GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Bachelor's Thesis

Untersuchungen der assoziierten Higgs-Photon-Produktion in Elektron-Positron-Kollisionen am FCC-ee

Studies of Higgs-boson production in association with a photon in electron-positron collisions at the FCC-ee collider

prepared by

Thomas W. Erlei

from Goslar

at the II. Physikalisches Institut

Thesis number: II.Physik-UniGö-BSc-2026/06

Thesis period: 3rd April 2026 until 13th July 2026

First referee: Prof. Dr. Arnulf Quadt

Second referee: PD Dr. Ralf Bernhard

Zusammenfassung

Diese Arbeit untersucht die assoziierte Produktion eines Higgs-Bosons mit einem Photon in Elektron-Positron-Kollisionen am FCC-ee. Der betrachtete Signalprozess ist $e^+e^- \to H\gamma$ bei einer Schwerpunktsenergie von $\sqrt{s} = 160\,\mathrm{GeV}$, gefolgt vom vollhadronischen Zerfall $H \to WW^* \to 4q$. Nach einer losen Vorselektion wird ein Gradient-Boosting-Entscheidungsbaum verwendet, um das seltene Signal von den deutlich größeren Photon-plus-Jets-Untergründen zu trennen. Ein enger Schnitt bei $s_{\mathrm{BDT}} > 0.988$ dient als Zählbenchmark und ergibt für $19.2\,\mathrm{ab}^{-1}$ eine erwartete Signalausbeute von $S = 11.07$ bei einem Untergrund von $B = 9.64 \times 10^4$. Für die statistische Abschätzung wird stattdessen ein eventbasierter Formfit im lockereren Bereich $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$ verwendet. Aus 1000 feinen Hilfsbins werden 20 Bins mit annähernd gleicher erwarteter Signalausbeute gebildet. Der statistische Fit mit festgehaltenem Untergrund ergibt $\mu = 1.0^{+21.9}_{-21.8}$ und eine erwartete Asimov-Signifikanz von $Z_{\mathrm{A}} = 0.0458$. Das Ergebnis zeigt, dass die Analysekette funktioniert, die vollhadronische $H \to WW^*$-Endzustandsanalyse in dieser Form aber stark untergrundlimitiert ist.

Stichwörter: Higgs-Boson; FCC-ee; Higgs-Photon-Produktion; Standardmodell; IDEA-Detektor

Abstract

This thesis studies Higgs-boson production in association with a photon in electron–positron collisions at FCC-ee. The signal process is $e^+e^- \to H\gamma$ at a centre-of-mass energy of $\sqrt{s} = 160\,\mathrm{GeV}$, followed by the fully hadronic decay $H \to WW^* \to 4q$. After a loose preselection, a gradient-boosted decision-tree classifier is used to separate the rare signal from the much larger photon-plus-jets background. A tight requirement of $s_{\mathrm{BDT}} > 0.988$ is used as a counting benchmark and gives, for $19.2\,\mathrm{ab}^{-1}$, an expected signal yield of $S = 11.07$ with a background yield of $B = 9.64 \times 10^4$. The statistical estimate instead uses an event-based shape fit in the looser range $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$. One thousand fine starting bins are merged into 20 bins with approximately equal expected signal yields. With the background fixed, the statistical fit gives $\mu = 1.0^{+21.9}_{-21.8}$ and an expected Asimov significance of $Z_{\mathrm{A}} = 0.0458$. The result shows that the analysis chain works technically, but that the fully hadronic $H \to WW^*$ final state is strongly background-limited in the present analysis.

Keywords: Higgs boson; FCC-ee; Higgs-photon production; Standard Model; IDEA detector

Contents

1. Introduction **1**

2. The Standard Model of Particle Physics **3**
2.1. Higgs sector and electroweak symmetry breaking 3
2.2. Higgs production in e^+e^- collisions 4
2.3. Higgs-boson production in association with a photon 5
2.4. Higgs-boson hadronic decay 6

3. Future Circular Collider **7**

4. IDEA Detector FCC-ee **8**
4.1. Inner detector 8
4.1.1. Vertex detector 8
4.1.2. Drift Chamber 9
4.1.3. Silicon wrapper 9
4.2. Electromagnetic calorimeter 9
4.3. Hadronic calorimeter 10
4.4. Muon system 10
4.5. Detector components most relevant for the signal 10

5. Study the Signal **12**
5.1. Analysis Strategy 13
5.2. Signal Definition 13
5.3. Background Samples 14
5.4. Event Weights and Expected Yields 16
5.5. Reconstructed Observables 17
5.6. Baseline Comparison of Signal and Background 18
5.6.1. Cut-Based Preselection 19
5.6.2. Multivariate Analysis 27
5.6.3. Shape-Based Likelihood Fit 33
5.6.4. Discussion 38

6. Conclusion and Outlook **40**

A. Detailed Yield and Shape-Fit Tables **41**

B. BDT Input-Variable Appendix Material **47**

C. Additional Images **55**

List of Figures

1.1. Representative one-loop Feynman diagrams for $e^+e^- \to H\gamma$. 2

2.1. Schematic Higgs potential. 3
2.2. Schematic diagrams for Higgsstrahlung and WW fusion. 4

4.1. Three-dimensional view of the IDEA detector concept. 9
4.2. Schematic detector response of different particle types. 11

5.1. Baseline comparison of selected reconstructed observables for signal and background events with physical event-weight normalization. The leading-photon energy and the recoil mass are directly connected to the $e^+e^- \to H\gamma$ signal topology. The lepton and jet observables help to identify the zero-lepton hadronic final state. 20
5.2. Normalized BDT-score distributions for the training and validation samples. The broad agreement shows that the classifier response is stable, while small residual differences in the low- and high-score tails indicate mild overtraining. The held-out validation sample is used for the quoted classifier performance and for the score-shape study. 30
5.3. ROC curve of the BDT classifier evaluated on the held-out validation sample. The validation AUC is 0.9901. 31
5.4. Event-based, physics-weighted BDT-score distribution in the high-score region after the preselection. The backgrounds are stacked by process group. The signal line is multiplied by 10^4 for visibility, and the open points show the unscaled Asimov expectation $S + B$. The 50 equal-width display bins are not used in the likelihood fit. 32
5.5. Expected sensitivity as a function of the BDT-score cut in the stable scan region with at least one expected signal event. Both $S/\sqrt{B}$ and the Asimov significance are shown. The tight counting benchmark is $s_{\mathrm{BDT}} > 0.988$. . . 33

5.6. Event-based templates in the 20 variable-width bins used by the likelihood. The background categories are stacked and are summed in the statistical model. The signal is multiplied by 10^5 for visibility; the open points show the unscaled Asimov expectation $S + B$. The nearly constant scaled signal height follows from the approximately equal signal yield per bin. 35

5.7. Expected statistical-only likelihood scan for the event-based 20-bin BDT-shape fit. The scan is extended below the physical boundary $\mu = 0$ to display both unconstrained one-standard-deviation crossings. In the physical fit, the lower interval ends at $\mu = 0$. 37

A.1. Expected signal yield in each of the 20 final likelihood bins, ordered from lower to higher BDT scores. The dashed line shows the mean of approximately 1.32 expected signal events per bin and confirms the intended equal-signal construction. 46

C.1. Cross sections for $e^+e^- \to H\gamma$ and $e^+e^- \to ZH$. 55

List of Tables

3.1. Main FCC-ee running points and centre-of-mass energies. 7

5.1. Monte Carlo samples, generator setup, and nominal cross sections. 15
5.2. Main groups of reconstructed observables used to distinguish signal and background. 18
5.3. Summary of the cut-based preselection used before the multivariate analysis. 25
5.4. Expected process-group yields after the cut-based preselection. The signal yield corresponds to the truth-selected fully hadronic $H \to WW^* \to 4q$ final state. The uncertainties are MC-statistical only and are calculated using Equation 5.26. 26
5.5. Main settings of the gradient-boosted decision-tree classifier. 28
5.6. Input variables used for the BDT training and their definitions. 28
5.7. Expected signal and background yields after the preselection and after the tight counting benchmark $s_{\mathrm{BDT}} > 0.988$. The background is shown by process group. The uncertainties are MC-statistical only. A zero entry means that no simulated event from that group passed the tight threshold. 34
5.8. Summary of the event-based adaptive BDT-score binning used in the shape likelihood. 35

A.1. Composition and principal reproducibility settings of the final BDT training. Counts are unweighted simulated entries after the per-process input limit; physical weights are retained for expected yields. 41
A.2. Expected yields by process group before and after the tight BDT counting benchmark. The quoted uncertainties are MC-statistical uncertainties. . . . 42
A.3. Per-process yields before and after the tight BDT counting benchmark. The preselection column corresponds to the selected sample before applying the BDT score requirement. The BDT column uses the tight counting benchmark $s_{\mathrm{BDT}} > 0.988$. Uncertainties are MC-statistical only. 43

A.4. Process-group composition of the event-based background template in the shape-fit range $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$. The groups are filled separately for the stacked figures and then summed to form B_k in the likelihood. 44

A.5. Event-based adaptive BDT-score bins used in the statistical-only shape likelihood. The bins cover $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$ and contain approximately equal expected signal yields. The final column gives the effective weighted background statistics, $N_{\mathrm{eff}}^{B} = (\sum w)^2 / \sum w^2$, as a diagnostic of template stability. 45

B.1. Grouping of BDT input variables used to organize the requested appendix distributions. 47

Nomenclature

Abbreviations

Abbreviation	Meaning
SM	Standard Model of particle physics
FCC	Future Circular Collider
FCC-ee	Electron–positron stage of the Future Circular Collider
IDEA	Innovative Detector for Electron–positron Accelerators
ECAL	Electromagnetic calorimeter
HCAL	Hadronic calorimeter
BDT	Boosted decision tree
MVA	Multivariate analysis

Frequently used symbols

Symbol	Meaning	Unit
$\sqrt{s}$	Centre-of-mass energy	GeV
m_H	Higgs-boson mass	GeV
m_W	W-boson mass	GeV
m_Z	Z-boson mass	GeV
E_γ	Photon energy	GeV
$\mathcal{L}$	Integrated luminosity	ab^{-1}
S	Expected signal yield after a selection	events
B	Expected background yield after a selection	events
μ	Signal-strength parameter relative to the Standard Model prediction	–
v	Higgs vacuum expectation value	GeV

1. Introduction

On a rainy Sunday morning, while walking through fields and with your eyes catching a field of crops, birds, and clouds, it is easy to forget that everything we see, feel, and hear is built from the same deeper construct, from the same fundamental ingredients. The striving of humanity is, and always will be, to discover the universe in all that there is to discover and to try to understand how we are built. We had to go a long way to get to today's understanding of the deepest constructs. Since the discovery of the electron in 1897, physicists have gradually uncovered ever smaller constituents of matter. This development ultimately led to the formulation of the Standard Model of particle physics, which is currently the most successful theory describing the fundamental building blocks of nature. The Standard Model contains 17 elementary particles: 12 fermions, which make up matter, 4 gauge bosons, which mediate the fundamental forces, and the Higgs boson, which is associated with the mechanism that gives mass to elementary particles [1]. The experimental discovery of the Higgs boson in 2012 marked a major confirmation of the Standard Model because it was predicted by the model and completed the particle content [2, 3]. As the theoretical foundation of modern particle physics, the Standard Model is of central importance to this thesis, as it provides the framework for understanding and interpreting the phenomena relevant to this research topic. The Standard Model predicts fundamental processes and parameters with high precision, but there are still phenomena that cannot be explained by this model. The most well-known examples are galaxy rotation curves in astrophysics. They are incompatible with the predictions of the Standard Model, which suggests the existence of a hitherto unknown or undetected particle or field, commonly referred to as dark matter [4]. To verify the existence of dark matter, a comprehensive and holistic understanding of elementary particles is required. Here, the Higgs boson is of particular importance because it is the only elementary scalar particle in the Standard Model and plays a central role in electroweak symmetry breaking [1]. Several theories beyond the Standard Model predict additional Higgs bosons or modified Higgs interactions [5]. Studying such extensions could reveal new phenomena and may help to clarify possible connections between the Higgs sector and unresolved questions such as dark matter. This bachelor thesis analyses the associated production of a Higgs boson and

a photon at the FCC-ee, the proposed electron–positron stage of the Future Circular Collider. The focus lies on electron–positron collisions at a centre-of-mass energy of 160 GeV, which corresponds to the W^+W^- production threshold [6]. The process $e^+e^- \to H\gamma$ is loop-induced in the Standard Model and therefore sensitive to the Higgs–photon–photon ($H\gamma\gamma$) and Higgs–photon–Z-boson ($H\gamma Z$) interactions [5, 7, 8]. The Standard Model prediction for the $e^+e^- \to H\gamma$ cross section is small, which makes the process experimentally challenging. However, because the process is loop-induced, additional particles or modified Higgs interactions could alter the loop contributions. This channel therefore offers an especially sensitive probe of possible deviations from the predictions of the Standard Model. Higgs production in association with a photon provides a complementary way to examine the structure of the Higgs and to search for possible indications of new physics. At the same time, the clean experimental environment of electron–positron collisions at the FCC-ee makes this collider suitable for precision studies of rare Higgs processes such as $H\gamma$ production [6].

Representative one-loop diagrams are shown in Fig. 1.1. The triangle diagrams are related to the loop-induced $H\gamma\gamma$ and $H\gamma Z$ interactions, whereas the box diagrams contribute directly to the complete production amplitude.

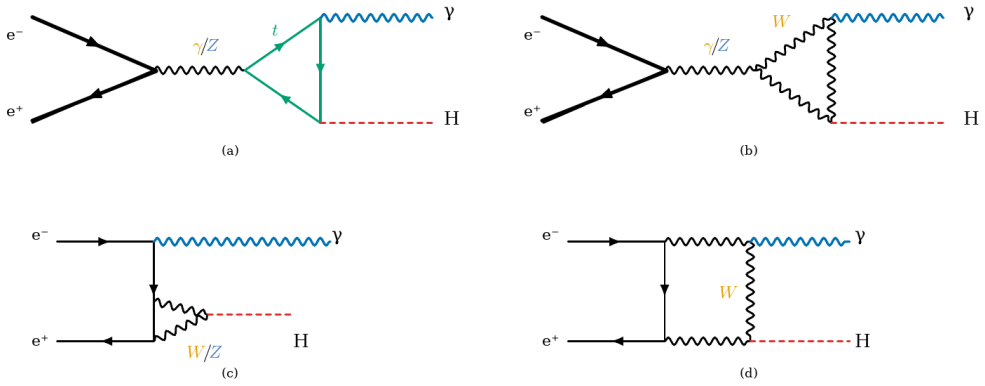


Figure 1.1.: Representative one-loop Feynman diagrams contributing to $e^+e^- \to H\gamma$ in the Standard Model. Own illustration based on Refs. [7, 8].

In this thesis, the signal process is studied in the fully hadronic decay channel $e^+e^- \to H\gamma$, with $H \to WW^* \to q\bar{q}q\bar{q}$. The central aim is to determine whether this rare signal can be separated from the much larger Standard Model backgrounds at a centre-of-mass energy of 160 GeV and to estimate the resulting expected statistical sensitivity to the $e^+e^- \to H\gamma$ production rate. To this end, a cut-based preselection is combined with a boosted decision tree that exploits differences between the reconstructed kinematic properties of signal and background events.

2. The Standard Model of Particle Physics

The Standard Model is a gauge theory based on $SU(3)_C \times SU(2)_L \times U(1)_Y$ [1]. Its fermions are arranged in three generations of quarks and leptons. Ordinary atomic matter is composed of particles from the first generation, while the heavier generations are unstable and decay into lighter states. The photon, the W and Z bosons, and the gluons mediate the electromagnetic, weak, and strong interactions, respectively. The electroweak $SU(2)_L \times U(1)_Y$ structure unifies the electromagnetic and weak interactions [9]. The Higgs field completes the model by enabling electroweak symmetry breaking: the W and Z bosons acquire mass, and fermion masses arise through Yukawa couplings to the Higgs field [1].

2.1. Higgs sector and electroweak symmetry breaking

Through the mechanism of electroweak symmetry breaking, elementary particles acquire mass; this is described by the Higgs sector. In the electroweak theory, the electromagnetic and weak forces are described in a unified way. This symmetry cannot be seen in nature at low energies. The symmetry is broken when the Higgs field has a non-zero vacuum expectation value, $v \simeq 246$ GeV [1], shown schematically in Figure 2.1. Consequently, the weak gauge bosons $W^{\pm}$ and Z obtain mass,

$$M_W = \frac{gv}{2} \qquad \text{and} \qquad M_Z = \frac{v}{2}\sqrt{g^2 + g'^2}, \quad (2.1)$$

where g is the gauge coupling of the weak isospin group $SU(2)_L$ and g' of the weak hypercharge group

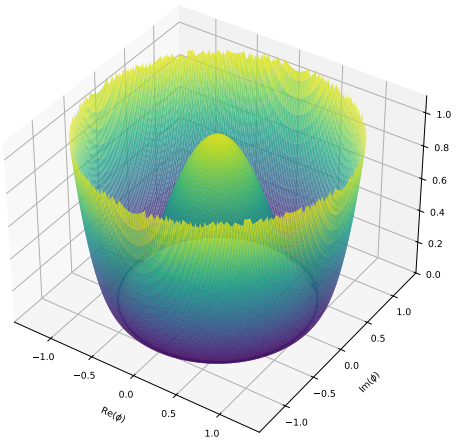


Figure 2.1.: Schematic Higgs potential, generated from $V(\phi) \propto \left(|\phi|^2 - v^2\right)^2$. The ring shows the degenerate minima after spontaneous symmetry breaking [1, 10].

$U(1)_Y$ [1]. The photon remains without mass. That is essential for the long-range nature of the electromagnetic force. Fermions also acquire mass through their Yukawa couplings to the Higgs field, according to the relation

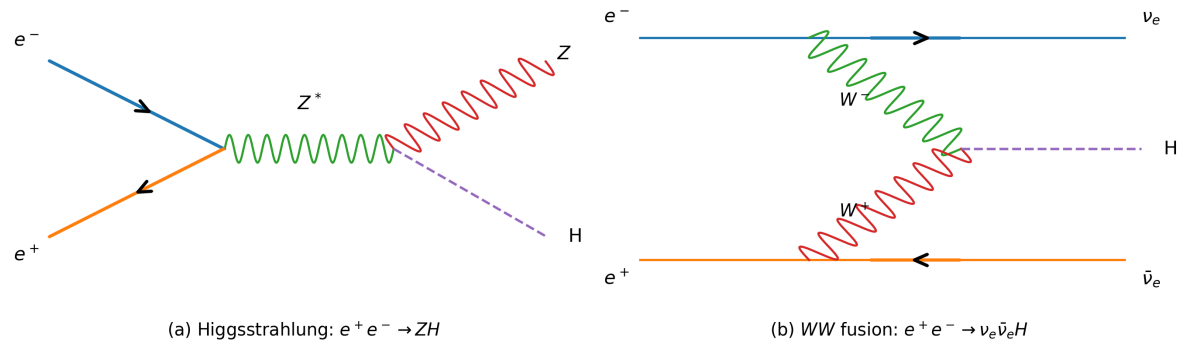
$$m_f = \frac{y_f v}{\sqrt{2}}, \tag{2.2}$$

where y_f is the Yukawa coupling of a fermion f [1]. After symmetry breaking, one scalar particle remains in the theory. That is the Higgs boson, which is directly connected to the Higgs field and is especially relevant for this thesis.

2.2. Higgs production in e^+e^- collisions

In the FCC-ee collider, the Higgs boson can be produced through many mechanisms. The most important ones at the collider are Higgsstrahlung $e^+e^- \to ZH$ and vector-boson fusion, through WW fusion $e^+e^- \to \nu_e\bar{\nu}_e H$ [6]. Electron–positron collisions have a much cleaner experimental environment and a well-defined initial state compared to hadron colliders. Thus it is possible to reconstruct Higgs events with high precision and even inclusive measurements are possible [6]. In addition to these dominant channels, rarer processes are also of great interest, such as the associated production of a Higgs with a photon, $e^+e^- \to H\gamma$. This is because this process is loop-induced in the Standard Model and therefore can be used as a sensitive probe of Higgs interactions and possible variation from the SM predictions [7, 8].

