DOCTORAL THESIS

Measurement of Higgs Boson Properties in Leptonic Final States using ML-methods

Laurits Tani

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2024

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
25/2024

Measurement of Higgs Boson Properties in Leptonic Final States using ML-methods

LAURITS TANI

TALLINN UNIVERSITY OF TECHNOLOGY
School of Science
Department of Cybernetics

The dissertation was accepted for the defence of the degree of Doctor of Philosophy (Applied Physics) on 11 March 2024

Supervisor: Prof. Dr. Mario Kadastik,
High Energy and Computational Physics,
National Institute of Chemical Physics and Biophysics
Tallinn, Estonia

Co-supervisor: Dr. Christian Veelken,
High Energy and Computational Physics,
National Institute of Chemical Physics and Biophysics
Tallinn, Estonia

Opponents: Professor Dr. Arnaud Ferrari,
Uppsala University,
Uppsala, Sweden

Professor Dr. Veronika Zadin,
University of Tartu,
Tartu, Estonia

Defence of the thesis: 7 June 2024, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been submitted for any academic degree elsewhere.

Laurits Tani ______________________
signature

ISSN 2585-6898 (publication)
ISBN 978-9916-80-140-6 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9916-80-141-3 (PDF)
DOI https://doi.org/10.23658/taltech.25/2024
Printed by Koopia Niini & Rauam

Tani, L. (2024). *Measurement of Higgs Boson Properties in Leptonic Final States using ML-methods* [TalTech Press]. https://doi.org/10.23658/taltech.25/2024

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
25/2024

Higgsi bosoni omaduste mõõtmine leptoneid sisaldavates kanalites kasutades masinõppe meetodeid

LAURITS TANI

TALTECH KIRJASTUS

Contents

List of Publications

The present Ph.D. thesis is based on the following publications that are referred to in the text by Roman numbers.

I L. Tani *et al.*, "Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics," *Eur. Phys. J. C*, vol. 81, no. 2, p. 170, 2021. DOI: 10.1140/epjc/s10052-021-08950-y. arXiv: 2011.04434 [hep-ex]

II CMS collaboration, "Search for Higgs boson pairs decaying to WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 07, p. 095, 2023. DOI: 10.1007/JHEP07(2023)095. arXiv: 2206.10268 [hep-ex]

III L. Tani and C. Veelken, "Comparison of Bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning applications in high energy physics," *Comput. Phys. Commun.*, vol. 294, p. 108 955, 2024. DOI: 10.1016/j.cpc.2023.108955. arXiv: 2201.06809 [physics.data-an]

IV T. Lange *et al.*, "Tau lepton identification and reconstruction: A new frontier for jet-tagging ML algorithms," *Comput. Phys. Commun.*, vol. 298, p. 109 095, 2024. DOI: 10.1016/j.cpc.2024.109095. arXiv: 2307.07747 [hep-ex]

Author’s Contributions to the Publications

I I continued the hyperparameter optimization topic started by Diana Rand in her Master’s thesis [5]. As her thesis focused purely on genetic algorithm, I introduced particle swarm optimization for comparison. For this paper I re-implemented genetic algorithm to introduce supplementary heuristics.

Furthermore, a set of simpler algorithms for the comparison and the overall framework to evaluate these algorithms was developed by me.

As the main author in this paper all the results and figures were prepared in addition to writing the manuscript by me.

II With prior knowledge in machine learning, my task was to develop a framework for everything machine learning related. This included implementing strategies for tuning the hyperparameters, selecting the features, preparing the data and validating training results. This all was done such that the machine learning part for each channel would be done in a consistent and automatized manner.

In addition to the machine learning development, I was responsible for two analysis channels, $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$, where I constructed a set of discriminating variables to help machine learning model to better distinguish HH signal from the background processes.

Furthermore, I assisted in system administration for the Tallinn Tier-2 data center, where the vast majority of the computations done for this analysis as well as for my other publications.

III This paper is the follow-up for my first paper on hyperparameter optimization, where the implementation of particle swarm optimization algorithm from the previous paper was in a large extent reused. The novel part in this paper was the Bayes optimization algorithm that I implemented for the comparison with the particle swarm optimization.

In addition to writing the manuscript, I was responsible for preparing the results and figures.

IV I produced the training samples and developed the code base for this analysis. Additionally, I was responsible for preparing all the plots and results published in the paper. Finally, I contributed substantially to writing the manuscript.

Approbation

I presented the results of the thesis at the following conferences:

1. **L. Tani**. 'Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics', The 2nd CERN Baltic Conference (CBC 2022): 10–12 October 2022, Vilnius

2. **L. Tani**. 'Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics', The 21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2022): 23–28 October 2022, Villa Romanazzi Carducci in Bari, Italy.

Abbreviations

AdaBoost Adaptive Boosting
ADASYN Adaptive Synthetic sampling
AF acquisition function
APDs silicon avalanche photodiodes
ASHA Asynchronous Successive Halving
ATLAS A Toroidal LHC Apparatus
AUC area under curve

bagging bootstrap aggregating
BDT boosted decision tree
BM benchmark
BO Bayesian optimization
BPIX barrel pixel detector
BR branching ratio
BSM Beyond Standard Model

c.o.m center-of-momentum
CKM Cabibbo–Kobayashi–Maskawa
CL confidence level
CMS Compact Muon Solenoid
CP charge and parity symmetry
CSC cathode strip chambers

DA data augmentation
DNN deep neural network
DT drift tube

EB ECAL barrel region
ECAL electromagnetic calorimeter
EE ECAL endcap region
EFT Effective Field Theory
EW electroweak

FE feature engineering
FPIX forward pixel detector

GA genetic algorithm
GAN generative adversarial networks
GD gradient descent
Geant4 GEometry ANd Tracking
GGF gluon gluon fusion
GS grid search
GSF Gaussian-sum filter

HB hadron barrel calorimeter
HBC ATLAS Higgs boson machine learning challenge
HCAL hadronic calorimeter
HE hadron endcap calorimeter
HEFT Higgs Effective Field Theory
HEP high energy physics
HF hadron forward calorimeter
HL-LHC High-Luminosity LHC
HLT High Level Trigger
HO hadron outer calorimeter
HPS hadron-plus-strips

IP interaction point

KF Kalman Filter

L1T Level-1 Trigger
LBN Lorentz Boost Network
LEP Large Electron Positron Collider
LF loss function
LHC Large Hadron Collider
LINAC linear accelerator
LO leading-order

MC Monte Carlo
MET missing transverse energy
ML machine learning
MVA multivariate analysis

NLO next-to-leading-order
NN neural network

OF objective function

PCA principal component analysis
PDF parton density function
PF Particle Flow
PMNS Pontecorvo–Maki–Nakagawa–Sakata
POG physics object group
pp proton-proton
PS proton synchrotron
PSB proton synchrotron booster
PSO particle swarm optimization

QCD quantum chromodynamics
QED quantum electrodynamics
QFT quantum field theory

RDA real data augmentation
RF random forest
RFE Recursive Feature Elimination
RMS root mean square
ROC receiver operating characteristic
RS random search

SDA synthetic data augmentation
SF surrogate function
SHAP SHapley Additive exPlanations
SL supervised learning
SM Standard Model
SMOTE Synthetic Minority Over-sampling Technique
SPS Super Proton Synchrotron
SR signal region
SUSY supersymmetry

TF target function

VAE variational autoencoder
VBF vector boson fusion
VEV vacuum expectation value
VPTs vacuum phototriodes

WIMP weakly interacting massive particle
WLS wavelength shifting
WP working point

XGBoost eXtreme Gradient Boosting

Introduction

The most successful theory of particle physics is the Standard Model (SM). It describes all known fundamental particles and the forces between them — all except for the gravity which is instead explained by the theory of general relativity. In 2012 the last missing link of SM, the Higgs boson, was discovered by the Compact Muon Solenoid (CMS) and A Toroidal LHC Apparatus (ATLAS) collaborations [6, 7] at the LHC experiment [8], marking the moment when all the particles predicted by the SM have been found. However, during the next more than a decade, no new discoveries have been made.

While SM manages to describe a majority of physics phenomena, several astrophysical observations fall beyond it's reach, thereby motivating theories explaining new physics beyond the SM. These astrophysical phenomena include for example the offset of the center of the total mass from the center of the baryonic mass in case of the Bullet Cluster [9] and the discrepancies between the predicted and measured galactic rotational curves [10]. Both of these phenomena indicate the existence of a new type of matter referred to as dark matter [11], which seems to interact with SM particles only via gravitational force.

As dark matter has never been directly observed, a scrutinous testing of the SM predictions is necessary. Being that the newly discovered Higgs boson interacts only with massive particles, it can be anticipated that also massive particles beyond SM would interact with the Higgs field, thereby motivating the study of the Higgs boson and the as of yet unmeasured properties of it. Higgs boson self-coupling (λ) represents one of these quantities and can be directly accessed via the Higgs boson pair (HH) production. However, measuring this process constitutes a major challenge — with a production cross section of 31 fb at the center-of-mass energy of 13 TeV, producing HH is $\mathcal{O}(1000)$ times rarer than single Higgs. A significant mismatch of the observed values and SM prediction would suggest new physics beyond SM.

One possible scenario for the excess of HH signal events may originate from the decays of unknown new heavy particles, which are postulated by various theoretical models predicting new physics. These models include for example two-Higgs-doublet models [12], Higgs portal models [13], composite-Higgs models [14] or even models inspired by extra dimensions [15]. The latter class of the listed theoretical models features new heavy particles having a spin 0 (radion-like) or spin 2 (graviton-like), which are two of physics cases of studied in the current thesis.

Also, the fact that besides the Higgs boson no new particles have been found at the LHC, motivates theories featuring new heavy particles that are too heavy to be produced at the energies the LHC is operating on. Still, these heavy particles can be detected inderectly by considering loop contribution to given Feynman diagrams. As the amount of theories featuring such heavy particles decaying into two Higgs bosons is essentially infinite, the contributions of such heavy particles could be approximated as contact interactions with the Higgs boson. This approach is referred to as Effective Field Theory (EFT) and is used for the physics searches in the $\mathrm{HH} \to \mathrm{multilepton}$ analysis in this thesis.

The HH analysis studied in the context of this thesis targets the decay channels, where the two Higgs bosons decay into either only vector bosons (VVVV), vector bosons and τ leptons (VV$\tau\tau$), or only τ leptons ($\tau\tau\tau\tau$) in the final states featuring multiple electrons, muons and hadronically decaying τ leptons. Major contributions in this analysis have been made to two analysis channels targeting the $\tau\tau\tau\tau$ decay mode — $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$, where the symbol τ_h denotes hadronically decaying taus. As $\tau\tau\tau\tau$ decay channel has a very small branching ratio, it features very few signal events. Therefore, since the main background in both $0\ell+4\tau_h$ and $1\ell+3\tau_h$ channels originates from electrons and jets that are incorrectly identified as τ_h, a sizable gain in sensitivity to the HH signal can be attained with better τ reconstruction and identification. Various novel machine learning (ML) methods for τ reconstruction and identification were studied in this thesis. Data used for the $\mathrm{HH} \to \mathrm{multilepton}$ analysis corresponds to the proton-proton collisions happening at the LHC at the center-of-momentum (c.o.m) energy of 13 TeV every 25 ns. The data was recorded by the CMS detector in the years 2016-2018.

In order to discriminate HH signal from background processes, special analysis and machine learning techniques were employed. The present document focuses primarily on the machine learning aspect of the $\mathrm{HH} \to \mathrm{multilepton}$ analysis. In addition to choosing a a suitable algorithm and preparing the data for the training, one needs to specify algorithm-inherent parameters, referred to as hyperparameters. The choice of these parameters significantly influences the performance of the trained model. Hence, various hyperparameter optimizationg algorithms were explored, including the suitability of evolutionary algorithms for this task.

This thesis is structured as follows: In Chapter 2 an overview of the most important aspects of the SM is given; Chapter 3 briefly describes the CMS detector, as well as particle reconstruction and identification methods; Chapter 4 covers the basic principles of machine learning; Chapter 5 gives an overview of the $\mathrm{HH} \to \mathrm{multilepton}$ analysis and provides further details on the $0\ell+4\tau_h$, $1\ell+3\tau_h$ analysis channels.

1 Standard Model

SM of particle physics is a theory describing fundamental particles in the nature and the forces between them, namely the interactions via gauge bosons and Higgs interactions. It describes three out of the four known forces — the electromagnetic, weak and strong force. The only known force it does not cover is the graviational force which is instead described by general theory of relativity.

SM is consistent with most of observations, with the recent precision tests establishing its validity up to the electroweak scale.

This chapter aims to describe the general details of the SM of particle physics in Sec. 1.1, while giving a more in depth view of the role of Higgs boson within the model in Sec. 1.2. An overview of the Higgs boson pair production process is given in Sec. 1.2.2.

1.1 Overview of the Standard Model

Standard Model consists of several theoretical ideas that are combined in a way that would reproduce the experimental data. The different ingredients describe different aspects of the theory — fermions are described by the Dirac equation of relativistic quantum mechanics, the fundamental description of the particles and their interactions are given by quantum field theory (QFT), the exact nature of the interactions is determined by the local gauge principle, and finally the masses for the particles are created by the Higgs mechanism by breaking the electroweak symmetry.

In total there are 26 free parameters in the SM, 14 of which are related to the Higgs field (12 Yukawa couplings of fermions to the Higgs field, the vacuum expectation value of the Higgs potential and the mass of the Higgs boson), 8 with the flavor sector (The mixing angles of the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) and Cabibbo–Kobayashi–Maskawa (CKM) matrices. Additional (26th) degree of freedom that is often taken to be 0 is the CP violating phase θ_{CP}.) and 3 related to the gauge interactions.

SM is based on the local gauge group $\mathrm{SU}(3)_\mathrm{c} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$, where the $\mathrm{SU}(3)_\mathrm{c}$ and $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ represent the gauge groups for strong interactions and electroweak interactions respectively. Each gauge group prescribes how the corresponding quantum field (spin-0 scalar field or spin-1 vector field) transforms (space rotations and Lorentz boosts) without affecting the equations of motion, thereby leading to conservation of charge as per Noether's theorem.

However, in order for these transformations to be local, one needs to introduce vector fields (i.e., gauge fields), the excitations of which correspond to the force-mediating particles and therefore to the generators of the symmetry group. The interaction strength depends on the charge corresponding to the associated gauge boson - in case of the electromagnetic force this charge corresponds to the electric charge. In case of weak interaction and strong force the corresponding charges are weak isospin and the color charge, respectively. In the SM the four spin-1 particles, also known as vector bosons, are the Z, $W^{\pm}$ for the weak force, photon (γ) for the electromagnetic force and gluon (g) for the strong force.

All the fundamental particles in SM are shown in Fig. 1.1. The quarks and the gluon are the only color charged particles and are thus participating in strong interactions. Due to color-confinement [16] color charged particles cannot exist freely in the nature — they form composite particles called baryons, which include both mesons (formed from quark and anti-quark) and hadrons (consisting of three quarks or anti-quarks). Attempting to separate colored particles requires so much energy, that it becomes energetically more favorable to produce new hadrons.

Matter consists of spin-$\frac{1}{2}$ fermions, which includes both quarks and leptons. Fermions come in three generations with each subsequent one being more massive than the previous one. All fermions with the exception of neutrinos are electrically charged and will thus interact electromagnetically by exchanging a photon. Additionally, all leptons interact via the weak force — by the exchange of $W^{\pm}$ or a Z boson. Incidentally, neutrinos participate only in weak interactions. Electrons together with up and down quarks form the stable matter surrounding us.

Unlike the vector bosons, that all have spin 1, Higgs boson is the only known (scalar) boson with a spin equal to 0. For Higgs interactions the interaction strength depends on the mass of the particle and the vacuum expectation value (VEV) of the Higgs field. Higgs boson and its properties will further be discussed in the following sections.

SM is consistent with almost all experimental observations. Still, despite being the most successful theory for it's purposes, there are several things it does not predict. These include:

- **Neutrino oscillations and neutrino masses:** Proton-proton chain is one of the two main fusion reactions in the Sun. In this process only electron neutrinos are created. When measuring the neutrino flavor composition of these solar neutrinos, one would also expect it to consists of only electron neutrinos. However, this is not the case, as the interaction eigenstates (i.e., the flavor eigenstates) differ from the mass eigenstates that are the ones that traverse the space giving rise to flavor mixing [17]. However, according to the SM neutrinos are massless, since introducing masses would require the presence of both left- and right-handed neutrinos and so far no evidence of right-handed neutrinos has been found. Although the mass differences ($\Delta m_{12}, \Delta m_{23}, \Delta m_{13}$) of the lepton generations have been measured, the masses (and the mass-hierarchy) of the neutrinos are unknown, while being estimated to be $\leq 1\text{eV}$.

- **Existence of dark matter:** From astronomical observations we see that the rotational speeds of galaxies do not match with what one would expect from visible matter [18]. This and several other observations motivate studies of new hypothetical particles such as weakly interacting massive particles (WIMPs) [19] and extensions to the SM where for example heavy resonance decays into a pair of Higgs bosons [20].

- **Matter antimatter asymmetry:** Matter and antimatter is produced and annihilated in pairs. However, we see matter in much bigger quantities around us than antimatter, suggesting there is an asymmetry. This asymmetry is not explained by the SM [21].

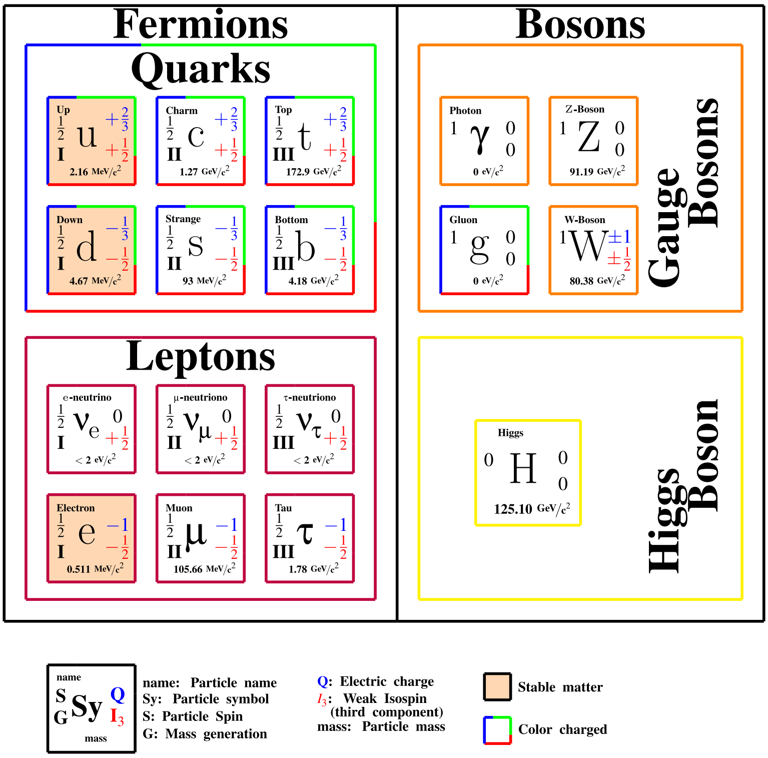


Figure 1.1: Fundamental particles in SM. Multicolored frame indicates that the particle interacts via strong force. Numbers given in red and blue denote the electric and weak isospin charges respectively [22].

1.2 Higgs boson

In this section a brief overview of the role of the Higgs boson in the SM is given and is mostly based on Ref. [23]. A more detailed discussion can be found in Refs. [23, 24].

The principle of local gauge invariance is supported by high-precision electroweak measurements. However, local gauge invariance can only be satisfied if the gauge boson of an interaction is massless, as introducing masses for the gauge boson would spoil the local gauge invariance. Indeed, for quantum electrodynamics (QED) and quantum chromodynamics (QCD) this criterium is fulfilled as their gauge bosons are massless - but we know that $W^{\pm}$ and Z are not massless.

In order to generate masses for gauge bosons, one needs to break the $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$ local gauge symmetry — this is achieved via the *Higgs mechanism*. The idea is to introduce a scalar field ϕ that gives rise to the mass terms.

In the simplest Higgs model, there are four degrees of freedom stemming from the two complex scalar fields that are placed in a weak isospin dublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.1}$$

where the scalar potential of the field is:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.2}$$

Having $\mu^2 > 0$ results in the minimum of the potential to be at $\phi = 0$. However, choosing $\mu^2 < 0$, an infinite set of degenerate minima will satisfy $\phi^\dagger\phi = -\frac{\mu^2}{2\lambda}$, resulting in a potential in a shape of a *Mexican hat* as depicted on Fig. 1.2:

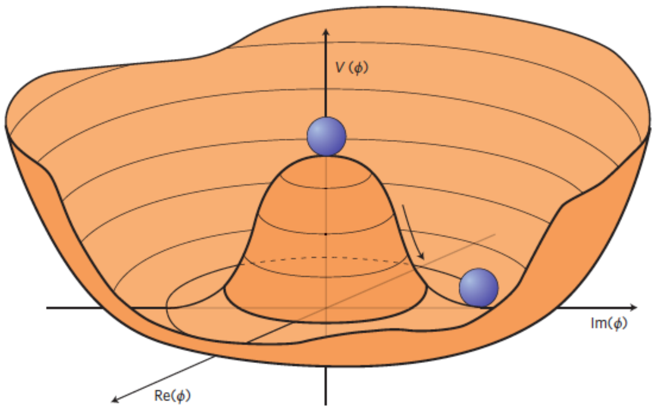


Figure 1.2: The Mexican hat shaped Higgs potential [25].

As the photon needs to remain massless after symmetry breaking, the minimum of the Higgs potential must have a non-zero VEV only for the neutral scalar field ϕ^0 in Eq. 1.1.

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.3}$$

The field ϕ can be expanded around this minimum, resulting in 3 massless Goldstone bosons, g_1, g_2 and g_4, giving us the longitudinal degrees of freedom corresponding to the W^+ W^- and Z that are associated to the $\mathrm{SU}(2)_L$, and a massive scalar boson h(x):

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} g_1(x) + \mathrm{i}g_2(x) \\ v + h(x) + \mathrm{i}g_4(x) \end{pmatrix} \xRightarrow{\text{unitary gauge}} \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.4}$$

In SM the mass of the Higgs boson is a free parameter and is given by $m_H = 2\lambda v^2$, where $v = 246$ GeV is the VEV of the Higgs field that sets the mass scale for the electroweak (EW) and Higgs bosons. The EW gauge boson masses are given by $m_W = m_Z\cos(\theta_W) = \frac{1}{2}g_W v$, where g_W denotes the weak coupling constant, and θ_W the weak mixing angle.

The coupling g_{HVV} of the Higgs boson to the vector bosons (V) is proportional to their mass m_V, and is given by: $g_{HVV} = g_V m_V$.

In addition to creating masses for the gauge bosons, Higgs mechanism is responsible for generating the masses of the fermions. The strength of the Higgs coupling to fermions g_f (i.e., the Yukawa coupling) is proportional to the fermions' mass and are given by Eq.1.5:

$$g_f = \sqrt{2}\frac{m_f}{v}\ , \tag{1.5}$$

where m_f denotes the mass of the fermion and v the vacuum expectation value of the Higgs field.

1.2.1 Single Higgs boson production

In the context of proton-proton (pp) colliders there exist four dominant mechanisms for producing a single Higgs boson. These processes, also depicted on Fig. 1.3 include:

(a) Higgs boson production via gluon gluon fusion (GGF) through a massive fermion loop;

(b) Higgs boson production via vector boson fusion (VBF);

(c) Higgs boson production in association with a gauge boson (Higgs strahlung);

(d) Higgs boson production in association with top quarks (ttH).

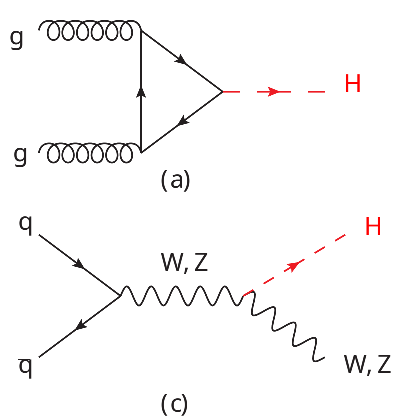


Figure 1.3: Feynman diagrams for (a) Higgs boson production via GGF through a top loop; (b) Higgs boson production via VBF; (c) Higgs boson production in association with a gauge boson (Higgs strahlung); (d) Higgs boson production in association with top quarks (ttH) [26].

Both the single Higgs boson production via GGF and Higgs production in association with top quarks as shown in Fig. 1.3 (a) and (d) respectively feature top Yukawa coupling (y_t), while the diagrams (b) and (c) on Fig. 1.3, the Higgs boson production via vector boson fusion and the Higgs strahlung, feature Higgs boson couplings to gauge bosons, the coupling strength of which is referred to as c_V. The most dominant single Higgs boson production channel at the LHC happens via GGF (depicted in Fig. 1.3(a)). The fermion loop depicted in Fig. 1.3(a) is needed, since Higgs does not couple to the massless gluons directly and as Higgs boson couples preferentially to the most massive particles, the fermion loop is preferentially features virtual top quarks.

Second most dominant single Higgs boson production channel is the VBF as depicted on Fig. 1.3(b). Despite having an order of magnitude smaller cross section (see Fig. 1.4), it is easier to identify as this production channel features only the Higgs decay products and two forward jets originating from the breaking up of the colliding protons.

