated top quark pair production with a heavy boson: differential cross sections at NLO + NNLL accuracy*, *The European Physical Journal C* **80** (2020) .
- [175] M. van Beekveld, A. Kulesza and L. Moreno Valero, *Threshold Resummation for the Production of Four Top Quarks at the LHC*, *Physical Review Letters* **131** (2023) .
- [176] “Standard model cross sections for cms at 13 tev.” TWiki: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV>.
- [177] J.M. Campbell, R.K. Ellis and C. Williams, *Vector boson pair production at the LHC*, *Journal of High Energy Physics* **2011** (2011) .
- [178] A. Karlberg, J. Malcles, B. Mistlberger, R. Di Nardo, S.P. Jones, H. Abidi et al., *Ad interim recommendations for the Higgs boson production cross sections at $\sqrt{s} = 13.6$ TeV*, Tech. Rep. LHCHWG-2024-001, CERN, Geneva (2024).
- [179] A. Kulesza, L. Motyka, D. Schwartländer, T. Stebel and V. Theeuwes, *Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy*, *Eur. Phys. J.* **C79** (2019) 249 [1812.08622].
- [180] “Standard model cross sections for cms at 13.6 tev.” TWiki: https://twiki.cern.ch/twiki/bin/viewauth/CMS/MATRIXCrossSectionsat13p6TeV_r12.