Representative diagrams for Higgsstrahlung and WW fusion are shown in Figure 2.2; the loop-induced $H\gamma$ diagrams are shown in Figure 1.1.



(a) Higgsstrahlung: $e^+e^- \to ZH$ (b) WW fusion: $e^+e^- \to \nu_e\bar{\nu}_e H$

Figure 2.2.: Schematic diagrams for Higgsstrahlung and WW fusion. Own illustration based on the production mechanisms described in Ref. [6].

2.3. Higgs-boson production in association with a photon

The process studied in this thesis is the associated production of a Higgs boson with a photon in electron–positron collisions,

$$e^+e^- \to H\gamma, \tag{2.3}$$

at a centre-of-mass energy of $\sqrt{s} = 160$ GeV. This energy lies close to the W^+W^- production threshold [6] and below the on-shell Higgsstrahlung threshold. The latter is set by

$$m_Z + m_H \simeq 216 \text{ GeV}, \tag{2.4}$$

with $m_H \simeq 125.2$ GeV the mass of the Higgs boson and $m_Z \simeq 91.2$ GeV the mass of the Z boson [1]. The fact that $\sqrt{s} = 160$ GeV lies below the on-shell Higgsstrahlung threshold removes the dominant on-shell tree-level Higgs-production channel $e^+e^- \to ZH$, as illustrated in Figure C.1, but it does not eliminate other signal-like backgrounds. As already mentioned, the production of a Higgs in association with a photon is especially rare, because in the Standard Model it does not occur at tree level. There are no tree-level $H\gamma\gamma$ or $HZ\gamma$ couplings; hence the reaction starts only at one-loop order, as shown in Figure 1.1. The amplitude is generated by quantum corrections and has important contributions from W-boson loops, top-quark loops and box diagrams [7, 8]. At energies above the Higgsstrahlung threshold, the Higgsstrahlung process $e^+e^- \to ZH$ has a much higher production rate, as shown in Figure C.1. That makes the measurement challenging; at the same time, this infrequency makes this process sensitive to deviations, as already discussed.

For the two-body final state in the centre-of-mass frame, energy conservation gives

$$E_\gamma + E_H = \sqrt{s}. \tag{2.5}$$

Momentum conservation requires $|\vec{p}_H| = |\vec{p}_\gamma| = E_\gamma$, while the Higgs energy fulfils $E_H^2 = E_\gamma^2 + m_H^2$. Solving these relations gives

$$E_\gamma = \frac{s - m_H^2}{2\sqrt{s}}. \tag{2.6}$$

With $\sqrt{s} = 160$ GeV and the Higgs-boson mass m_H, that corresponds to about 31 GeV and is a characteristic signature [5, 7].

2.4. Higgs-boson hadronic decay

The fully hadronic Higgs-boson decay studied in this thesis is

$$H \to W^+W^{-*} \to q_u\bar{q}_d\, q'_d\bar{q}'_u, \tag{2.7}$$

where $q_u, q'_u \in \{u, c\}$ denote up-type quarks and $q_d, q'_d \in \{d, s, b\}$ denote down-type quarks. Since $m_H = 125.20 \pm 0.11$ GeV is smaller than $2m_W \simeq 160.74$ GeV, two on-shell W bosons are kinematically forbidden and one must be off shell [1]. Nevertheless, $H \to WW^*$ remains one of the dominant Standard Model Higgs-boson decay modes, with a branching ratio of 2.14×10^{-1} [11]. Combined with the hadronic W-boson branching fraction, $\mathrm{BR}(W \to \text{hadrons}) = (67.41 \pm 0.27)\%$ [1], this gives about 9.7% for the fully hadronic final state. After hadronisation, the two quark–antiquark pairs form a nominal four-jet topology, providing a fully visible Higgs-boson final state for $e^+e^- \to H\gamma$.

The branching fraction entering the selected signal channel can be written explicitly as

$$\begin{aligned} \mathrm{BR}(H \to WW^* \to 4q) &= \mathrm{BR}(H \to WW^*)\,[\mathrm{BR}(W \to \text{hadrons})]^2 \\ &\simeq 0.214 \times (0.6741)^2 \simeq 9.7 \times 10^{-2}. \end{aligned} \tag{2.8}$$

The virtual W^* boson does not have a fixed invariant mass. If the other W boson is close to its mass shell, the mass available to the virtual boson is bounded approximately by $m_{W^*} \lesssim m_H - m_W \simeq 45$ GeV. Consequently, the four-quark final state does not generally form two dijet systems that both peak at m_W. One pair can be compatible with an on-shell W boson, whereas the second has a lower and broader invariant-mass spectrum. The quarks from the virtual W^* can also be softer or less well separated, so hadronisation and jet clustering may merge them or reconstruct fewer than four separate jets. The parton-level four-quark state must therefore be distinguished from an experimentally reconstructed exact four-jet topology [1, 11].

At $\sqrt{s} = 160$ GeV, the Higgs boson recoils against the photon with a momentum of about 31 GeV and an energy

$$E_H = \frac{s + m_H^2}{2\sqrt{s}} \simeq 129 \text{ GeV}. \tag{2.9}$$

Apart from neutrinos produced in secondary hadron decays, the hard final state contains no invisible particles. It therefore consists of an isolated photon balanced by a hadronic system with an invariant mass close to m_H, little genuine missing transverse momentum, and no prompt isolated charged lepton.

3. Future Circular Collider

The FCC-ee (Future Circular Collider – electron–positron stage) is the proposed first stage of CERN's FCC programme. The feasibility study describes a new circular tunnel of about 90.7 km in the Geneva region and a high-luminosity e^+e^- collider operating as a Higgs, electroweak, and top-quark factory [6, 12, 13]. Its principal running points are listed in Table 3.1. Precision measurements of the Higgs boson, the W and Z bosons, and the top quark would provide stringent tests of the Standard Model and indirect sensitivity to new physics. In its resolution of 22 May 2026, the CERN Council identified FCC-ee as the preferred option for CERN's next flagship collider. The resolution also makes clear that implementation and construction remain subject to separate Council decisions [14].

Table 3.1.: Main FCC-ee running points and their characteristic centre-of-mass energies [6].

Running point	Main physics process / goal	$\sqrt{s}$ [GeV]
Z pole	Precision studies of the Z boson	91.2
WW threshold	Precision studies of W-pair production	160
ZH peak	Higgsstrahlung, $e^+e^- \to ZH$	240
$t\bar{t}$ threshold	Top-quark pair production	365

4. IDEA Detector FCC-ee

To detect and ultimately examine the resulting products from the e^+e^- collision, different layers of complex detector components are needed. The most vital task of the detector is to detect the position of the passing particle, as well as the momentum and the energy of the particles. There are different kinds of particles which interact differently with other materials, because of the different applicable forces. Therefore, the detector has different layers, which will be explained in the following sections.

4.1. Inner detector

The IDEA detector is organised in layers around the collision point. Closest to the beam line lies the silicon pixel vertex detector, which grants precise measurements of track origins and secondary vertices. It is surrounded by a large and light drift chamber. This is the main tracking device of IDEA, which measures the momentum of charged particles with high precision. Outside the drift chamber, a silicon wrapper is the last tracking device [6, 15].

These detectors measure the paths of charged particles as they pass through the detector. The charged particle leaves a small signal, for instance by producing ionisation in a gas detector or charge in a silicon sensor. The detector reconstructs the particle's trajectory as a track, with the help of different layers. Furthermore, the detectors can measure the momentum and charge sign, because of the magnetic field [16].

4.1.1. Vertex detector

The vertex detector is mainly constructed for the reconstruction of the primary and secondary vertices, which is essential for b-tagging. $H \rightarrow b\bar{b}$ is the main decay of the Higgs, so the reconstruction of the b vertex is important [1, 15]. The detector is therefore not only a tracking device; it can also give information about the lifetime and event topology [15].

4.1.2. Drift Chamber

The drift chamber is a low-mass cylindrical chamber operated with a helium–isobutane gas mixture. Charged-particle trajectories curve in the detector magnetic field, and the measured curvature is used to determine the momentum and charge sign. The small material budget reduces multiple scattering, photon conversions, and hadronic interactions, which is especially important for precise tracking [15, 17]. The chamber also contributes to particle identification. Charged particles ionise the gas along their trajectories, and counting the individual ionisation clusters improves the separation of particle species compared with a conventional truncated-mean dE/dx measurement [15, 17].

4.1.3. Silicon wrapper

The silicon wrapper, the last layer, improves the momentum resolution, especially for particles with high momentum and in the forward region, where the drift chamber gets fewer hits. It also helps the hermeticity of the detector, as shown in Figure 4.1 [15].

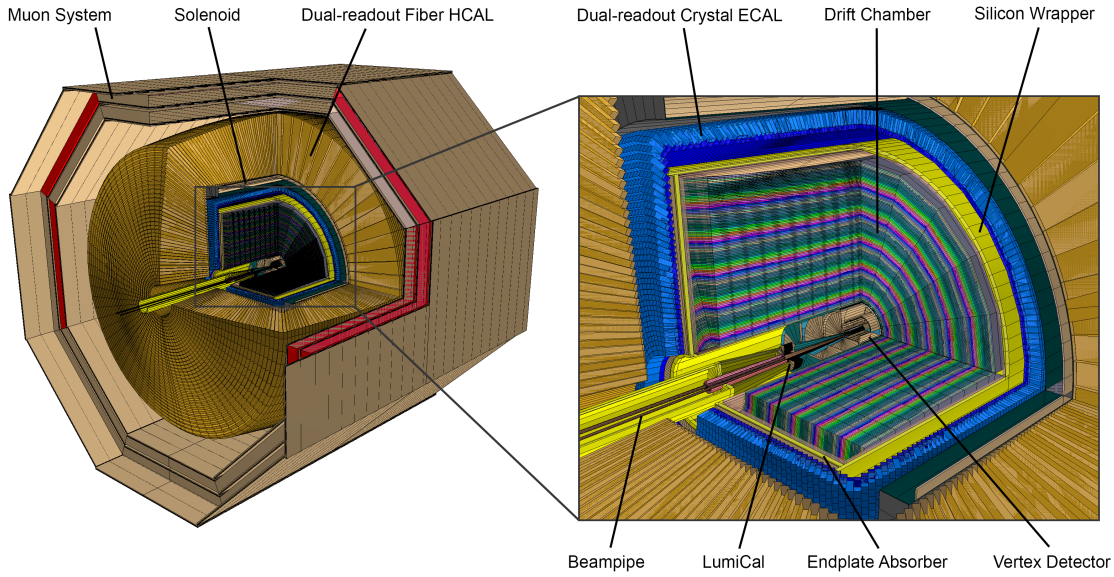


Figure 4.1.: Three-dimensional view of the IDEA detector concept. Reproduced from Figure 1 of Ref. [15] under the Creative Commons Attribution 4.0 licence.

4.2. Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is a dual-readout crystal electromagnetic calorimeter placed between the tracking system and the solenoid. The ECAL detector

measures the energy of a particle, which interacts with the electromagnetic field, by absorbing it in the crystal material. As the particle passes through, it creates a shower of secondary particles. These deposit their energy step by step in the detector [16]. This energy produces a measurable signal that will be read out with the use of scintillation and Cherenkov light. The crystal technology provides an excellent electromagnetic resolution; furthermore, the fine granularity and longitudinal segmentation improve the shower reconstruction [15].

4.3. Hadronic calorimeter

Because of the lower probability of hadronic interactions, hadrons can pass through the ECAL. Hence, to detect hadrons, another calorimeter, the so-called hadronic calorimeter, surrounds the ECAL, as shown in Figure 4.1. The HCAL and ECAL are similar in many respects. The hadronic calorimeter is a fibre-sampling dual-readout calorimeter with scintillating and Cherenkov fibres. Its main role is the measurement of hadrons and, therefore, the reconstruction of jets. That is an important point for this bachelor's thesis because it will study the decay of the Higgs into two W bosons, which can decay into four quarks [15].

4.4. Muon system

The muon system is the outermost part of the detector and is used to identify muons. They can pass through the inner detector and calorimeters much more easily than most other particles. This makes it important for distinguishing muons from hadrons and for reconstructing final states that contain muons. In the context of this thesis, however, the muon system is included mainly for completeness, since it is not one of the most relevant detector components for the analysis, where the inner detector and especially the electromagnetic calorimeter play the central role [15].

4.5. Detector components most relevant for the signal

For the signal process studied in this thesis, $e^+e^- \to H\gamma$ with $H \to WW^* \to 4q$, the most relevant detector components are the electromagnetic calorimeter, the tracking system, and the hadronic calorimeter. The photon is reconstructed as an isolated electromagnetic

shower in the ECAL, while the charged decay products of the hadronic final state are measured in the vertex detector and drift chamber, as can be seen in Figure 4.2 [18]. Together with the HCAL, this allows the reconstruction of the four-jet final state from the Higgs-boson decay. Therefore, the IDEA concept is well suited for this analysis, since it combines precise photon measurements with efficient tracking and jet reconstruction [6, 15].

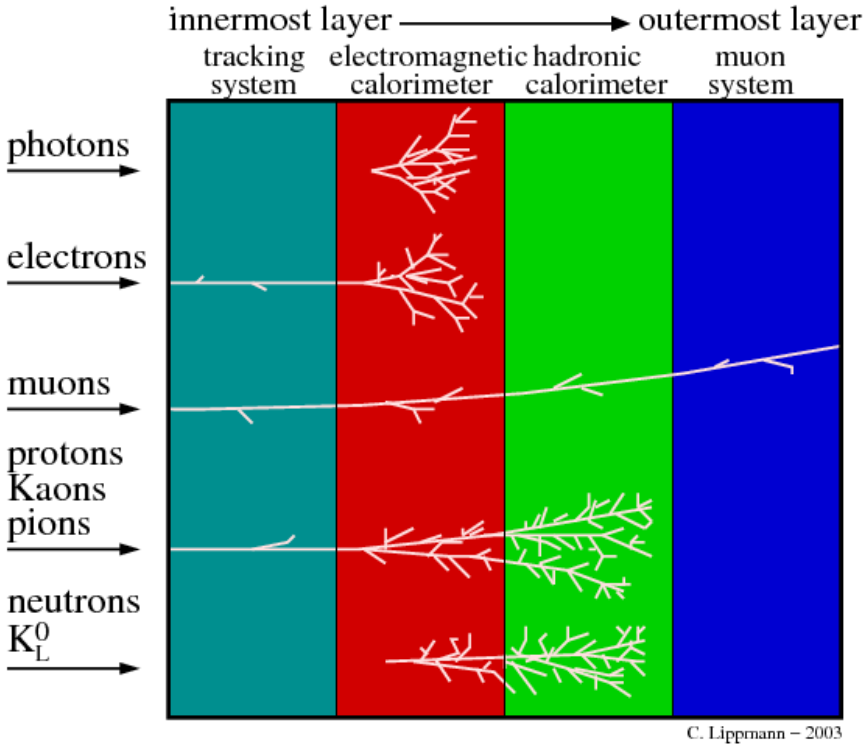


Figure 4.2.: Schematic response of different particle types in the tracking system, electromagnetic calorimeter, hadronic calorimeter, and muon system. Reproduced from Figure 1 of Ref. [18].

5. Study the Signal

After introducing the physics motivation and the experimental environment, this chapter describes how the signal process is selected in the simulated event samples. The focus is on the channel

$$e^+e^- \to H\gamma \tag{5.1}$$

at a centre-of-mass energy of

$$\sqrt{s} = 160 \text{ GeV}, \tag{5.2}$$

where the Higgs boson decays through the fully hadronic channel

$$H \to WW^* \to q\bar{q}'\, q''\bar{q}'''. \tag{5.3}$$

Here the primes only distinguish the quark flavours in the two hadronic W decays: each W boson decays into a quark and an antiquark of the corresponding weak-isospin partner type. The experimentally visible final state therefore consists of one photon and hadronic activity from the Higgs-boson decay. In a simplified picture, this corresponds to

$$e^+e^- \to H\gamma \to WW^*\gamma \to 4q + \gamma. \tag{5.4}$$

The most characteristic object in the event is the photon produced together with the Higgs boson. For this two-body process, the energy of the photon in the laboratory frame is fully determined by the precisely known e^+e^- centre-of-mass energy and the precisely known Higgs-boson mass; the energy of the Higgs boson is fixed by the same kinematic relation. As discussed in Section 2.3, the expected photon energy is approximately

$$E_\gamma = \frac{s - m_H^2}{2\sqrt{s}} \approx 31 \text{ GeV}. \tag{5.5}$$

The photon is therefore one of the central objects used to distinguish signal from background. The Higgs boson recoiling against the photon decays into two W bosons, where one of them is off-shell. Since both W bosons are required to decay hadronically, the final signal region is expected to contain jets and no isolated charged leptons.

5.1. Analysis Strategy

The analysis is built in several steps. First, signal and background samples are read from simulated ROOT files [19]. A cut-based preselection then removes events that are clearly incompatible with the expected signal topology. After this preselection, a multivariate analysis improves the separation between signal and background. The BDT output is used both for a tight counting benchmark and, over a looser score range, for a binned shape-likelihood fit that connects the selected sample to the expected signal cross section.

The full analysis chain can be summarized as

$$\begin{aligned}&\text{Monte Carlo samples} \rightarrow \text{ROOT trees} \rightarrow \text{event preselection}\\ &\rightarrow \text{training data frame} \rightarrow \text{BDT training and validation} \rightarrow \text{BDT score distribution}\\ &\rightarrow \text{adaptive score binning} \rightarrow \text{shape-likelihood fit.}\end{aligned} \tag{5.6}$$

The preselection is not intended to give the final signal region by itself. Instead, it removes events that are clearly outside the expected topology and keeps a sample on which the BDT can still learn the remaining differences between signal and background. The multivariate analysis combines several event variables into a single classifier score. Events with high scores are more signal-like, while events with low scores are more background-like. A tight score requirement is useful for quoting the rejection power of the classifier in one number, but the final statistical estimate retains the score shape in a wider interval rather than reducing it to one event count.

5.2. Signal Definition

A clear signal definition is essential for the analysis. In this thesis, only the fully hadronic Higgs-boson decay channel is counted as signal:

$$e^+e^- \rightarrow H\gamma, \qquad H \rightarrow WW^* \rightarrow q\bar{q}'\, q''\bar{q}'''. \tag{5.7}$$

All other Higgs-boson decay modes are not part of the signal definition used for the final multivariate analysis. This is important because the process $e^+e^- \rightarrow H\gamma$ can also occur with different Higgs-boson decays, for example $H \rightarrow b\bar{b}$, $H \rightarrow \tau^+\tau^-$, or $H \rightarrow \mu^+\mu^-$. These final states have different experimental signatures and are therefore not included as signal in the fully hadronic $H \rightarrow WW^*$ analysis.

The signal topology can be summarized as follows:

- one energetic photon with an energy close to the value expected from two-body

kinematics,

- a recoil system compatible with the Higgs-boson mass,
- hadronic activity from the decay $H \to WW^* \to 4q$,
- no isolated charged leptons in the final signal region.

This defines the zero-lepton hadronic channel. Events with isolated electrons or muons are rejected in the signal region, because they are more compatible with semi-leptonic or leptonic W decays and with several important background processes.

5.3. Background Samples

The expected rate of the signal process is small compared to the dominant Standard Model background processes. Therefore, the main challenge of the analysis is not the identification of the photon alone, but the suppression of the much larger background contributions. Background events can have a similar reconstructed topology if they contain a photon and hadronic activity. In particular, ordinary fermion production and four-fermion processes can enter the selected event sample.

The largest background class is radiative fermion-pair production. Events of the type $e^+e^- \to q\bar{q}\gamma$ contain a prompt photon together with hadronic activity and can therefore imitate the visible signature of the signal. They are especially important because their cross section is many orders of magnitude larger than the $H\gamma$ signal cross section. Radiative charged-lepton samples are also included. They are strongly suppressed by the lepton veto and jet requirements, but they can still enter if leptons are outside the reconstruction acceptance, are not isolated, or if tau decays produce hadrons.

Photon-associated four-fermion production gives another relevant background source. The $Z\gamma$ and $W\gamma$ categories can contain a photon together with hadronic jets from vector-boson decays. Depending on the decay mode, they may also contain neutrinos or isolated leptons; these parts are reduced by the missing-transverse- momentum requirement and the lepton veto, but hadronic or partially reconstructed events can remain. Finally, other $H\gamma$ decays are treated as background to this specific measurement. They share the same production photon and recoil kinematics as the target process, but the Higgs-boson decay is not the fully hadronic WW^* decay used for the signal definition.