The Higgs coupling strength to other particles is proportional to their mass and therefore it decays preferentially to the most massive particles that are kinematically accessible. This means that Higgs can decay via $H \to f\bar{f}$ only if $m_H > 2m_f$ and in order for a Higgs boson decay into W^+W^- or ZZ, one of the vector bosons in the decay needs to be virtual. Decays into massless particles, such as photons can only proceeds through a massive boson or fermion loop. As a consequence of the Higgs coupling strength being proportional

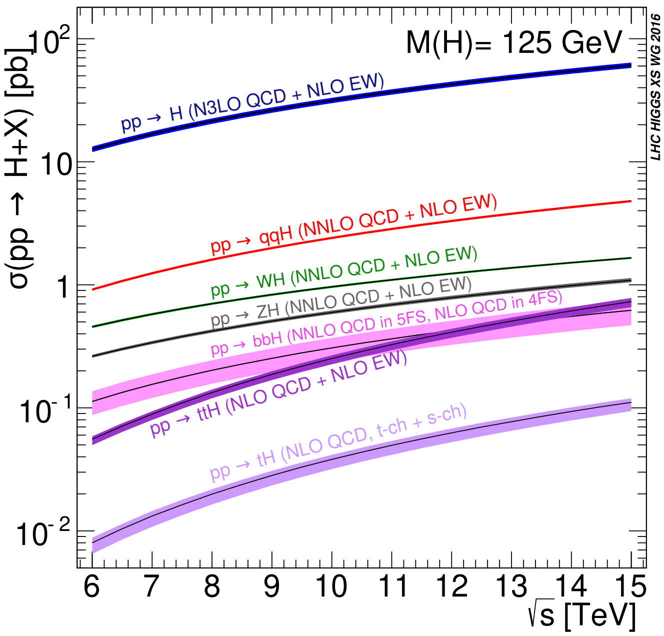


Figure 1.4: Cross section for producing a SM Higgs with a mass of 125 GeV for different production modes [28].

to the mass of the particle, the decays into more massive particles have also larger branching ratio (BR) — as Higgs boson decay into a pair of top quarks is kinematically not allowed the largest BR of a SM Higgs boson of a mass of $m_H = 125.09 \pm 0.21(\text{stat.}) + 0.11(\text{syst.})$ GeV [27] is into $b\bar{b}$ as depicted on Fig. 1.5.

1.2.2 Higgs boson pair production

Although Higgs boson was discovered at the LHC already in 2012 by the CMS and ATLAS collaborations, there are still open questions regarding the nature of this particle.

Most of the Higgs boson properties have been established with a precision of 10% or better [29], such as the Higgs boson mass for which the precision of the mass measurement is on the level of permille [30]. One of the properties of the Higgs boson that has not been well constrained as of yet is the Higgs boson trilinear self-coupling (λ) which directly modifies the shape of the Higgs potential (see Sec. 1.2).

Although the Higgs trilinear self-coupling could be measured indirectly in single Higgs processes by the means of next-to-leading-order (NLO) contributions, the modifications of λ will introduce only minor deviations to the Higgs boson production and decay rates [31]. Stronger bounds on λ can be obtained by the means of measuring the Higgs boson pair (commonly referred to as di-Higgs or HH) production cross section. However, the production rate of di-Higgs is relatively small in comparison to the single Higgs boson production rate, having a SM cross section of a mere 31 fb [32], being a factor $\mathcal{O}(1000)$ smaller than the production cross section of single Higgs boson.

The dominant Higgs boson pair production mechanisms in the LHC are GGF-like and VBF-like production, double Higgs strahlung emitted by vector boson and double Higgs strahlung emitted by top quark. Leading order diagrams for the di-Higgs production are

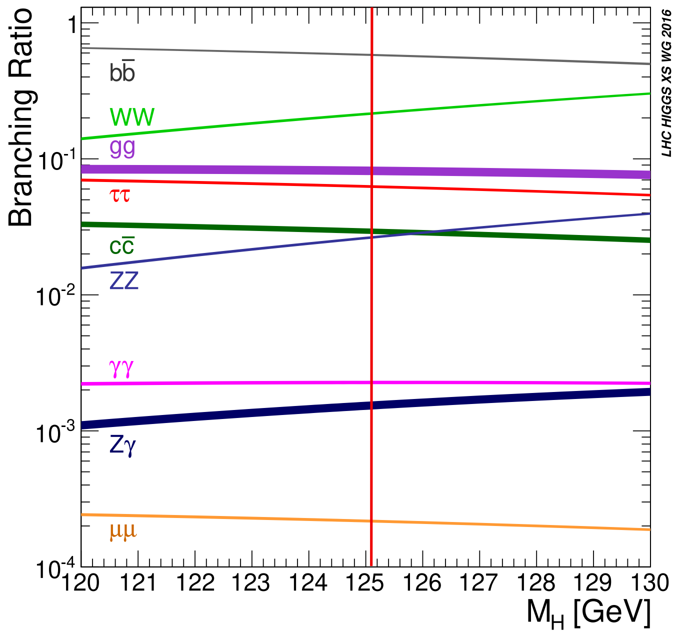


Figure 1.5: Higgs boson branching ratio vs. Higgs boson mass, where the vertical red line indicates the measured Higgs boson mass, which decays most often to a pair of b quarks followed by the decay into W^+W^- [28].

given on Fig. 1.6.

The diagrams where Higgs is emitted from top quark are affected by the top Yukawa coupling strength parameter y_t and are shown on Figs. 1.6 (a) and on diagrams on the bottom row of Fig. 1.6 (c) as indicated by the magenta dot. The diagrams where a Higgs boson is emitted by a vector boson are affected by the Higgs boson coupling strength to vector boson, c_V, parameter as shown on diagrams on Fig. 1.6 (b) and on diagrams on the top row of Fig. 1.6 (c). The c_V is highlighted by the orange dot.

The Higgs boson pair production via VBF production mode is also affected by the Higgs coupling strength to two vector bosons, c_{2V} as depicted on the Fig. 1.7 and indicated by the green dot. Similarly to Higgs self-coupling λ, c_{2V} does not affect the single Higgs production directly, but does contribute to the Higgs boson production cross section via higher order diagrams. Higgs self-coupling λ affects all the double Higgs production channels and is denoted as a red dot on the Figs. 1.6 in the right hand side column.

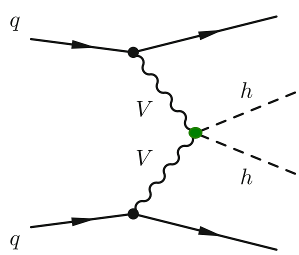


Figure 1.7: Higgs boson pair production via VBF is affected by the Higgs boson coupling strength to two vector bosons. The green dot indicates Higgs boson coupling to two vector bosons, $\mathbf{c_{2V}}$ [34].

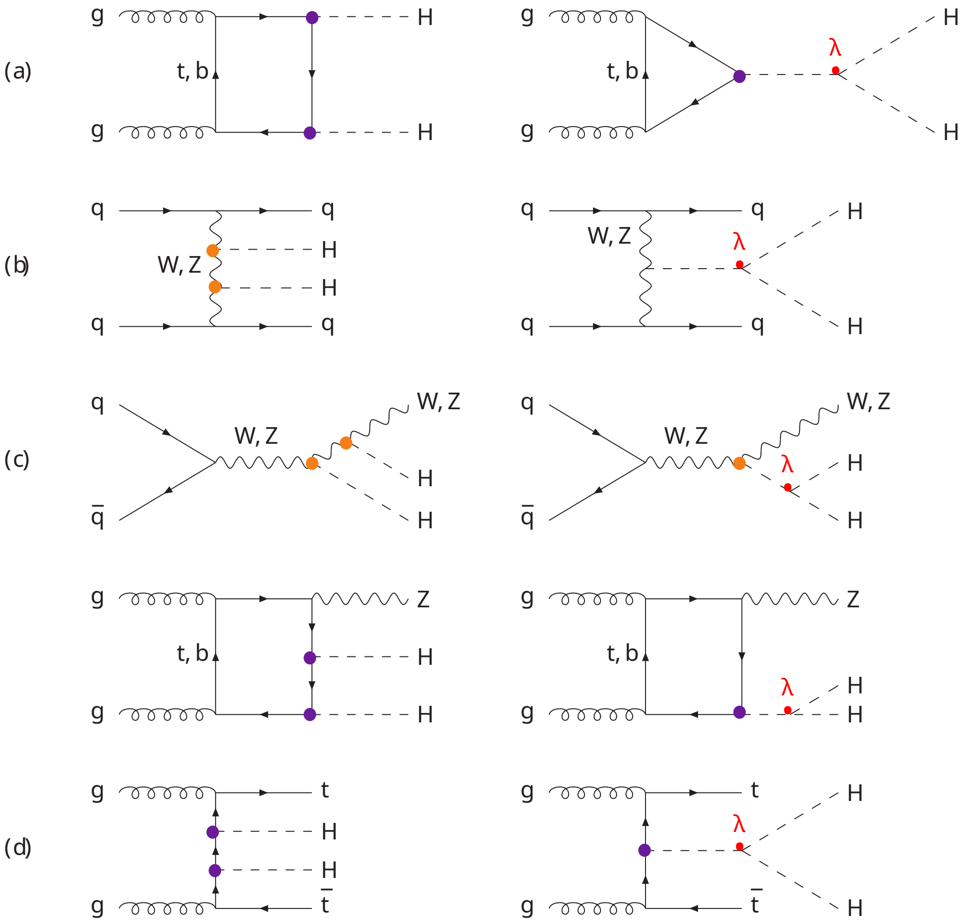


Figure 1.6: Feynman diagrams for (a) di-Higgs production via gluon-gluon fusion; (b) di-Higgs production vector-boson fusion; (c) di-Higgs production from double Higgs boson strahlung; (d) di-Higgs production from double Higgs strahlung emitted by top quarks. The red dots on the diagrams denote the Higgs self-coupling λ, orange dots the Higgs boson coupling to vector bosons c_V and magenta dots the Higgs boson coupling to top quarks y_t [33].

As can be seen from Fig. 1.8, the most dominant Higgs boson pair production mode is again the GGF, having a cross section of 31.05 fb at c.o.m energy of 13 TeV, followed by VBF that is a factor ~ 18 times less abundant and having a cross section of 1.73 fb. The two main contributions for the GGF process are depicted on Fig. 1.6 (a1) and Fig. 1.6 (a2) and are referred to as the *box* ($\square$) and *triangle* ($\triangle$) diagrams. The $\square$ diagram is sensitive the top Yukawa coupling y_t^2, while the $\triangle$ one depends on the Higgs self coupling parameter both on the y_t and λ.

HH production cross section via GGF production mode is dominated by the $\square$ diagram, but is reduced due to the destructive interference between the $\square$ and $\triangle$ diagrams, meaning that higher values of λ do not necessarily cause a bigger production cross section. While the events from $\triangle$ diagram mostly contribute to the lower values of the m_{HH} spectrum, starting from the kinematic threshold of $2m_H$, events from the $\square$ diagram reside mostly near the medium values of m_{HH}. These thresholds come from the fact that in order to produce Higgs boson pair via the $\triangle$ diagram, the virtual Higgs needs to exceed the kinematic threshold of ~ 250 GeV only barely, while in case of the $\square$ diagram the two Higgs bosons are created independently, meaning that not only is the production cross section affected, but also the kinematic properties of the Higgs bosons. The peak of the $\square$ as well as interference term peak around the virtual $t\bar{t}$ threshold of $2m_t$, which is even further enhanced for negative values of κ_λ, as the sign of the interference term will flip. Maximum interference is achieved at $\kappa_\lambda = 2.45$ [35].

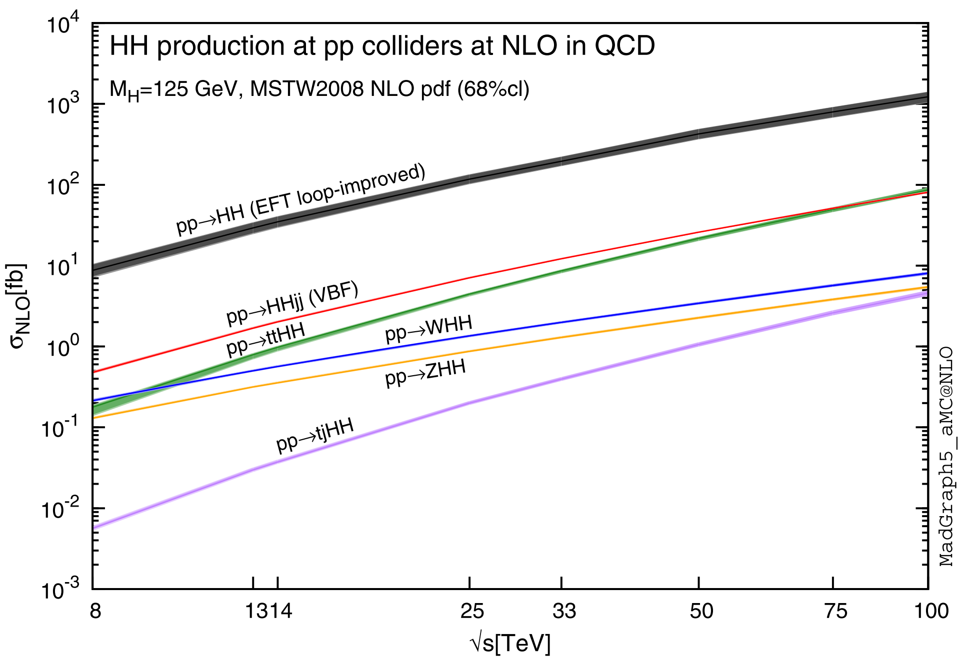


Figure 1.8: Cross section of various double Higgs production modes for a a range of center-of-mass values. The most dominant double Higgs production mode is the GGF [36].

In the next sections two methods for producing a Higgs boson pair beyond the SM are described.

Resonant di-Higgs production

A pair of Higgs bosons is not necessarily produced physics within the SM. Extensions of the SM, called Beyond Standard Model (BSM) physics, also allow two Higgs bosons to be produced via a heavy resonance χ as depicted on Fig. 1.9. The decay of this new particle χ into a pair of Higgs bosons is referred to as resonant HH production. This kind of new heavy particle is postulated by various theoretical models for new physics, including two-Higgs-doublet models [12], Higgs portal models [13], composite-Higgs models [14] and models inspired by extra dimensions [15]. In this thesis, we are looking at the latter model, where heavy resonances corresponding to heavy spin-0 (radion-like) and spin-2 (graviton-like) particles with masses ranging from 250 GeV to 1 TeV.

The resonant HH production here is not model specific and the resonance is assumed to have a narrow width in m_{HH} relative to the experimental resolution.

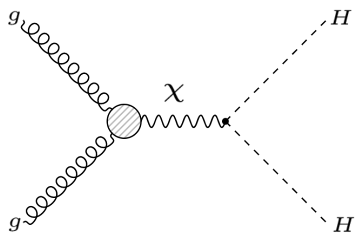


Figure 1.9: Higgs boson pair production via a heavy resonance χ [2].

Non-resonant di-Higgs production

After the discorvery of the Higgs boson in 2012 no new particles have been found at the LHC The reason for no evidence of these new particles might be due to the high mass of the(se) new particle(s), which is beyong the energies the LHC offers. However, these new heavy particles would still contribute indirectly via loops, thereby providing a handle for their discovery. When studying such BSM physics, the amount of possible theoretical models for producing two Higgs bosons when considering loop contributions from such new particles is essentially limitless. Therefore, it makes sense to group a set of physics models together and look only at the contact interactions or the effective coupling behavior with the Higgs boson using the EFT approach. In the context of this thesis 5 couplings for the EFT approach are used. These include couplings such as c_2, c_{2g}, c_g (depicted on Fig. 1.10), describing the interactions between Higgs bosons and $t\bar{t}$. The EFT representation used in this thesis is Higgs Effective Field Theory (HEFT) which is one of two most commonly used formalisms found in the literature [37].

In this thesis EFT scenarios are studied in 5D parameter space consisting of the three EFT couplings, c_2, c_{2g}, c_g, and possibly modified values for the λ and y_t, which are condensed into a set of EFT scenarios, trying to cover the different kinematic scenarios. In the context of this thesis, two sets of EFT scenarios, JHEP03 and JHEP04, are tested. The parameter values for these can be found in Ref. [38] and Ref. [35] respectively.

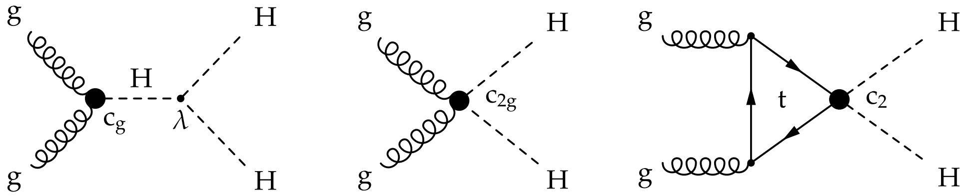


Figure 1.10: BSM non-resonant Higgs boson pair production with the effective couplings c_g, c_{2g} and c_2 [2].

2 The CMS experiment at LHC

CMS [6] and ATLAS [7] are the two multipurpose detectors operating at the 26.7 km long superconducting proton-proton collider named LHC [8, 39] (shown in Fig. 2.1).

CMS, the detector by which the data for the purposes in this thesis is collected, has a cylindrical design that is centered around the interaction point which features a 4 T solenoid magnet surrounded by a massive iron yoke housing the muon system. This superconducting magnet in the CMS experiment spans 12.5 meters and is the central feature of CMS, that with a free bore radius of 3.15m, it is large enough to fit the tracker and both electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL). This detector layout minimizes the amount of material and energy loss in front of the calorimeters, which is beneficial for the Particle Flow (PF) reconstruction [40], which is described later in Sec. 2.7. The CMS detector is 21.6 m in length, has a diameter of 14.6 m and weighs 12 500 tons. Even with these kinds of dimensions the layout is very compact when considering the capabilities and comparing to the design of ATLAS as the tracking and calorimeter systems are encapsulated in the solenoid volume.

In the following sections the supporting apparatus and the subsystems of the CMS detector are described: In Sec. 2.1 the accelerator responsible for accelerating and colliding particles is described, followed by a brief description of the coordinate system used by the CMS experiment in Sec.2.2. In Sec.2.3 an overview of the tracker is given. Both calorimeter systems are described in Sec.2.4, while Sec.2.5 gives an outline of the workings of the muon system. Finally in Sec.2.7 introduction to the reconstruction methods featuring the PF algorithm are presented.

2.1 Accelerator

The 26.7 km long LHC was built into the tunnels previously used by the Large Electron Positron Collider (LEP) [41] and was designed to study SM in energy range where new phenomena could be seen - this included the validation of the Higgs mechanism (Sec. 1.2), supersymmetry (SUSY) and charge and parity symmetry (CP) violation.

As the tunnels have a diameter of only 3.8 meters, the installation of two separate rings for particle beams is not feasible, meaning that both rings are integrated into a single magnetic structure that includes two sets of coils housed within a common yoke and cryostat, allowing for the utilization of the limited space.

In order to reach the designed luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for proton-proton collisions at the center-of-mass energy of 13 TeV a series of pre-accelerators (depicted in Fig. 2.2) are required.

In the first stage particles are extracted from a source, such as hydrogen gas, and then accelerated using a linear accelerator (LINAC), followed by proton synchrotron booster (PSB), proton synchrotron (PS) and Super Proton Synchrotron (SPS). After the last stage of pre-acceleration with the SPS, the beams are injected to LHC at an energy of 450 GeV.

However, in order to be able to bend particle at those energies, the LHC magnets are cooled with pressurized superfluid helium to a temperature of 1.9 K.

Each detector is located in one of the interaction points, where two proton beams (also heavy ions such as Pb) collide.

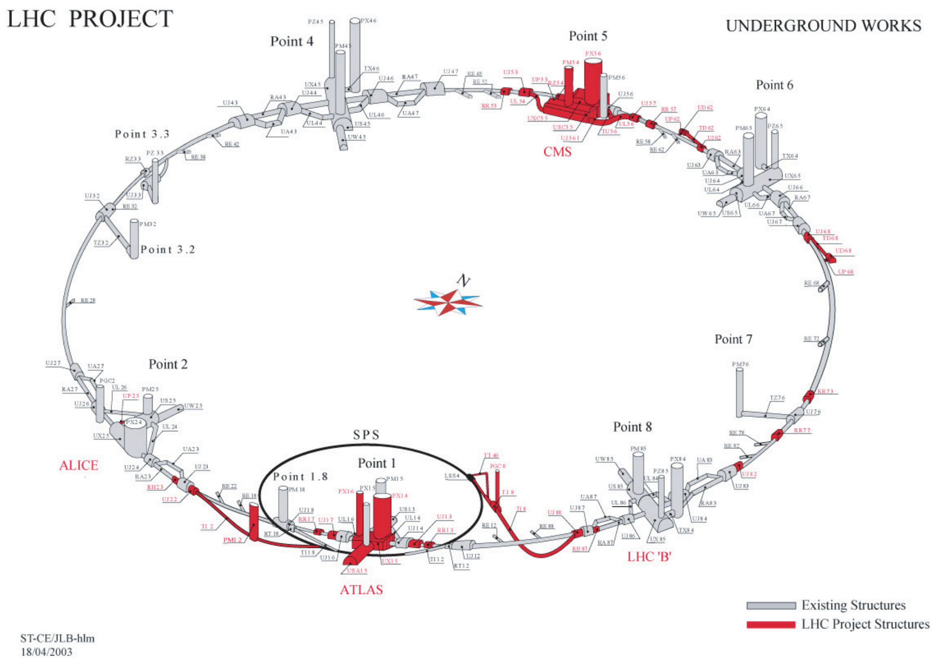


Figure 2.1: Layout of the LHC experiment [8].

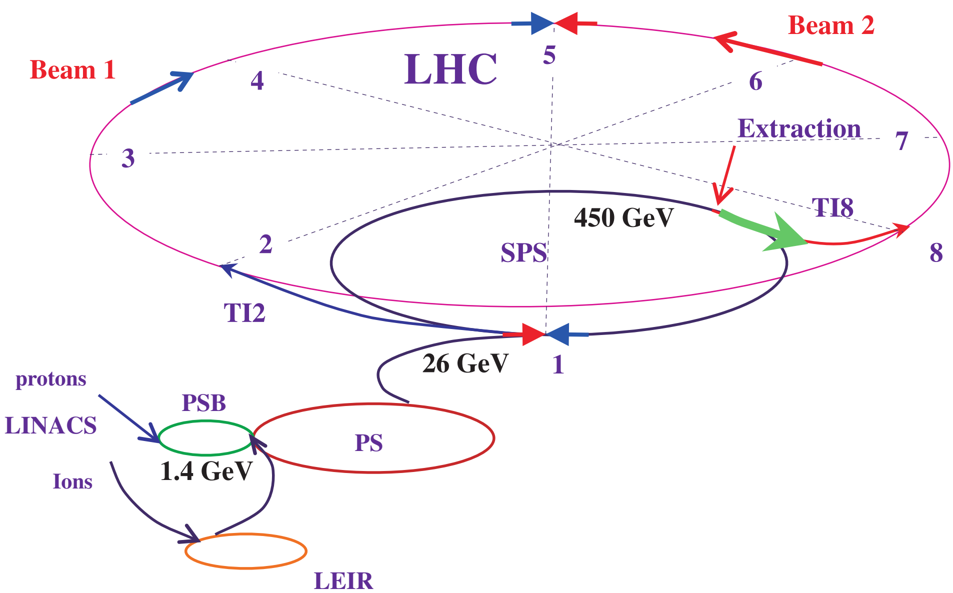


Figure 2.2: Schematic view of the acceleration steps undergone at the LHC [39].

The number of events per second generated by beam-beam collisions is given by $N = \mathscr{L} \cdot \sigma$, where σ and $\mathscr{L}$ denote the cross-section of a process and the luminosity respectively.

For a Gaussian beam profile luminosity $\mathscr{L}$ is given by:

$$\mathscr{L} = \frac{N_b^2 \cdot n \cdot f_r \cdot \gamma}{4 \cdot \pi \cdot \varepsilon_n \cdot \beta^*},$$

where N_b denotes the number of particles in a bunch, n the number of bunches per beam, f_r the frequency of revolution, γ the Lorentz factor, ε_n the normalized transverse emittance and β^* the β function at the collision point [8].

However, as in order to avoid unwanted effects caused by the very nonlinear forces in the beam-beam interaction, one has to reduce the crossing angle of the two colliding beams. Small crossing angles between two beams are highly preferred as can be seen in the definition of the geometrical luminosity reduction factor F in Eq. 2.1:

$$F = \left(1 + \frac{\theta_c \cdot \sigma_z}{2 \cdot \sigma^*}\right)^{-\frac{1}{2}}. \tag{2.1}$$

Here θ_c denotes the crossing angle at the interaction point (IP), σ_z the root mean square (RMS) of the bunch length and σ^* the transverse RMS beam size at the crossing point

2.1.1 Phenomenology of LHC

Circular pp colliders provide high interaction rates together with a high center of mass energy, thereby accommodating excellent potential for new physics discoveries at high energy scales.

The protons that are collided are bound states of two up-quarks and a down quark, referred to as valence quarks, that are held together by gluons. However, depending on the energy the proton is probed at, also the apparent structure will change — for example it is possible to resolve the three valence quarks and the gluon exchanges therein at high energies.

Still, gluons can also be virtual, meaning they can split into quark and anti-quark pairs, thereby giving rise to a multitude of virtual particles called sea quarks that thereafter can combine again. This means that in addition to the valence quarks, also sea quarks and gluons can initiate a hard scattering process by randomly coming in and out of existance. The proton constituents initiating a hard process are known as partons.