The Monte Carlo samples used in the analysis are listed in Table 5.1. The table gives the exact sample names used by the analysis code, because these names define which simulated processes enter the signal and background categories. The sample identifiers, nominal

cross sections, k-factors, and Winter2023 Key4HEP release were checked against the FCC-ee `EventProducer` production dictionary at the fixed Git commit cited in Ref. [20]. A byte-identical copy of the dictionary and its SHA-256 checksum are included with the thesis source and documented in the accompanying metadata README. The software chain is documented through the primary references for Key4HEP, MG5_aMC@NLO, PYTHIA, and WHIZARD [21–25]. The production dictionary assigns a nominal k-factor of one to all listed samples and specifies the campaign, software release, and generator chain. Patch-level generator versions are not provided in the available metadata and are therefore omitted from Table 5.1.

In the supervised BDT training, the class label is denoted by y. The label $y = 1$ is reserved for the target signal, namely the $H\gamma$ sample corresponding to the fully hadronic $H \to WW^* \to q\bar{q}'\,q''\bar{q}'''$ final state. All other samples are treated as background and are assigned the label $y = 0$. Therefore, an $H\gamma$ sample with a Higgs-boson decay mode outside the target channel is not just "not counted as signal"; it enters the training and the yield calculation with the background label $y = 0$. The label is an analysis category and does not imply that all label-zero samples have the same physical origin.

Table 5.1.: Monte Carlo samples, generator setup, and nominal cross sections used in the analysis [1, 20].

Category	Monte Carlo sample names	Physics content	Generator, shower, and production version	Used $\sigma \times$ BR [pb]
Signal source	`mgp8_ee_ha_ecm160_hww`	$e^+e^- \to H\gamma$ with $H \to WW^*$. The target signal is the fully hadronic component $H \to WW^* \to q\bar{q}'\,q''\bar{q}'''$.	MG5_aMC@NLO + Pythia 8 winter2023, Key4HEP 2022-12-23 LO matrix element	4.07×10^{-6} inclusive $H \to WW^*$; 1.85×10^{-6} after $WW^* \to 4q$
Other $H\gamma$ decays	`mgp8_ee_ha_ecm160_hbb`	$e^+e^- \to H\gamma$ with Higgs-boson decay modes other than the target fully hadronic WW^* channel.	MG5_aMC@NLO + Pythia 8 winter2023, Key4HEP 2022-12-23 LO matrix element	$b\bar{b}$: 1.10×10^{-5}
	`mgp8_ee_ha_ecm160_hcc`			$c\bar{c}$: 5.46×10^{-7}
	`mgp8_ee_ha_ecm160_hgg`			gg: 1.55×10^{-6}
	`mgp8_ee_ha_ecm160_htautau`			$\tau\tau$: 1.19×10^{-6}
	`mgp8_ee_ha_ecm160_hzz`			ZZ^*: 4.99×10^{-7}
	`mgp8_ee_ha_ecm160_haa`			$\gamma\gamma$: 4.29×10^{-8}
	`mgp8_ee_ha_ecm160_hmumu`			$\mu\mu$: 4.11×10^{-9}
	`mgp8_ee_ha_ecm160_hss`			$s\bar{s}$: 4.54×10^{-9}
	`mgp8_ee_ha_ecm160_hdd`			$d\bar{d}$: 5.67×10^{-12}
	`mgp8_ee_ha_ecm160_huu`			$u\bar{u}$: 5.67×10^{-12}
Radiative two-fermion backgrounds	`wzp6_ee_qqa_ecm160`	Radiative fermion-pair production, separated by flavour in the listed samples.	WHIZARD + Pythia 6 winter2023, Key4HEP 2022-12-23 LO matrix element	$q\bar{q}\gamma$: 24.188
	`wzp6_ee_cca_ecm160`			$c\bar{c}\gamma$: 7.445
	`wzp6_ee_bba_ecm160`			$b\bar{b}\gamma$: 8.061
	`wzp6_ee_eea_ecm160`			$e^+e^-\gamma$: 920.5
	`wzp6_ee_mumua_ecm160`			$\mu^+\mu^-\gamma$: 2.818
	`wzp6_ee_tautaua_ecm160`			$\tau^+\tau^-\gamma$: 2.598

Table 5.1.: Monte Carlo samples, generator setup, and nominal cross sections used in the analysis (continued).

Category	Monte Carlo sample names	Physics content	Generator, shower, and production version	Used $\sigma \times$ BR [pb]

Four-fermion electroweak backgrounds	`wzp6_ee_aqqZ_ecm160`
	`wzp6_ee_allZ_ecm160`
	`wzp6_ee_aqqW_ecm160`
	`wzp6_ee_alnuW_ecm160`
		Photon-associated electroweak four-fermion production represented by the listed $\gamma q\bar{q}Z$, $\gamma\ell^+\ell^-Z$, $\gamma q\bar{q}W$, and $\gamma\ell\nu W$ samples.	WHIZARD + Pythia 6 winter2023, Key4HEP 2022-12-23 LO matrix element	$\gamma q\bar{q}Z$: 9.243×10^{-1}
				$\gamma\ell\ell Z$: 1.837×10^{-1}
				$\gamma q\bar{q}W$: 2.328×10^{-2}
				$\gamma\ell\nu W$: 1.387×10^{-2}

The signal sample for $H \to WW^*$ is inclusive in the decay modes of the two W bosons. Since this analysis targets the fully hadronic final state, the generator-level decay history is used to split this sample. An event is counted as signal only if both W bosons from the Higgs-boson decay into quark–antiquark pairs, corresponding to $H \to WW^* \to q\bar{q}'q''\bar{q}'''$. Events from the same inclusive $H \to WW^*$ sample with at least one leptonic W decay remain in the analysis as background with $y = 0$. With this event-by-event truth selection, the hadronic WW^* decay fraction enters through the selected fraction of the inclusive sample when the physical weights are summed.

The classifier is trained to learn differences between the reconstructed properties of the events with $y = 1$ and the events with $y = 0$. The same classification is used when summing the expected signal and background yields later in the analysis.

5.4. Event Weights and Expected Yields

The number of simulated events in a Monte Carlo sample is not identical to the number of events expected for a given integrated luminosity. For this reason, event weights are used. Each simulated event receives a physical weight that converts the simulated sample into the corresponding expected event yield.

For a process with an expected yield N_{expected} and a number of simulated events $N_{\text{simulated}}$, the event weight is defined as

$$w_i = \frac{N_{\text{expected}}}{N_{\text{simulated}}}. \tag{5.8}$$

Here w_i is the physical weight assigned to simulated event i, $N_{\text{expected}} = \sigma\mathcal{L}$ is the expected number of events for the process cross section σ and integrated luminosity $\mathcal{L}$, and $N_{\text{simulated}}$ is the number of generated Monte Carlo events for the same process. All yields quoted in the analysis section are normalized to $\mathcal{L} = 19.2$ ab^{-1}. All events from the same process

receive the same physical weight in this simplified treatment. The expected signal yield S and expected background yield B after a given selection are then obtained by summing the event weights:

$$S = \sum_{i \in \mathcal{S}} w_i, \qquad B = \sum_{i \in \mathcal{B}} w_i. \tag{5.9}$$

In Equation 5.9, $\mathcal{S}$ denotes the set of selected signal events and $\mathcal{B}$ denotes the set of selected background events according to the signal definition in Section 5.2.

Two simple quantities are used to monitor the performance of a selection. The first is the signal-to-background ratio,

$$\frac{S}{B}, \tag{5.10}$$

which measures the purity of the selected sample. The second is the approximate statistical sensitivity,

$$\frac{S}{\sqrt{B}}. \tag{5.11}$$

In these expressions, S and B are the weighted yields defined in Equation 5.9. The quantity $S/\sqrt{B}$ is useful for comparing different cuts and classifier score thresholds. However, it is only an approximate measure. The final statistical estimate is therefore obtained later from a binned Poisson likelihood using the shape of the BDT-score distribution and a fluctuation-free Asimov data set [26, 27].

5.5. Reconstructed Observables

The separation of signal and background is based on reconstructed event quantities. The most important observables describe the photon candidate, the hadronic activity, the recoil system, and the global event topology. The photon candidate is especially important because the signal process contains a characteristic photon from the two-body production process $e^+e^- \to H\gamma$.

The recoil mass m_{recoil} against the photon is defined by

$$m^2_{\text{recoil}} = \left(p_{e^+} + p_{e^-} - p_\gamma\right)^2. \tag{5.12}$$

Here p_{e^+} and p_{e^-} are the four-momenta of the incoming positron and electron, and p_γ is the four-momentum of the selected photon candidate. In the centre-of-mass frame of the initial electron–positron system, Equation 5.12 can be written as

$$m^2_{\text{recoil}} = \left(\sqrt{s} - E_\gamma\right)^2 - E_\gamma^2 = s - 2E_\gamma\sqrt{s}. \tag{5.13}$$

In Equation 5.13, E_γ is the energy of the signal photon, which is identical to the absolute value of the three-momentum of the photon because the photon is massless, and $\sqrt{s}$ is the centre-of-mass energy of the initial electron–positron system. For signal events, the system recoiling against the photon is the Higgs boson. Therefore, the recoil mass is expected to be consistent with the Higgs-boson mass of 125 GeV, within experimental resolutions.

The main groups of reconstructed variables used in the analysis are summarized in Table 5.2.

Table 5.2.: Main groups of reconstructed observables used to distinguish signal and background.

Variable group	Example variables	Physical purpose
Photon variables	Photon energy, photon polar angle, photon isolation	Identify the photon from $e^+e^- \to H\gamma$
Recoil variables	Recoil mass against the photon	Test whether the system recoiling against the photon is compatible with a Higgs boson
Jet and hadronic variables	Jet multiplicity, jet energies, visible hadronic activity	Describe the fully hadronic Higgs-boson decay $H \to WW^* \to 4q$
Lepton-veto variables	Number of isolated electrons and muons	Reject leptonic and semi-leptonic final states
Global event variables	Visible energy, missing transverse momentum, angular variables, event-shape quantities	Separate signal-like event topology from background topology

The variables in Table 5.2 are used in different parts of the analysis. Some variables are used directly in the cut-based preselection, while others are used as input variables for the multivariate classifier. Variables such as the photon energy and the recoil mass are useful because they are directly connected to the signal topology. They must be treated carefully because they are also strongly correlated with each other through Equation 5.13. The following section describes the cut-based preselection built from these observables.

5.6. Baseline Comparison of Signal and Background

Before applying the preselection that is defined in the course of this section, the reconstructed observables are compared for signal and background events. These comparisons give a first indication of which variables provide separation power and help to motivate

both the cut-based preselection and the choice of input variables for the multivariate classifier. In this context, "baseline" means the reconstructed event sample before applying the cut-based preselection defined in Section 5.6.1.

The histograms in Figure 5.1 are shown with the physical event-weight normalization. Keeping the relative normalization is useful here, because it shows not only the shape differences between signal and background, but also which background categories dominate the selected phase space.

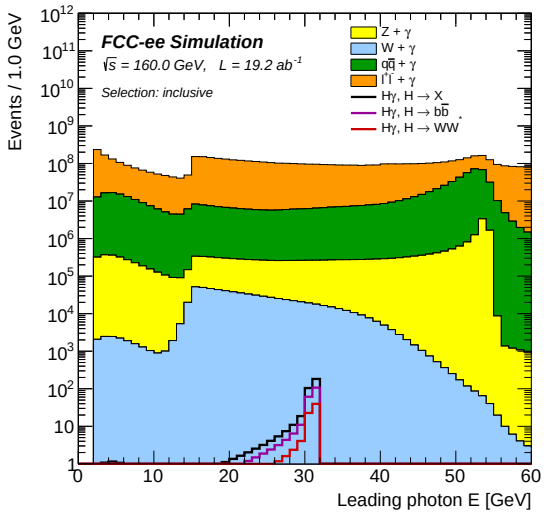
Figure 5.1 shows representative signal-background comparisons for important reconstructed observables. The leading-photon energy is one of the most characteristic variables of the signal process, since the two-body kinematics of $e^+e^- \to H\gamma$ predicts a photon energy close to 31 GeV at $\sqrt{s} = 160$ GeV. The recoil mass against the photon is also important, because for signal events the system recoiling against the photon should be consistent with the Higgs-boson mass of about 125 GeV within the experimental resolution. The number of isolated electrons and muons and the jet multiplicity are used to identify the zero-lepton hadronic topology. Photon isolation, which is not displayed in the figure, is used separately to suppress photon candidates embedded in hadronic activity.

The comparison shows that no single variable is sufficient to isolate the signal completely. The photon energy and recoil mass provide strong handles, but the recoil mass is almost fully determined by the photon energy in the centre-of-mass frame, as shown in Equation 5.13. It therefore does not provide an independent selection handle of the same size as a genuinely separate observable. The hadronic and lepton-veto variables further improve the separation, but the expected background remains large. Therefore, the following selection first applies a loose cut-based preselection and then uses a multivariate classifier to combine several observables simultaneously.

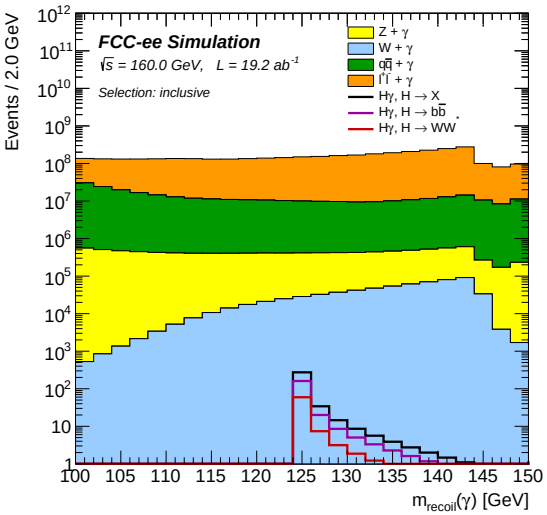
5.6.1. Cut-Based Preselection

After comparing the basic signal and background distributions, a loose cut-based preselection is applied before the multivariate analysis. The aim is not to define the final signal region already at this stage, but to remove events that are clearly incompatible with the expected topology while retaining enough signal events for training the BDT. The target channel is the zero-lepton hadronic final state introduced in Section 5.2.

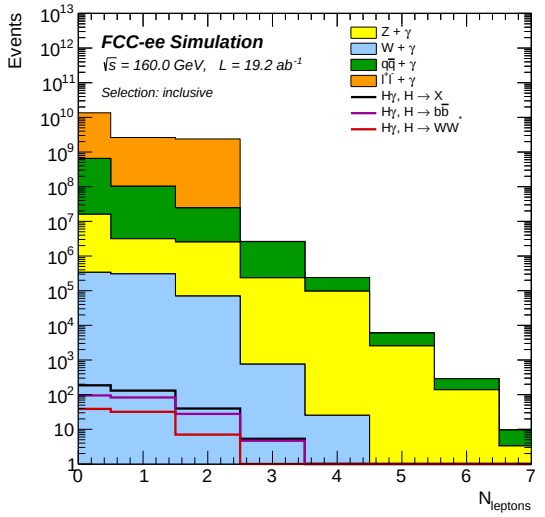
The preselection is kept deliberately loose, since the signal process is rare and only a limited number of signal events remains available after the baseline requirements. A very tight selection at this stage would therefore risk removing events that could still be useful for the later multivariate classification. All requirements that define the preselection are given explicitly in the following subsections and summarized in Table 5.3.



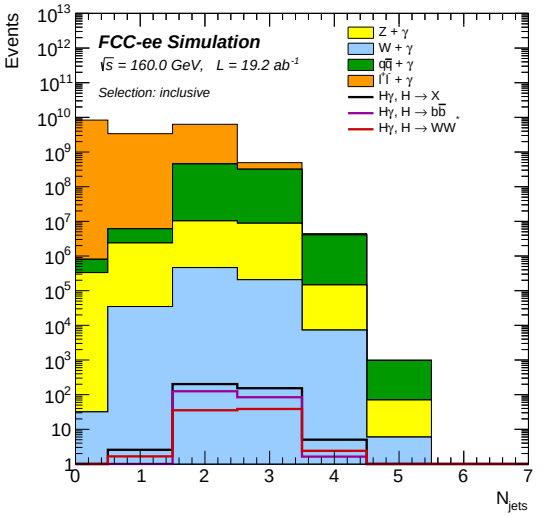
(a) Leading-photon energy



(b) Recoil mass against the photon



(c) Number of isolated veto leptons



(d) Jet multiplicity

Figure 5.1.: Baseline comparison of selected reconstructed observables for signal and background events with physical event-weight normalization. The leading-photon energy and the recoil mass are directly connected to the $e^+e^- \to H\gamma$ signal topology. The lepton and jet observables help to identify the zero-lepton hadronic final state.

Object Definitions

The reconstructed objects are taken from the Winter2023 IDEA fast-simulation samples, which use the DELPHES framework within the Key4HEP software stack [21, 28]. The object efficiencies and jet settings quoted below were checked against the IDEA detector card at the fixed Git commit cited in Ref. [29]. A byte-identical copy of that card and its SHA-256 checksum are included with the thesis source and documented in the accompanying metadata README.

The photon candidate is built from reconstructed photon objects. For a reconstructed candidate c, the relative cone-isolation variable used in this analysis is

$$I_{c,R} = \frac{1}{|\vec{p}_c|} \sum_{\substack{i \in \text{Reco} \\ 0.01 < \Delta R_{ic} < R}} |\vec{p}_i|, \qquad \Delta R_{ic} = \sqrt{(\eta_i - \eta_c)^2 + (\Delta\phi_{ic})^2}. \tag{5.14}$$

Here $\vec{p}_c$ and $\vec{p}_i$ are the three-momenta of the candidate and reconstructed particle i, respectively, R is the outer cone radius, η is the pseudorapidity, and $\Delta\phi$ is the periodic azimuthal-angle difference. The inner boundary removes the candidate itself. For the selected photon, relative isolation values with $R = 0.3$ and $R = 0.4$ are stored, together with the unnormalized cone-momentum sum for $0.01 < \Delta R < 0.4$. The preselection uses $I_{\gamma,0.4}$.

The lepton veto starts from the reconstructed `Electron` and `Muon` collections in the Winter2023 IDEA analysis input. The detector-card efficiency step gives non-zero efficiency only for candidates with energy of at least 2 GeV and $|\eta| \leq 3.0$. A reconstructed electron or muon is counted as a veto lepton if $|\vec{p}_\ell| > 5$ GeV and $I_{\ell,0.4} < 0.25$, where $\vec{p}_\ell$ is the lepton three-momentum and the isolation is defined by Equation 5.14 with $c = \ell$ and $R = 0.4$. The number of veto leptons is

$$n_{\text{veto}\,\ell} = n_{e,\text{iso}} + n_{\mu,\text{iso}}, \tag{5.15}$$

where $n_{e,\text{iso}}$ and $n_{\mu,\text{iso}}$ are the corresponding isolated-electron and isolated-muon multiplicities. No explicit veto on reconstructed hadronic τ candidates is applied; τ leptons enter the veto only if they decay into an isolated electron or muon, plus the corresponding neutrinos.

The jet multiplicity n_{jets} is the number of reconstructed jets used to describe the hadronic activity in the event. These jets are taken from the stored `Jet` collection of the Winter2023 IDEA samples. The collection is produced in the DELPHES reconstruction from particle-flow candidates using the e^+e^- generalized k_T jet algorithm with $p = -1$, radius parameter $R = 1.5$, and $p_T^{\text{jet}} > 1$ GeV [28–30]. No additional analysis-level jet-

energy or polar-angle requirement is applied before counting the jets in this preselection. The variable n_b denotes the number of these jets with a positive b-tag flag in the same reconstructed jet collection. The missing-transverse-momentum magnitude is read from the reconstructed `MissingET` collection and is denoted by

$$E_T^{\text{miss}} \equiv |\vec{p}_T^{\,\text{miss}}| = \sqrt{(p_x^{\text{miss}})^2 + (p_y^{\text{miss}})^2}.$$

Here p_x^{miss} and p_y^{miss} are the transverse Cartesian components of the missing-momentum vector. The visible energy is defined separately as $E_{\text{vis}} = \sum_i E_i$, where E_i is the energy of reconstructed visible particle i.

Photon Candidate Selection

The photon energy is one of the most useful handles in this analysis. For the two-body process $e^+e^- \to H\gamma$, it is fixed in the initial e^+e^- centre-of-mass system by the centre-of-mass energy and the Higgs-boson mass,

$$E_\gamma = \frac{s - m_H^2}{2\sqrt{s}}. \tag{5.16}$$

In Equation 5.16, E_γ is the photon energy in this frame, $s = (\sqrt{s})^2$, and m_H is the Higgs-boson mass. At $\sqrt{s} = 160$ GeV, this gives an expected photon energy of approximately 31 GeV. The photon energy distribution is therefore one of the most important variables for separating signal and background.

Events are required to satisfy

$$n_\gamma \geq 1. \tag{5.17}$$

where n_γ is the number of reconstructed photons passing the photon object requirements. If more than one photon is reconstructed in an event, the highest-energy photon is used as the candidate photon. The photon energy is required to lie in a window around the expected signal region,

$$20 \text{ GeV} < E_{\gamma,\text{cand}} < 33 \text{ GeV}. \tag{5.18}$$

In Equation 5.18, $E_{\gamma,\text{cand}}$ is the energy of the selected photon candidate. The lower boundary removes soft photons from background processes and detector activity, while the upper boundary suppresses high-energy photon backgrounds. This is especially useful because the signal photon is expected around 31 GeV and does not extend far above this region.

In addition, the photon is required to lie inside the detector acceptance,

$$|\cos\theta_{\gamma,\text{cand}}| < 0.99. \tag{5.19}$$

where $\theta_{\gamma,\text{cand}}$ is the polar angle of the selected photon candidate with respect to the beam axis. This removes photons that are very close to the beam direction, where the reconstruction is more difficult and where radiative background contributions are enhanced.

The photon is also required to be isolated from the hadronic activity in the event. This requirement uses the photon isolation variable $I_{\gamma,0.4}$ defined in Equation 5.14:

$$I_{\gamma,0.4} < 0.20. \tag{5.20}$$

This cut mainly rejects photons embedded in hadronic activity, for example photons from neutral-meson decays inside jets, while keeping isolated prompt-photon candidates from the hard process.

Recoil Mass Requirement

The recoil mass against the photon is computed from Equation 5.12. For signal events, the system recoiling against the photon is the Higgs boson, so the recoil mass is expected to be consistent with the Higgs-boson mass of 125 GeV within experimental resolutions. A recoil-mass window is applied:

$$123\ \text{GeV} < m^{\gamma}_{\text{recoil}} < 140\ \text{GeV}. \tag{5.21}$$

Here $m^{\gamma}_{\text{recoil}}$ denotes the recoil mass calculated with respect to the selected photon candidate. This requirement suppresses events in which the photon does not recoil against a system compatible with a Higgs boson. Since $m^{\gamma}_{\text{recoil}}$ is strongly anti-correlated with $E_{\gamma,\text{cand}}$ through Equation 5.13, the recoil-mass cut is mainly a consistency requirement and not an independent handle of the same strength as the photon-energy selection.

Hadronic Signal Region

The lepton-veto requirement is

$$n_{\text{veto}\,\ell} = 0. \tag{5.22}$$

Here $n_{\text{veto}\,\ell}$ is the number of isolated reconstructed electron and muon candidates defined in Equation 5.15, using the object definition stated in Section 5.6.1. The veto therefore acts on the stored reconstructed lepton collections, with the additional isolation criterion used

in this analysis. It is applied for two reasons. First, the target signal defined in Section 5.2 does not contain isolated electrons or muons from the hard final state, while several background processes do. Second, the zero-lepton selection makes this analysis orthogonal to analyses targeting semi-leptonic or fully leptonic Higgs-boson decay topologies.

The hadronic character of the selected events is enforced by requiring reconstructed jet activity:

$$n_{\text{jets}} \geq 2. \tag{5.23}$$

Here n_{jets} is the number of jets in the stored `Jet` collection defined in Section 5.6.1. For the Winter2023 IDEA samples, this collection is obtained from particle-flow candidates with the e^+e^- generalized k_T algorithm, using $p = -1$, $R = 1.5$, and $p_T^{\text{jet}} > 1$ GeV [28–30]. No further analysis-level jet-energy or polar-angle cut is imposed before the requirement in Equation 5.23.

Although the target parton-level final state contains four quarks, an exact four-jet requirement is not imposed at this stage. This is because the reconstructed jet multiplicity can differ from the parton multiplicity due to jet merging, soft radiation, detector effects, and the presence of an off-shell W boson in $H \to WW^*$. The looser requirement of at least two reconstructed jets keeps events with clear hadronic activity without removing too much signal before the multivariate analysis.

In addition, events with large missing transverse momentum are rejected:

$$E_T^{\text{miss}} < 40 \text{ GeV}. \tag{5.24}$$

Here E_T^{miss} is defined in Section 5.6.1. The fully hadronic signal is expected to be largely visible in the detector, whereas large missing transverse momentum is more typical of events with neutrinos in the final state.

A loose requirement on the number of b-tagged jets is also applied:

$$n_b \leq 2. \tag{5.25}$$

Here n_b is the number of reconstructed jets with a positive b-tag flag, as defined in Section 5.6.1. This requirement is only a loose rejection of events with a large number of b-tagged jets. It reduces some heavy-flavour dominated backgrounds without imposing a strict anti-b-tag selection, which would risk rejecting signal events because of mistags.

Summary of the Preselection

The full cut-based preselection is summarized in Table 5.3. The cuts are chosen to select events with an isolated photon, a Higgs-like recoil mass, and a hadronic zero-lepton final state.

Table 5.3.: Summary of the cut-based preselection used before the multivariate analysis.

Requirement	Selection cut
Photon multiplicity	$n_\gamma \geq 1$
Photon energy	$20\ \text{GeV} < E_{\gamma,\text{cand}} < 33\ \text{GeV}$
Photon angular acceptance	$\lvert\cos\theta_{\gamma,\text{cand}}\rvert < 0.99$
Photon isolation	$I_{\gamma,0.4} < 0.20$
Recoil mass	$123\ \text{GeV} < m^{\gamma}_{\text{recoil}} < 140\ \text{GeV}$
Lepton veto	$n_{\text{veto}\,\ell} = 0$
Jet multiplicity	$n_{\text{jets}} \geq 2$
Missing transverse momentum	$E_T^{\text{miss}} < 40\ \text{GeV}$
b-jet requirement	$n_b \leq 2$

The preselection is optimized to keep a sufficiently large number of signal events. This is important because the later BDT training requires enough signal events to learn the signal topology. At the same time, the cuts already remove a large fraction of events that are clearly incompatible with the fully hadronic $H\gamma$ signal.

Expected Yields after Preselection

The expected signal and background yields after the preselection are calculated by summing the physical event weights. For the inclusive $H \to WW^*$ sample, events are assigned to the signal category only when the generator-level decay chain contains two hadronically decaying W bosons from the Higgs boson. All other selected events, including non-hadronic $H \to WW^*$ decays and other $H\gamma$ decay modes, enter the background category.

The MC-statistical uncertainty of a weighted yield is calculated from the same selected events as

$$\delta N_{\text{MC}} = \sqrt{\sum_i w_i^2}, \tag{5.26}$$

where the sum runs over the selected simulated events in the corresponding category. The resulting preselection yields are summarized in Table 5.4. The rows use the same process grouping as the signal–background comparison plots and the later BDT background-composition table. The corresponding per-sample yields are given in Appendix A. The

total background uncertainty is obtained by adding the per-process MC-statistical uncertainties in quadrature.

Table 5.4.: Expected process-group yields after the cut-based preselection. The signal yield corresponds to the truth-selected fully hadronic $H \to WW^* \to 4q$ final state. The uncertainties are MC-statistical only and are calculated using Equation 5.26.

Process group	Preselection yield	Fraction of total B
Signal $H\gamma$, $H \to WW^* \to 4q$	31.22 ± 0.02	–
Radiative $q\bar{q}\gamma$ production	$(5.24 \pm 0.004) \times 10^7$	79.8%
Radiative charged-lepton backgrounds	$(1.14 \pm 0.013) \times 10^7$	17.4%
$Z\gamma$-associated backgrounds	$(1.68 \pm 0.00085) \times 10^6$	2.57%
$W\gamma$-associated backgrounds	$(7.71 \pm 0.0027) \times 10^4$	0.12%
Other $H\gamma$ decay modes	232.6 ± 0.1	$< 0.01\%$
Non-hadronic $H \to WW^*$ in the $H\gamma$ sample	9.380 ± 0.009	$< 0.01\%$
Total background	$(6.56 \pm 0.014) \times 10^7$	100%

The dominant background after the preselection is radiative $q\bar{q}\gamma$ production, which has the same basic photon-plus-hadronic-activity signature as the signal but a much larger cross section. Radiative charged-lepton production is the next-largest contribution at this stage, mainly because of its very large inclusive rate. The lepton veto and hadronic requirements suppress these samples, but a sizeable residual contribution remains after the loose preselection. The $Z\gamma$ and $W\gamma$ process groups contribute smaller but still non-negligible yields through photon-associated four-fermion topologies. The other $H\gamma$ decay modes and the non-hadronic part of the inclusive $H \to WW^*$ sample are kept as background because they are not part of the fully hadronic signal definition in Section 5.2.

The total preselection yields are therefore

$$S = 31.22, \qquad B = 6.56 \times 10^7, \tag{5.27}$$

which correspond to

$$\frac{S}{B} = 4.76 \times 10^{-7}, \qquad \frac{S}{\sqrt{B}} = 3.86 \times 10^{-3}. \tag{5.28}$$

The background yield remains very large after the cut-based selection. Therefore, the preselection alone is not sufficient for a precise measurement, and a multivariate classifier is introduced in the next section.

Although the signal-to-background ratio is still very small after the preselection, this step is useful because it defines a more signal-like phase-space region. The following multivariate analysis then combines several reconstructed observables to improve the signal-background separation beyond what is possible with simple one-dimensional cuts.

5.6.2. Multivariate Analysis

The preselection described in the previous section selects events with the basic properties expected from the signal: a photon, a Higgs-like recoil mass, and a hadronic final state without isolated charged leptons. However, the expected background yield is still much larger than the signal yield. A further separation based only on a few one-dimensional cuts is therefore not sufficient.

For this reason, a multivariate classifier is used. The aim is to combine several kinematic and event-level observables into one discriminating variable. In this analysis, the classifier output is called the BDT score. Events with a score close to one are more signal-like, while events with a score close to zero are more background-like. The score is not interpreted as a direct probability. It is used as a ranking variable to decide how signal-like an event is.

BDT Classifier and Training Setup

The classifier used in this thesis is a gradient-boosted decision-tree classifier. In practice, the model is implemented with the `HistGradientBoostingClassifier` from scikit-learn [31]. A boosted decision tree is well suited for this kind of analysis because it can combine several weakly separating variables and also make use of correlations between them.

The training sample is built from the events passing the loose preselection. After limiting each simulated process to at most 100 000 entries, the MVA input contains 1 808 540 unweighted simulated events: 76 997 truth-selected fully hadronic $H \to WW^* \to 4q$ signal events and 1 731 543 background events. A stratified 70/30 split gives 53 898 signal and 1 212 080 background events for training, and 23 099 signal and 519 463 background events for a held-out validation sample. This sample is not used to fit the classifier, but it is used to evaluate its performance and to optimize the score working point; it therefore does not constitute an independent test sample.

The model settings are summarized in Table 5.5. The parameters were chosen to keep the model flexible enough to separate signal and background, but not so flexible that it simply learns statistical fluctuations in the training sample.

Because the background yield is much larger than the signal yield, the training cannot only follow the absolute physical event weights. Otherwise, the loss function would be dominated by background events and the classifier would have little incentive to learn the rare signal topology. For the training itself, class balancing is therefore used: the total training weight of the signal class and the total training weight of the background class are rescaled to be equal. This rescaling is used only while fitting the classifier. For all

Table 5.5.: Main settings of the gradient-boosted decision-tree classifier.

Parameter	Value	Meaning
Number of boosting steps	350	Maximum number of trees added to the model
Learning rate	0.045	Size of the correction made by each tree
Maximum leaf nodes	31	Limits the complexity of each tree
Minimum samples per leaf	40	Prevents very small leaves and reduces overtraining
L2 regularization	0.1	Penalizes overly strong model weights
Early stopping	enabled	Stops the training if the validation performance no longer improves
Validation fraction	0.15	Internal validation fraction used during the training procedure

quoted yields, efficiencies, significances, and likelihood inputs, the original physical event weights are used again.

Input Variables

The BDT uses reconstructed quantities describing the selected photon candidate, the recoil system, the hadronic system, the event balance, missing transverse momentum, and object multiplicities. The variables used in the training are listed explicitly in Table 5.6. The selected photon candidate is the highest-energy reconstructed photon passing the photon-object requirements introduced in Section 5.6.1. The hadronic system is built from the four-momentum sum of all reconstructed jets in the stored jet collection. It is therefore not a constrained four-jet mass and it is still defined for events with fewer or more than four reconstructed jets.

Table 5.6.: Input variables used for the BDT training and their definitions.

Variable	Definition
`gamma_cand_e`	Photon candidate energy $E_{\gamma,\mathrm{cand}}$.
`gamma_cand_pt`	Photon candidate transverse momentum $p_{T,\gamma,\mathrm{cand}}$.
`gamma_cand_absCosTheta`	Absolute photon polar-angle variable $\lvert\cos\theta_{\gamma,\mathrm{cand}}\rvert$.
`gamma_cand_iso03`	Relative photon isolation $I_{\gamma,0.3}$ defined by Equation 5.14.
`gamma_cand_iso04`	Relative photon isolation $I_{\gamma,0.4}$ defined by Equation 5.14.
`gamma_cand_coneP04`	Unnormalized scalar momentum sum $\sum_i \lvert\vec{p}_i\rvert$ for $0.01 < \Delta R_{i\gamma} < 0.4$.
`n_photons`	Number of reconstructed photon candidates passing the photon-object requirements.

Table 5.6.: Input variables used for the BDT training and their definitions (continued).

Variable	Definition
`recoil_m_gamma`	Recoil mass against the selected photon, $m^{\gamma}_{\text{recoil}}$.
`n_electrons`	Number of reconstructed electron candidates.
`n_muons`	Number of reconstructed muon candidates.
`n_leptons`	Total reconstructed electron and muon multiplicity used as an event-level lepton-count variable.
`recoil_absdiff_125`	Distance of the photon recoil mass from the nominal Higgs-boson mass, $\lvert m^{\gamma}_{\text{recoil}} - 125\ \text{GeV}\rvert$.
`gamma_e_absdiff_signal`	Distance of the photon energy from the two-body signal value, $\lvert E_{\gamma,\text{cand}} - E^{\text{sig}}_{\gamma}\rvert$, with $E^{\text{sig}}_{\gamma} = (s - m_H^2)/(2\sqrt{s})$.
`had_m_absdiff_125`	Distance of the hadronic invariant mass from the nominal Higgs-boson mass, $\lvert m_{\text{had}} - 125\ \text{GeV}\rvert$.
`visible_energy_absdiff_160`	Distance of the visible energy from the centre-of-mass energy, $\lvert E_{\text{vis}} - 160\ \text{GeV}\rvert$.
`met_over_visible_energy`	Ratio $E_T^{\text{miss}}/E_{\text{vis}}$.
`met_over_ht_jets`	Ratio E_T^{miss}/H_T.
`gamma_pt_over_e`	Photon transverse-momentum fraction, $p_{T,\gamma,\text{cand}}/E_{\gamma,\text{cand}}$.
`had_m_recoil_absdiff`	Absolute difference between hadronic mass and photon recoil mass, $\lvert m_{\text{had}} - m^{\gamma}_{\text{recoil}}\rvert$.
`had_e_plus_gamma_e_absdiff_160`	Energy-closure variable, $\lvert E_{\text{had}} + E_{\gamma,\text{cand}} - 160\ \text{GeV}\rvert$.
`dphi_gamma_had_absdiff_pi`	Back-to-backness variable, $\lvert\pi - \Delta\phi(\gamma_{\text{cand}}, \text{had})\rvert$.
`gamma_had_pt_balance`	Vector transverse-momentum imbalance, $\lvert\vec{p}_{T,\gamma,\text{cand}} + \vec{p}_{T,\text{had}}\rvert$.
`gamma_had_pt_balance_over_sumpt`	Vector transverse-momentum imbalance divided by the scalar transverse-momentum sum, $\lvert\vec{p}_{T,\gamma,\text{cand}} + \vec{p}_{T,\text{had}}\rvert/(p_{T,\gamma,\text{cand}} + p_{T,\text{had}})$.
`had_pt_over_gamma_pt`	Ratio $p_{T,\text{had}}/p_{T,\gamma,\text{cand}}$.
`had_pt_minus_gamma_pt_abs`	Absolute difference $\lvert p_{T,\text{had}} - p_{T,\gamma,\text{cand}}\rvert$.
`n_jets`	Number of reconstructed jets in the stored jet collection.
`n_tracks`	Number of reconstructed tracks.
`n_charged_reco`	Number of reconstructed charged particle-flow objects.
`n_bjets`	Number of reconstructed jets with a positive b-tag flag.
`leading_jet_pt`	Transverse momentum of the highest-p_T reconstructed jet.
`had_m`	Invariant mass of the hadronic system formed from the vector sum of all reconstructed jets.
`ht_jets`	Scalar jet activity $H_T = \sum_j p_{T,j}$.
`dphi_gamma_had`	Azimuthal separation between the selected photon and the hadronic system, folded into $[0, \pi]$.
`had_e`	Energy of the hadronic system, built from the vector sum of all reconstructed jets.
`MET`	Magnitude of the reconstructed missing transverse momentum, $E_T^{\text{miss}} = \lvert\vec{p}_T^{\text{miss}}\rvert$.
`visible_energy`	Total reconstructed visible energy of the event.
`cosTheta_miss`	Absolute cosine of the polar angle of the missing-momentum vector, $\lvert\cos\theta_{\text{miss}}\rvert = \lvert p_z^{\text{miss}}\rvert/\lvert\vec{p}^{\text{miss}}\rvert$.

The input list is chosen from the reconstructed objects and from the expected signal topology. No separate feature-importance study is used to rank the variables in this thesis. Therefore, terms such as "important" refer to the physics motivation of a variable, not to a formal importance ordering from the trained classifier. Variables such as the photon energy and the recoil mass are included because they reflect the two-body process $e^+e^- \to H\gamma$. They are nevertheless interpreted with care, because they can strongly

shape the selected event sample.

Classifier Performance

The trained classifier is first checked on the held-out validation sample. The normalized BDT-score distributions for the training and validation samples are shown in Figure 5.2. Signal events are shifted towards high BDT scores, while background events are concentrated mainly at lower scores.

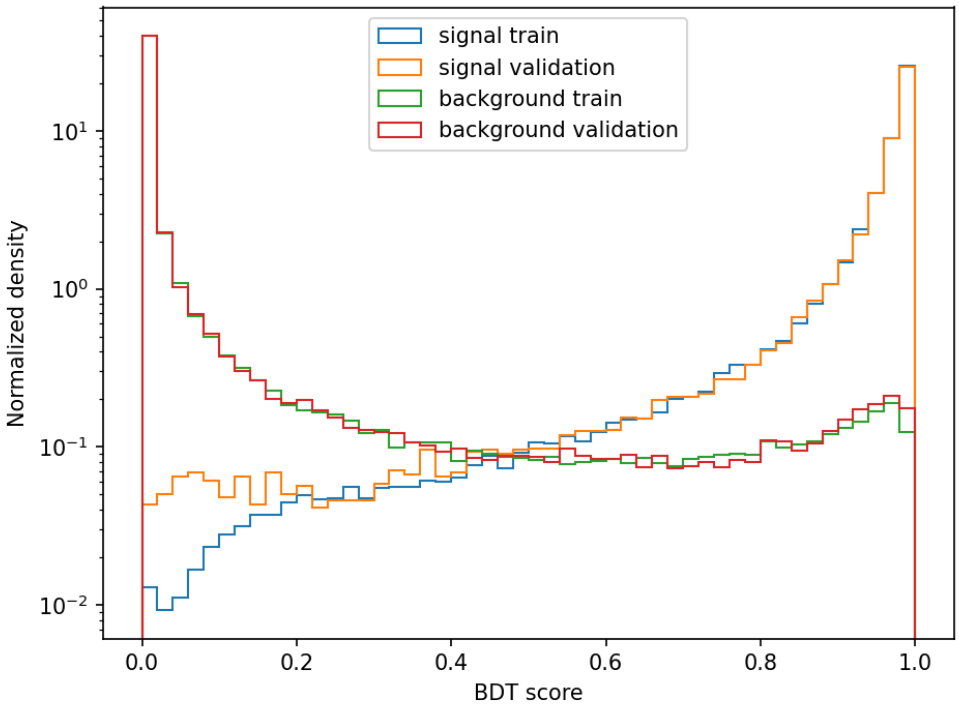


Figure 5.2.: Normalized BDT-score distributions for the training and validation samples. The broad agreement shows that the classifier response is stable, while small residual differences in the low- and high-score tails indicate mild overtraining. The held-out validation sample is used for the quoted classifier performance and for the score-shape study.