However, at the energy scales LHC is operating, the hard scattering process is more likely to be initiated by gluons than valence quarks as only a fraction of the energy carried by the proton (given by the Bjorken variable x) is needed to initiate the process. Therefore, in order to determine the kinematics of a given process in pp colliders, one needs in addition to the matrix element calculation also information regarding the probability for a given particle type to carry a fraction of the protons energy (referred to as the parton density function (PDF)). The NNLO proton PDFs at energy scales $10\,\mathrm{GeV}^2$ and $10^4\,\mathrm{GeV}^2$ are given on left and right side of Fig. 2.3 respectively. On the right side plot we see that at the center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ that the LHC is operating on the gluon scattering is the dominant process, especially on the low x. As most of the collisions at the LHC are happening also at low x, then it makes the LHC effectively a gluon collider. However, there does not exist a way to calculate PDFs from the theory and they need to be measured experimentally by fitting phenomenological models to the data.

Finally, in addition to the hard scattering process pp interactions give rise also to other types of hadronic activity called underlying event, which together with the presence of multiple simultaneous pp interactions, referred to as pile-up, adding another level of difficulty to event reconstruction.

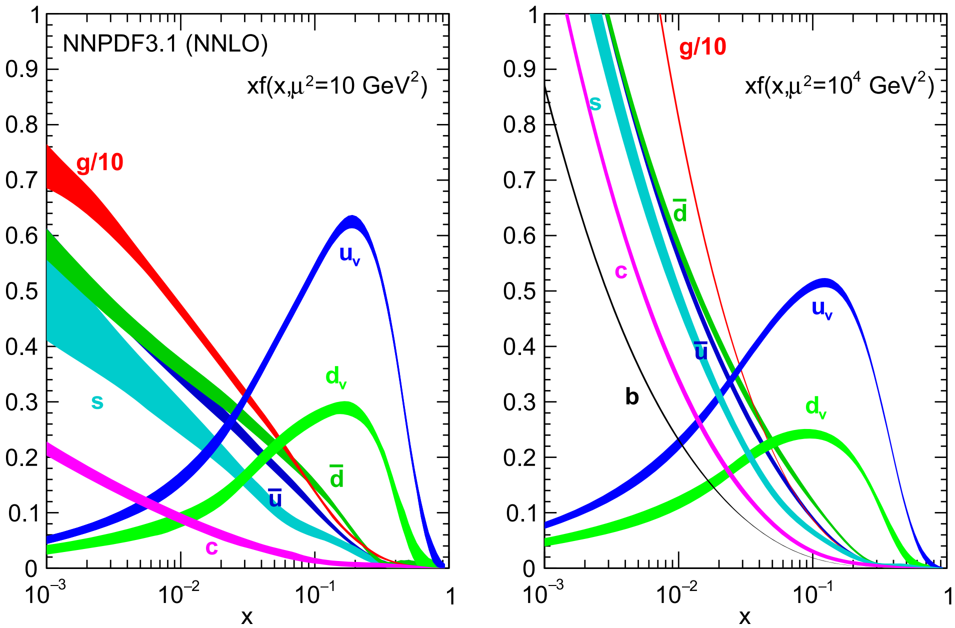


Figure 2.3: The NNLO proton PDFs at squared energy scales $10\,\mathrm{GeV}^2$ *(left) and* $10^4\,\mathrm{GeV}^2$ *(right) [42].*

2.2 Coordinate system and detector layout

The CMS experiment uses the coordinate system depicted in Fig.2.4, where the z-axis points in the direction of the counter-clockwise proton beam when looking from above, y-axis points vertically upwards and x-axis points in the direction of the center of the LHC. The nominal collision point is the center point of the coordinate system.

The azimuthal angle ϕ is the angle on the x-y plane and is measured with respect to the x-axis. The radial component on this plane is denoted with r, while the polar angle θ is the angle between the particle's momentum and the positive direction of the z-axis. A commonly used Lorentz invariant quantity in high energy physics (HEP) that describes the angle of the particle relative to the positive direction of the z-axis is pseudorapidity η as defined by $\eta = -\ln\tan\left(\frac{\theta}{2}\right)$. When the mass of the particle with respect to the momentum is negligible it converges to rapidity $y = \frac{1}{2}\ln\left(\frac{E+p_L}{E-p_L}\right)$, but is more convenient to use as it only depends on the polar angle θ with high $|\eta|$ values correspond to the *forward/endcap* direction, while small $|\eta|$ values correspond to *barrel* region.

Angular distances in this coordinate system are given by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

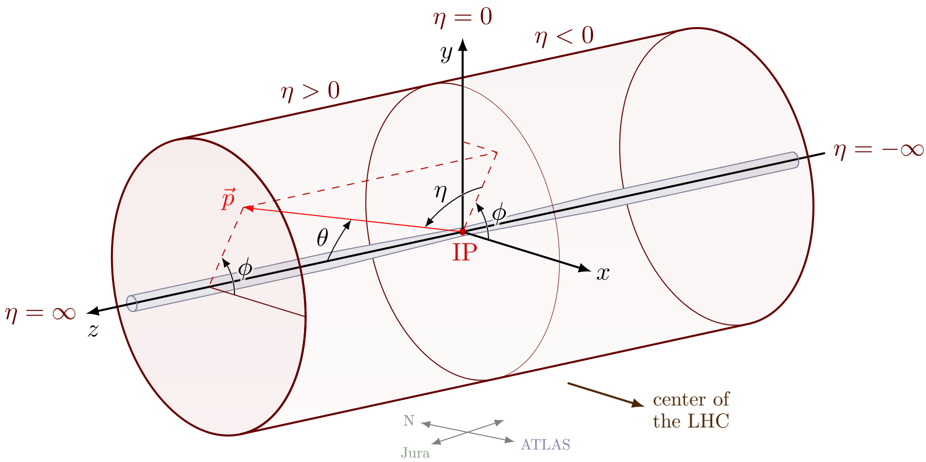


Figure 2.4: CMS coordinate system [43].

2.3 Tracking system

Efficient and precise reconstruction of the trajectories of charged particles with transverse momentum above 1 GeV in the pseudorapidity range $|\eta| < 2.5$ is a part of the LHC physics program. However, at the designed LHC luminosity of $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ there will be on average 1000 particles from 23-32 overlapping proton-proton interactions traversing the tracker at each bunch crossing happening every 25 ns. Therefore, in addition to the intense particle flux that could cause severe radiation damage to the tracking system, the tracker should offer high granularity and fast response in order to reliably identify trajectories and attribute these to the correct bunch crossing.

The CMS tracking system depicted in Fig. 2.5 has a cylindrical volume with a length of 5.8 meters and a diameter of 2.6 meters and has 10 layers of silicon microstrip detectors. Additionally 3 layers of silicon pixel detectors, which make up the innermost subdetector, are placed close to the interaction region which improves the measurement of impact parameters of tracks, i.e., the the charged particles, and the position of secondary vertices, corresponding to the position where a particle decayed. Inner tracking system provides a precise and efficient measurement of high number of tracks that are associated with each collision.

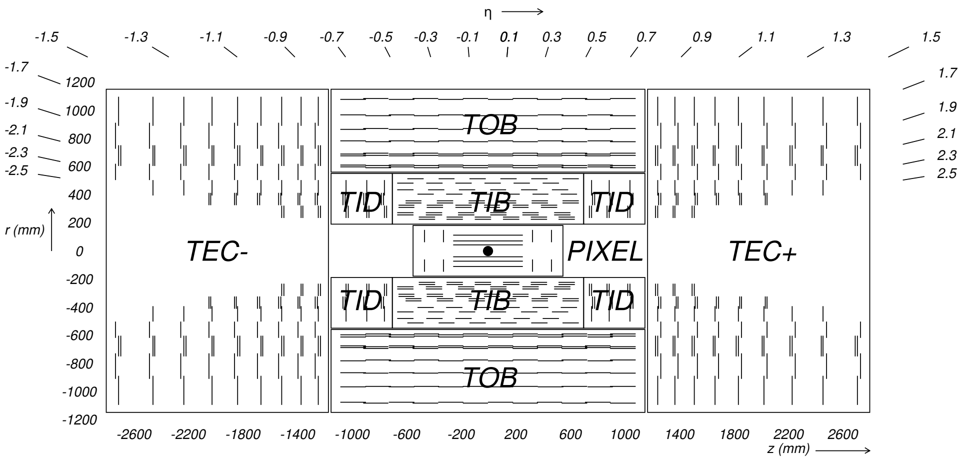


Figure 2.5: Schematic of the CMS tracker cross section. Each singular line corresponds to a detector module and a double line indicates back-to-back modules that deliver stereo hits. Figure taken from Ref. [6]

Good vertex finding is necessary for pileup rejection as it allows us to infer the primary vertex more accurately and gives us better track reconstruction ability, thereby allowing us to reconstruct tracks and identify particles more accurately with a better momentum resolution.

Furthermore, a precise measurement of secondary vertices and impact parameters is necessary for the efficient identification of heavy flavors (such a b-quarks) which are produced in many of the interesting physics channels. Also, the reconstruction of hadronically decaying τ leptons, as described in more details in Sec. 2.7.4, benefit from the precise assessment of the secondary vertices as they are a signature in several discovery channels and need to be reconstructed in one-prong and three-prong decay topologies.

This fine-grained tracker together with the ECAL and muon system allows to identify electrons and muon and provides a pure and efficient track reconstruction in jets with p_T up to ~ 1 TeV.

Tracking information is heavily used in high level trigger in CMS, which allows to reduce the event rate from 40 MHz to 100 Hz.

2.3.1 Pixel detector

In order to deal with the harsher conditions such as higher instantaneous luminosity and pileup that more than doubled in comparison to the design values, the CMS Phase-1 pixel detector was replaced with an improved pixel system [44] in the year-end technical stop of the LHC in 2016-2017.

With the upgrade, the new pixel detector consists of three disks at each end and four barrel layers, with the very first one being brought closer to the interaction point as the beampipe was replaced with a thinner one. The hit coverage is up to pseudorapidity range of $|\eta| < 2.5$ and the total pixel detector silicon area was increased more than 70%. The comparison of the old and new configuration is given in Fig. 2.6.

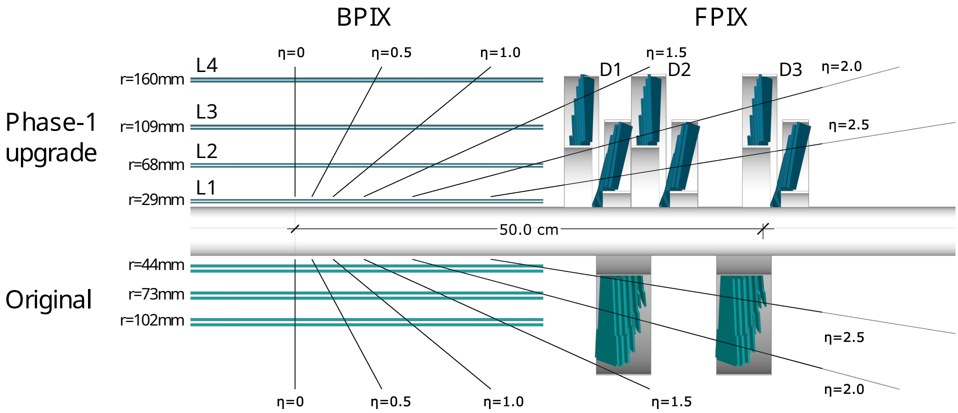


Figure 2.6: Comparison of the old and new tracker layout in the longitudinal view. The pixel detector consists of two mechanically and electrically independent parts, the barrel pixel detector (BPIX) and forward pixel detector (FPIX) disks [44].

As a consequence and better tracking performance (due to added redundancy to avoid hit losses, and therefore a better track and pattern recognition) was achieved together with more robust tracking and improved radiation tolerance while allowing higher rates.

These significantly higher data rates are now possible because of the new readout chip design and the increased bandwidth of the digital data transmission via optical links.

The upgraded pixel detector is designed to last until the end of Run-3 (2023-2025) after which yet another upgrade is needed in order to facilitate the High-Luminosity LHC (HL-LHC) run.

2.4 Calorimeters

2.4.1 Electromagnetic calorimeter

ECAL is meant to measure the energies of particles that are interacting mainly via the electromagnetic force such as electrons, positrons and muons. The key criteria in the design of the ECAL was the ability to detect the decay of the Higgs boson into two photons. In order to do these measurements ECAL system consisting out of lead tungstate ($PbWO_4$) crystals that have a coverage in pseudorapidity of $|\eta| < 3.0$ is placed within the solenoid volume.

The measurements are done using scintillation light produced in the crystals is detected by silicon avalanche photodiodes (APDs) in the ECAL barrel region (EB) region ($|\eta| < 1.479$) and by vacuum phototriodes (VPTs) in the ECAL endcap region (EE) ($1.479 < |\eta| < 3.0$) region. The thickness of ECAL in radiation lengths is bigger than $25.8X_0$[1]

Additionally, in order to resolve the photons from π^0 decays so as to discriminate them from prompt (i.e., produced close to the beamline) photons, a preshower system is installed in front of the EE. This is much finer-grained detector that consists of two layers, each comprising a lead radiator followed by a plane of silicon strip sensors which can indi-

[1] X_0 denotes the radiation length, so the distance a particle needs to travel to decrease it's energy by a factor of $1/e$.

cate the presence of a photon or an electron in the ECAL by requiring an associated signal in the preshower.

This design of combining active detector with passively absorbing material facilitates a very precise energy resolution of:

$$\frac{\sigma}{E[\mathrm{GeV}]} = \sqrt{{\frac{2.8\%}{\sqrt{E[\mathrm{GeV}]}}}^{2} + {\frac{12\%}{E[\mathrm{GeV}]}}^{2} + 0.3\%^{2}} \tag{2.2}$$

2.4.2 Hadronic calorimeter

The idea of HCAL is to measure the energies of particles that mainly interact via the strong nuclear force. This means that HCAL system is particularly important for the measurement of hadron jets and neutrinos or exotic particles resulting in apparent missing transverse energy.

Both ECAL and the brass/scintillator sampling HCAL have a coverage in pseudorapidity of $|\eta| < 3.0$. In CMS HCAL depicted on Fig. 2.7 consists of several layers of brass absorbers and plastic scintillator tiles. Scintillation light is converted by wavelength shifting (WLS) fibers embedded into the scintillator tiles and thereafter channeled via clear cables to photodedtectors[2] where they are detected.

The central part of HCAL is complemented by the hadron outer calorimeter (HO) called the a *tail-catcher* that resides in the barrel region and is placed outside of the solenoid. With this addition hadronic showers are ensured to be sampled with nearly 11 λ_I[3].

The forward iron/quartz-fibre calorimeters ensure coverage up to $|\eta| < 5.2$ for the measurement of transverse energy E_T in the event, with the thickness of HCAL is in the range of 7-11 λ_I (10-15 λ_I with HO included) depending on η.

By using high density crystals a design of a calorimeter that fulfills all important characteristics (being fast, having fine granularity and being radiation resistant) in the LHC environment were made possible.

The energy resolution of the ECAL and HCAL in the barrel region according to Ref. [45] is measured to be:

$$\frac{\sigma}{E[\mathrm{GeV}]} = \frac{84.7\%}{\sqrt{E}} + 7.4\%\,. \tag{2.3}$$

2.5 Muon system

The outermost subdetector is the muon system depicted in Fig. 2.8. It consists of several layers of aluminum drift tubes in the barrel region and cathode strip chambers in the endcap part.

Outside the solenoid coil, the magnetic flux is returned through a yoke consisting of three layers of steel interleaved with four muon detector planes. Drift tube (DT) chambers and cathode strip chambers (CSC) detect muons in the regions $|\eta| < 1.2$ and $0.9 < |\eta| < 2.4$, respectively, and are complemented by a system of resistive plate chambers (RPC) covering the range $|\eta| < 1.6$.

[2]These hybrid photodiodes (HPDs) provide gain and operate in high axial magnetic fields.

[3]λ_I denotes the hadronic interaction length, meaning the distance that a particle travels before it undergoes inelastic collision

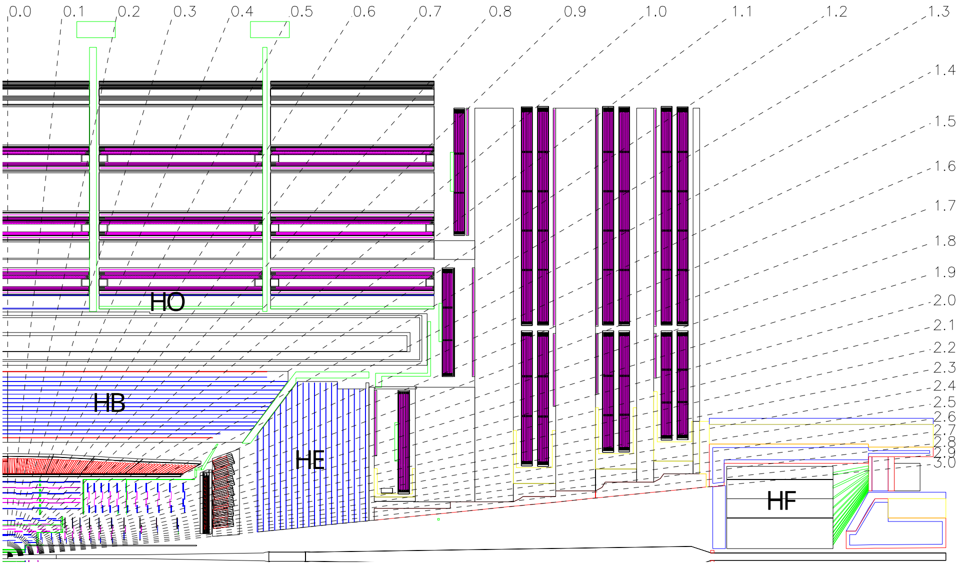


Figure 2.7: Cross section of the HCAL system with its subsystems - hadron barrel calorimeter (HB), hadron endcap calorimeter (HE), HO and hadron forward calorimeter (HF) [6].

Precise and robust measurements of muons has been one of the goals of CMS design from the start. Detecting muons plays an important role in CMS as it provides a powerful tool to recognize signatures of interesting processes even with very high background rates.

The three main purposes the muon system serves are muon identification, measuring the momentum and triggering. A good muon momentum resolution and trigger capability are enabled by the solenoidal magnet and its flux-return yoke, which acting as an absorber prevents the hadronic showers from reaching the muon system.

As the CMS muon system is designed to have the capability of reconstructing the momentum and charge of muons over the the entire kinematic range of the LHC.

In addition to the muon system, tracker provides independent muon momentum measurements, which are mostly used for particles with a transverse momentum of $p_T < 200\mathrm{GeV}$. This is coming from the fact the calorimeters and the solenoid coil represent a large amount of material ($> 16\lambda_I$) before the muon system, which consequently induces multiple scattering.

2.6 Triggers

The peak instantaneous luminosity of 2×10^{34} cm^{-2}s^{-1} and an average pileup of ~ 30 means having thousands of particles crossing the CMS detector in each bunch crossing, happening every 25 ns. Unfortunately saving this amount of data is not within the current system capabilities - one single event in the AOD takes 500 kb of disk space, meaning one would need to write $500\mathrm{kB} \cdot \frac{1}{25\mathrm{ns}} = 20\frac{\mathrm{TB}}{\mathrm{s}}$ to the disk.

In order to satisfy the system limitations one can save events with a maximum event rate of 1 kHz while keeping as much of interesting events as possible for the offline analysis. This is achieved by the two-level trigger, where the first, the Level-1 Trigger (L1T), is firmware based and reduces the event rate from 40 MHz to 100 kHz [46]. The L1T is based on the information from ECAL, HCAL and muon chambers. With this information it

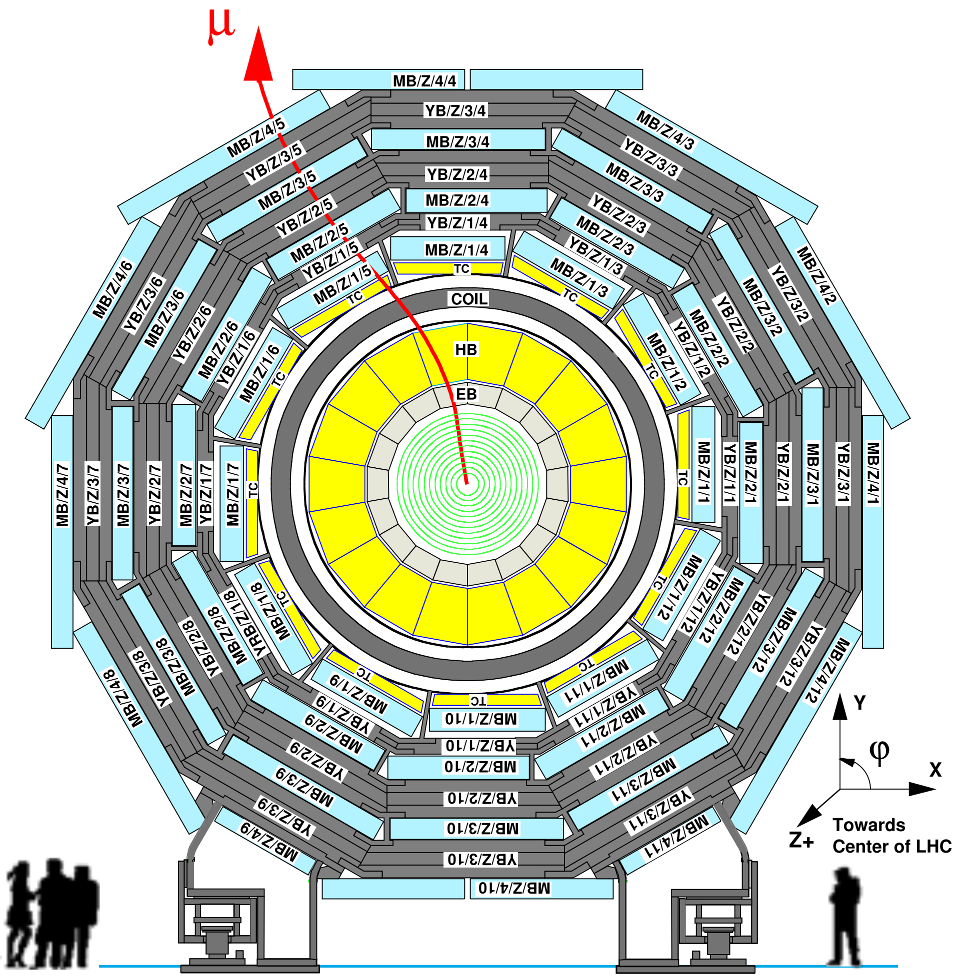


Figure 2.8: Cross section of the muon chamber layout [6].

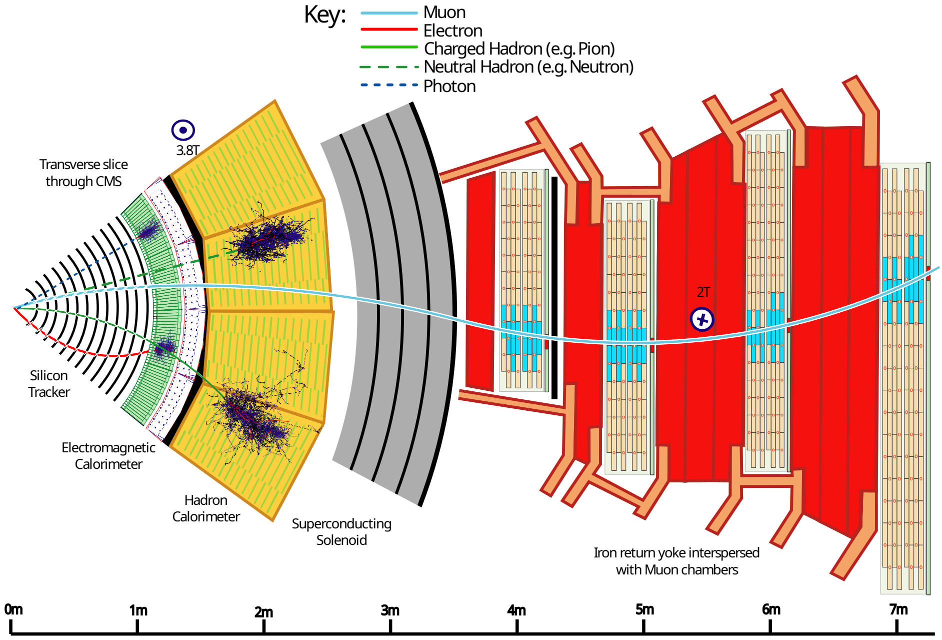


Figure 2.9: A slice of the CMS detector [40].

is possible to perform complex selections and high-level quantity computations (such as invariant mass of a pair of objects) in the last stage of the L1T trigger, the global trigger.

The second trigger, software-based High Level Trigger (HLT), reduces the rate further to the target value of 1 kHz. In principle HLT is just a trimmed down version of the offline reconstruction software that runs on a computer farm. The HLT menu is composed of more than 400 different HLT paths[4] For the reconstruction purposes PF algorithm as described in Sec. 2.7 is employed, improving the energy resolution of the trigger objects. Consequently the online reconstruction and selection is much closer to the one in the analyses.

2.7 Particle reconstruction

A cross-section of the CMS detector together with all detector elements described in the previous sections is shown on Fig.2.9.

All the information from various subdetectors is taken as an input by the PF [40] algorithm that reconstructs the measured particles and determines their properties. This is done by combining corresponding measurements from the basic elements (tracks[5] and clusters[6]) from all detector layers to identify the final-state particle. Different species of neutral and charged hadrons will not be separated in the PF reconstruction algorithm.

The first subdetector the resulting particles from the collision enter is the tracker in which tracks and vertices from hits in the sensitive layers are reconstructed. The mea-

[4] sequence of reconstruction and filtering modules.