The training and validation curves agree over most of the score range. Small but visible differences remain for the signal at very low scores and for the background in the most signal-like tail. These differences are evidence of mild overtraining and the agreement is therefore not described as perfect. The effect is small compared with the overall signal–background separation, and the performance values quoted below are obtained from the held-out validation sample rather than from the training curve. The residual differences are nevertheless a limitation of the present classifier and should be monitored when the training or input-variable set is changed.

The agreement is also quantified using the receiver-operating-characteristic (ROC) curve. The ROC curve is obtained by scanning a cut on the BDT score and plotting the signal efficiency against the background false-positive rate. The area under this curve, denoted AUC, summarizes the ranking performance of the classifier. An AUC of 0.5 corresponds to random ordering, whereas an AUC of 1 corresponds to perfect separation. The obtained values are

$$\mathrm{AUC}_{\mathrm{train}} = 0.9923, \qquad \mathrm{AUC}_{\mathrm{validation}} = 0.9901. \tag{5.29}$$

The small reduction from the training to the validation AUC is consistent with the mild overtraining visible in Figure 5.2. The validation AUC is nevertheless high, so the classifier has strong ranking power. The AUC is used only as a classifier-quality check; the expected physics sensitivity is evaluated from the physically weighted score distribution and the likelihood fit.

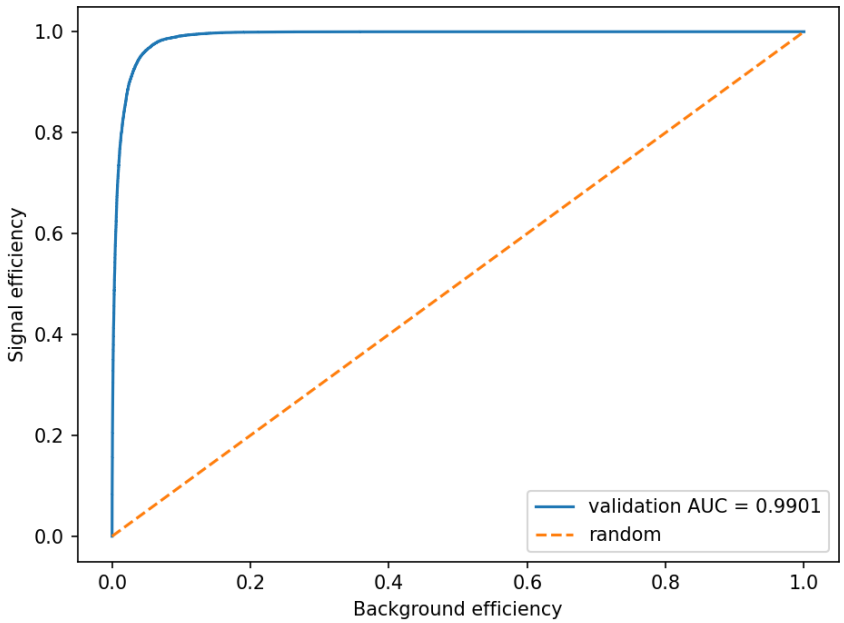


Figure 5.3.: ROC curve of the BDT classifier evaluated on the held-out validation sample. The validation AUC is 0.9901.

The normalized curves in Figure 5.2 do not show the large difference between the physical signal and background normalizations. The high-score part of the event-based application output is therefore shown in Figure 5.4. Each preselected event contributes with its stored physical weight. The background is separated into the major process groups, while the signal is multiplied by 10^4 only to make its shape visible on the same logarithmic scale. The 50 equal-width bins between 0.90 and 1.00 are used for visualization

and are not the bins of the likelihood fit.

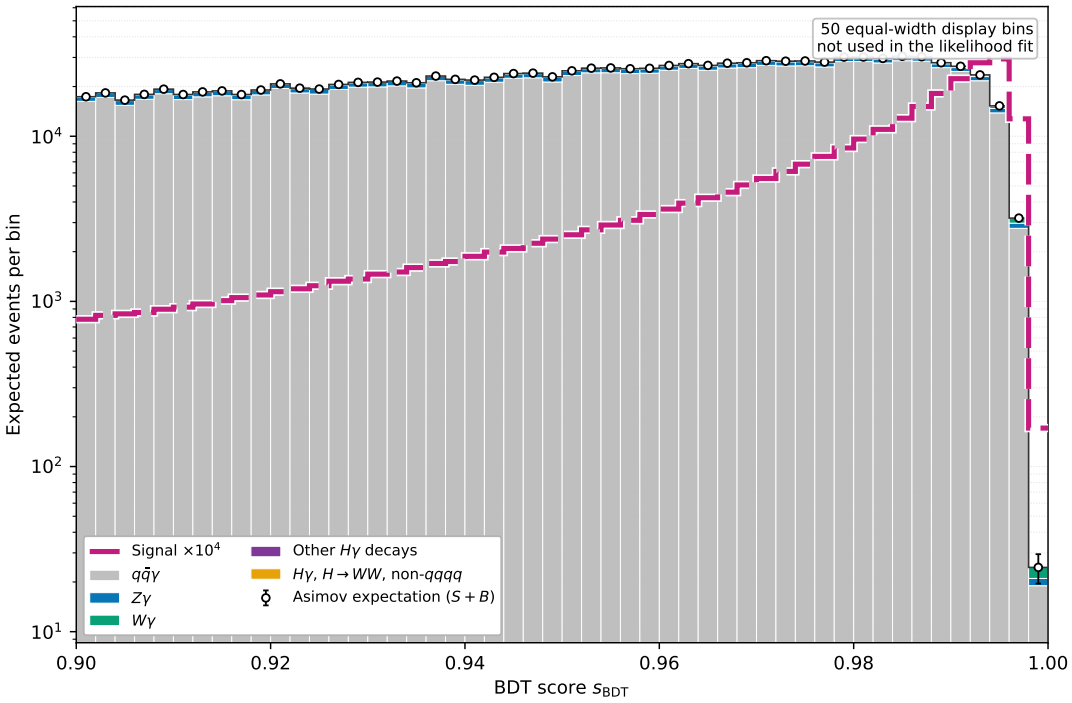


Figure 5.4.: Event-based, physics-weighted BDT-score distribution in the high-score region after the preselection. The backgrounds are stacked by process group. The signal line is multiplied by 10^4 for visibility, and the open points show the unscaled Asimov expectation $S + B$. The 50 equal-width display bins are not used in the likelihood fit.

The weighted distribution illustrates why a high AUC does not automatically lead to a sensitive measurement. The signal fraction increases towards high scores, but the much larger radiative $q\bar{q}\gamma$ component remains dominant throughout the displayed range. The event-level output also makes it possible to keep the background categories separate. Their integrated yields in the fit range are listed in Table A.4.

Tight Counting Benchmark

To summarize the rejection power of the BDT with one number, different lower cuts on the score are scanned. For each cut value, the expected signal yield S, the expected background yield B, the signal-to-background ratio S/B, and the approximate significance $S/\sqrt{B}$ are calculated. Cut points with less than one expected signal event are excluded from this benchmark because they give an unstable counting interpretation. The stable part of the scan is shown in Figure 5.5.

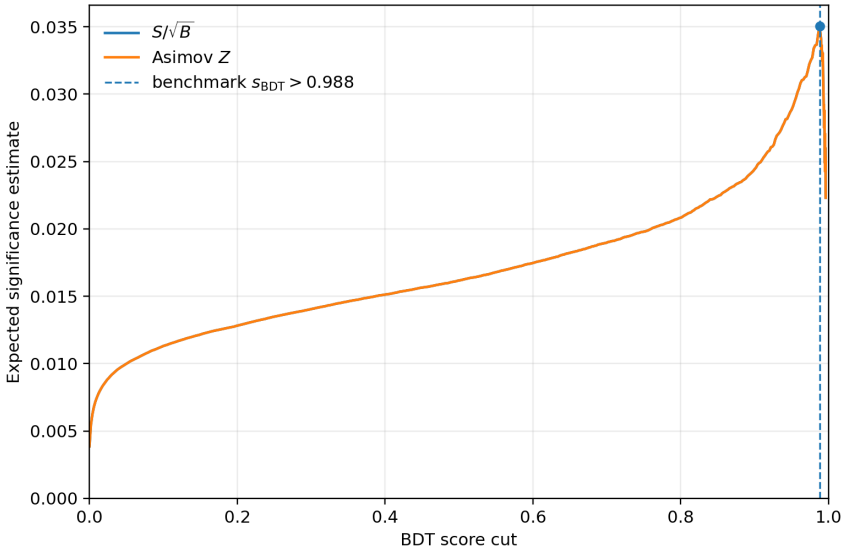


Figure 5.5.: Expected sensitivity as a function of the BDT-score cut in the stable scan region with at least one expected signal event. Both $S/\sqrt{B}$ and the Asimov significance are shown. The tight counting benchmark is $s_{\mathrm{BDT}} > 0.988$.

The threshold used for the tight counting benchmark is

$$s_{\mathrm{BDT}} > 0.988. \tag{5.30}$$

Very tight cuts can give unstable scan points when only a tiny expected signal yield remains. The value in Equation 5.30 lies inside the stable part of the scan and gives the largest $S/\sqrt{B}$ there with the physical event weights. Starting from the preselection, it keeps about 35.5% of the signal and about 1.47×10^{-3} of the background.

This tight threshold is used to quote a one-bin benchmark and to display the remaining background composition. It is not used as the lower boundary of the shape-likelihood fit in Section 5.6.3, because such a fit should retain the information in a wider part of the BDT-score distribution.

The tight benchmark improves the signal-to-background ratio by roughly a factor of 2.4×10^2 and $S/\sqrt{B}$ by a factor of about 9.2. The dominant remaining background is radiative $q\bar{q}\gamma$ production, which accounts for approximately 92% of the background above the tight threshold.

5.6.3. Shape-Based Likelihood Fit

The tight counting benchmark is useful for comparing selections, but reducing the BDT output to the single region $s_{\mathrm{BDT}} > 0.988$ discards information from events at lower scores.

Table 5.7.: Expected signal and background yields after the preselection and after the tight counting benchmark $s_{\mathrm{BDT}} > 0.988$. The background is shown by process group. The uncertainties are MC-statistical only. A zero entry means that no simulated event from that group passed the tight threshold.

Process category	After preselection	After $s_{\mathrm{BDT}} > 0.988$
Signal $H\gamma$, $H \to WW^* \to 4q$	31.22 ± 0.02	11.07 ± 0.01
Radiative $q\bar{q}\gamma$ production	$(5.24 \pm 0.004) \times 10^7$	$(8.90 \pm 0.16) \times 10^4$
$Z\gamma$-associated backgrounds	$(1.68 \pm 0.00085) \times 10^6$	5201 ± 49
$W\gamma$-associated backgrounds	$(7.71 \pm 0.0027) \times 10^4$	2137 ± 5
Other $H\gamma$ decay modes	232.6 ± 0.1	23.50 ± 0.02
Non-hadronic $H \to WW^*$ in the $H\gamma$ sample	9.380 ± 0.009	0.2670 ± 0.0014
Radiative charged-lepton backgrounds	$(1.14 \pm 0.013) \times 10^7$	0
Total background, B	$(6.56 \pm 0.014) \times 10^7$	$(9.64 \pm 0.16) \times 10^4$
S/B	4.76×10^{-7}	1.15×10^{-4}
$S/\sqrt{B}$	3.86×10^{-3}	3.56×10^{-2}

The statistical interpretation therefore uses a binned shape likelihood in the looser interval

$$0.90 \leq s_{\mathrm{BDT}} \leq 1.00. \tag{5.31}$$

The input is taken directly from the event-level scored ROOT files produced in the final BDT-application step. For every preselected event, these files contain the BDT score, the physical event weight, the signal label, and a process-group identifier. In the inclusive $H \to WW^*$ sample, the signal label is assigned event by event using `is_hww_qqqq`; consequently, no additional $\mathrm{BR}(W \to \mathrm{hadrons})^2$ factor is applied to the signal template. The normalized training histogram is not used as the fit template.

The adaptive binning is constructed directly from the weighted signal events. First, the range in Equation 5.31 is divided into 1000 fine bins of width 10^{-4}. Beginning at $s_{\mathrm{BDT}} = 1$, adjacent fine bins are merged towards lower scores until the target yield of approximately one twentieth of the total signal is reached. The same final boundaries are then used for the signal and every background group. Each of the 20 final bins contains at least one simulated signal event and at least one simulated background event. Therefore, no neighbouring bins need to be combined further to avoid empty template bins. The exact boundaries and yields are given in Table A.5.

The resulting likelihood templates are shown in Figure 5.6. The varying widths are intentional: each bin contains approximately the same expected signal yield. The signal line therefore appears almost flat after multiplication by 10^5, whereas the background falls substantially towards the highest scores.

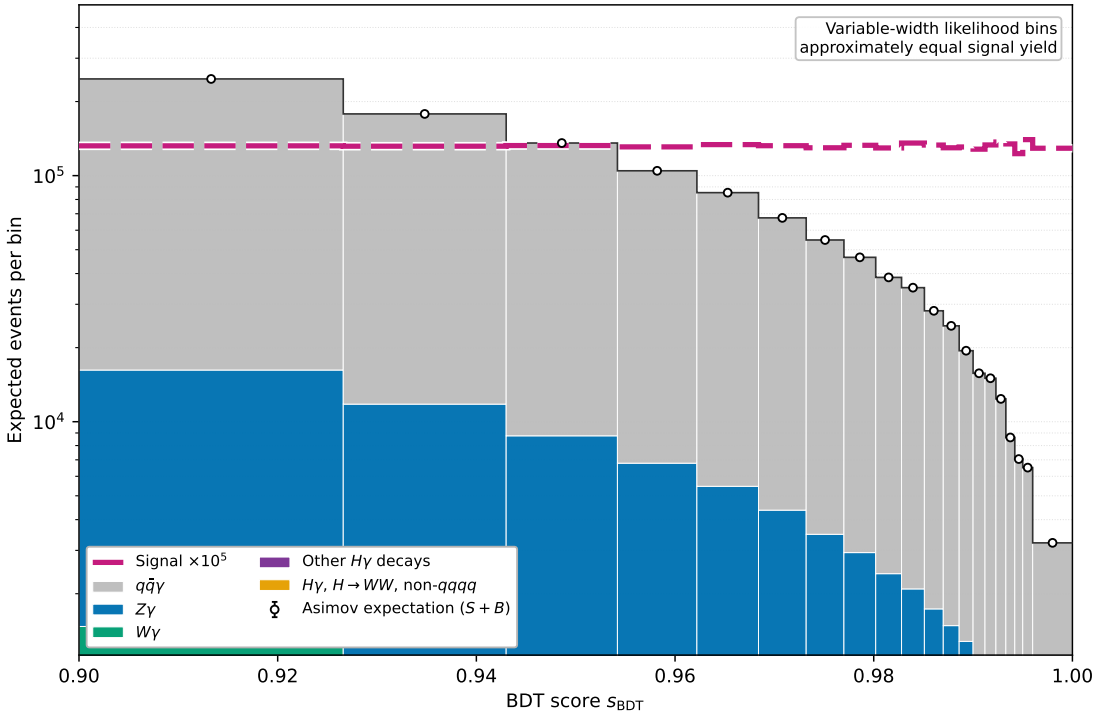


Figure 5.6.: Event-based templates in the 20 variable-width bins used by the likelihood. The background categories are stacked and are summed in the statistical model. The signal is multiplied by 10^5 for visibility; the open points show the unscaled Asimov expectation $S + B$. The nearly constant scaled signal height follows from the approximately equal signal yield per bin.

The main properties of the likelihood input are summarized in Table 5.8. The score range contains an expected signal yield of 26.40 and an expected background yield of 1.13×10^6. The average signal yield per final bin is 1.32. Radiative $q\bar{q}\gamma$ production accounts for 93.4% of the background in this interval.

Table 5.8.: Summary of the event-based adaptive BDT-score binning used in the shape likelihood.

Quantity	Value
BDT-score range	$0.90 \leq s_{\mathrm{BDT}} \leq 1.00$
Initial number of fine bins	1000
Final number of adaptive bins	20
Total expected signal yield	26.40
Total expected background yield	1.13×10^6
Target signal yield per bin	1.32
Total S/B in the fit range	2.33×10^{-5}

The signal-strength parameter is defined as the normalization of the selected signal

relative to the Standard Model expectation,

$$\mu = \frac{\sigma(e^+e^- \to H\gamma)\,\mathrm{BR}(H \to WW^* \to 4q)}{[\sigma(e^+e^- \to H\gamma)\,\mathrm{BR}(H \to WW^* \to 4q)]_{\mathrm{SM}}}. \tag{5.32}$$

If the Higgs- and W-boson branching fractions are fixed to their Standard Model values, this is equivalent to the production-cross-section ratio. For bin k, the expected yield is

$$\nu_k(\mu) = \mu S_k + B_k, \qquad B_k = \sum_g B_{k,g}, \tag{5.33}$$

where g denotes the background groups shown in Figure 5.6. The binned likelihood is

$$L(\mu) = \prod_{k=1}^{20} \mathrm{Pois}(N_k \,|\, \mu S_k + B_k)\,. \tag{5.34}$$

For $\mu = 1$, a fluctuation-free Asimov data set fixes each bin count to its expected value:

$$N_{k,\mathrm{A}} = S_k + B_k. \tag{5.35}$$

Although the components are retained separately for the templates and the stacked plots, the present statistical-only likelihood uses their fixed sum. No background-normalization, shape, or Monte Carlo statistical nuisance parameters are included.

The one-standard-deviation interval for the single fit parameter is obtained from

$$-2\Delta \ln L(\mu) = 1. \tag{5.36}$$

The resulting scan is shown in Figure 5.7.

The Asimov data set gives the expected best-fit value

$$\hat{\mu} = 1. \tag{5.37}$$

Without imposing the physical boundary, the one-standard-deviation interval is

$$\mu = 1.0^{+21.9}_{-21.8}. \tag{5.38}$$

The corresponding unconstrained intersections of the likelihood scan with the line $-2\Delta \ln L(\mu) = 1$ are at approximately -20.8 and 22.9. With the physical constraint $\mu \geq 0$, the lower side reaches the boundary at zero. The expected relative statistical

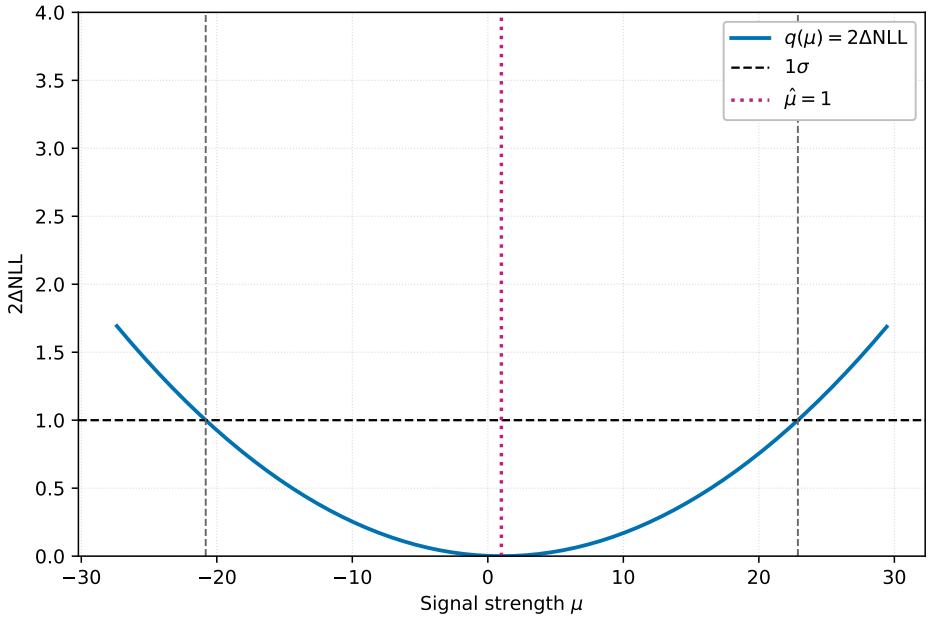


Figure 5.7.: Expected statistical-only likelihood scan for the event-based 20-bin BDT-shape fit. The scan is extended below the physical boundary $\mu = 0$ to display both unconstrained one-standard-deviation crossings. In the physical fit, the lower interval ends at $\mu = 0$.

uncertainty is therefore still of order

$$\frac{\Delta\sigma}{\sigma} \simeq \Delta\mu \approx 21.9, \tag{5.39}$$

or about $2.2 \times 10^3\,\%$ when the branching fractions are treated as fixed.