[5] formed from hits in the tracker

[6] energy deposits in a calorimeter

surement of the momenta and the electric charge of the charged particles in the tracker is made possible by the magnetic field that bends the trajectories of the said charged particles. Track reconstruction, based on Kalman Filter (KF), is done in three distinct steps. In the first step an initial seed is generated such that few hits from a charged-particle trajectory would be compatible with it. Next, the trajectory is built by gathering hits from all the tracker layers along the trajectory of the charged-particle that is used for final fitting to determine the properties of the charged particle in the third step. The properties include measures such as the origin, direction and transverse momentum.

In the next subdetector, ECAL, electrons and photons are absorbed. The clusters of energy recorded in the neighboring cells resulting from the electromagnetic showers are used to infer the energy and the direction of the particles. The hadronic showers that might be initiated already in the ECAL by the charged and neutral hadrons are then fully absorbed in the HCAL, where again the created clusters are used to estimate the direction and the energy of the particle. However, if a charged hadron is missed by the tracking algorithm and is detected in the calorimeters, then it will be identified as a neutral hadron. This measurement will be with a reduced efficiency and very degraded energy resolution because of the biased direction caused by the bent trajectory in the magnetic field.

The two particles that leave the calorimeters with little to no interaction are muons and neutrinos, the latter of which escape undetected. Muons however do produce hits in the muon system, which together with the inner tracking assist in identifying muons determining their properties.

In order to reconstruct decaying particles, basic particles, such as electrons, photons, muons, and charged and neutral hadrons need to be reconstructed by the PF first.

In the following sections a brief overview of the reconstruction of electrons (Sec. 2.7.1), muons (Sec. 2.7.2), jets (Sec. 2.7.3) and taus (Sec. 2.7.4) is given.

2.7.1 Electron reconstruction

Electrons are reconstructed using the track information from the tracker and the cluster properties from the ECAL, with a requirement that the momentum-energy ratio is compatible with unity[7] while not being connected to a HCAL cluster.

There are two distinct electron seeding approaches: ECAL-based approach and tracker based approach.

The ECAL-based approach considers ECAL clusters with $E_T > 4$ GeV and uses the cluster energy and position to infer the locations of hits in the tracker layers. Due to the thickness of the tracker, most of the electrons will lose a large fraction of their energy before ECAL via bremsstrahlung. This, means that in order to get good reconstruction performance, one needs to include the bremsstrahlung photons in the electron reconstruction. This is done by collecting additional ECAL energy in a window that is narrow in η and wide in ϕ around the electron direction in order to account for the bend in the electron trajectory.

However, this approach is not straight forward for electrons in jets due to overlapping contributions for other particle deposits in the supercluster[8] leading to large inefficiencies. Also, as the backward propagation from the supercluster is not that accurate, it is likely to include many hits from other particles in the inner tracker layers, which causes large misreconstruction rates. Furthermore, the tracks of electrons with small p_T are sig-

[7]This means, that the energy from the ECAL should not be bigger than the amount inferred from the track.

[8]Cluster of clusters with a spread in the ϕ direction

nificantly bent in the magnetic field such that the supercluster region cannot include all deposits.

In the tracker based approach tracks with their p_T exceeding 2 GeV are used as potential seeds and allows to reconstruct also electrons with a $p_T < 4$ GeV, thereby effectively selecting electrons and positrons also from conversions in the tracker material. When the electron radiates only small amounts of energy, the track can be reconstructed across the whole tracker propagated to the ECAL surface and then be matched with the closest ECAL cluster. Even if there is some soft photon emission, most of the hits along the electron trajectory will still be collected with pattern recognition.

Based on the χ^2 and the number of hits of a track, a preselection is applied. The tracks passing the preselection are refitted with a Gaussian-sum filter (GSF). Finally, a final selection based on the score of a boosted decision tree (BDT) that takes as input the χ^2, the number of hits of a track, the energy lost along the GSF track, the distance between the extrapolation of the track to the ECAL inner surface and the closest ECAL cluster, and the ratio of GSF and KF track fits, is done.

After reconstructing the electron candidate, various selection criteria are applied for different working points (WPs). These include for example contraints on the electron kinematics, track criteria, multivariate analysis (MVA) cuts (as defined by EGamma and JetMET physics object groups (POGs)), which are given in Table. 2.1.

2.7.2 Muon reconstruction

Based on the track forming method, there are three different types of muons that compose the final muon collection: standalone muons, global muons and tracker muons. Standalone muons are seeded by the DT and CSC hits when forming a track. Tracker muons are formed by the *inner tracks*[9] that have a p_T larger 0.5 GeV and a total momentum bigger than 2.5 GeV. A tracker muon track is formed if any of the muon segment is compatible with the extrapolation of that track. Global muons are formed by the standalone muons that are matched to the inner tracks

As the inner track and muon segment reconstruction are done with a high degree of efficiency, $\sim 99\%$ of the muons produced within the muon system's acceptance are reconstructed as a global muon, tracker muon or even both. This efficiency can be attributed to the precise measurements by the inner tracker and to the high purity of the calorimeters that are absorbing all other particles[10] before they might reach the muon system.

Standalone muons are rarely used by themselves as they are often mixed with cosmic muons and have worse momentum resolution.

Similarly to the electron reconstruction, also for muons selection criteria are applied for different WPs. A summary table of muon selection criteria is given in Table. 2.2.

2.7.3 Jet reconstruction

Cascades of charged and neutral particles are referred to as jets. They are formed due to hadronization as according to color confinement, particles with a color charge cannot exist freely in the wild.

In CMS, jets are clustered from PF candidates by a jet clustering algorithm such as the anti-k$_T$ algorithm, such that jets with a cone size of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ are used as standard jets. These standard jets are also known as "AK4". Jets with a cone size of $\Delta R = 0.8$ are referred to as "AK8" jets and are used for cases where the two subjets are

[9]Track in the inner tracker

[10]with the exception of neutrinos

Table 2.1: *Loose, fakeable and tight selection criteria for electrons. A hyphen (−) indicates selection criteria that are not applied.*

Electrons			
Observable	Loose	Fakeable	Tight
Cone-p_T	> 7 GeV	> 10 GeV	> 10 GeV
$\|\eta\|$	< 2.5	< 2.5	< 2.5
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_e	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
$\sigma_{i\eta i\eta}$	−	$<$ { 0.011 / 0.030 }[1]	$<$ { 0.011 / 0.030 }[1]
H/E	−	< 0.10	< 0.10
1/E - 1/p	−	> -0.04	> -0.04
Conversion rejection	−	✓	✓
Missing hits	≤ 1	$= 0$	$= 0$
EGamma POG MVA	>WP-loose[2]	>WP-90 (>WP-loose)[2,†]	>WP-loose[2]
Deep Jet of nearby jet	−	<WP-tight[3] (<WP-medium[3])	<WP-medium[3]
Jet relative isolation[4]	−	< 0.7 (−) †	−
Prompt-e MVA	−	< 0.30 (> 0.30)	> 0.30

[1] Barrel / endcaps.
[2] WPs as defined by EGamma POG.
[3] WPs as defined by JetMET POG.
[4] Defined as 1/p_T^{ratio}-1 if the electron is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with ΔR=0.4 otherwise.
† Fails (passes) the requirement prompt-e MVA > 0.30.

more difficult to separate. The examples of "AK8" jets could include for example jets for W boson decays.

The energy or jets consist of only hadrons and photons, can be measured by only the calorimeters such that individual jet particles are not needed to be separated. These jets can therefore be reconstructed without any contribution from the tracker and the muon detectors [40].

As reconstructed jets originate from various sources, such as b-quark hadronization, a deep neural network (DNN) named DeepJet [47] is used in order to discriminate between jets stemming from b-quarks, c-quarks, gluons and light-quarks[11].

2.7.4 Tau reconstruction

Tau leptons is the heaviest particle in the SM theory that is a key component in multiple measurements like electroweak interactions, lepton flavor universality, production and CP properties of the Higgs boson via its Yukawa coupling to fermions as well as several BSM models.

[11] u,d,s quarks

Table 2.2: Loose, fakeable and tight selection criteria for muons. A hyphen (–) indicates selection criteria that are not applied.

Muons			
Observable	Loose	Fakeable	Tight
p_T	> 5 GeV	> 10 GeV	> 10 GeV
$\|\eta\|$	< 2.4	< 2.4	< 2.4
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_μ	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
PF muon	>WP-loose[1]	>WP-loose[1]	>WP-medium[1]
Deep Jet of nearby jet	–	<WP-interp. (<WP-medium)[2]	<WP-medium[2]
Jet relative isolation[3]	–	<0.8 (–) †	–
Prompt-μ MVA	–	< 0.5 (> 0.5)	> 0.5

[1] WPs as defined by Muon POG.
[2] Upper cut on the Deep Jet score defined with a linear interpolation from Deep Jet WP-medium at cone-p_T 20 GeV to Deep Jet WP-loose at cone-p_T 45 GeV, taking the Deep Jet WPs as defined by JetMET POG.
[3] Defined as 1/`jetPtRatio`-1 if the muon is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with ΔR=0.4 otherwise.
† Fails (passes) the requirement prompt-μ MVA > 0.5.

Tau lepton has a mass of $m_\tau = (1776.86 \pm 0.12)$ MeV and is thus the heaviest known lepton, which can decay both leptonically and hadronically. Average lifetime of tau lepton is $t_\tau = (2.903 \pm 0.005) \cdot 10^{-13}$ s, which for a 30 GeV tau lepton corresponds to a decay length of:

$$\lambda_\tau = \mathrm{c} \cdot t_\tau \cdot \gamma\beta = (3 \cdot 10^{11}\mathrm{mm/s}) * (2.9 \cdot 10^{-13}\mathrm{s}) \cdot (30\mathrm{GeV}/1.78\mathrm{GeV}) \approx 1.5\mathrm{mm}\,,$$

with c as the speed of light and $\gamma\beta = p/m$, with momentum p and rest mass m written in natural units. Since the innermost layer of the silicon tracker in CMS is at a distance of ~ 3 cm from beamline, then the fraction of prompt τ leptons that decay after reaching the innermost layer of the detector is negligible.

Tau decays always involve at least one charged particle, referred to as *prong*. Leptonic decays are classified as 1-prong decays, while hadronic decays are mostly mediated by mesonic resonances leading to final states with 1-, 3- or 5-prong decays, though the latter amounts only about 0.1% of the total decays.

As shown of Fig. 2.10, $\tau^\pm$ decays feature either a charged lepton ($e^\pm$ or $\mu^\pm$) and corresponding 2 neutrinos or a few hadrons and one neutrino. Hadronic τ decays, denoted by τ_h can be separated from quark and gluon jets by considering the multiplicity of the particles contained within the jet and the jet radius.

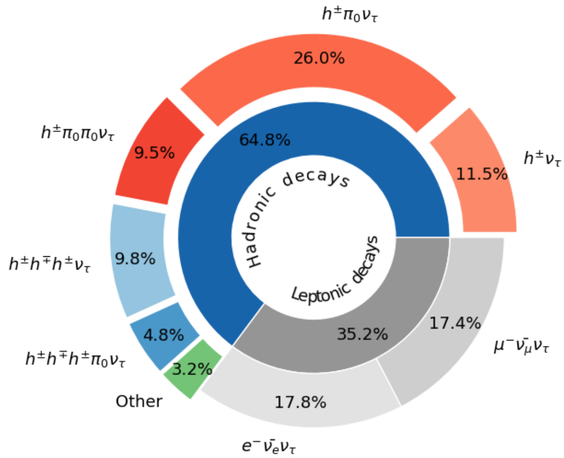


Figure 2.10: Tau decaymodes. Two thirds of tau lepton decay hadronically

Reconstruction of hadronically decaying tau lepton As τ leptons that are decaying leptonically are already reconstructed as electrons and muons, then in the context of HEP only hadronically decaying tau leptons τ_h are considered.

In CMS τ_h are reconstructed using hadron-plus-strips (HPS) algorithm [48] that takes PF candidates as input. As can be seen from Fig.2.10 τ_h always has a odd number of charged hadrons (prongs) and can include 0 or more π^0. The *strip* in the name refers to π^0 decay products, so the electrons and muions that are confined within a dynamically defined phasespace in the $\Delta\eta \times \Delta\phi$ plane. The strips are seeded by the highest p_T electron or photon that isn't yet included in a strip. Next, any photon (electron) that is within the strip and particle p_T dependent area will be merged with the strip and the new weighed p_T of the strip is calculated. This process continues until there are no electrons and photons included in a strip.

In order for a charged hadron to be considered in the τ_h reconstruction, it needs to have a $p_T > 0.5$ GeV and originate from a location near to the primary vertex. From τ travel distance calculated earlier, this distance is set to $d_{xy} < 0.1$ cm.

The initial τ_h identification was done based on the isolation of the τ_h reconstructed by the HPS. However, this approach was superseded by a multivariate convolutional deep neural network named DeepTau [49] in order to increase efficiency. This approach divides the CMS detector into a grid of cells with $\Delta\eta \times \Delta\phi = 0.02 \times 0.02$ within the signal cone of size $\Delta R = 0.1$ around the τ_h axis. Within the isolation cone of size $\Delta R = 0.5$ the grid is divided into cells of size $\Delta\eta \times \Delta\phi = 0.02 \times 0.02$.

For each PF candidate that falls into one of the two grid, up to 37 variables are in addition to the reconstructed τ_h properties are used by the network. This approach increases the identification efficiency for τ_h by approximately 20%

With the ongoing Run3 and the upcoming HL-LHC, most probably the currently used DeepTau algorithm needs to be revisited. For the purposes of tau reconstruction already some novel approaches have been proposed, such as graph-based tau reconstruction [50] or even transformers like ParticleTransformer [51] that initially was proposed for jet tagging. The performance of various approaches for tau reconstruction and identification was studied in the context of this thesis also in [4].

Table 2.3: The DeepTau discriminant WPs. Summary table based on Ref. [49] is taken from Ref. [52].

		VVTight	VTight	Tight	Medium	Loose	VLoose	VVLoose	VVVLoose
$D^{\tau_h}_{jet}$	efficiency (%)	40	50	60	70	80	90	95	98
	misidentification rate (%)	$0.2-0.5$	$0.4-0.8$	$0.6-1$	$1-2$	$2-4$	$4-9$	$7-10$	$10-20$
$D^{\tau_h}_{e}$	efficiency (%)	60	70	80	90	95	98	99	99.5
	misidentification rate (%)	0.01	0.03	0.07	0.2	0.5	1	3	7
$D^{\tau_h}_{\mu}$	efficiency (%)	✗	✗	99.5	99.8	99.9	99.95	✗	✗
	misidentification rate (%)			0.03	0.04	0.06	0.2		

3 Machine learning

With the ever increasing amount of data produced in the LHC and the number of events happening each second, the task of analyzing all the data and simulating events has become a difficult task to solve. Writing a program in the traditional way, giving it some conditions to filter out events, is not feasible as being able to cover all the different cases is improbable and maintaining this long list of complex rules would be hard to maintain. This is exactly where various machine learning teachniques shine — finding rare signal in huge amounts of data, dealing with fluctuating environments, speeding up simulations and reconstructing objects via pattern recognition are only some areas ML are known to solve well.

In general, machine learning algorithms can divided into categories based on how they learn — what kind of supervision the system receives during training, can it learn incrementally on the fly and does the model compare the similarity between known instances or does it create a model based on the instances and make predictions based on that model.

Based on the amount of supervision the system receives, one categorizes the ML algorithms as supervised, semisupervised, unsupervised or as using reinforcement learning. In case of supervised learning one provides labels of all the instances to the model, while in cases of semisupervised (unsupervised) training labels are provided only for a subset of samples (are not provided for any sample). Similarly to unsupervised learning, also reinforcement learning does not use any labeled data — instead it acts based on the feedback it received from the environment after taking a given action. The feedback from the environment can be interpreted as a reward that the agent tries to maximize.

Training a model incrementally, possibly in a live environment, is referred to as online learning, while training on the full dataset all in one go is called batch learning. Further categorization can be done whether the predictions the trained model makes are based on the similarities between the know instances or based on the model created on the known instances.

Many ideas introduced in this chapter can be found in most of classic ML textbooks such as Ref. [53].

In the context of the $\mathrm{HH} \to \mathrm{multilepton}$ analysis as described in Sec. 4, we used supervised training with the trained BDT discriminators being given the full training dataset with labels all in one go. Consequently, only supervised learning techniques, out of these categories will be discussed in more details.

3.1 Supervised learning

In supervised learning (SL) the desired target values for all training instances, i.e *labels*, are provided to the training algorithm. Therefore, for each given input object in $x_i \in X$, there is a corresponding target value $y_i \in Y$ that the machine learning model tries to predict using a mapping ϕ:

$$\phi : X \mapsto Y\,, \tag{3.1}$$

where the symbol X refers to the set of input vectors and Y the corresponding labels. Using these labels one mostly aims to solve either regression, classification or ranking tasks.

In order to let the model know how far off is its prediction from the desired one, an objective function (OF) needs to be defined. A typical OF (denoted as Ω) consists out of two parts - the loss function (LF) denoted as Λ and a regularization term denoted as ρ:

$$\Omega(\theta) = \Lambda(\theta) + \rho \tag{3.2}$$

The goal of training loss is to measure how predictive our model is with respect to the data we are training on, whereas regularization control the amount of complexity of our model. The value of the training loss is usually a function of the predicted values ($\hat{y}$) and the target values y. Common loss functions include for example mean-squared-error (MSE) for regression and cross-entropy loss for classification tasks. A complex model, i.e., model with a large amount of training parameters, is prone to overfit[1], thus not making good predictions on the new instances.

Generally a simple and predictive model is preferred. The trade-off of a model being predictive and simple is referred to as the *bias-variance trade-off* [54] in machine learning.

Most notable SL algorithms employed in HEP analyses are neural networks and decision trees that will be discussed in the following sections in context of supervised learning.

3.2 Neural networks

Neural network (NN) is a wide class of ML algorithms used for supervised, unsupervised, semisupervised and reinforcement learning. They frequently outperform other ML techniques on very large and complex problems. However owing to their complex structure, they generally work best if millions of training samples are available as if only a few training samples are available per feature, the neural net is prone to overfit the data.

The complexity of the NN means also that they feature a lot of hyperparameters (see Sec. 3.5.2), which means that it is very difficult to find the optimal configuration for the training.

A depiction of a NN is shown in Fig. 3.1. This NN takes three features as input and predicts two values. The layer consisting out of the input neurons is referred to as the *input layer* and is highlighted by a dashed green box in the Fig. 3.1, while the layer consisting out of the output neurons (i.e., the predictions) is referred to as the *output layer* and is highlighted by a dashed red box. The layers between the input and output layer are called *hidden layers*, of which there are three in the example NN shown in Fig. 3.1 and

[1]Instead of approximate the underlying function, the models starts to fit the statistical fluctuations in the training dataset, which, however, might not be present in the validation and testing datasets (see Sec. 3.5.3.

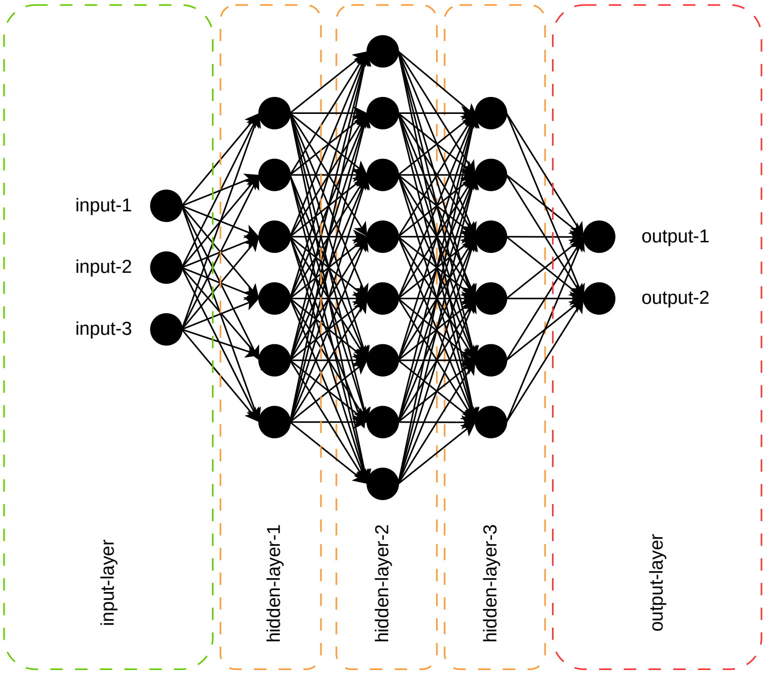


Figure 3.1: A possible layout of a neural network that takes three features as inputs and predicts two values. The input nodes are highlighted in a dashed green box, while the output nodes are in a dashed red box. The three hidden layers are shown in orange boxes. Each black dot represents a node and each black line represents a weight.

that are highlighted by a dashed orange box. All the hidden layers in this example are *fully connected* aka. *dense*, meaning that all neurons in a given layer are connected to every neuron in the previous layer (i.e., its input neurons).

The training of a neural network done using a process called *backpropagation* [55], where the differences between target values and predictions made in a *forward pass* are computed and the error gradient[2] is propagated from the output layer to the input layer. In other words, backpropagation is the process of updating the weights of the network in order to reduce the error of its prediction. The error gradient is a function of the model parameters (weights and biases) that the chosen optimizer will try to minimize. Commonly used optimizers include gradient descent [56], Adam (Adaptive Moment Estimation) [57], LBFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno) [58] or even evolutionary algorithms described in Sec. 3.5.2. In essence, backpropagation is just a way to determine the error gradient in a non-analytical manner, with the connection weights of the network are tweaked using the computed error gradients.

During the forward pass segment and in case of a fully connected layer the input value for each neuron is computed as:

$$h_{\mathbf{W},\mathbf{b}}(\mathbf{X}) = \phi(\mathbf{X}\mathbf{W} + \mathbf{b}) \tag{3.3}$$

The input features for a given layer is denoted by the vector $\mathbf{X}$. The weight matrix $\mathbf{W}$ has a row of weights corresponding to each neuron. The $\mathbf{b}$ denotes the bias vector that contains all the connection weights between the bias neuron and artificial neuron. In Eq. 3.3 the

[2] difference of predicted value and the expected value for all nodes at each layer.

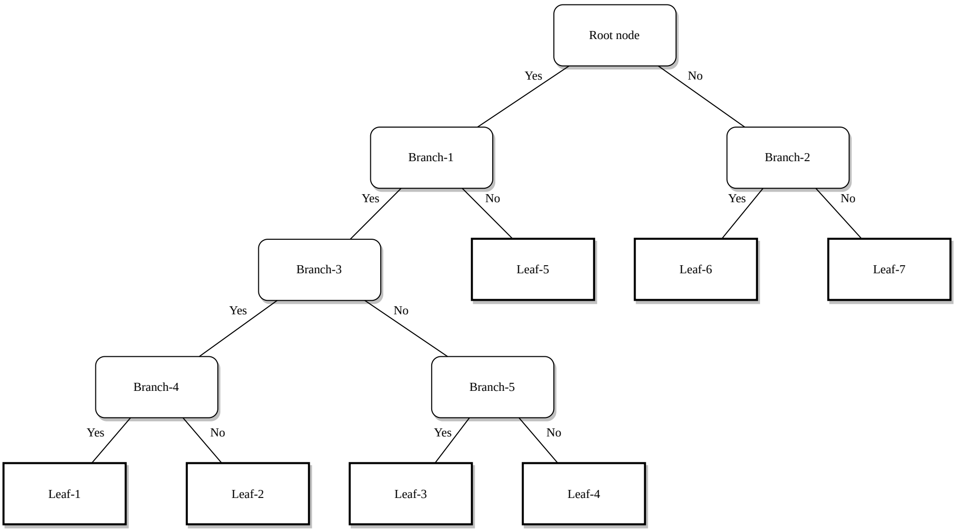


Figure 3.2: A decision tree starts with a root node, *where based on a given feature a split is made. These splits on feature values will be done until a maximum depth or target purity in a sample has been achieved. Each line in the figure corresponds to an answer to a question - this is called a* branch. *For each* leaf node *one can simply look at the predicted class.*

function ϕ is the *activation function* that introduces non-linearity to the backpropagation, as chaining together two linear functions ($\mathbf{XW}+\mathbf{b}$) without the added non-linearity, one would end up again with a linear function when using the chain rule.

When using MLP for regression one shouldn't use any activation functions in the output layer, making it then possible for these to output any range of values. This only for cases when the range is not set previously.

3.3 Decision trees

Decision trees are non-parametric[3] algorithms mostly used for supervised learning, that make very few assumptions about the data. This means that unlike for example linear models, no assumption about the data (being linear) is made, thus making it possible to fit the data very closely. Additionally, very little data preparation is needed and no *feature scaling*[4] is required.

Despite NNs being more flexible, by default decision trees perform well on structured data (such as tabular data).

An example decision tree is shown in Fig. 3.2.

Owing to the fact of being simpler, decision trees can work well with relatively small amounts of data. Still, they are powerful algorithms able to fit complex datasets and are able to perform classification and regression tasks both for single- and multi-output cases.

[3]The number of parameters is not determined prior to training. This allows the model structure to be flexible to fit the data closely. A parametric model would be for example a linear model, where the bias and slope are the two parameters to be determined. However by limiting the degrees of freedom one reduces the risk of overfitting.

[4]the process of normalizing the range of feature values in data

The two most important downsides of decision trees are its that they are prone to creating orthogonal decision boundaries, which makes them sensitive to rotations in the training set and that they are very sensitive to small variations in the training data. The former can be solved by employing principal component analysis (PCA) as it often will result in better orientation of the training data, while the instability issues can be limited by using for example *ensemble methods* [59] such as *bootstrap aggregating (bagging)* [60], *boosting* [61] and *stacking* [62] that average over several trees.