For fixed background expectations, the multi-bin Asimov discovery significance is

$$Z_{\mathrm{A,shape}} = \left[2\sum_{k=1}^{20}\left((S_k + B_k)\ln\left(1 + \frac{S_k}{B_k}\right) - S_k\right)\right]^{1/2} = 0.0458. \tag{5.40}$$

Treating the whole interval $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$ as one counting bin gives $Z_{\mathrm{A}} = 0.0248$, so the score shape improves the expected sensitivity by a factor of 1.85. Relative to the tight $s_{\mathrm{BDT}} > 0.988$ counting benchmark, whose corresponding Asimov value is approximately 0.0357, the gain is about 28%. Nevertheless, the expected sensitivity remains very small because the background is much larger than the signal in every bin. Since systematic uncertainties are omitted, the quoted precision should be interpreted as an optimistic statistical-only estimate.

5.6.4. Discussion

The fully hadronic $H \to WW^* \to q\bar{q}'\, q''\bar{q}'''$ channel is experimentally challenging. The visible signature is simple in principle: an isolated photon from the production process and hadronic activity from the Higgs-boson decay. In practice, the signal rate is very small, while several Standard Model processes produce photon-plus-jets final states with similar reconstructed properties.

The signal label is defined from the generator-level Higgs-boson and W-boson decay history. Only the fully hadronic $H \to WW^*$ component is labelled as signal; other $H\gamma$ decays and non-hadronic $H \to WW^*$ events are treated as background. This keeps the training labels, event-level application labels, physical weights, and statistical interpretation consistent.

The BDT has strong ranking power, with a validation AUC of 0.9901. The training–validation comparison nevertheless shows a small difference in the highest-score tail, indicating mild overtraining. The held-out validation sample is used for the quoted AUC and classifier-quality checks. The physical yields and the final likelihood templates are obtained from the event-level scored application files.

After the cut-based preselection, the expected yields are

$$S = 31.22, \qquad B = 6.56 \times 10^7, \tag{5.41}$$

which gives

$$\frac{S}{B} = 4.76 \times 10^{-7}, \qquad \frac{S}{\sqrt{B}} = 3.86 \times 10^{-3}. \tag{5.42}$$

At the tight counting benchmark $s_{\mathrm{BDT}} > 0.988$, the corresponding values are

$$S = 11.07, \qquad B = 9.64 \times 10^4, \tag{5.43}$$

and

$$\frac{S}{B} = 1.15 \times 10^{-4}, \qquad \frac{S}{\sqrt{B}} = 3.56 \times 10^{-2}. \tag{5.44}$$

The signal-to-background ratio therefore improves by about 2.4×10^2, and the approximate counting sensitivity improves by a factor of about 9.2. Radiative $q\bar{q}\gamma$ production still contributes approximately 92% of the background above the tight threshold.

The final statistical estimate retains the event-level score information over $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$. The exact adaptive 20-bin likelihood gives

$$\mu = 1.0^{+21.9}_{-21.8} \tag{5.45}$$

before the physical lower boundary is imposed, and an expected Asimov significance of

$$Z_{\mathrm{A}} = 0.0458. \tag{5.46}$$

The shape information increases the expected significance by a factor of 1.85 relative to using the same loose score range as a single bin, but the expected cross-section uncertainty remains of order $2.2 \times 10^3\%$. The channel is therefore still strongly background limited. With the present selection and classifier, this channel alone does not provide a meaningful statistical constraint on the Standard Model signal strength at the assumed luminosity.

The event-based implementation uses the stored scores directly and provides process-resolved templates. Important limitations remain. The current likelihood fixes the total background and includes neither systematic uncertainties nor component-specific nuisance parameters. It also does not propagate the finite weighted Monte Carlo statistics as bin-by-bin nuisance parameters. All 20 bins have non-zero simulated support, and the effective background statistics remain above about 1.7×10^2 events per bin. A final analysis should nevertheless propagate these finite weighted Monte Carlo statistics through bin-wise nuisance parameters. In addition, a separately generated or untouched application sample would make the final sensitivity estimate independent of the samples used during classifier development.

The most promising physics improvement is a more detailed reconstruction of the hadronic Higgs-boson decay. Variables based on jet pairing, reconstructed dijet masses, angular separations, and the compatibility of the two jet pairs with an on-shell and an off-shell W boson could provide additional separation. A dedicated classifier or control region for radiative $q\bar{q}\gamma$ production may also be more effective than treating all backgrounds inclusively. The process-resolved templates produced here provide the starting point for a profile-likelihood model with normalization and shape constraints.

6. Conclusion and Outlook

This thesis studies $e^+e^- \to H\gamma$ at $\sqrt{s} = 160$ GeV in the fully hadronic channel $H \to WW^* \to 4q$. A generator-level requirement on the Higgs- and W-boson decay chain defines the target signal, while other $H\gamma$ decay modes and non-hadronic $H \to WW^*$ events are treated as background. For 19.2 ab^{-1}, the cut-based preselection retains an expected 31.22 signal events together with 6.56×10^7 background events.

A gradient-boosted decision-tree classifier was trained with reconstructed photon, recoil, hadronic, event-balance, and missing-transverse-momentum observables. Its validation AUC is 0.9901. The tight counting benchmark $s_{\mathrm{BDT}} > 0.988$ retains $S = 11.07$ and $B = 9.64 \times 10^4$, improving S/B by about a factor of 2.4×10^2 and $S/\sqrt{B}$ by about a factor of 9.2 relative to the preselection. Radiative $q\bar{q}\gamma$ production nevertheless remains the dominant background.

The final statistical estimate uses the event-level score distribution rather than a single tight score threshold. In the range $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$, 1000 fine starting bins are merged into 20 variable-width bins with approximately equal expected signal yields. The statistical-only shape likelihood gives $\mu = 1.0^{+21.9}_{-21.8}$ before imposing the physical boundary and an expected Asimov significance of $Z_{\mathrm{A}} = 0.0458$. Although the score shape improves the sensitivity by a factor of 1.85 relative to a one-bin use of the same range, the expected relative cross-section uncertainty remains of order $2.2 \times 10^3\%$.

The analysis therefore establishes a reproducible event-level BDT and shape-fit chain, but the present fully hadronic channel is strongly background limited. A realistic continuation should profile background-normalization and score-shape uncertainties, include bin-wise Monte Carlo statistical nuisance parameters, and reserve an untouched application sample. More detailed reconstruction of the two hadronic W candidates, together with dedicated control regions or rejection methods for $q\bar{q}\gamma$, offers the clearest route to improved sensitivity and to a future combination with other $H\gamma$ decay channels.

A. Detailed Yield and Shape-Fit Tables

This appendix collects the numerical inputs used for the preselection, the tight counting benchmark, and the final BDT-shape likelihood. The preselection and $s_{\mathrm{BDT}} > 0.988$ tables are taken from the final MVA application output. Their quoted uncertainties are MC-statistical uncertainties calculated with Equation 5.26; systematic and generator-normalization uncertainties are not included.

The shape-fit material is obtained directly from the event-level scored ROOT files. Each entry provides the BDT score, physical weight, signal label, and process group. The adaptive boundaries are therefore derived from the actual weighted signal distribution, not from an interpolation of cumulative cut yields. The signal already consists of the truth-selected $H \to WW^* \to 4q$ events, so no additional hadronic W-decay branching factor is applied. The 20 final bins all contain simulated signal and background events. The composition and main reproducibility settings of the final BDT training are summarized in Table A.1.

Table A.1.: Composition and principal reproducibility settings of the final BDT training. Counts are unweighted simulated entries after the per-process input limit; physical weights are retained for expected yields.

Quantity	Value
Integrated luminosity	19.2 ab^{-1}
Input construction	At most 100 000 entries per simulated process; 1 808 540 events in total, comprising 76 997 signal and 1 731 543 background events
Training sample (70%)	1 265 978 events: 53 898 signal and 1 212 080 background
Held-out validation sample (30%)	542 562 events: 23 099 signal and 519 463 background
Classifier setup	37 input variables; random seed 12345; scikit-learn 1.4.0 `HistGradientBoostingClassifier`
Fitting procedure	Equal total signal and background training weights; internal early-stopping validation fraction of 15% of the training sample

Table A.2.: Expected yields by process group before and after the tight BDT counting benchmark. The quoted uncertainties are MC-statistical uncertainties.

Process group	Label	Preselection yield	BDT yield	BDT efficiency
Signal $H\gamma, H \to WW^* \to 4q$	1	31.22 ± 0.02	11.07 ± 0.01	0.355
Radiative $q\bar{q}\gamma$ production	0	$(5.24 \pm 0.004) \times 10^7$	$(8.9 \pm 0.16) \times 10^4$	1.70×10^{-3}
$Z\gamma$-associated backgrounds	0	$(1.68 \pm 0.00085) \times 10^6$	5201 ± 49	3.09×10^{-3}
$W\gamma$-associated backgrounds	0	$(7.71 \pm 0.0027) \times 10^4$	2137 ± 5	0.028
Other $H\gamma$ decay modes	0	232.6 ± 0.1	23.50 ± 0.02	0.101
Non-hadronic $H \to WW^*$ in $H\gamma$ sample	0	9.380 ± 0.009	0.2670 ± 0.0014	0.028
Radiative charged-lepton backgrounds	0	$(1.14 \pm 0.013) \times 10^7$	0	0
Total signal	1	31.22 ± 0.02	11.07 ± 0.01	0.355
Total background	0	$(6.56 \pm 0.014) \times 10^7$	$(9.64 \pm 0.16) \times 10^4$	1.47×10^{-3}

Table A.3.: Per-process yields before and after the tight BDT counting benchmark. The preselection column corresponds to the selected sample before applying the BDT score requirement. The BDT column uses the tight counting benchmark $s_{\mathrm{BDT}} > 0.988$. Uncertainties are MC-statistical only.

Process	Group	Preselection yield	BDT yield	BDT efficiency
`mgp8_ee_ha_ecm160_hww`	Signal $H\gamma, H \to WW^* \to 4q$	31.22 ± 0.02	11.07 ± 0.01	0.355
`wzp6_ee_qqa_ecm160`	Radiative $q\bar{q}\gamma$ production	$(3.15 \pm 0.0038) \times 10^7$	$(5.14 \pm 0.15) \times 10^4$	1.63×10^{-3}
`wzp6_ee_cca_ecm160`	Radiative $q\bar{q}\gamma$ production	$(1.1 \pm 0.0013) \times 10^7$	$(1.99 \pm 0.053) \times 10^4$	1.81×10^{-3}
`wzp6_ee_bba_ecm160`	Radiative $q\bar{q}\gamma$ production	$(9.91 \pm 0.0039) \times 10^6$	$(1.77 \pm 0.017) \times 10^4$	1.79×10^{-3}
`wzp6_ee_aqqZ_ecm160`	$Z\gamma$-associated backgrounds	$(1.47 \pm 0.00084) \times 10^6$	4951 ± 49	3.36×10^{-3}
`wzp6_ee_allZ_ecm160`	$Z\gamma$-associated backgrounds	$(2.09 \pm 0.0011) \times 10^5$	250.0 ± 3.8	1.19×10^{-3}
`wzp6_ee_aqqW_ecm160`	$W\gamma$-associated backgrounds	$(6.35 \pm 0.0025) \times 10^4$	2100 ± 5	0.033
`wzp6_ee_alnuW_ecm160`	$W\gamma$-associated backgrounds	$(1.35 \pm 0.00077) \times 10^4$	36.72 ± 0.40	2.71×10^{-3}
`mgp8_ee_ha_ecm160_hbb`	Other $H\gamma$ decay modes	182.6 ± 0.1	11.18 ± 0.02	0.061
`mgp8_ee_ha_ecm160_hgg`	Other $H\gamma$ decay modes	26.21 ± 0.03	10.58 ± 0.02	0.404
`mgp8_ee_ha_ecm160_hzz`	Other $H\gamma$ decay modes	6.413 ± 0.008	1.087 ± 0.003	0.169
`mgp8_ee_ha_ecm160_hcc`	Other $H\gamma$ decay modes	9.183 ± 0.010	0.6333 ± 0.0026	0.069
`mgp8_ee_ha_ecm160_haa`	Other $H\gamma$ decay modes	0.1722 ± 0.0004	0.007856 ± 0.000080	0.046
`mgp8_ee_ha_ecm160_hss`	Other $H\gamma$ decay modes	0.07664 ± 0.00008	0.007332 ± 0.000025	0.096
`mgp8_ee_ha_ecm160_hdd`	Other $H\gamma$ decay modes	$(1.56 \pm 0.0017) \times 10^{-4}$	$(1.47 \pm 0.0051) \times 10^{-5}$	0.094
`mgp8_ee_ha_ecm160_huu`	Other $H\gamma$ decay modes	$(3.11 \pm 0.0033) \times 10^{-5}$	$(2.93 \pm 0.01) \times 10^{-6}$	0.094
`mgp8_ee_ha_ecm160_htautau`	Other $H\gamma$ decay modes	7.935 ± 0.013	0	0
`mgp8_ee_ha_ecm160_hmumu`	Other $H\gamma$ decay modes	$(5.61 \pm 0.21) \times 10^{-5}$	0	0
`mgp8_ee_ha_ecm160_hww`	Non-hadronic $H \to WW^*$ in $H\gamma$ sample	9.380 ± 0.009	0.2670 ± 0.0014	0.028
`wzp6_ee_mumua_ecm160`	Radiative charged-lepton backgrounds	$(1.26 \pm 0.026) \times 10^4$	0	0
`wzp6_ee_tautaua_ecm160`	Radiative charged-lepton backgrounds	$(1.74 \pm 0.0029) \times 10^6$	0	0
`wzp6_ee_eea_ecm160`	Radiative charged-lepton backgrounds	$(9.67 \pm 0.13) \times 10^6$	0	0

Table A.4.: Process-group composition of the event-based background template in the shape-fit range $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$. The groups are filled separately for the stacked figures and then summed to form B_k in the likelihood.

Background group	Expected yield	Fraction of B
Radiative $q\bar{q}\gamma$ production	1.06×10^6	93.44%
$Z\gamma$-associated backgrounds	64 355	5.68%
$W\gamma$-associated backgrounds	9 896	0.873%
Other $H\gamma$ decay modes	85.1	0.0075%
Non-hadronic $H \to WW^*$ in the $H\gamma$ sample	1.828	0.00016%
Total background	1.13×10^6	100.00%

Table A.5.: Event-based adaptive BDT-score bins used in the statistical-only shape likelihood. The bins cover $0.90 \leq s_{\mathrm{BDT}} \leq 1.00$ and contain approximately equal expected signal yields. The final column gives the effective weighted background statistics, $N^B_{\mathrm{eff}} = (\sum w)^2 / \sum w^2$, as a diagnostic of template stability.

Bin	BDT-score interval	S_k	B_k	S_k/B_k	$N^B_{\mathrm{eff},k}$
1	$[0.9000, 0.9266)$	1.321	2.47×10^5	5.35×10^{-6}	8 305
2	$[0.9266, 0.9430)$	1.317	1.78×10^5	7.39×10^{-6}	5 955
3	$[0.9430, 0.9542)$	1.325	1.36×10^5	9.77×10^{-6}	4 541
4	$[0.9542, 0.9622)$	1.309	1.05×10^5	1.25×10^{-5}	3 489
5	$[0.9622, 0.9684)$	1.337	85 267	1.57×10^{-5}	2 847
6	$[0.9684, 0.9732)$	1.323	67 340	1.96×10^{-5}	2 271
7	$[0.9732, 0.9770)$	1.298	54 752	2.37×10^{-5}	1 869
8	$[0.9770, 0.9802)$	1.328	46 554	2.85×10^{-5}	1 596
9	$[0.9802, 0.9828)$	1.295	38 585	3.36×10^{-5}	1 308
10	$[0.9828, 0.9851)$	1.356	35 051	3.87×10^{-5}	1 173
11	$[0.9851, 0.9870)$	1.331	28 228	4.72×10^{-5}	957
12	$[0.9870, 0.9886)$	1.298	24 521	5.29×10^{-5}	835
13	$[0.9886, 0.9900)$	1.307	19 430	6.73×10^{-5}	672
14	$[0.9900, 0.9912)$	1.281	15 731	8.14×10^{-5}	555
15	$[0.9912, 0.9923)$	1.332	15 016	8.87×10^{-5}	515
16	$[0.9923, 0.9933)$	1.370	12 379	1.11×10^{-4}	440
17	$[0.9933, 0.9942)$	1.345	8 628	1.56×10^{-4}	322
18	$[0.9942, 0.9950)$	1.228	7 051	1.74×10^{-4}	263
19	$[0.9950, 0.9960)$	1.401	6 507	2.15×10^{-4}	249
20	$[0.9960, 1.0000]$	1.292	3 218	4.02×10^{-4}	168
Total	$[0.90, 1.00]$	26.40	1.13×10^6	2.33×10^{-5}	–

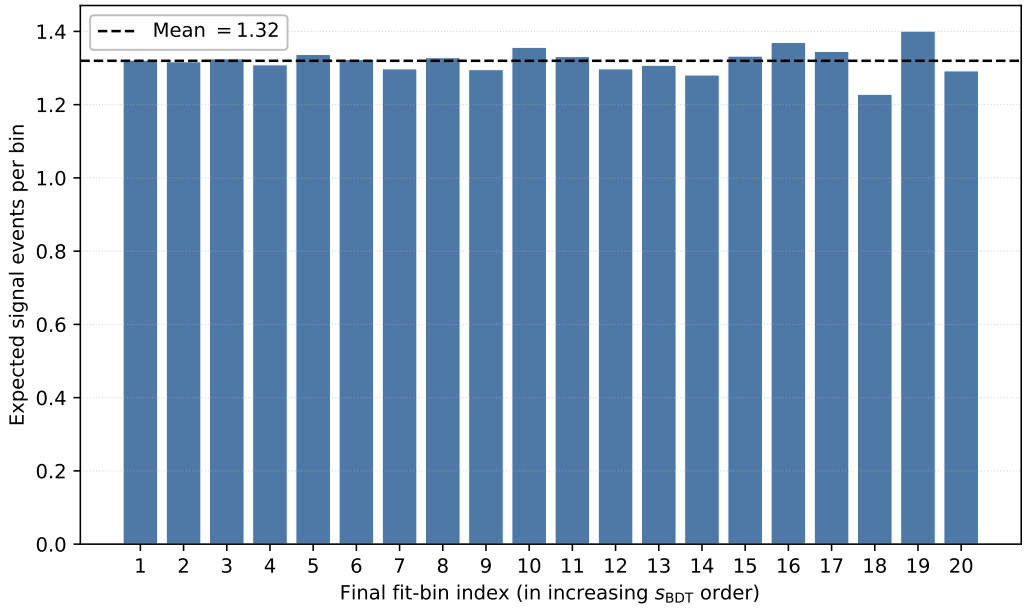


Figure A.1.: Expected signal yield in each of the 20 final likelihood bins, ordered from lower to higher BDT scores. The dashed line shows the mean of approximately 1.32 expected signal events per bin and confirms the intended equal-signal construction.