Two of the most widely used boosting methods are Adaptive Boosting (AdaBoost) [63] and gradient boosting [64]. In essence, both of these represent the same model, with the difference arising from the method they are trained. Both of these models are trained sequentially, with each ensuing one trying to correct the mistakes of the previous one. However while in the case of AdaBoost the sample weights of misclassified events will be increased for each new predictor, then in case of gradient boosting each new predictor will be trained on the residual errors made by the previous predictor.

As one might expect, using tree ensembles also comes with a downside - decision trees are called also *white box models* because in contrast to neural networks or *tree ensembles*, they are very intuitive and easy to interpret. In case of neural networks and tree ensembles it is very difficult to say why a certain prediction, despite potentially very accurate, was made. Decision trees are fundamental components of random forests (RFs)[5] and BDTs[6], acting as the *weak learners* that are combined to create a strong predictive model. Despite being BDTs being considered as black boxes, they are still more transparent compared to DNNs as they are easier to explain and feature usually less hyperparameters.

In order for an ensemble method to perform well, each predictor should be as independent from the others as possible. Aggregating different predictors can reduce both bias and variance, though in general the resulting model will be with a similar bias but lower variance than a single predictor that was trained on the original training set.

A flavor of BDT used extensively in this thesis is eXtreme Gradient Boosting (XGBoost) [65] (see Fig 3.3). As the training continues XGBoost becomes more precise as the errors are corrected every time the enseble grows.

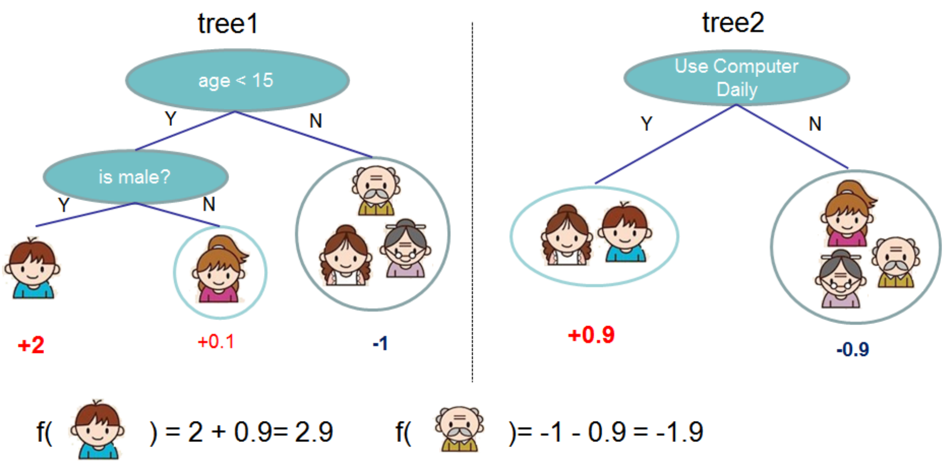


Figure 3.3: Two decision trees are trained sequentially, with the latter one aiming to correct the errors of the first one by taking the difference on the prediction and the truth values as input [65].

[5]decision trees can be trained in parallel and are combined using bagging

[6]decision trees are trained sequentially using boosting with their predictions are combined

The training process starts by making an initial prediction that serves as the base learner or the root of a given tree. If the base learner is too strong, learning can be negligible in the following rounds, thus the gains will be minimal. This is why for example a neural network can not be used as the "weak learner". Using the prediction by the first tree gradients and hessians with respect to these predicted values will be calculated for each instance of training data. Trees are added to the enseble sequentially such that each new tree aims to correct the mistakes that were made by the previous trees. This is achieved by recursively partitioning the data based on the feature values such that they minimize the loss function.

Each subsequently added tree does not contribute to the overall prediction equally, as with the *shrinkage* parameter (similar to learning rate in neural networks) the impact of additional trees will be scaled down, thus allowing the model to converge more gradually and reducing the risk of overfitting.

As one adds an additional split to the tree being constructed, a pruning score is calculated that takes into account how much the objective function improved and how much the loss reduced. Additionally, the L1 (Ridge) and L2 (Lasso) regularization terms are added in order to control the complexity of the model and to prevent overfitting. If the pruning score is below a given threshold, thus showing that removal of the node does not result in a significant increase in the loss function, that given node (and thus the rest of the subtree) is removed. Performing pruning results in deeper trees that are also more optimized.

This process continues until a given stopping criteria are met, after which a prediction is made by adding the predictions (adjusted by the shrinkage parameter) of all the trees in the enseble together. The stopping criterion can be a predefined number of trees, good enough performance or detection of overfitting.

Additionally, XGBoost incorporates other levers to combat overtraining[7], such as column subsampling (colsample-bytree) and row-subsampling (subsample), with the prior selects a subset of features to be used when creating a new tree and the latter the subset of training data instances to be used when finding the split.

As an aside, XGBoost handles missing values automatically and can deal with sparse data and provides a way to measure the importance of the features

3.4 Data augmentation

In order to train a state-of-the-art BDT or a NN in HEP usually a huge amount of data is needed. However, generating this data is often very time consuming and demands a lot of computing resources.

This problem can be combated by either increasing the size of the dataset artificially by adding new instances without the resource-heavy data generation process or by augmenting the already existing data with some more descriptive variables. This kind of procedures are known as data augmentation (DA) methods.

The first method is done on the dataset level and can in general be divided into two categories: real data augmentation (RDA) and synthetic data augmentation (SDA). In case of RDA only minor changes are made to the real data before augmenting the data set. These minor changes could be among plethora of other methods rotations (in case of rotational symmetry of the data) of an event or zooming [66].

More traditional sampling methods like oversampling, undersampling [67] and more complex generative models like generative adversarial networks (GAN) [68] and varia-

[7]A scenario where during training the model starts fitting already the noise

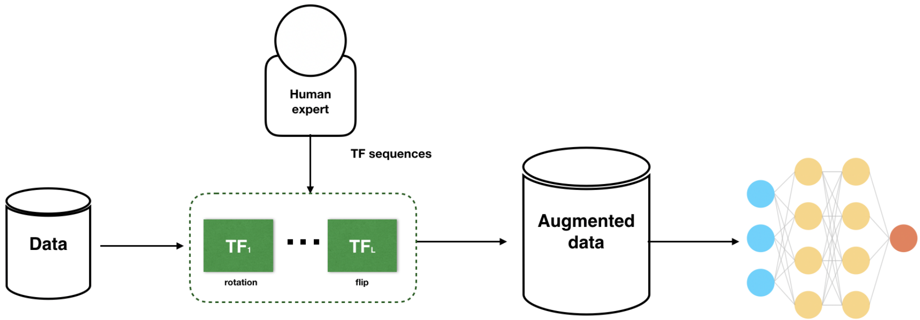


Figure 3.4: Augmenting data includes adding features and instances to the final dataset. Figure taken from [73].

tional autoencoder (VAE) [69] that could also be used in fast simulation which is a notable bottleneck in HEP analysis workflows [70, 71] are examples of SDA and are described in more detail in Sec. 3.4.1.

The second method called *feature engineering* is described in a bit more detail in Sec. 3.4.2.

Introducing noise into data to form additional data points improves the learning ability of several models which otherwise would perform relatively poorly [72], indicating that using DA one can create variations that the model might see also in the real world. Preprocessing data with DA usually results in superior training outcomes due to the face that it acts as a regularizer by reducing overfitting during training.

So in general employing DA methods to ones ML workflows results in improvement of model prediction precision as the data is not as scarce, thus enabling rare event prediction, reducing cost of data collection and labeling, overcoming class imbalance problems, improving overall model prediction precision and reduction of overfitting by creating data variability and thus also generalizing the model.

3.4.1 Sampling

An event (such as the production of a Higgs boson) that represents less than 5% of the dataset is considered a rare event [74]. When using such an imbalanced dataset, it is difficult to get a model that does accurate and meaningful predictions due to lack of information of this rare event [67]. To overcome this issue, the class distributions of the dataset can be adjusted by the means of various sampling techniques described in this section.

The most common sampling strategies employed in the ML pipelines are over- and undersampling [67]. In case of oversampling minority class samples will be randomly duplicated. However, this approach might lead to overfitting, as exact copies of the minority samples will be made.

Also undersampling is often a good solution to this problem as it tries to alleviate the class imbalance by discarding majority class samples by which potentially important majority class samples might be lost.

Oversampling and undersampling are in essence equivalent but opposite techniques. The advantage of both approaches is that they are straight-forward to implement.

A more complex oversampling technique that makes use of the creation of artificial data points is Synthetic Minority Over-sampling Technique (SMOTE) [75, 76]. The artificial data points for the minority class are created by a more involved oversampling. That is, synthetic samples[8] are created along the line segments joining any or all of the k nearest neighbors in the minority class. Doing this causes the classifier to create decision regions that are larger and less specific, making the classifier to learn more general regions rather than being subsumed by the samples from majority class around them. As a result by the use of SMOTE the classifier learning bias is shifted towards the minority class, thereby allowing the model to generalize better.

Yet another possibility is to use Adaptive Synthetic sampling (ADASYN) [76], with the idea being to use a weighted distribution for each minority class. The weight is dependent on the level of difficulty in learning that given class. This means more synthetic data would be generated for harder to learn minority classes. ADASYN learns adaptively the the weight distributions. Consequently, in addition to providing a balanced representation of the data distribution, the learning algorithm is forced to focus on learning difficult examples more. As per Ref. [76] ADASYN improves accuracy for both minority and majority classes and does not sacrifice one class in preference for another.

3.4.2 Feature engineering

Feature engineering focuses on creating instance level input features that allow the algorithm to learn the patterns more easily. In the context of HEP analyses these features are often physics inspired and derived from first principles.

Feature engineering (FE) is one of the key concepts in ML workflows and is used to make the training more effective. Although a sufficiently complex network can learn all patterns in the data by solely utilizing "low level" variables like four-momenta, the utilization of "high level" variables can often result in more advanced models. Models using both low and high level variables have been shown to outperform models using only one type of variable as shown in [77].

An example from HEP for feature engineering is Lorentz Boost Network (LBN) [77]. As illustrated on Fig. 3.5 LBN acts as the first stage in a two-level model by exploiting exclusively the four-momenta of the final-state particles by employing Lorentz transformations. This allows the network to exploit and uncover structures in particle collision events.

The three LBN hyperparameters depicted on Fig.3.5 are the number of final-state particles N, number of combinations M and features to be created F are to be chosen appropriately with regards to the research question (e.g the decay channel studied).

[8] artificially created samples that do not need to be physical

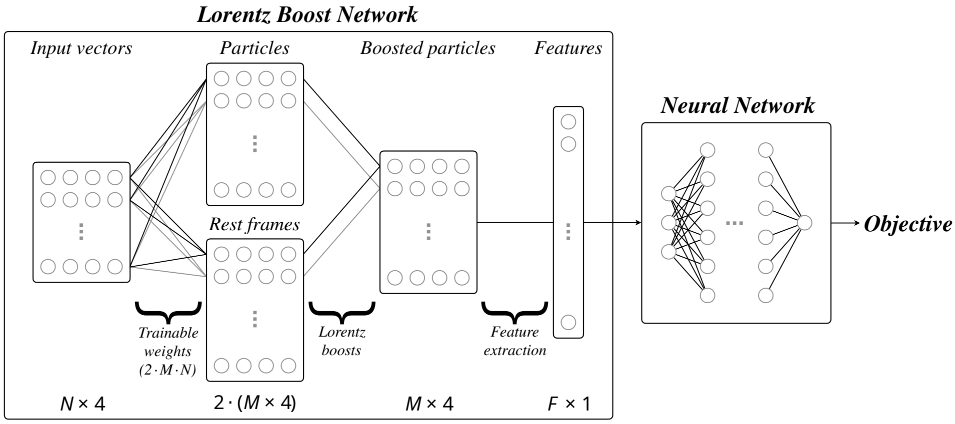


Figure 3.5: Lorentz Boost Network taking final state particles four-vectors as input in order to create new features by combining input particles and doing Lorentz transformations. Figure taken from [77].

LBN combines particles in two different ways - creating composite particles and forming appropriate rest frames. These composite particles are boosted into these rest frames using Lorentz transformations such that the properties of the parent particle could be inferred directly. Examples of these inferred properties could include for example masses or angles.

Finally, this information derived by the LBN is propagated to the second network that can be replaced depending on the specific analysis task like signal classification or mass regression.

3.5 Model optimization

Choosing an appropriate model and preparing data is only a part of the process of training an optimally performing model. In order to optimize the performance of the model one can for example choose to use only the most important features (Sec. 3.5.1) or do hyperparameter tuning (Sec. 3.5.2). After one has finished with model optimization a cross-validation (Sec. 3.5.3) needs to be done in order to ensure the model behaves as expected. The latter of these was studied in more details in the context of this thesis.

3.5.1 Feature importance

As already indicated in the introduction to the model optimization, feature importance plays an important role in both creating a robust model and by helping one to interpret the model. Features that do not help to distinguish between signal and background or that might be considered as noise make the model prone to overfitting, meaning the model starts to fit the noise. Additionally, including a plethora of features will increase the computational time it takes to train the model.

There exist various methods of evaluating the importance of features. The *permutation importance* [78] can be considered as the most general one of these, as it is agnostic to the choice of the model. The idea behind permutation importance is to shuffle the feature values or setting them to the mean value of the feature and then measuring drop in performance - the higher the drop the bigger the impact.

For tree ensembles measuring the feature importance can be done also by looking at the splits in the trees and how often a feature was used. A more frequent use of a feature in a split to reduce for example impurity or entropy would indicate a bigger importance of a feature.

Other examples of strategies for figuring out feature importances include for example SHapley Additive exPlanations (SHAP) [79–81], which could give more stable results by assigning each feature an importance value for a particular prediction.

As features can be in some extent correlated, removing a feature can increase the importance of another one. Removing features iteratively with a given step size can be done using for example Recursive Feature Elimination (RFE) [78] algorithm.

Having now a smaller amount of features for the model now allowes our model to train faster and using the information from the feature importances gives us a further insight into the data.

3.5.2 Hyperparameter optimization

In addition to finding the suitable observables (see Sec. 3.5.1) to the chosen ML algorithm to use as the input, as set of algorithm-specific parameters (called hyperparameters) need to be specified. The choice of the hyperparameters has significant impact on the performance of the model and thus they need to be chosen carefully. In contrast to the parameters that are autonomously learned during training, hyperparameters have to be specified prior to the training. The choice of hyperparameters is not trivial and is often done manually. This, however, requires expert knowledge of both the data and the ML method, takes a lot of human time, and makes the experiment not repeatable. Even if the hyperparameters are tuned using a given algorithm, the choice of a well-performing optimization algorithm is unclear.

The hyperparameters to be optimized in case of a neural network (see Sec. 3.2) could include for example the number of hidden layers, dropout rate, loss function and the number of epochs as shown in Fig. 3.6 among plethora of other parameters.

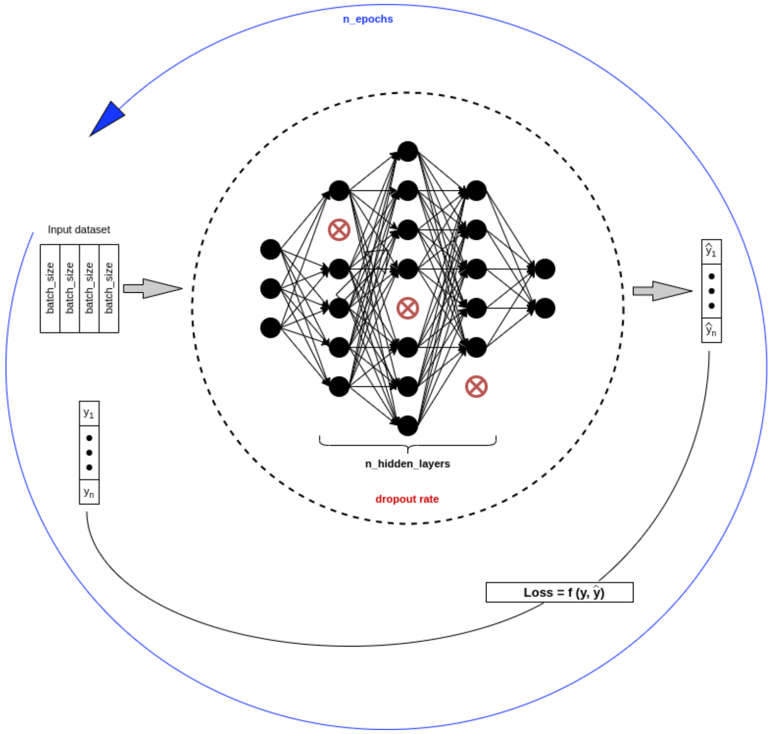


Figure 3.6: A set of hyperparameters that could be optimized could include for example the number of hidden layers (n_hidden_layers), the dropout rate, batch size, loss function and number of epochs (n_epochs).

During the hyperparameter optimization, machine learning model can be considered as a black box that takes the hyperparameters as input and gives a *fitness score*[9] corresponding this set as the output. This means that the hyperparameter optimization task can be expressed differently as a function minimization or maximization problem, where a point h in the hyperparameter space $\mathscr{H}$, corresponding to one possible solution to the optimization problem, is mapped to a *fitness* or a *score* value $s(h)$. Therefore this value quantifies the performance of the ML algorithm using the parameters h for a given task.

Casting non-floating type hyperparameters to a suitable encoding in the hyperparameter space $\mathscr{H}$ corresponds to the N-dimensional Eucledian space $\mathbb{R}^N$, where N denotes the number of hyperparameters. Formally, the optimal hyperparameters found in context of the optimization, denoted by the symbol $\hat{h}$, need to satisfy the condition:

$$\hat{h} = \arg\max_{h \in \mathscr{H}} s(h)\,, \tag{3.4}$$

where $s : \mathscr{H} \to \mathbb{R}$ is the objective function that maps the point h in hyperparameter space $\mathscr{H}$ to the score $s(h)$. This formulation allows one to compare the performance of the task of hyperparameter optimization using different methods.

[9] A score describing how good a model is when trained using the hyperparameters given as the input

The three distinct hyperparameter optimization algorithm studied in the context of this thesis were the particle swarm optimization [82], genetic algorithm [83] and Bayesian optimization [84–88]. Both the particle swarm optimization and Bayesian optimization are inspired by nature - the former mimicing the movement of a *swarm of particles* or a *flock of birds* and the latter aiming to imitate the evolution of genes. This is in contrast to the Bayesian optimization that makes use of the Bayesian statistics. A brief introduction to all these algorithms are given in the following sections and are described in more details in the two papers written during the thesis - [1] and [3], the former of which studies the performance of particle swarm optimization and genetic algorithm and compares these to two simpler models such as gradient descent (GD), random search (RS) and grid search (GS), while the latter compares the Bayesian optimizatio and the particle swarm optimization algorithms.

An alternative algorithm that was not studied in the context of this thesis is Asynchronous Successive Halving (ASHA) [89]. Still, this is a popular choice among the ML community in HEP for optimizing hyperparameters.

Particle swarm optimization

Particle swarm optimization (PSO) [82] is a computational method for the purposes of optimizing continuous nonlinear functions. As the name suggests, the function maximization is done using a *swarm of particles* that traverses the hyperparameter space $\mathscr{H}$. The position of each particle in the swarm corresponds to a set of hyperparameters h. PSO is inherently a highly parallel algorithm as the multitude of particles in the swarm explore multiple solutions simultaneously.

The evolution of the swarm is an iterative process - the position, x_i^k, and momentum, p_i^k, of every particle will be updated every iteration according to the equations Eq. 3.5 and Eq. 3.6 respectively

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + w \cdot \mathbf{p}_i^k + \mathbf{F}_i^k \tag{3.5}$$

$$\mathbf{p}_i^{k+1} = \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \tag{3.6}$$

The $\mathbf{F}_i^k$ in Eq. 3.5 represents an attractive *force* causing the particles to move in the direction of the previously discovered extremums[10] and is defined as:

$$\mathbf{F}_i^k = c_1 \cdot r_1 \cdot (\hat{\mathbf{x}}_i^k - \mathbf{x}_i^k) + c_2 \cdot r_2 \cdot (\hat{\hat{\mathbf{x}}}^k - \mathbf{x}_i^k) \ , \tag{3.7}$$

where the coefficients c_1 and c_2 are referred to as the *cognitive* and the *social* weights [90]. The symbols r_1 and r_2 represent random numbers, which are drawn from an uniform distribution in the interval [0,1]. $\hat{\hat{x}}^k$ and $\hat{x}^k$ in Eq. 3.7 denote the global and the personal best locations visited respectively.

Further details of the implementation of the PSO studied in the context of this thesis can be found in [1] and [3].

Genetic algorithm

Genetic algorithm (GA) is an evolutionary algorithm that draws inspiration from the natural selection. It features a set of possible solutions to the optimization problem that evolve during multiple generations in order to produce the optimal solution.

[10] The extremum of a subset of the swarm and the extremum a single particle has seen. The information of the extremum found by a subset of the swarm (with a size of N_{info}) is exchanged by the means of *espionage*.

A possible solution is referred to as a chromosome and represents one point in the hyperparameter space $\mathscr{H}$. As GA evolves multiple chromosomes simultaneously during each generation, it explores multiple possible solutions in parallel.

Each chromosome consists out of genes with the number of genes in a chromosome corresponding to the dimension of the hyperparameter space $\mathscr{H}$. A graphical depiction of a chromosome can be seen on Fig. 3.7.

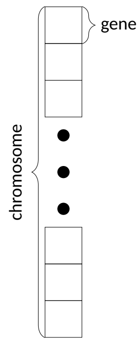


Figure 3.7: Multiple genes make up a chromosome. The number of genes corresponds to the dimension of the hyperparameter space $\mathscr{H}$. Figure taken from [1].

Similarly to PSO evolving towards the optimum is iterative and ends when a selected stopping criteria are reached. The stopping criteria could include the number of generations, not having improved the fitness in a given time or a acceptable performance of the model. Each iteration (i.e., generation) is a multi-step process. The three distinct stages in a given iteration are the *selection* of parents, the *crossover* of the genes, and the *mutation*.

The choice of the parent can be done using various strategies, such as the *tournament method* [91–93] or *roulette wheel selection* [94]. The genes of the two selected parents will spawn a offspring chromosome for the next generation by crossover [95, 96]. Again, multiple choices for the crossover strategy exist. These include for example *k-point crossover* and *uniform crossover*. Finally, the genes of the offspring will be mutated [83] in order to add some randomness or diversity to the population. This allowes the population to explore parts of the hyperparameter space $\mathscr{H}$ that weren't populated by the previous generation. Additionally, this increases the chances of not getting stuck in the local minima.

Further details of the exact implementation and strategy choices can be found in [1].

Bayesian optimization

Bayesian optimization (BO) is an optimization algorithm designed to perform numerical approximations of target functions (TFs) $s(h)$ that are time consuming to evaluate and for which the analytical form and derivatives are not necessarily known. This is made possible by not performing the maximization directly on the TF, but on an approximation of it called the surrogate function (SF), which is chosen to be fast to evaluate, with known derivatives and analytic form.

The numerical maximization of the TF is an iterative procedure. Each iteration can be divided into two main parts - choosing a point h in the hyperparameter space $\mathscr{H}$ on which to perform the time consuming TF evaluation, and updating the SF with the TF value at h, thus improving the accuracy of the approximation. This update to the SF after each TF evaluation is the origin of the name for the BO - SF is the *prior function* before choosing the next point to evaluate and the *posterior function* after evaluation.

A depiction of this process at iterations 7 and 8 is shown on the left and right side plot of Fig. 3.8 respectively. At iteration 7 acquisition function (AF) returns the parameter set (shown as a yellow circle on the bottom part of Fig. 3.8) to be evaluated on the TF at iteration 8. Shaded blue area denotes the uncertainty around the predicted mean shown in dotted black line. Solid blue line denotes the TF that the AF aims to approximate.

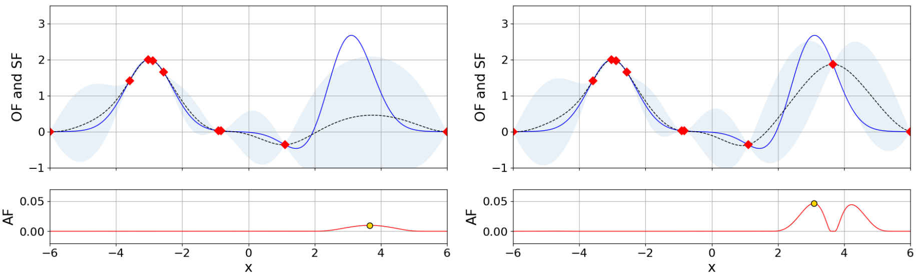


Figure 3.8: SF resembles TF more accurately after evaluating the TF at h and updating the SF with the returned value. Red markers on the upper part of each plot denote values already evaluated on the TF and yellow circle on the bottom part of the plot the location where to probe the TF in the next iteration. Shaded blue area denotes the uncertainty around the predicted mean shown in dotted black line. Solid blue line denotes the TF that the AF aims to approximate. Figure taken from [3].

The point to evaluate on the TF in the first part of each iteration is found using an AF. The most popular choices for the AF include metrics such as expected improvement [97] and probability of improvement [98]. All these methods inherently offer a trade-off between *exploration*[11] and *exploitation*[12].

Further details regarding the choice of parameters and design choices of the implementation details can be found in [3].

Choosing a suitable algorithm

The performance of all these algorithms was evaluated on two benchmark tasks - the Rosenbrock function [99] and the ATLAS Higgs boson machine learning challenge (HBC) [100], with the former being a well known benchmark task used for the purposes of evaluating the performance function minimization algorithms. The latter one serves as a more realistic case for hyperparameter optimization in the context of ML in HEP.