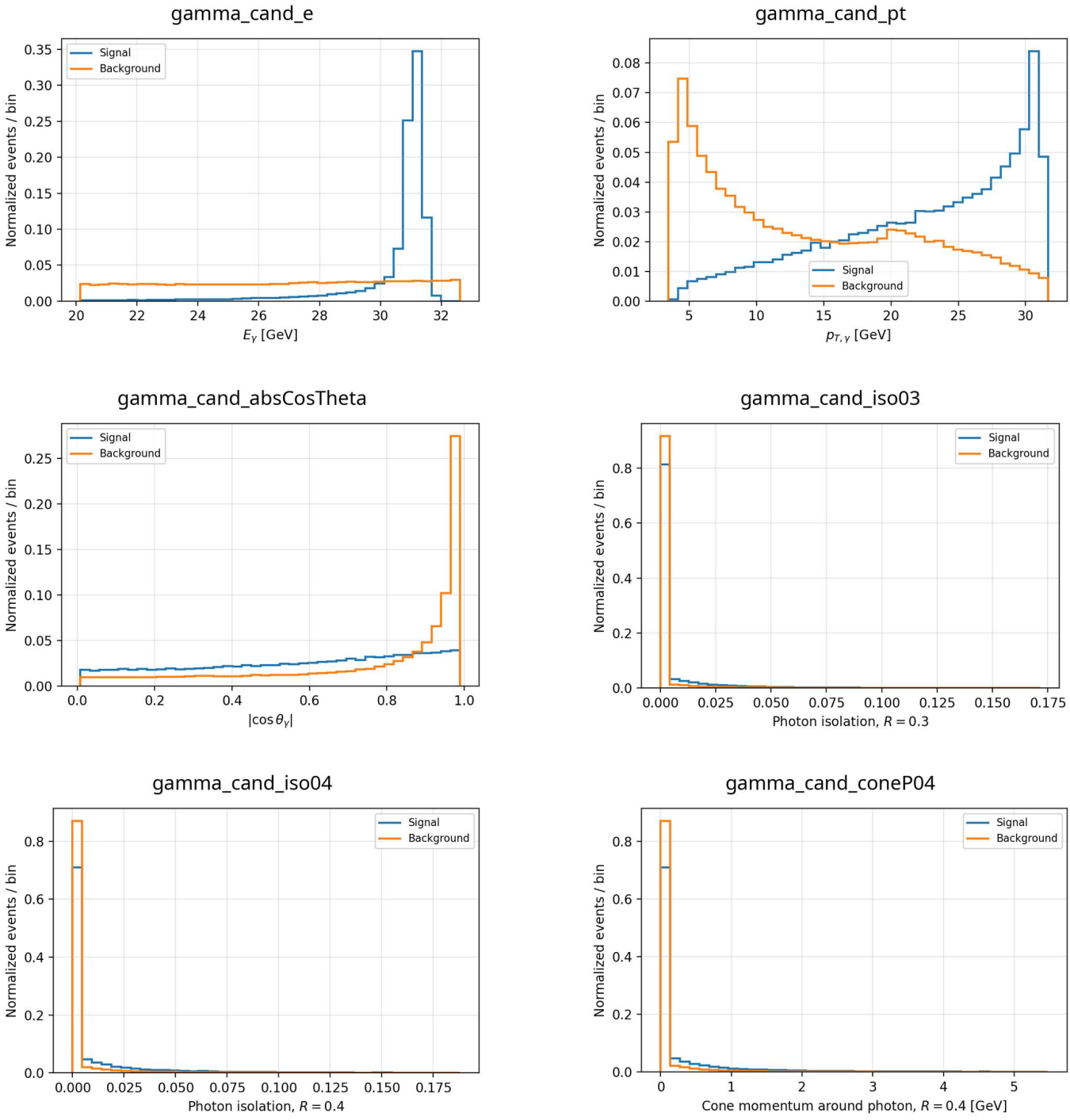
B. BDT Input-Variable Appendix Material

The BDT input variables are grouped in Table B.1. This is the grouping used for the signal–background input-variable distributions in the appendix: photon-candidate variables, recoil and mass-compatibility variables, hadronic-system variables, event-balance variables, and missing-transverse-momentum or object- multiplicity variables. The plots following the table show the distributions after the preselection and before the BDT score requirement. They are normalized to unit area in order to compare the shapes of signal and background.

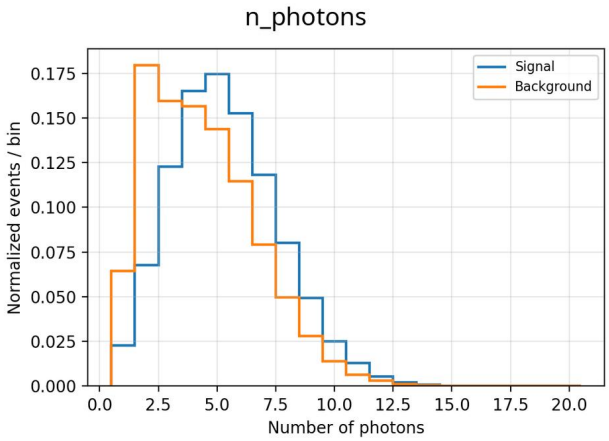
Table B.1.: Grouping of BDT input variables used to organize the requested appendix distributions.

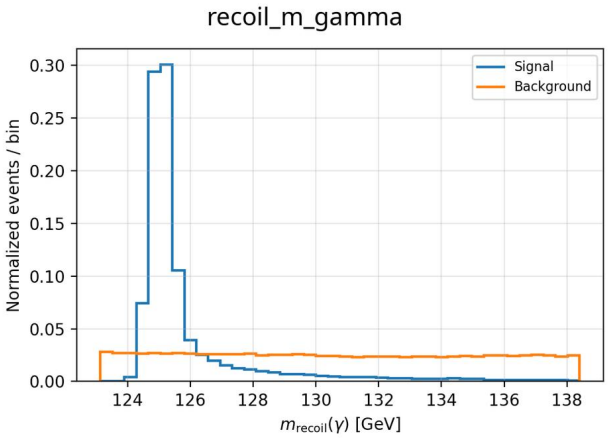
Group	Variables
Photon candidate	`gamma_cand_e`, `gamma_cand_pt`, `gamma_cand_absCosTheta`, `gamma_cand_iso03`, `gamma_cand_iso04`, `gamma_cand_coneP04`, `gamma_e_absdiff_signal`, `gamma_pt_over_e`, `n_photons`
Recoil and Higgs-mass compatibility	`recoil_m_gamma`, `recoil_absdiff_125`, `had_m`, `had_m_absdiff_125`, `had_m_recoil_absdiff`
Hadronic system	`n_jets`, `n_bjets`, `leading_jet_pt`, `ht_jets`, `had_e`
Event balance	`visible_energy`, `visible_energy_absdiff_160`, `had_e_plus_gamma_e_absdiff_160`, `dphi_gamma_had`, `dphi_gamma_had_absdiff_pi`, `gamma_had_pt_balance`, `gamma_had_pt_balance_over_sumpt`, `had_pt_over_gamma_pt`, `had_pt_minus_gamma_pt_abs`
Missing transverse momentum and lepton/object multiplicities	`MET`, `cosTheta_miss`, `met_over_visible_energy`, `met_over_ht_jets`, `n_tracks`, `n_charged_reco`, `n_electrons`, `n_muons`, `n_leptons`

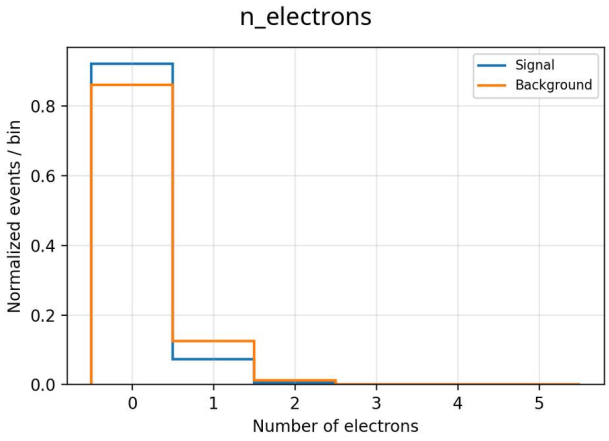
BDT input-variable distributions after preselection and before the BDT score cut

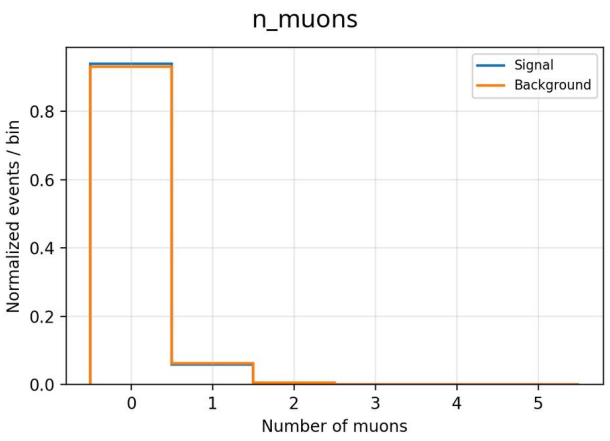


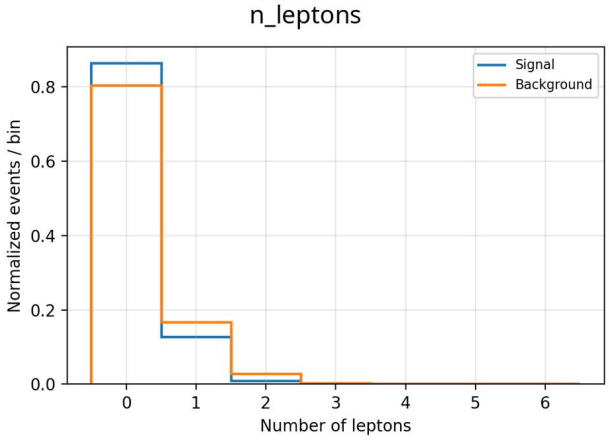
BDT input-variable distributions after preselection and before the BDT score cut

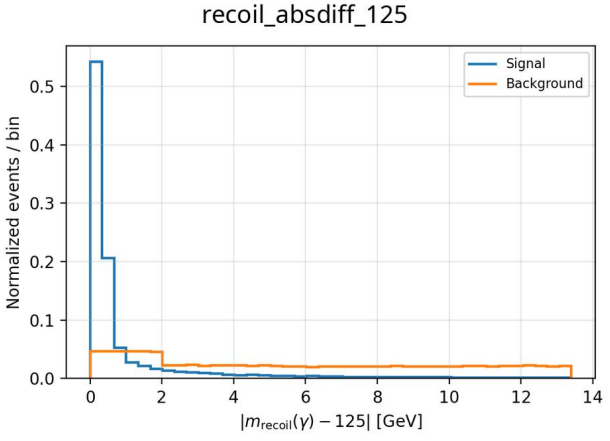




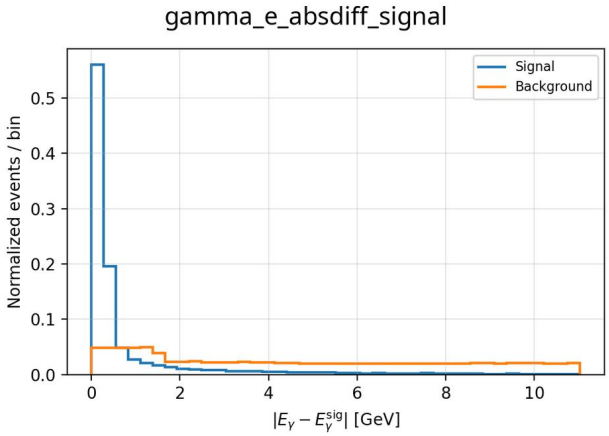


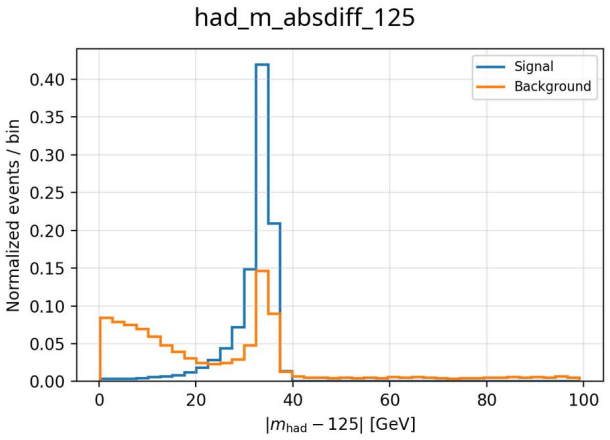


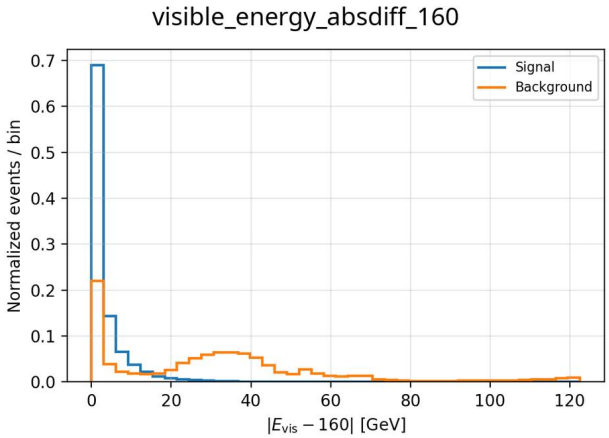


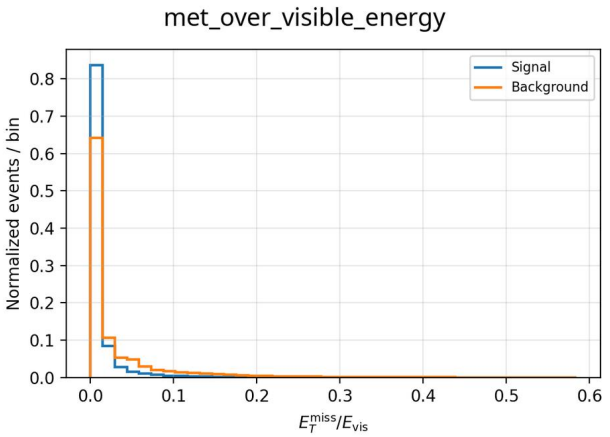


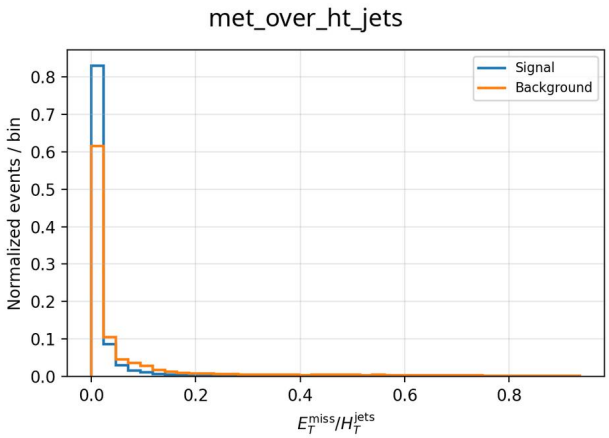
BDT input-variable distributions after preselection and before the BDT score cut