As can be seen from Fig. 3.9 then both evolutionary algorithms, GA and PSO, perform noticeably better than more simpler methods like GS, GD and RS.

[11]Giving more emphasis on exploring the full parameter space.

[12]Giving more emphasis on exploring near the minimum/maximum found so far during the optimization.

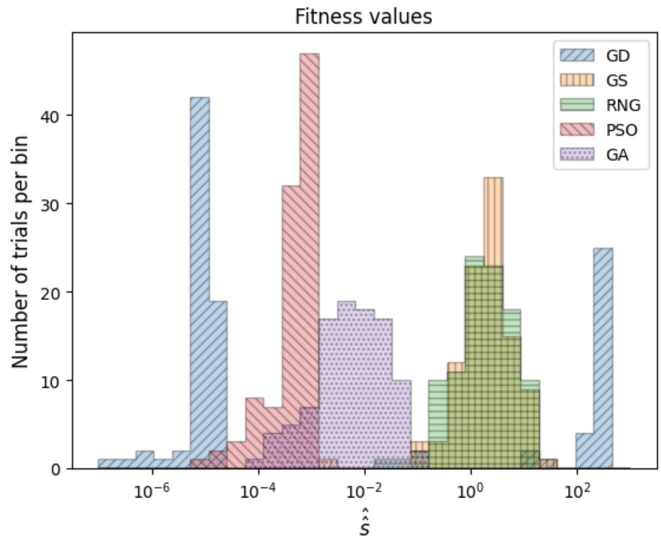


Figure 3.9: Gradient descent (GD), grid search (GS), random search (RNG), particle swarm optimization (PSO) and genetic algorithm (GA) were used to find the minimum of the Rosenbrock function 100 times. Both evolutionary algorithms, PSO and GA, perform steadily well.

BO on the other hand converges very quickly in the early stages, but is surpassed in performance of the PSO after 1000 evaluations[13] (Fig. 3.10). This is not surprising, as BO is aimed to minimize functions that are time consuming to evaluate, and is thus not expected to run for such a many iterations.

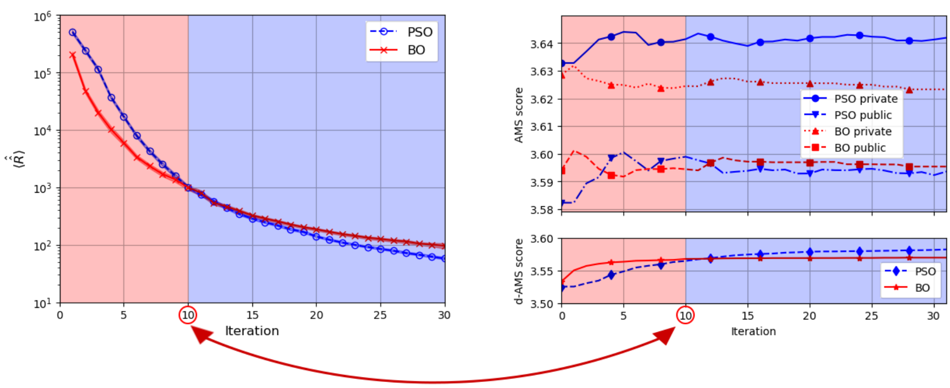


Figure 3.10: Bayesian optimization converges quickly in the beginning, but is surpassed in performance by particle swarm optimization after 10th iteration.

This means that based on the task to be optimized and computational resources available, different algorithm might be best suited for the task. BO would be a great choice for the purposes of tuning the hyperparameters of a ML model whose training takes days

[13] 100 solutions are evaluated in parallel for 10 iterations.

or even weeks, while if the model training takes less than an hour, PSO would give most probably better results.

3.5.3 Cross-validation

Despite a model performing well on a given subsample of the dataset, it is not enough to say it will also perform on an independent subsample of the same set. For these purposes the full dataset is divided usually into three distinct parts - *training*, *validation and test datasets*. Each of these datasets serves a different purpose and should not be allowed to mix with others also known as data snooping [101].

Training dataset is the one used exclusively for training, meaning fitting the parameters, of initial machine learning model. After each iteration the fitted model is evaluated on the validation dataset, which will provide an unbiased evaluation of the model. This evaluation information is often used to calculate losses.

When doing hyperparameter optimization, the training set is often the one that is split into yet another training (train-1) and validation dataset (val-2) as shown in Fig. 3.11. This means that the original dataset is split into four subsets: train-1, val-1, val-2 and test. This is done in order to avoid tuning the hyperparameters only for a specific set of data, which causes us to overestimate the performance of the model.

Figure 3.11: Splitting a dataset into training, validation and testing subsets. When doing hyperparameter optimization training dataset can be split into another smaller training and validation dataset in order to avoid biases.

Finally, the test dataset, also know as the *hold out set*, is used to evaluate the final performance of the model.

In order to get a more accurate reading of the model's performance, cross-validation can be done multiple times by the means of *k-fold cross-validation* [102]. This is done by dividing the full dataset into k subsets called *folds*. Then the model is trained, validated and tested on different folds of the data each time, ensuring there is no data snooping for the evaluations.

4 HH $\rightarrow$ Multilepton analysis

The study of the Higgs boson pair production has only recently become feasible due to a very small production rate. Even with the integrated luminosity of $\sim 140\,\mathrm{fb}^{-1}$ by both CMS and ATLAS detectors during the full Run-2, the amount of expected events relative to the background barely enough to probe the SM sensitivity. In order to overcome this obstacle, a multitude of HH decay modes are studied such that the sensitivity for the SM HH production could be maximized.

The branching fractions for some HH decay channels are shown in Fig. 4.1. The HH decay channels included in already published CMS analyses include $\mathrm{HH} \rightarrow \mathrm{bb}\tau\tau$ [103, 104], $\mathrm{HH} \rightarrow \mathrm{bbbb}$ [105, 106], $\mathrm{HH} \rightarrow \mathrm{bb}\gamma\gamma$ [107, 108] and $\mathrm{HH} \rightarrow \mathrm{bbZZ}$ [109]. The combination of these analyses [29] together with the $\mathrm{HH} \rightarrow \mathrm{multilepton}$ [2] analysis reaches already a 95% confidence level (CL) upper limit on the cross section at 3.4 times the SM expectation.

This thesis focuses on the $\mathrm{HH} \rightarrow \mathrm{multilepton}$ analysis. It covers a composite HH final state of Higgs decays into τ leptons or vector bosons (V), with the final states being $\mathrm{HH} \rightarrow \mathrm{VVVV}/\mathrm{VV}\tau\tau/\tau\tau\tau\tau$. Similar to the $\mathrm{ttH} \rightarrow \mathrm{multilepton}$ [110] analysis, also $\mathrm{HH} \rightarrow \mathrm{multilepton}$ analysis focuses on the decay channels featuring ≥ 2 leptons (e, μ) or hadronically decaying taus (τ_h) in the final states. In contrast to other HH analyses featuring at least one $\mathrm{H} \rightarrow \mathrm{bb}$ decay, the $\mathrm{HH} \rightarrow \mathrm{multilepton}$ analysis makes up for the low BR with a relatively clean and low background lepton signature. The reason for the low background is due to the high number of final state objects and/or a signature not common for the background processes for most of the analysis channels, thereby heavily reducing the amount of background indistinguishable from the signal (i.e., *prompt background*). Furthermore, the leptonic signature offers good sensitivity at low energies giving the analysis and edge in the low mass regime in resonant theory interpretations or low $|\kappa_\lambda|$ scenarios in non-resonant interpretations Still, having such low background makes the contributions from backgrounds stemming from misidentified leptons or jets (i.e., *fake background*) and leptons with misidentified charge (i.e., *charge-flip background*) more important.

The $\mathrm{HH} \rightarrow \mathrm{multilepton}$ analysis focuses on three distinct decaymodes, $\mathrm{HH} \rightarrow \mathrm{WWWW}$, $\mathrm{HH} \rightarrow \mathrm{WW}\tau\tau$ and $\mathrm{HH} \rightarrow \tau\tau\tau\tau$, with smaller contributions from decay modes including Z boson: $\mathrm{HH} \rightarrow \mathrm{ZZZZ}$ and $\mathrm{HH} \rightarrow \mathrm{ZZ}\tau\tau$. The importance from $\mathrm{HH} \rightarrow \mathrm{ZZZZ}$ and $\mathrm{HH} \rightarrow \mathrm{ZZ}\tau\tau$ are suppressed as the BR of $H \rightarrow ZZ$ relative to the $H \rightarrow WW$ is an order of magnitude smaller and the and Z decays into 2 leptons are happening only $\sim 3\%$ of the time. Furthermore, in this analysis most of the analysis channels employ also a Z-veto and $H \rightarrow ZZ$-veto, causing the contributions from $\mathrm{HH} \rightarrow \mathrm{ZZZZ}$ and $\mathrm{HH} \rightarrow \mathrm{ZZ}\tau\tau$ to be very small.

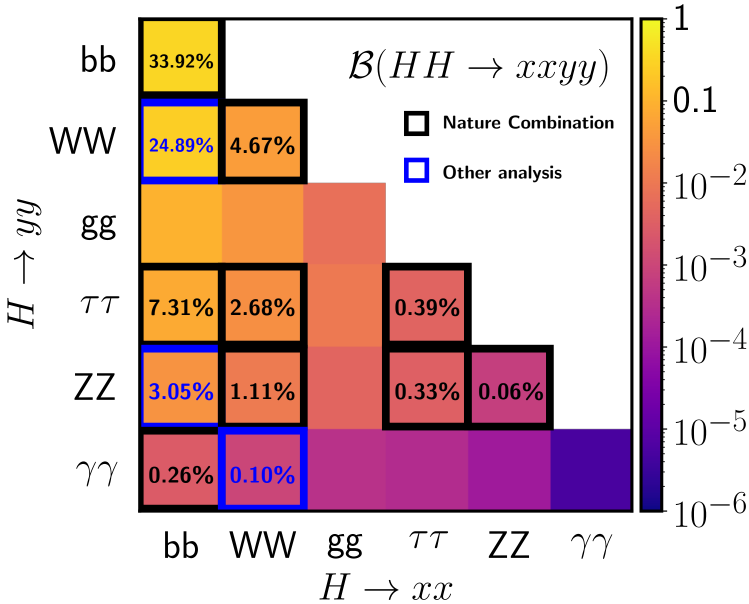


Figure 4.1: Branching fractions for the most notable di-Higgs decay channels, with the decaymodes included in the Nature combination being highlighted in black [22].

The $\mathrm{HH} \to \mathrm{multilepton}$ analysis was into 7 sub-categories, referred to as *analysis channels*, based on the multiplicity of leptons and hadronically decaying taus: $0\ell+4\tau_\mathrm{h}$, $1\ell+3\tau_\mathrm{h}$, 4ℓ, $2\ell(ss)+0/1\tau_\mathrm{h}$ [1], $2\ell+2\tau_\mathrm{h}$, $3\ell+0\tau_\mathrm{h}$, $3\ell+1\tau_\mathrm{h}$. In the context of this thesis the $0\ell+4\tau_\mathrm{h}$ and $1\ell+3\tau_\mathrm{h}$ channels were studied in more details and will therefore be covered in more depth. These seven channels cover the majority of the leptonic final states for the decay modes $\mathrm{HH} \to 4\mathrm{V}$, $\mathrm{HH} \to 2\mathrm{V}2\tau$ and $\mathrm{HH} \to 4\tau$.

Three distinct physics scenarios are studied in the $\mathrm{HH} \to \mathrm{multilepton}$ analysis: resonant spin-0, resonant spin-2 and non-resonant Higgs production. As described in Sec. 1.2.2, both resonant physics scenarios postulate the existence of a new heavier particle that decays into two Higgs bosons. As the kinematics are different for the hypothesized radion like spin-0 and for a graviton like spin-2 particle then the analysis is optimized for these two cases separately. Furthermore, the masses for these hypothezised particles are unknown, and are therefore scanned in the range 250 GeV to 1 TeV.

The non-resonant physics scenario described in Sec. 1.2.2 is aimed at studying the effects of non-resonant BSM phenomena as well as the SM Higgs boson pair production. As the number of theoretical BSM models is large, then the collection of these theoretical models together with SM-like scenarios with modified SM couplings are tested by the means of EFT. The idea behind EFT is to reduce the new physics to point-like interactions that are described by the effective coupling that modify the behavior of the Higgs boson, allowing the reduction of BSM scenarios to effective scenarios. Models with similar signatures are grouped together, making it possible to exclude (or find hints of) a generalized signature, which allows one to evaluate the legitimacy of variety of models simultaneously. From hereon each such grouping is referred to as benchmark (BM).

In order to ensure an optimal performance of the analysis, dedicated BDTs were trained for each of the physics scenarios in all 7 analysis channels such that the BDTs are parametrized

[1] Two same-signed (ss) leptons with zero or one hadronic taus

according to the mass or the BM for resonant and non-resonant cases respectively. As a part of this thesis a major contribution was made to the development and implementation of the ML routines, including for example data preparation, implementation of the optimization algorithms and automation. An overview of the ML choices, techniques and implementations for this analysis is given in Sec. 4.3.

For the analysis the data recorded by the CMS experiment during the full LHC Run2[2] was used amounting to a luminosity of 137.6 fb^{-1} at $\sqrt{s} = 13$ TeV for this period. For the purposes of training BDTs as well as for the signal and background modeling and overall optimization of the analysis events from Monte Carlo (MC) simulation were used. Further details concerning the simulated samples can be found in Ref. [2].

The chapter is structured as follows: in Sec. 4.1 the event selection criteria for the $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$ channels; Sec. 4.2 describes the datasets used in the analysis; Sec. 4.3 gives an overview of the ML methods used. The results of the analysis are described in the Sec. 4.4.

4.1 Event selection

Each of the 7 channels in the $\mathrm{HH} \to \mathrm{multilepton}$ analysis has a different set of selection criteria for an event to be considered in the given channel. The channels are defined by the multiplicity of Tight ID leptons (ℓ) and hadronically decaying tau leptons (τ_h), with all reconstructed objects being passed through filtering algorithms in order to reduce the number unwanted events that are influenced by effects such as detector noise or miscalibration. The working points for these objects are described in Tables. 2.1, 2.2 and 2.3 respectively. Analysis channels featuring a small object multiplicity and thus a high expected background rate while having a relatively small signal contribution are not used. With this choice of the 7 channels most of the leptonic W decay modes are covered.

In order to ensure orthogonality with other di-Higgs analyses all channels featured in the $\mathrm{HH} \to \mathrm{multilepton}$ analysis reject events that contain one or more b-tagged[3] jets. Additionally, vetoing such events helps to reject background events that feature top quark decays.

The two channels studied in more details in this thesis are the $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$. A more detailed description of the event selection for these two channels is given in the following sections, while the details regarding the event selection of the remaining 5 channels and two control regions used in the signal extraction are given in Ref. [2].

4.1.1 $0\ell + 4\tau_\mathrm{h}$

The $0\ell + 4\tau_\mathrm{h}$ is aimed at selecting HH signal events in the $\tau\tau\tau\tau$ decay mode, where all of the four tau leptons decay hadronically. Such events are obtained by requiring zero electrons and zero muons that pass the tight object selection to be present. As shown in Table. 4.1 then for an event to be considered in the $0\ell + 4\tau_\mathrm{h}$ channel, all four taus are required to pass the Tight identification criteria and have a p_T above 40/40/20/20 GeV for leading/sub-leading/third/fourth τ_h respectively. Additionally, a total charge of 0 is required.

[2] 2016-2018

[3] b-tagging is done by the means of the DeepJet [47] algorithm. If the b-tagging score exceeds a given threshold (at a given working point) a jet is referred to as a b-tagged jet.

Table 4.1: Event selection criteria for the channels $0\ell+4\tau_\mathrm{h}$ and $1\ell+3\tau_\mathrm{h}$.

Criterion	$0\ell+4\tau_\mathrm{h}$	$1\ell+3\tau_\mathrm{h}$
Tight ID leptons	0	1
Lepton p_T	–	$>20(15)$ GeV for an electron (muon)
Tight ID τ_h	≥ 4	3
τ_h p_T	$\geq 40/40/20/20$ GeV	$>40/30/20$ GeV
τ_h $\lvert\eta\rvert$ veto + DeepTau vs. Ele WP-VLoose if $m_Z - 20$ GeV $< m^{OS}_{l\tau_h} < m_Z + 10$ GeV	–	$\lvert\eta\rvert \leq 1.460 \vee \lvert\eta\rvert \geq 1.558$
Low mass resonance veto	–	–
Z mass veto	–	–
b-jet veto	✓	✓
Charge sum	$\sum Q(\tau_h) = 0$	$\sum Q(\tau_h) + Q(\ell) = 0$

All the above listed criteria were used in order to maximize the expected sensitivity of this channel.

Applying this event selection for the $0\ell+4\tau_\mathrm{h}$ channel results in the background composition shown in Fig. 4.2 showing that the background stemming from fake or misidentified τ_h makes up two thirds of the expected background, while background from ZZ and single Higgs processes contributed 21.2 % and 12.1 % respectively. Almost all HH signal events originate from the $H \to \tau\tau\tau\tau$ decay mode.

As fakes from jets make up the largest portion of the backgrounds, $0\ell+4\tau_\mathrm{h}$ would benefit the most from a better reconstruction and identification rather than from better discriminating variables.

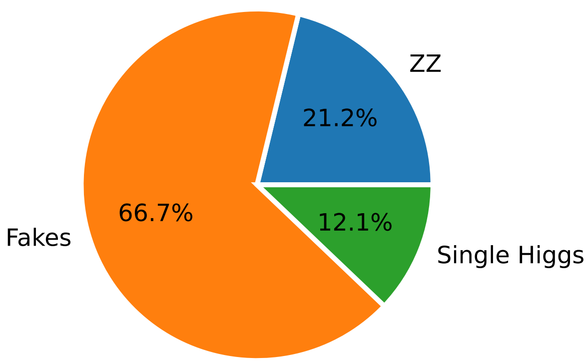


Figure 4.2: Background composition of the $0\ell+4\tau_\mathrm{h}$. As fakes make up two thirds of the background, better particle reconstruction and identification would benefit this channel greatly.

4.1.2 $1\ell + 3\tau_\mathrm{h}$

Similarly to $0\ell + 4\tau_\mathrm{h}$, the $1\ell + 3\tau_\mathrm{h}$ channel is aimed at selecting HH decays in the $\tau\tau\tau\tau$ decay mode with the distinction that one of the taus decays leptonically.

In order for an event to be considered in the $1\ell + 3\tau_\mathrm{h}$ channel it must have exactly one tight lepton within $|\eta| < 2.1$ with $p_T > 20$ in case of an electron or $p_T > 15$ in case of a muon. The three tau candidates required in this channel need to pass the tight criteria while passing $p_T > 40/30/20$ GeV for the leading/sub-leading/third tau candidate. The total charge sum of the three tau candidates is required to be ± 1, while the charge sum of the taus candidates plus lepton is 0.

Background rising from the events where a electron is identified as a τ_h, (fake τ_h) is not modeled using the data driven background estimation nor is properly described in the simulations. This is a source for mismatch between signal and background prediction at low τ_h energies for the channels $1\ell + 3\tau_\mathrm{h}$ and $3\ell + 1\tau_\mathrm{h}$. For this additional requirements on the τ_h are placed in both these channels. For $3\ell + 1\tau_\mathrm{h}$ crack-veto[4] is applied together with tightening the DeepTau working point for discriminating against electrons.

For the $1\ell + 3\tau_\mathrm{h}$ channel the main *fake* background arises from the $Z \to e^+e^-$ decays. However if one employ the crack veto and would tighten the DeepTau vs. electron WP cut, the signal efficiency would be noticeably reduced. This is avoided by applying these cuts only is the invariant mass of the oppositely charged τ_h and lepton pair is near the Z-mass peak: $m_Z - 20\mathrm{GeV} < m^{OS}_{\ell\tau_H} < m_Z + 10\mathrm{GeV}$.

Using the event selection summarized in Table. 4.1 results in a background composition shown in Fig. 4.3. Again, a the biggest contributor to the background is the fake background at 44.0%. The background rising from the ZZ is in the same ballpark at 38.0% followed by contributions from single Higgs and smaller backgrounds at 16.0% and 2.0% respectively. Roughly 80% of the HH signal events originate from the $H \to \tau\tau\tau\tau$ decay mode, and 20% from the $H \to WW^*\tau\tau$.

Similarly to the $0\ell + 4\tau_\mathrm{h}$, also $1\ell + 3\tau_\mathrm{h}$ would benefit from a better particle reconstruction and identification. However, as a big fraction of the expected background originates from ZZ, then a smart selection of variables for signal-background separation would help.

4.2 Datasets

The analyzed dataset consists of 137.6 fb^{-1} of pp collisions, with an average of 33 pp collisions taking place at each bunch crossing. In order to select the targeted final states for this analysis, a set of lepton triggers, tau triggers and lepton-tau cross-triggers was used.

For the purposes of analysis optimization, signal-background modeling and the training of the BDTs, MC event simulation was used. The MC event simulation comprises of the event generation, pile-up generation and the full detector simulation, with the latter using GEometry ANd Tracking (Geant4) [111] for this purpose. In order to further increase the agreement between the measured data and the MC simulation, a set of additional MC corrections are applied in the analysis.

Data for all of the three signal scenarios is produced using a looser event selection than the one used in the signal region (SR) of a given channel for the purposes of increasing selection efficiency while accepting a bigger fakerate. This is done by relaxing the DeepTau discriminator WP against jets (see Table. 2.3) to VVVLoose for all the channels that have

[4] Dead region i.e., *crack* is one of the reasons for an electron misidentified as a τ_h. If the τ_h is compatible with the main crack in the ECAL is at $1.460 < |\eta| < 1.558$, this τ_h is vetoed.

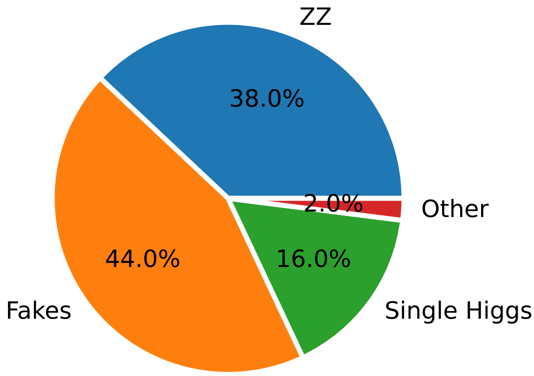


Figure 4.3: Background composition of the $1\ell+3\tau_h$ channel using the selection criteria from Table. 4.1.

τ_h in the final state and the lepton ID to `Loose` lepton definition (see Tabels 2.1 and 2.2) for all the channels that had a lepton in the final state.

4.2.1 Background samples

The background samples used in the analysis include single boson production (W, Z), diboson production ($\mathtt{WW}$, $\mathtt{WZ}$, $\mathtt{ZZ}$), triple-boson production ($\mathtt{WWW}$, $\mathtt{WWZ}$, $\mathtt{WZZ}$, $\mathtt{ZZZ}$, $\mathtt{WZ}\gamma$), processes with one or two top quarks ($\mathtt{t}$, $\mathtt{t\bar{t}}$, $\mathtt{t\bar{t}W}$, $\mathtt{t\bar{t}Z}$, $\mathtt{t\bar{t}WW}$) and single Higgs boson production ($\mathtt{ggH}$, $\mathtt{qqH}$, $\mathtt{tHq}$, $\mathtt{tHW}$, $\mathtt{t\bar{t}H}$, $\mathtt{ggH}$, $\mathtt{WH}$, $\mathtt{ZH}$), which are all modeled in simulation. Additionally, some backgrounds can arise from processes where lepton pairs emerge from photon conversions in the detector material. This set of backgrounds is modeled using a combination of $\mathtt{X}+\mathtt{jets}$ and $\mathtt{X}+\gamma$ for single boson and top production. More details regarding the exact MC generators and tunes used in this analysis can be found in Ref. [2].

4.2.2 Non-resonant HH production

Non-resonant HH production samples were generated both at leading-order (LO) and NLO accuracy in QCD. These samples cover the ggHH and qqHH production processes, where the Higgs bosons decay to either $\mathtt{WW}^*$, $\mathtt{ZZ}^*$ or $\tau\tau$.

NLO samples are used for extracting the HH signal from the data, while the LO samples are used to train the ML classifier because of the bigger sample size.

A total of 12 ggHH samples corresponding to the 12 BM scenarios in the HEFT approach and one corresponding to the SM values were simulated for training the BDT classifier. We label these scenarios as JHEP04 BM1-12 and SM, which is treated as BM0 here [35]. The benchmark JHEP04 BM8 is complemented by a modified version of it in Ref. [112] referred to as JHEP04 BM8a. Additionally, a total of 7 benchmark scenarios (JHEP03 BM1-7) from Ref. [38] were simulated only for the purposes of signal extraction.

As mentioned in Sec. 1.2.2, all these aforementioned BM scenarios represent different combinations of κ_λ, κ_t, c_g, c_{2g} and c_2 HEFT parameter values. The parameter values for all these combinations can be found in Table 3 of Ref. [2]. The choice parameter values in these combinations are chosen such that the 5-dimensional parameter space would be evenly populated.

For the purposes of increasing the number of simulated events for a given kinematic configuration or modelig configurations not explicitly generated (JHEP03 BM1-7), the ggHH samples are merged and the samples in this superset are reweighed according to the procedure introduced in Ref. [113] such that the $m_H H$ and $|\cos\theta^*|$[5] match the ones computed at NLO accuracy given in Ref. [112].

4.2.3 Resonant HH production

The resonant HH production was simulated at LO for both the spin-0 (radion) and spin-2 (graviton) cases. A total of 18 different generator m_X points were used, where m_X has values of 250, 260, 270, 280, 300, 350, 400, 350, 500, 550, 600, 650, 700, 750, 800, 850, 900 and 1000 GeV.

4.2.4 Triggers

For the purposes of choosing the events for a given analysis channel a different set of triggers was used. The dependence on the lepton multiplicity, p_T threshold and other quality criteria for each trigger used in this analysis is given in Table. 4.2. The triggers used for each analysis channel is given in Table. 4.3.

Table 4.2: The requirements on the electrons (e), muons (μ), and hadronically decaying tau leptons (τ_h) for the given triggers used in the $\mathrm{HH} \to \mathrm{multilepton}$ analysis. A range of values is given for a trigger if the value changed in time [2].

Trigger	Selection requirements for objects
Single e	$p_T(\mathrm{e}) > 27-35$ GeV
Single μ	$p_T(\mu) > 27-35$ GeV
Double e	$p_T(\mathrm{e}) > 23, 12$ GeV
e $+\mu$	$p_T(\mathrm{e}) > 23$ GeV, $p_T(\mu) > 8$ GeV
$\mu+$e	$p_T(\mu) > 23$ GeV, $p_T(\mathrm{e}) > 8-12$ GeV
Double μ	$p_T(\mathrm{e}) > 17, 8$ GeV
e $+\tau_\mathrm{h}$	$p_T(\mathrm{e}) > 24$ GeV, $p_T(\tau_\mathrm{h}) > 20-30$ GeV, $\lvert\eta(\mathrm{e},\tau_\mathrm{h})\rvert < 2.1$
$\mu+\tau_\mathrm{h}$	$p_T(\mu) > 19-20$ GeV, $p_T(\tau_\mathrm{h}) > 20-27$ GeV, $\lvert\eta(\mu,\tau_\mathrm{h})\rvert < 2.1$
Triple e	$p_T(\mathrm{e}) > 16, 12, 8$ GeV
Two e $+\mu$	$p_T(\mathrm{e}) > 12, 12$ GeV, $p_T(\mu) > 8$ GeV
Two $\mu+$e	$p_T(\mu) > 9, 9$ GeV, $p_T(\mathrm{e}) > 9$ GeV
Triple μ	$p_T(\mu) > 12, 10, 5$ GeV

[5] cosine of the polar angle of one Higgs boson with respect to the beam axis in the HH rest frame

Table 4.3: Triggers used for different analysis channels [22].

Trigger	$0\ell+4\tau_{\mathrm{h}}$	$1\ell+3\tau_{\mathrm{h}}$	$2\ell(ss)+\leq 1\tau_h$	$2\ell+2\tau_h$	$3\ell+0\tau_h$	$3\ell+1\tau_h$	4ℓ
Double τ_h trigger	✓	✓	✗	✗	✗	✗	✗
Lepton + τ_h cross-trigger	✗	✓	✗	✗	✗	✗	✗
Single lepton trigger	✗	✓	✓	✓	✓	✓	✓
Double lepton trigger	✗	✗	✓	✓	✓	✓	✓
Triple lepton trigger	✗	✗	✗	✗	✓	✓	✓

4.3 Machine learning

Once the events have been selected for each of the channels, the signal and background for each physics scenario (spin-0 resonant HH production, spin-2 resonant HH production, non-resonant HH production) is separated using parametrized BDTs.

For each scenario there are two dedicated BDTs trained - one where events with odd event number are used for training and events with even eventNumber are used for testing and the other vice versa. This means that a total of 3 (physics scenarios) $\times$ 2 (data halves) $\times$ 7 (analysis channels) = 42 BDT classifiers were trained in this analysis.

The data used for the training and the data preparation procedures used were described in the Sec. 4.3.1. In Sec. 4.3.2 the choice of the model is motivated and the optimization of the model input parameters and variables is described.

4.3.1 Data preparation

Prior to ML training, the following data preparation was done for all of the three HH signal scenario described above.

In order to ensure that the background statistics is associated with all of the different signal scenarios, a modified version of the vanilla oversampling (see Sec.3.4.1) was used. With this modification, all backgrounds are duplicated for each signal sample, which in contrast to having background samples randomly assigned to the signal, results in a uniform background distribution associated with all the signal samples. The procedure is different for the resonant and non-resonant scenarios: in case of the resonant signal scenarios additional feature (column) is added to the samples that corresponds to the mass point m_X of the associated signal sample. The features used for each event in the $0\ell+4\tau_{\mathrm{h}}$ and $1\ell+3\tau_{\mathrm{h}}$ channels are described in Sec. 4.3.3. For non-resonant signal samples, the procedure entails one-hot-encoding the BM scenario: creating an additional 13-dimensional vector, where all values corresponding to the different scenarios are set to 0 with the exception of the active scenario, which is set to one. Again the background events with the additional 13-dimensional feature vectors are duplicated for each signal scenario.

For both of the resonant HH production scenarios, spin-0 and spin-2, the input training variables are decorrelated from gen_{mHH}[6]. This is done by fitting the mean of each feature vs the gen_{mHH} using a polynomial that is often of a high order. The feature values are subsequently divided by the value of this function for a given mass for both signal and background samples. The effect of this procedure is shifting the signal distributions such that they overlay each other, while having no such effect on the background samples. This

[6] the generator (simulated) HH mass

procedure is aimed at aiding the ML training in the low gen_{mHH} cases. A good illustration of the results of this procedure can be found on Fig. 6.11 in Ref. [22].

All background events are being reweighted according to their expected yields in the signal region, thereby ensuring that their relative contributions remain the same. As a final data preparation step the sum of weights for both signal and background is normalized to 10^5 events, thereby giving the same importance to both of the classes in the training.

4.3.2 Model selection

For the task of classifying events into signal and background a parametrized BDT, namely XGBoost [65], was used. A parametrized BDT was used in order to treat several mass hypotheses for a wide range of mass values for the resonant case or a multiple sets of benchmark points for the non-resonant HH production scenario. This was needed due to the fact that a model trained on a low mass point has worse performance on high mass signal and vice versa. Although one could train a separate BDT for each mass point or benchmark scenario, the expected behavior is similar to training a single parametrized network that learns several scenarios at once. The latter case is slightly more preferable, as in addition to being simpler version, it uses the information from all the scenarios thereby learning the more general featured shared between the scenarios. A BDT over NN was preferred due to having only low statistics in the $\mathrm{HH} \to \mathrm{multilepton}$ analysis and thus allowing to use a more simple model without sacrificing performance.

The set of input features from the available data that is chosen prior to training will be described in Sec. 4.3.4, while the optimized set of hyperparameters used in the analyses will be described in Sec. 4.3.5.

During training the area under curve (AUC) of the receiver operating characteristic (ROC) curve was used as the objective function.

In order to evaluate whether the model will perform on the unseen "test" data, the input data used for training was split into two halves - "odd" and "even". The split was done based on a unique identifier called *event number* such that the model that was trained on the events with odd event numbers is validated on the events with even event numbers and vice versa.

4.3.3 Feature engineering

In order to increase the separation power of the BDTs, a custom set of features were created separately for each of the 7 analysis channels.

A description of the features used in the $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$ channels is given in the following sections. The overview of the features used in the other 5 analysis channels can be found in Ref. [52] and Ref. [22].

The strategy for the choice of variables used for the final training of the model will be given in Sec. 4.3.4.

$0\ell + 4\tau_\mathrm{h}$

As was shown in Fig. 4.2, the ZZ background constitutes the second largest contributor to the overall background, amounting to more than 20%. Therefore, a set of discriminating variables targeting Z-decays was calculated for each event. When constructing these variables, oppositely charged taus are paired.

For the first set of features the pair that has the invariant mass closest to the Z-mass peak is assigned to variables `Zee_bestTauHPair_*`, where the star (*) indicates a quantity such as mass m, difference of azimuthal angle ($\Delta\phi$), rapidity-azimuthal ($\eta - \phi$) distance

(ΔR), transverse momentum (p_T) or the difference of pseudorapidities ($\Delta\eta$).

Next, pair that has the smallest ΔR is associated to variables `dr_bestTauHPair_*`, while the pair with the largest p_T is associated to variables `pt_bestTauHPair_*`. Again the star (*) corresponds to the same set of quantities as before. For each of the variable `*_bestTauHPair_*` described above another set of features is created that corresponds to the respective quantities corresponding to the tau pair that was formed from the left over oppositely charged taus and is referred to as `*_secondTauHPair_*`.

Futhermore, for each event additional variables such as `HT`[7], `STMET`[8], `mht`[9], `met_LD`[10], `met`[11], `mTauTau`[12], `pt_HH_recoil`[13], `most_Zee_like`[14], `m_tau*_tau*`[15], `max_pt_pair_pt`[16], `min_dr_pair_dr`[17], `diHiggsVisMass`[18] together with the quantities such as p_T and η for each tau were calculated.

As mentioned in the Sec. 4 $0\ell+4\tau_{\mathrm{h}}$ is targeting the $\mathrm{HH}\to\tau\tau\tau\tau$ decay modes, then it only makes sense to use features that are aimed at reconstructing the $\mathrm{H}\to\tau\tau$ decays accurately. For these purposes the SVFIT [114] algorithm was used, that reconstructs the invariant mass of the di-Higgs system referred to as `diHiggsMass`.

$1\ell+3\tau_{\mathrm{h}}$

For the $1\ell+3\tau_{\mathrm{h}}$ channel the ZZ background plays a major role, however no special effort to introduce variables to discriminate between ZZ and signal was made.

Similarly to the $0\ell+4\tau_{\mathrm{h}}$ the p_T and η values for all leptons and τ_h were included, with the exception to use `conePt` instead of p_T in the case of the lepton.

Again features such as `met`, `mht`, `met_LD`, `HT`, `STMET`, `diHiggsVisMass`, `diHiggsMass` and `pt_HH_recoil` were made use of together with the quantities such as ΔR, mass and transverse momentum for different pairings of lepton and taus.

Features specific for this channel include the minimum and maximum $\Delta\phi$ of the possible oppositely charged lepton-tau and tau-tau pairs[19] and the minimal and maximal $\Delta\phi$ between particles in a pair[20].

---

[7] Scalar sum of p_T of all fakeable leptons, fakeable taus and selected jets in the event.

[8] A "transverse mass" like variable constructed from fakeable leptons, jets and MET. STMET = HT + MET.pt()

[9] Vector sum of pT of all fakeable leptons, selected jets and fakeable had-taus in the event. Has better stability against pileup but poor resolution.

[10] a specific linear combination between MET and MHT designed as a compromise between robustness against pileup as well as good enough resolution. met_LD = 0.6 MET.pt() + 0.4 MHT.pt() (coefficients obtained by fitting 2D distribution between MET and MHT)

[11] missing transverse energy in the event

[12] the mass of the tau pair if it is a valid solution for the SVFIT algorithm otherwise set to a default value of -1.

[13] the transverse momenta of the four taus and the MET system

[14] The mass difference of the Z mass and the mass of the tau pair that had mass closest to Z peak.

[15] the invariant mass of tau pair constructed from different combinations of taus

[16] The p_T value of tau pair that has the biggest p_T

[17] The ΔR value of the tau pair that has the smallest ΔR

[18] the mass of the *visible* part of the di-Higgs system, meaning the invariant mass of the four taus, without considering the contribution from MET.

[19] `dphi_[lep/tau]_tau_OS_pair_[max/min]`

[20] `dphi_HHvis_[min/max]`

4.3.4 Training variable selection

Before tuning the hyperparameters a subset of input training features need to be selected among plethora of available ones. Having fewer input variables without offering performance makes the trained model more interpretable and simple, which as per Occam's razor is highly preferred. As an additional bonus, having fewer variables should make the learning algorithm faster and reduce overtraining.

To choose a well performing subset of training variables an algorithm consisting out of two main steps was implemented. In the first step training variable from a pair of variables that are correlated above a given threshold ζ_{TH} is removed from the list of training variables. In our analysis we have chosen the threshold ζ_{TH} to be 0.8, meaning that if a pair of variables has a correlation that is above 80%, the variable with smaller feature importance will be discarded. The chosen BDT algorithm, XGBoost, as most other modern machine learning methods, handles multicollinearity well and the performance itself is not affected when having highly correlated input variables. Still, in case of a high degree of correlation between two variables, the feature importances will be affected. As is the case in general in boosting, also XGBoost ignores one of the variables in the fully correlated pair.

As one might expect the noise-dominated features will tend to be less correlated with other features than those that are correlated more with the target. Consequently the proportion of noise-dominated variables will increase after the first step.

After the removal of the highly correlated variables, the algorithm proceeds to the second step, where the less important training variables are iteratively dropped. At every step a new BDT is trained with the reduced list of input variables and the feature importance of the reduced list of input variables calculated. In our implementation we use the so called "weight" feature importance from the XGBoost that corresponds to the number of times a given variable was used to split the data across all trees. The "weight" metric is in principle just a simpler way to measure gain, which is a measure of how much improvement in accuracy is due to a given variable. At each step a given number ξ_{drop} of least important features will be dropped until a suitable number $N_{trainvars}$ of training variables is reached.

For both `0l_4tau` and `1l_3tau` ξ_{drop} was chosen to be 5 and $N_{trainvars}$ to be 9 in case of resonant spin-0 and spin-2 and 15 in case of nonresonant-HH production.

$0\ell + 4\tau_\mathrm{h}$

For the $0\ell + 4\tau_\mathrm{h}$ channel a different choice of discriminating variables to be used in the ML training were used for each physics scenario. The features used for a given physics scenario are given in Table. 4.4 and are denoted by ✓.

The distributions for each variable used for the final BDT training for a given background and different signal scenarios are shown in Figs. 4.4 - 4.22.

$1\ell + 3\tau_\mathrm{h}$

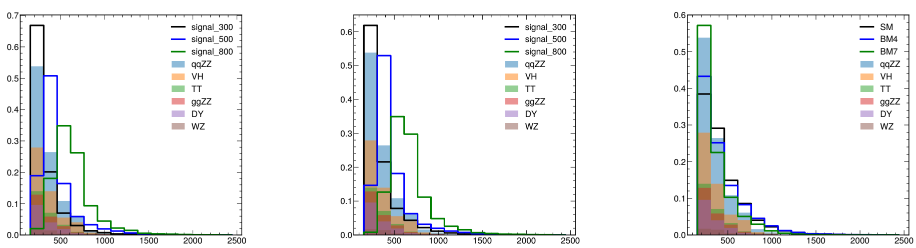
In case of the $1\ell + 3\tau_\mathrm{h}$ channel a the same set of discriminating variables for the BDT training was used for both of the resonant physics scenarios, as the performance change was negligible. This was not the case for the nonresonant scenario, where a different choice of the features to be used was made.

The features used for a given physics scenario are given in Table. 4.5 and are denoted by ✓.

The distributions for each variable used for the final BDT training for a given background and different signal scenarios are shown in Figs. 4.23 - 4.39.

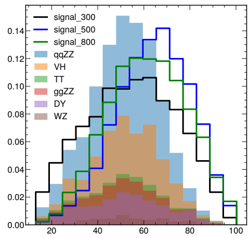
Table 4.4: *Discriminating variables used for the three physics scenarios in the $0\ell + 4\tau_h$ channel.*

	spin-0	**spin-2**	**non-resonant**
STMET	✗	✗	✓
Zee_bestTauHPair_m	✗	✗	✓
Zee_secondTauHPair_m	✗	✗	✓
deltaEta_tau2_tau3	✓	✗	✗
diHiggsMass	✓	✓	✓
diHiggsVisMass	✓	✗	✗
dr_bestTauHPair_dr	✗	✓	✓
dr_bestTauHPair_m	✓	✓	✓
dr_secondTauHPair_dr	✗	✗	✓
dr_secondTauHPair_m	✓	✗	✗
dr_tau1_tau3	✗	✓	✗
dr_tau2_tau4	✓	✓	✗
mTauTau	✗	✗	✓
m_tau2_tau3	✗	✗	✓
met_LD	✓	✗	✓
pt_bestTauHPair_m	✗	✓	✗
tau1_eta	✗	✗	✓
tau1_pt	✗	✓	✓
tau4_pt	✗	✗	✓

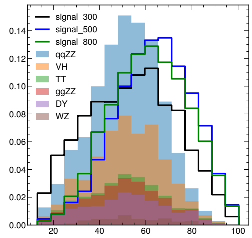


(a) spin-0 *(b) spin-2* *(c) nonres*

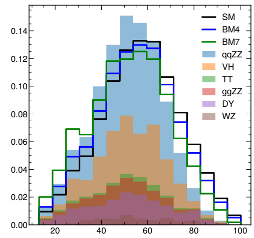
Figure 4.4: The distribution of `STMET` *for the three physics scenarios.*



(a) spin-0

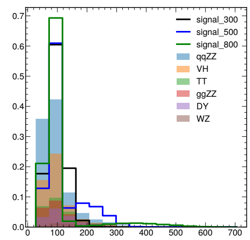


(b) spin-2

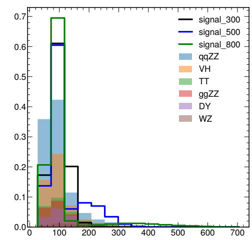


(c) nonres

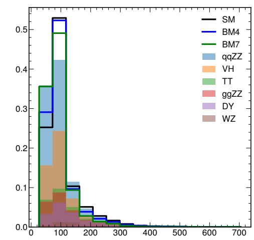
Figure 4.5: The distribution of `Zee_bestTauHPair_m` *for the three physics scenarios.*



(a) spin-0

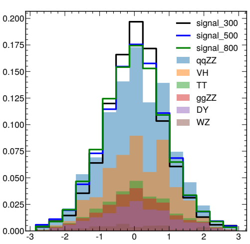


(b) spin-2

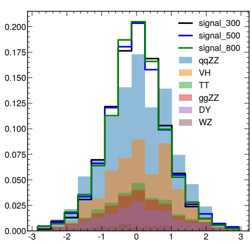


(c) nonres

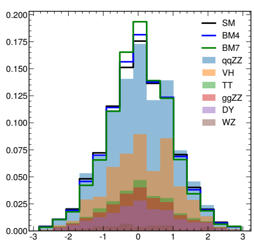
Figure 4.6: The distribution of `Zee_secondTauHPair_m` *for the three physics scenarios.*



(a) spin-0

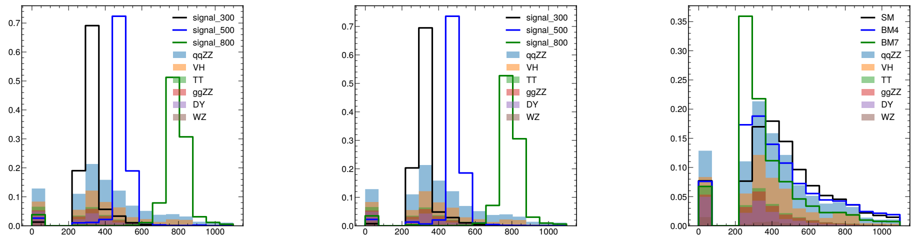


(b) spin-2



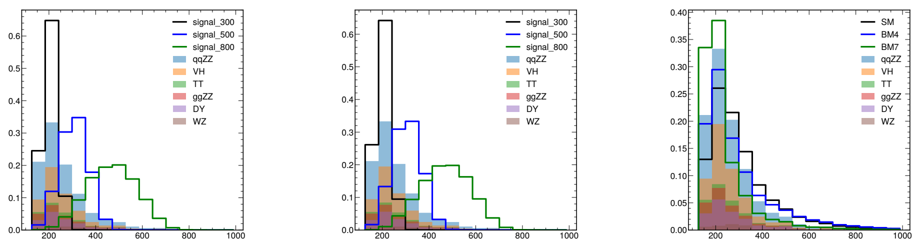
(c) nonres

Figure 4.7: The distribution of `deltaEta_tau2_tau3` *for the three physics scenarios.*



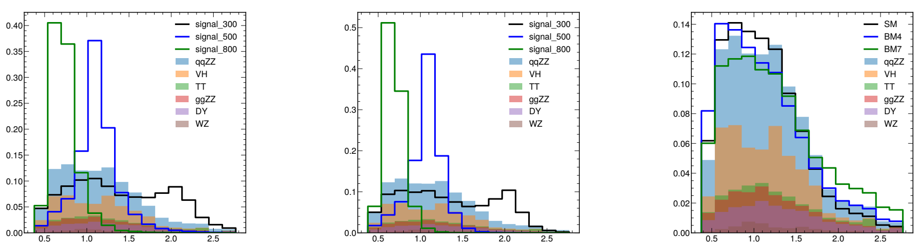
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.8: The distribution of `diHiggsMass` *for the three physics scenarios.*



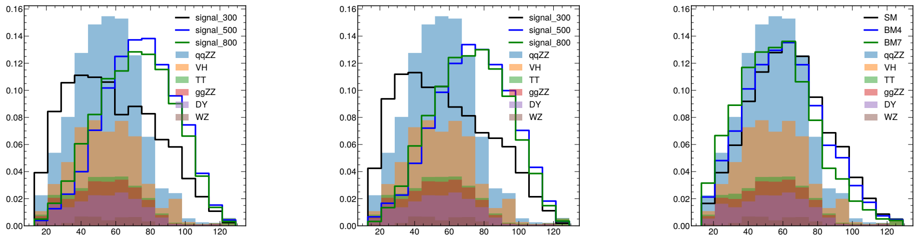
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.9: The distribution of `diHiggsVisMass` *for the three physics scenarios.*



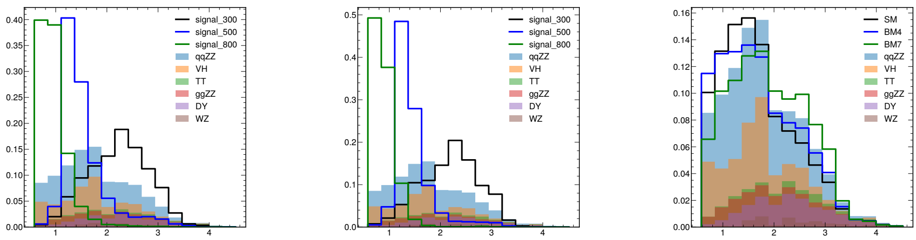
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.10: The distribution of `dr_bestTauHPair_dr` *for the three physics scenarios.*



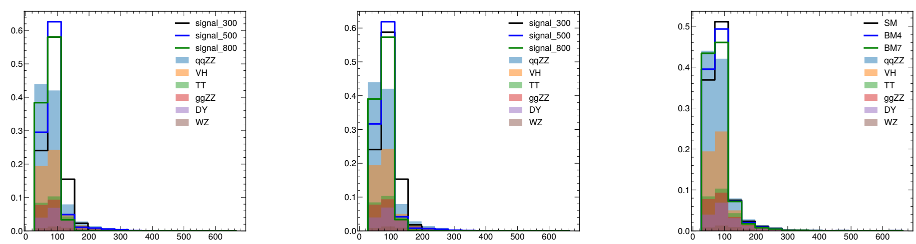
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.11: The distribution of `dr_bestTauHPair_m` *for the three physics scenarios.*



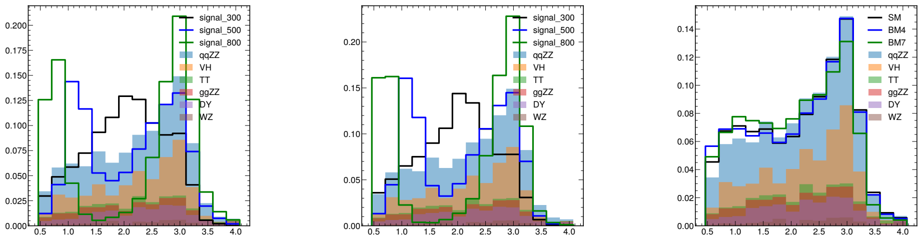
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.12: The distribution of `dr_secondTauHPair_dr` *for the three physics scenarios.*



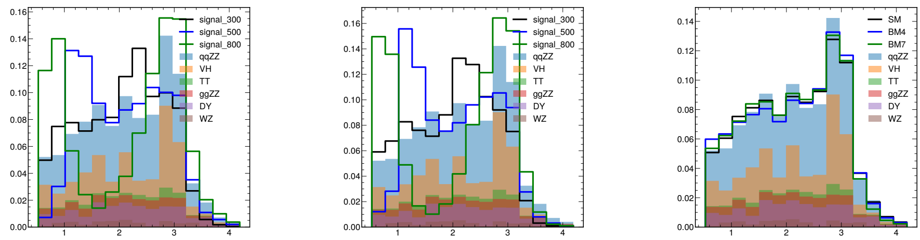
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.13: The distribution of `dr_secondTauHPair_m` *for the three physics scenarios.*



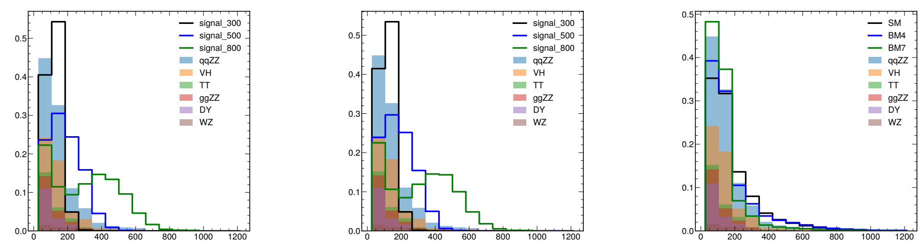
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.14: The distribution of `dr_tau1_tau3` for the three physics scenarios.