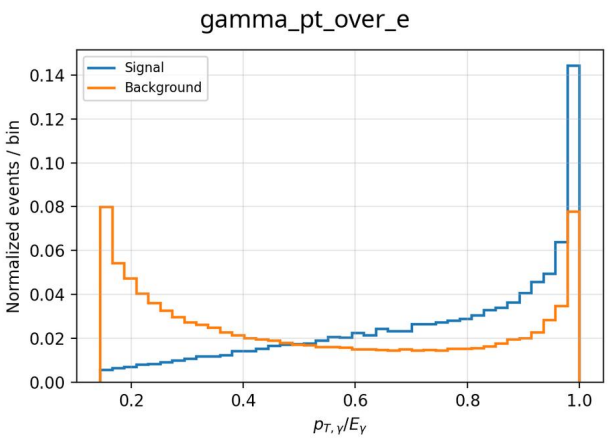






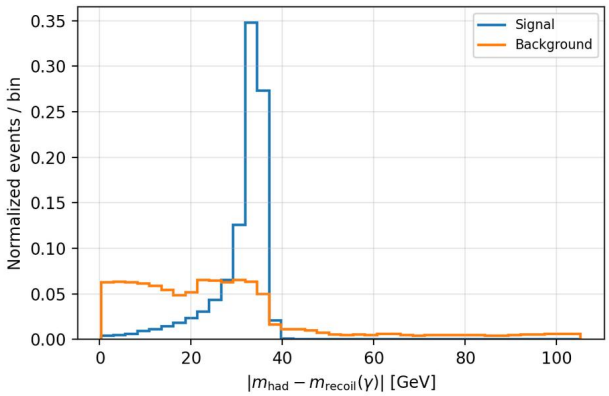




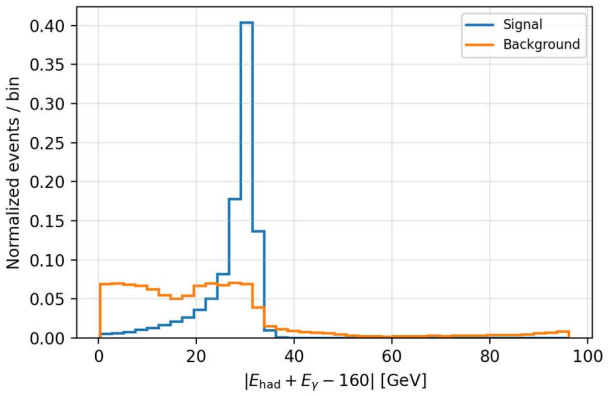


BDT input-variable distributions after preselection and before the BDT score cut

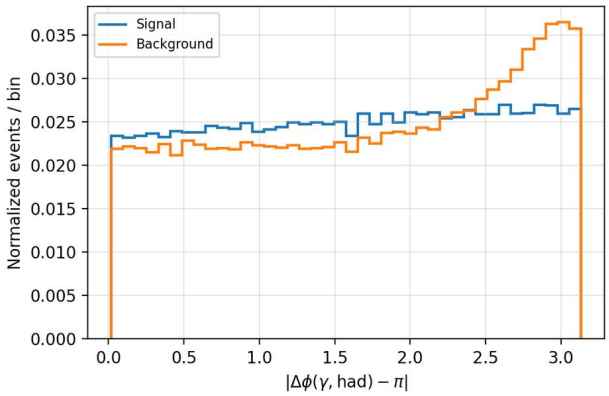
had_m_recoil_absdiff



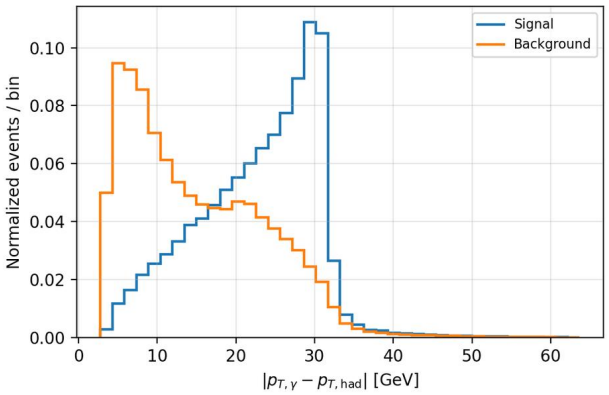
had_e_plus_gamma_e_absdiff_160



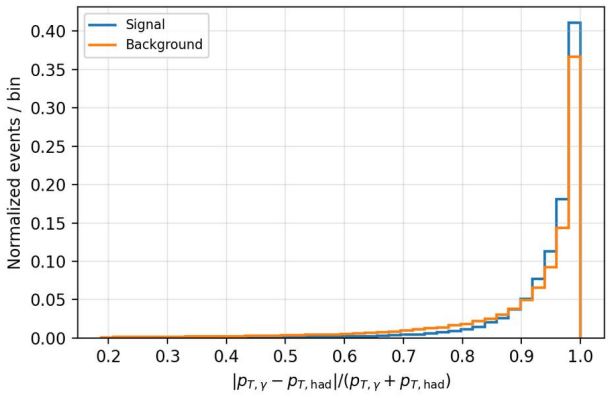
dphi_gamma_had_absdiff_pi



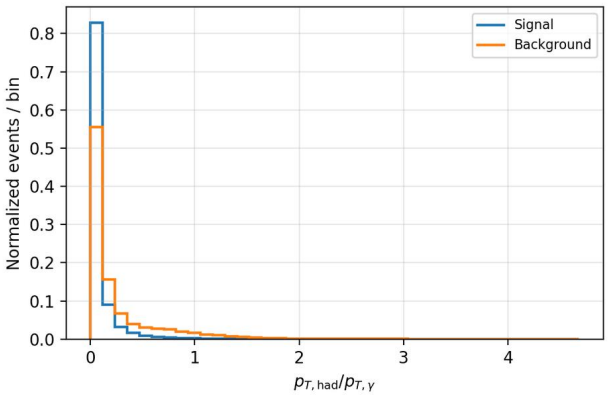
gamma_had_pt_balance



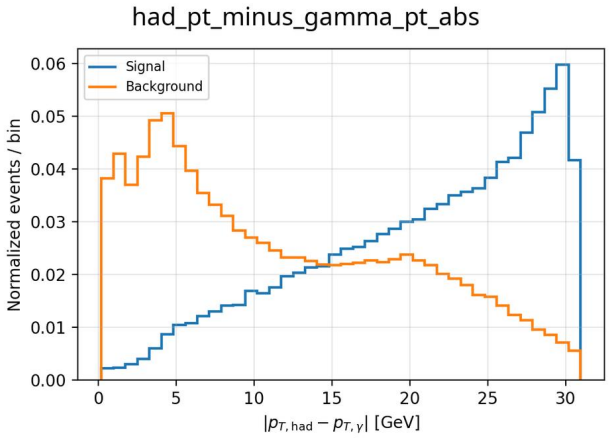
gamma_had_pt_balance_over_sumpt

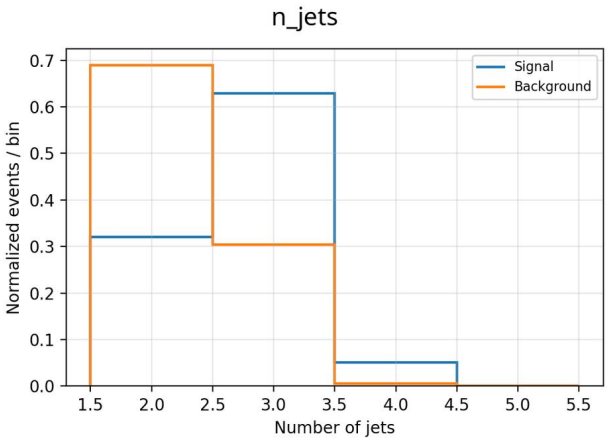


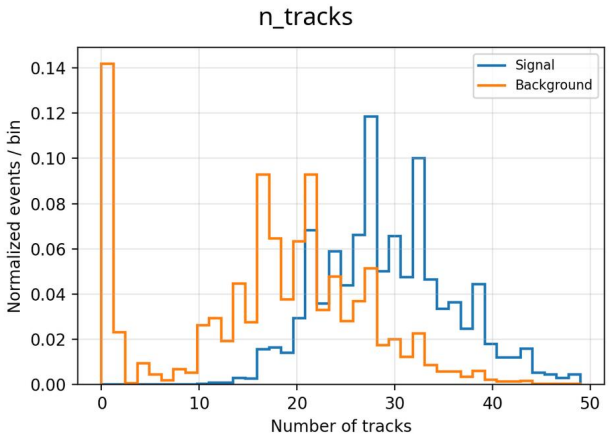
had_pt_over_gamma_pt

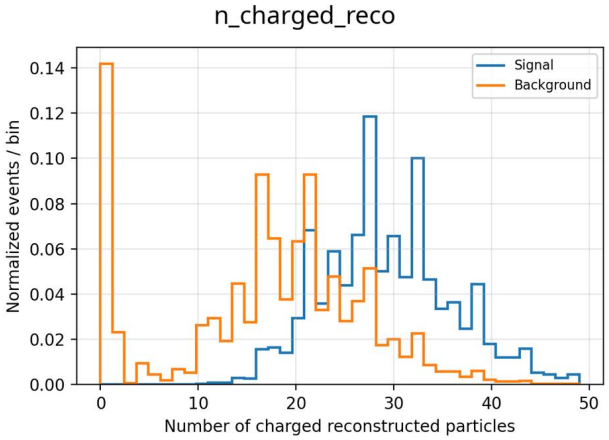


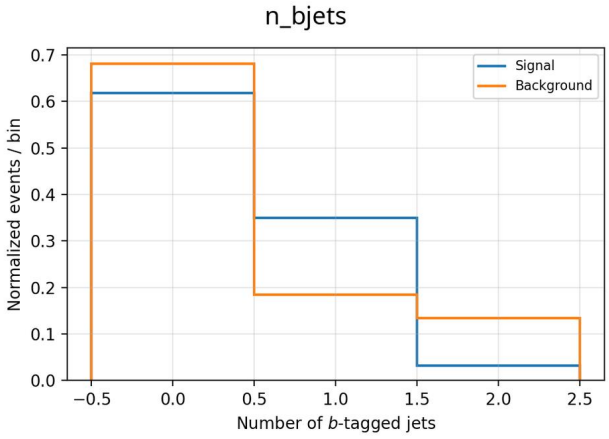
BDT input-variable distributions after preselection and before the BDT score cut

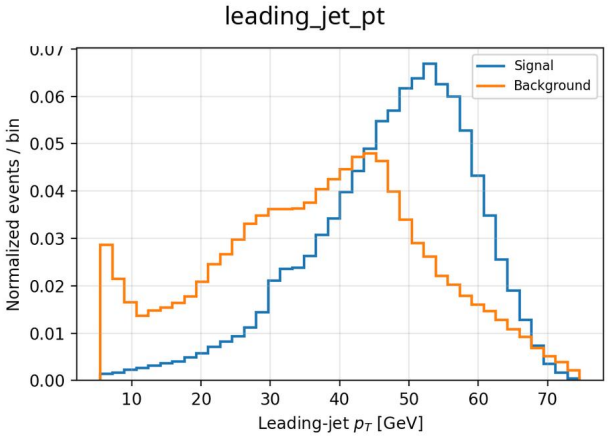






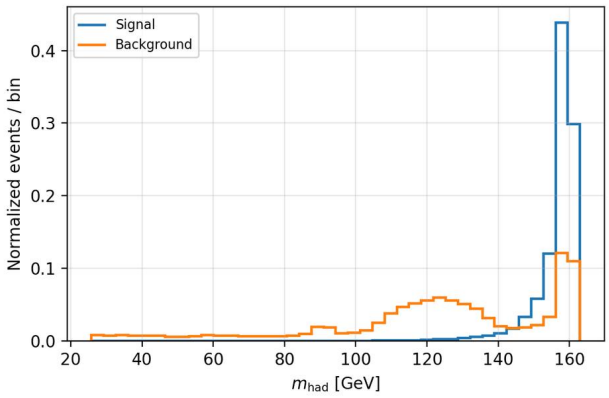




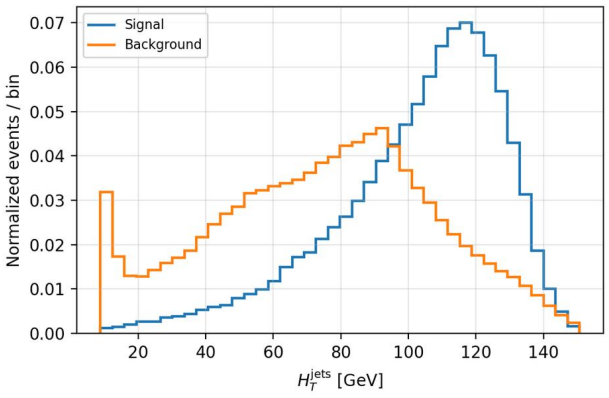


BDT input-variable distributions after preselection and before the BDT score cut

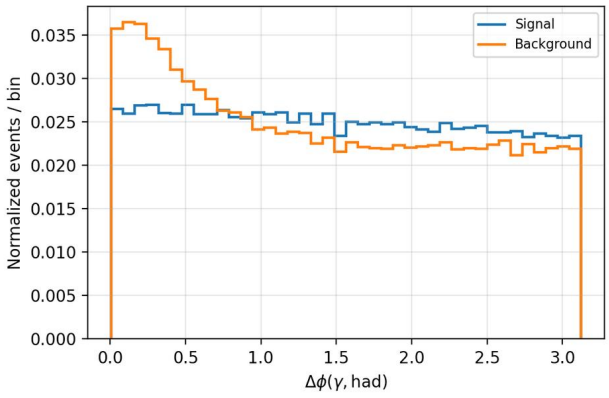
had_m



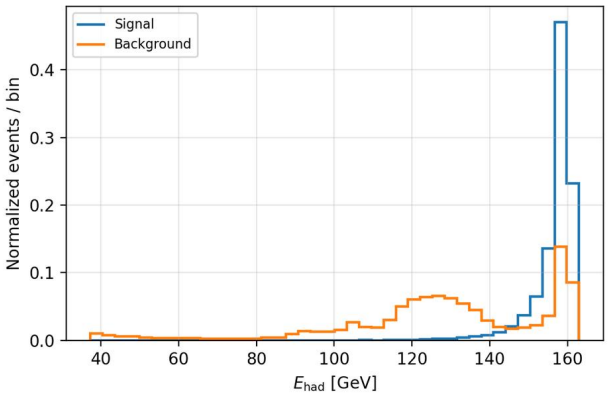
ht_jets



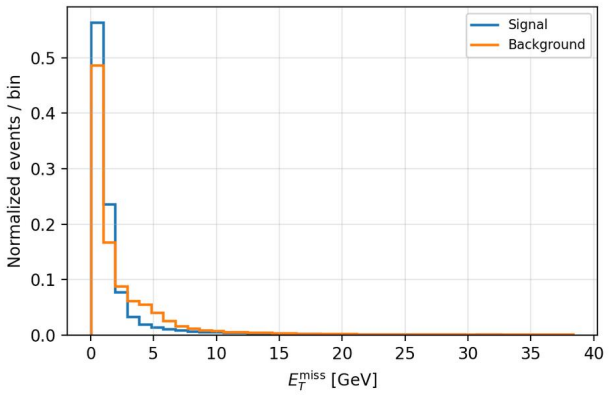
dphi_gamma_had



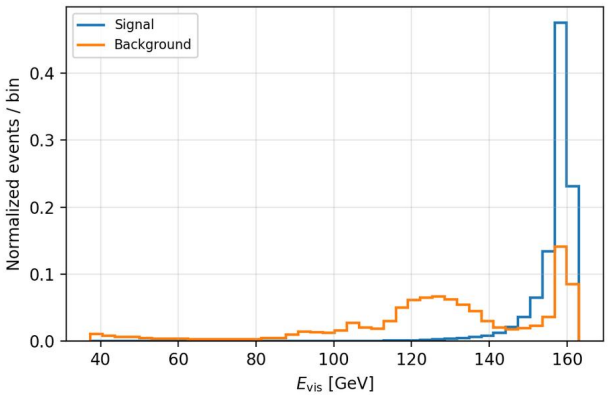
had_e



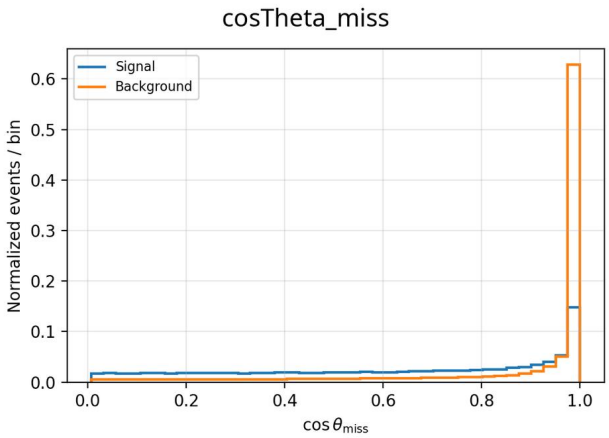
MET



visible_energy



BDT input-variable distributions after preselection and before the BDT score cut



C. Additional Images

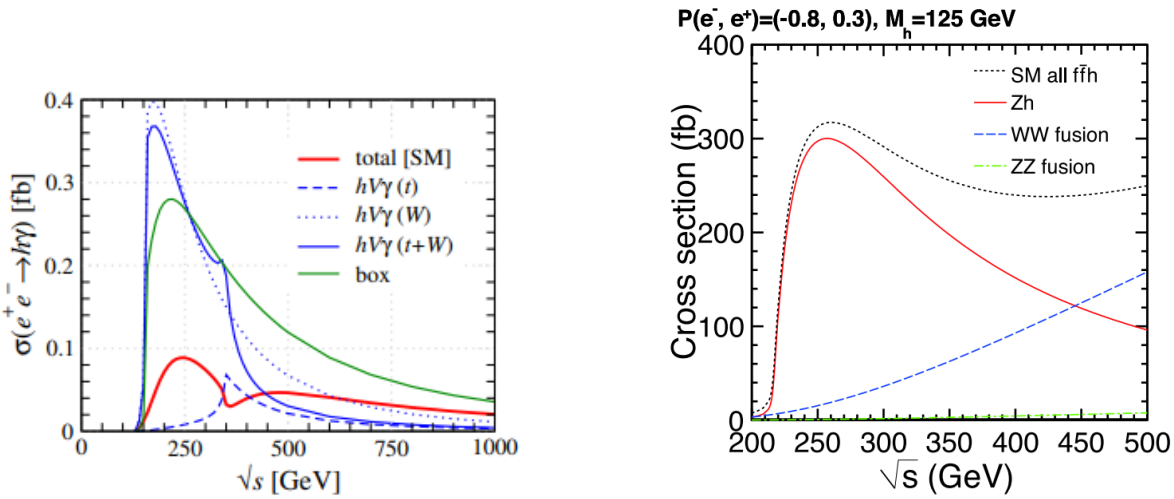


Figure C.1.: Centre-of-mass-energy dependence of the $e^+e^- \to H\gamma$ cross section (left), reproduced (cropped) from Figure 1 of Ref. [5] under the Creative Commons Attribution 4.0 licence, and of the $e^+e^- \to ZH$ cross section (right), reproduced (cropped) from Figure 2 of Ref. [32], where it is attributed to the ILC Technical Design Report [33].

Bibliography

[1] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **110(3)**, 030001 (2024)

[2] ATLAS Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, Phys. Lett. B **716(1)**, 1 (2012)

[3] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716(1)**, 30 (2012)

[4] G. Bertone, D. Hooper, *History of Dark Matter*, Rev. Mod. Phys. **90(4)**, 045002 (2018)

[5] S. Kanemura, K. Mawatari, K. Sakurai, *Single Higgs Production in Association with a Photon at Electron–Positron Colliders in Extended Higgs Models*, Phys. Rev. D **99(3)**, 035023 (2019)

[6] FCC Collaboration, *Future Circular Collider Feasibility Study Report: Volume 1, Physics, Experiments, Detectors*, Eur. Phys. J. C **85(12)**, 1468 (2025)

[7] A. Abbasabadi et al., *Higgs-Boson–Photon Associated Production at e^+e^- Colliders*, Phys. Rev. D **52(7)**, 3919 (1995)

[8] A. Djouadi et al., *Associated Production of Higgs Bosons and a Photon in High-Energy Electron–Positron Collisions*, Nucl. Phys. B **491(1–2)**, 68 (1997)

[9] S. L. Glashow, *Partial-Symmetries of Weak Interactions*, Nucl. Phys. **22(4)**, 579 (1961)

[10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13(16)**, 508 (1964)

[11] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, Technical Report CERN-2017-002-M, CERN, Geneva (2017)

[12] FCC Collaboration, *Future Circular Collider Feasibility Study Report: Volume 2, Accelerators, Technical Infrastructure and Safety*, Eur. Phys. J. Spec. Top. **234(19)**, 5713 (2025)

[13] FCC Collaboration, *Future Circular Collider Feasibility Study Report: Volume 3, Civil Engineering, Implementation and Sustainability*, Eur. Phys. J. Spec. Top. **234(17)**, 5113 (2025)

[14] CERN Council, *Council Resolution on the 2026 update of the European Strategy for Particle Physics*, Technical Report CERN/3997/RA, CERN, Geneva (2026)

[15] The IDEA Study Group, *The IDEA Detector Concept for FCC-ee*, Technical Report FERMILAB-PUB-25-0189-PPD, Fermi National Accelerator Laboratory, Batavia, IL (2025)

[16] H. Kolanoski, N. Wermes, *Particle Detectors: Fundamentals and Applications*, Oxford University Press, Oxford (2020)

[17] G. Chiarello et al., *The Tracking System for the IDEA Detector at Future Lepton Colliders*, Nucl. Instrum. Meth. A **936**, 503 (2019)

[18] C. Lippmann, *Particle Identification*, Nucl. Instrum. Meth. A **666**, 148 (2012)

[19] R. Brun, F. Rademakers, *ROOT: An Object-Oriented Data Analysis Framework*, Nucl. Instrum. Meth. A **389(1–2)**, 81 (1997)

[20] HEP-FCC Collaboration, *EventProducer: Fixed FCC-ee Production Dictionary and Campaign Configuration*, Git commit `b0acee4` (2026), a byte-identical snapshot, full commit hash, and SHA-256 checksum are included with the thesis source

[21] G. Ganis, C. Helsens, V. Völkl, *Key4hep, a Framework for Future HEP Experiments and Its Use in FCC*, Eur. Phys. J. Plus **137(1)**, 149 (2022)

[22] J. Alwall et al., *The Automated Computation of Tree-Level and Next-to-Leading Order Differential Cross Sections, and Their Matching to Parton-Shower Simulations*, JHEP **07**, 079 (2014)

[23] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, 159 (2015)

[24] T. Sjöstrand, S. Mrenna, P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05**, 026 (2006)

[25] W. Kilian, T. Ohl, J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at the LHC and ILC*, Eur. Phys. J. C **71**, 1742 (2011)

[26] G. Cowan et al., *Asymptotic Formulae for Likelihood-Based Tests of New Physics*, Eur. Phys. J. C **71**, 1554 (2011)

[27] G. Cowan et al., *Erratum to: Asymptotic Formulae for Likelihood-Based Tests of New Physics*, Eur. Phys. J. C **73**, 2501 (2013)

[28] J. de Favereau et al. (DELPHES 3 Collaboration), *DELPHES 3: A Modular Framework for Fast Simulation of a Generic Collider Experiment*, JHEP **02**, 057 (2014)

[29] HEP-FCC Collaboration, *FCC-config: Fixed Winter2023 FCC-ee IDEA DELPHES Card*, Git commit `cdc8bcd` (2022), a byte-identical snapshot, full commit hash, and SHA-256 checksum are included with the thesis source

[30] M. Cacciari, G. P. Salam, G. Soyez, *FastJet User Manual*, Eur. Phys. J. C **72**, 1896 (2012)

[31] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, J. Mach. Learn. Res. **12(85)**, 2825 (2011)

[32] J. Yan et al., *Measurement of the Higgs Boson Mass and $e^+e^- \to ZH$ Cross Section Using $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ at the ILC*, Phys. Rev. D **94(11)**, 113002 (2016)

[33] T. Behnke et al., *The International Linear Collider Technical Design Report, Volume 1: Executive Summary*, Technical Report ILC-REPORT-2013-040, International Linear Collider (2013)

Erklärung nach §13(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:
Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 23. August 2026

(Thomas W. Erlei)

Erklärung zur Nutzung von ChatGPT und vergleichbaren Werkzeugen im Rahmen von Prüfungen In der hier vorliegenden Arbeit habe ich ChatGPT oder eine andere KI wie folgt genutzt:

- ☐ gar nicht
- ☐ bei der Ideenfindung
- ☐ bei der Erstellung der Gliederung
- ☐ zum Erstellen einzelner Passagen, insgesamt im Umfang von _____% am gesamten Text
- ☐ zur Entwicklung von Software-Quelltexten
- ☒ zur Optimierung oder Umstrukturierung von Software-Quelltexten
- ☒ zum Korrekturlesen oder Optimieren
- ☒ Weiteres, nämlich: - Umformulieren einzelner Passagen zum flüssigeren Lesen

Ich versichere, alle Nutzungen vollständig angegeben zu haben. Fehlende oder fehlerhafte Angaben werden als Täuschungsversuch gewertet.

Göttingen, den 12.07.2026 _____
(Datum) *(Unterschrift)*