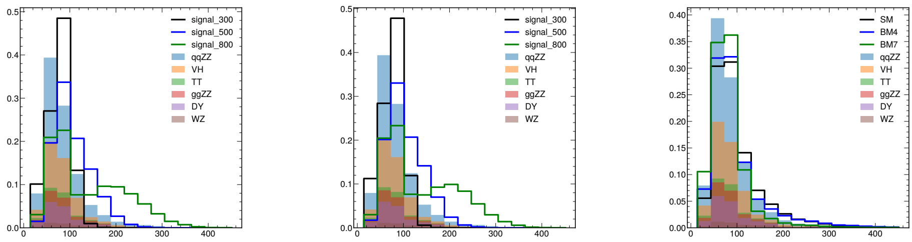
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.15: The distribution of `dr_tau2_tau4` for the three physics scenarios.



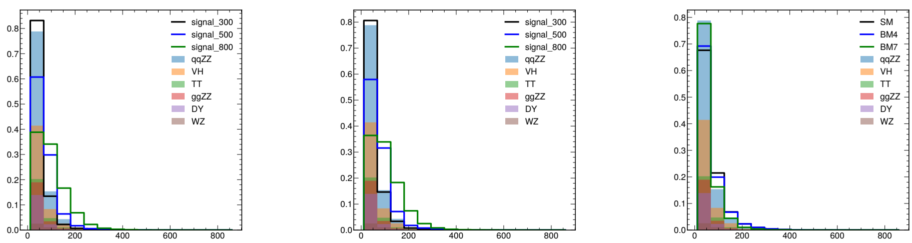
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.16: The distribution of `mTauTau` for the three physics scenarios.



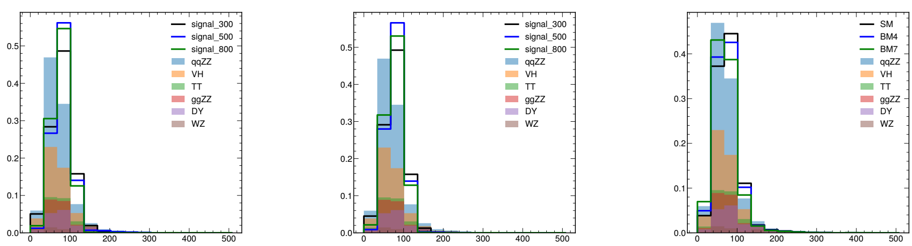
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.17: The distribution of `m_tau2_tau3` *for the three physics scenarios.*



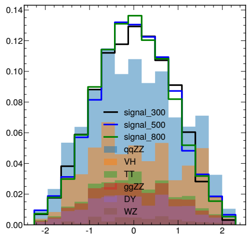
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.18: The distribution of `met_LD` *for the three physics scenarios.*

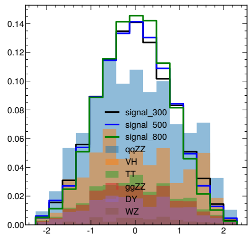


(a) spin-0 (b) spin-2 (c) nonres

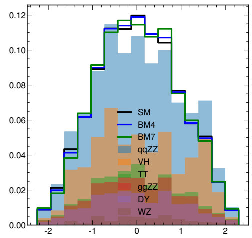
Figure 4.19: The distribution of `pt_bestTauHPair_m` *for the three physics scenarios.*



(a) spin-0

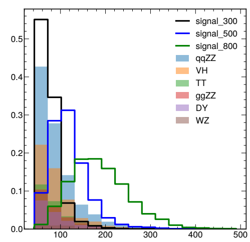


(b) spin-2

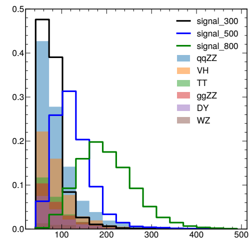


(c) nonres

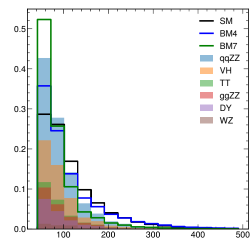
Figure 4.20: The distribution of `tau1_eta` *for the three physics scenarios.*



(a) spin-0

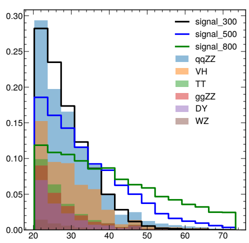


(b) spin-2

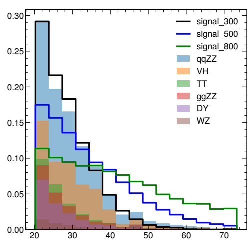


(c) nonres

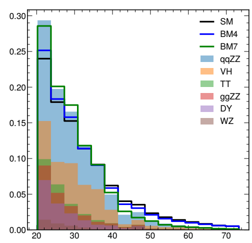
Figure 4.21: The distribution of `tau1_pt` *for the three physics scenarios.*



(a) spin-0



(b) spin-2

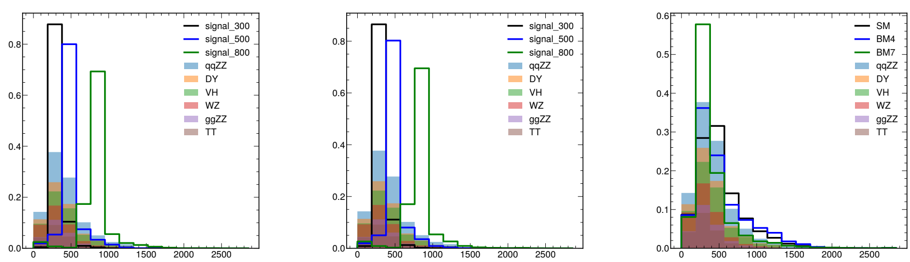


(c) nonres

Figure 4.22: The distribution of `tau4_pt` *for the three physics scenarios.*

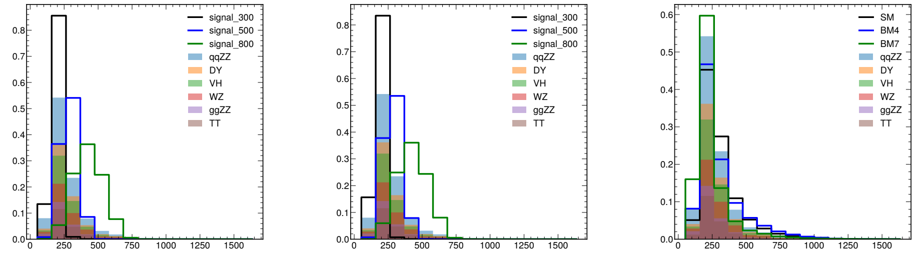
Table 4.5: *Discriminating variables used for the three physics scenarios in the $1\ell + 3\tau_\mathrm{h}$ channel.*

	spin-0	**spin-2**	**non-resonant**
diHiggsMass	✓	✓	✓
diHiggsVisMass	✓	✓	✓
dr_lep_tau1	✓	✓	✗
dr_lep_tau2	✓	✓	✓
dr_lep_tau3	✓	✓	✓
dr_tau1_tau2	✓	✓	✓
dr_tau1_tau3	✓	✓	✓
dr_tau2_tau3	✓	✓	✗
lep_eta	✗	✗	✓
lep_conePt	✗	✗	✓
mT_lep	✓	✓	✓
m_lep_tau1	✗	✗	✓
met	✗	✗	✓
mht	✗	✗	✓
pt_HH_recoil	✗	✗	✓
tau1_pt	✗	✗	✓
tau2_pt	✗	✗	✓



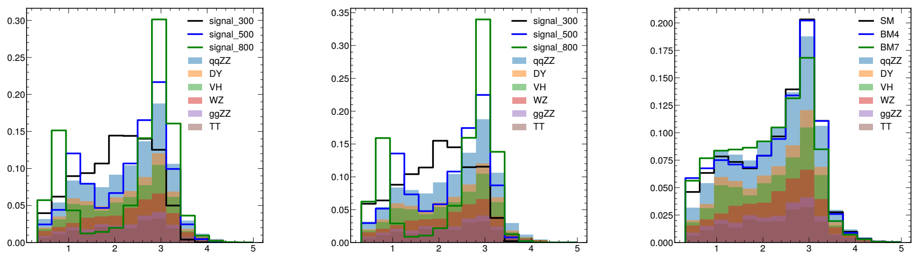
(a) spin-0 *(b) spin-2* *(c) nonres*

Figure 4.23: The distribution of `diHiggsMass` *for the three physics scenarios.*



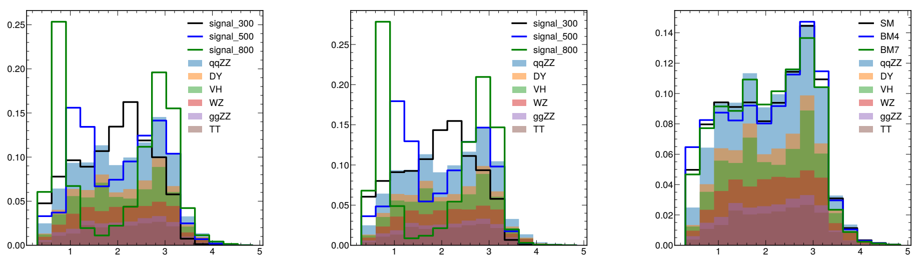
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.24: The distribution of `diHiggsVisMass` *for the three physics scenarios.*



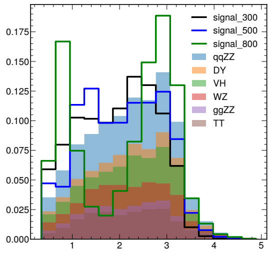
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.25: The distribution of `dr_lep_tau1` *for the three physics scenarios.*

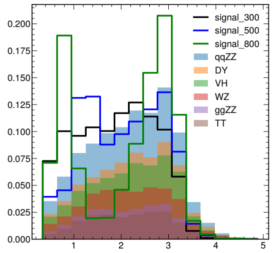


(a) spin-0 (b) spin-2 (c) nonres

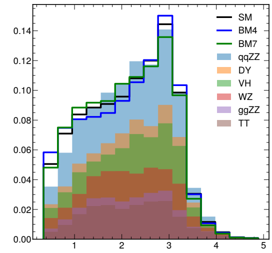
Figure 4.26: The distribution of `dr_lep_tau2` *for the three physics scenarios.*



(a) spin-0

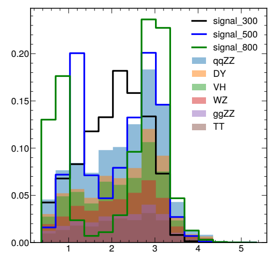


(b) spin-2

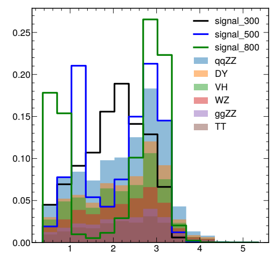


(c) nonres

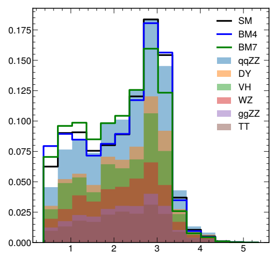
Figure 4.27: The distribution of `dr_lep_tau3` *for the three physics scenarios.*



(a) spin-0

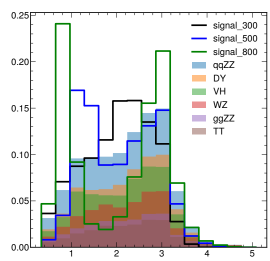


(b) spin-2

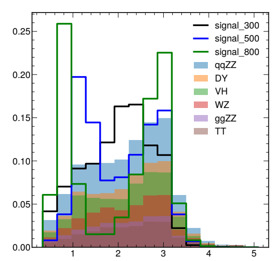


(c) nonres

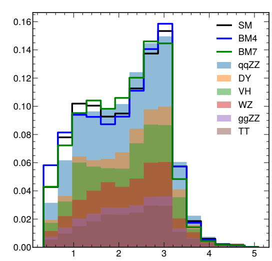
Figure 4.28: The distribution of `dr_tau1_tau2` *for the three physics scenarios.*



(a) spin-0

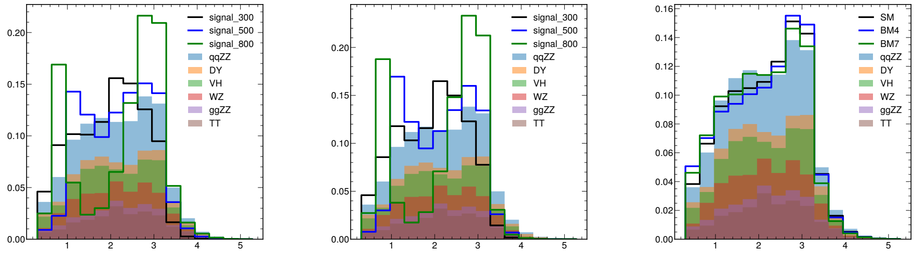


(b) spin-2



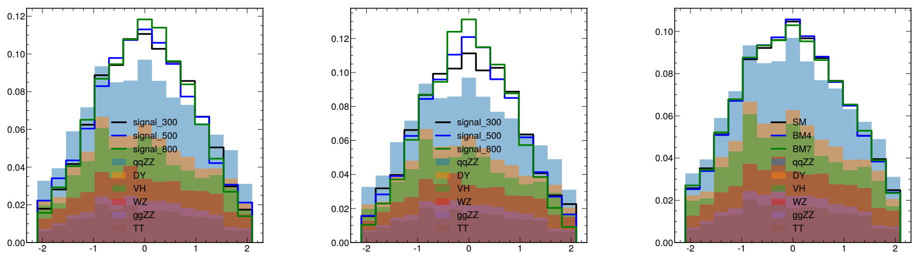
(c) nonres

Figure 4.29: The distribution of `dr_tau1_tau3` *for the three physics scenarios.*



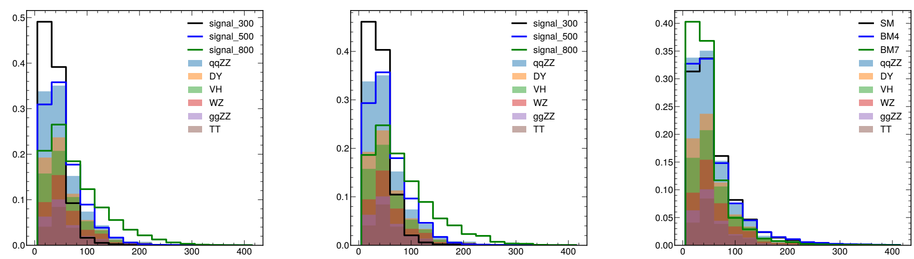
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.30: The distribution of `dr_tau2_tau3` for the three physics scenarios.



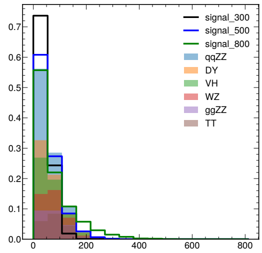
(a) spin-0 (b) spin-2 (c) nonres

Figure 4.31: The distribution of `lep_eta` for the three physics scenarios.

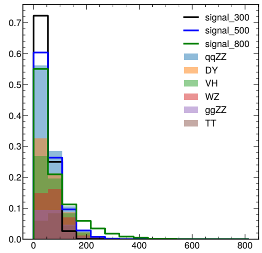


(a) spin-0 (b) spin-2 (c) nonres

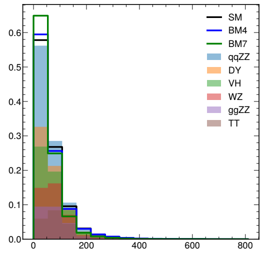
Figure 4.32: The distribution of `lep_pt` for the three physics scenarios.



(a) spin-0

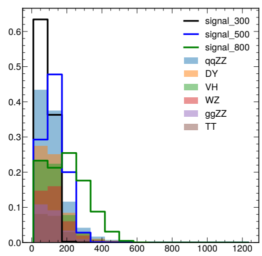


(b) spin-2

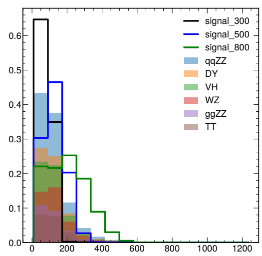


(c) nonres

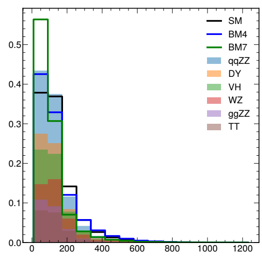
Figure 4.33: The distribution of `mT_lep` *for the three physics scenarios.*



(a) spin-0

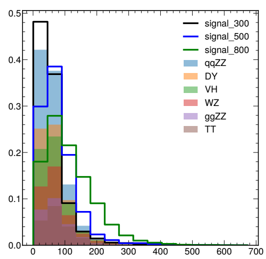


(b) spin-2

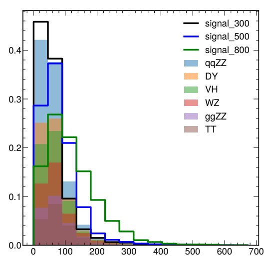


(c) nonres

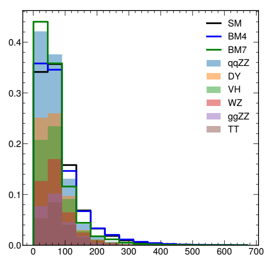
Figure 4.34: The distribution of `m_lep_tau1` *for the three physics scenarios.*



(a) spin-0

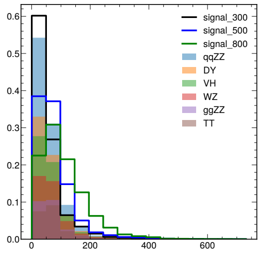


(b) spin-2

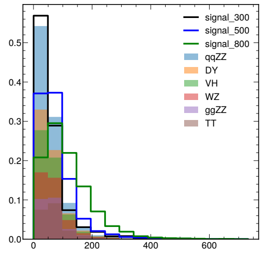


(c) nonres

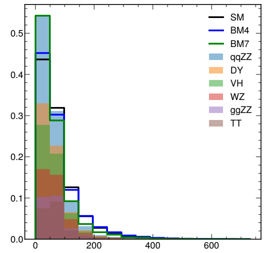
Figure 4.35: The distribution of `met` *for the three physics scenarios.*



(a) spin-0

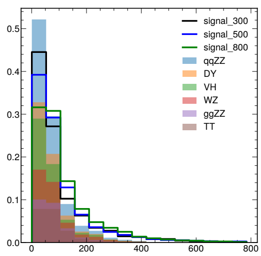


(b) spin-2

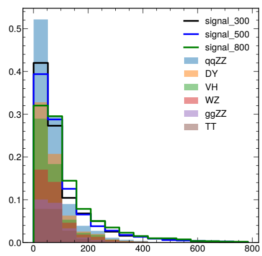


(c) nonres

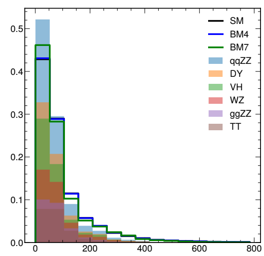
Figure 4.36: The distribution of `mht` *for the three physics scenarios.*



(a) spin-0

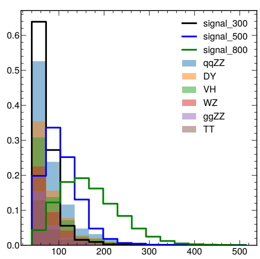


(b) spin-2

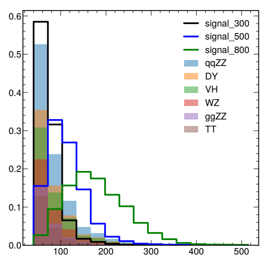


(c) nonres

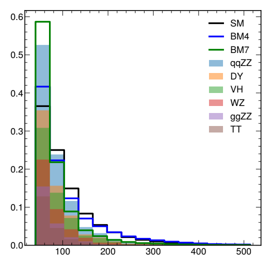
Figure 4.37: The distribution of `pt_HH_recoil` *for the three physics scenarios.*



(a) spin-0

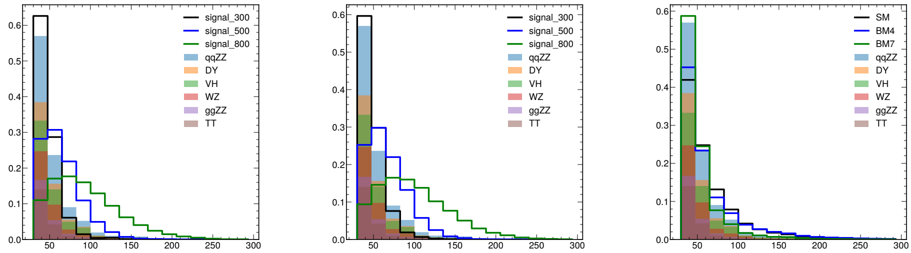


(b) spin-2



(c) nonres

Figure 4.38: The distribution of `tau1_pt` *for the three physics scenarios.*



(a) spin-0 *(b) spin-2* *(c) nonres*

Figure 4.39: The distribution of `tau2_pt` *for the three physics scenarios.*

4.3.5 Hyperparameter optimization

The analysis channels in the $\mathrm{HH} \to \mathrm{multilepton}$ analysis feature a small number of events (in comparison to the other analyses) and thus the time spent on the evaluation of the OF is relatively short, being only $\sim \mathcal{O}(30)$ minutes. As PSO is a great algorithm for exactly these kinds of scenarios, it was chosen to be the hyperparameter optimization algorithm in this analysis.

The *metaparameters*[21] chosen for the PSO that are shared for every analysis channel are given in Table. 4.6

In total 7 XGBoost algorithm hyperparameters that have the biggest impact on the training were chosen to be optimized. The bounds between which the hyperparameters were optimized are given in Table. 4.7.

Table 4.6: Parameter settings for the PSO algorithm [1].

Parameter	Value
N_{info}	10
c_1	1.62
c_2	1.62
w_{min}	0.4
w_{max}	0.8

Using PSO hyperparameter optimization algorithm resulted in models with a $\mathcal{O}(10\%)$ better AUC score in comparison to the models with manually tuned hyperparameters. The hyperparameter values for the $0\ell + 4\tau_{\mathrm{h}}$ and $1\ell + 3\tau_{\mathrm{h}}$ channels are given in Table. 4.8. The hyperparameters for the other 5 analysis channels can be found in Ref. [52].

[21]parameters needed to be chosen prior to the hyperparameter optimization that are inherent to every hyperparameter optimization algorithm.

Table 4.7: Minimum and maximum values of the hyperparameters for the HBC. The hyperparameters are detailed in Ref. [65].

Hyperparameter	min	max
n-estimators	1	500
learning-rate	10^{-5}	1
max-depth	1	6
gamma	0	5
min-child-weight	0	500
subsample	0.8	1
colsample-bytree	0.3	1

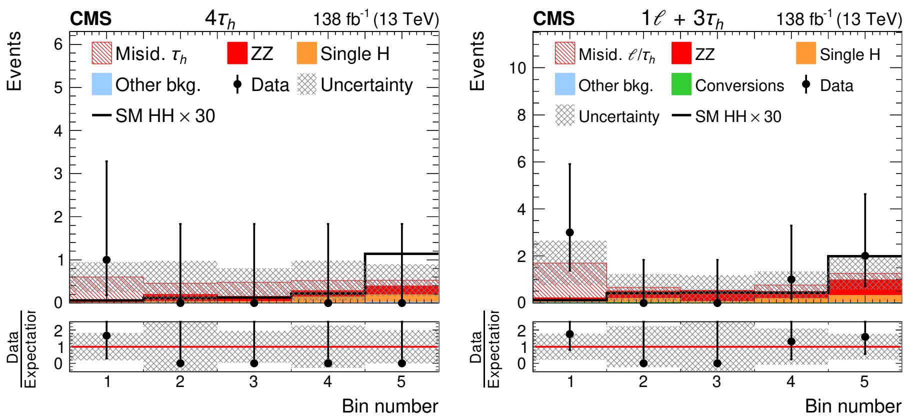
Table 4.8: Minimum and maximum values of the hyperparameters for the HBC. The hyperparameters are detailed in Ref. [65].

Hyperparameter	$0\ell+4\tau_\mathrm{h}$			$1\ell+3\tau_\mathrm{h}$		
	spin-0	spin-2	non-resonant	spin-0	spin-2	non-resonant
n-estimators	22	22	101	166	166	160
learning-rate	0.406	0.406	0.036	0.149	0.149	0.057
max-depth	4	4	4	4	4	4
gamma	1.75	1.75	5.75	3.35	3.35	2.91
min-child-weight	159	159	499	326	326	449
subsample	0.904	0.904	0.504	0.8	0.8	0.6
colsample-bytree	1.0	1.0	0.7	0.829	0.829	0.624

4.4 Results

In the following an overview of the $\mathrm{HH} \rightarrow \mathrm{multilepton}$ analysis results for both resonant as well as for the non-resonant physics scenarios is given, with the contributions by the $0\ell+4\tau_\mathrm{h}$ and $1\ell+3\tau_\mathrm{h}$ channels shown for every result.

In Fig. 4.40 the nonRes BDT output distribution for the $0\ell+4\tau_\mathrm{h}$ and $1\ell+3\tau_\mathrm{h}$ channels are shown. The postfit event yields corresponding to these can be found in Table. 4.9



(a) $0\ell + 4\tau_h$

(b) $1\ell + 3\tau_h$

Figure 4.40: Distribution of the output of the BDT classifier trained for the non-resonant HH production for the JHEP BM7 for the $0\ell + 4\tau_h$ and $1\ell + 3\tau_h$ channels. The cross section of the SM HH signal is scaled up by a factor of 30. Shaded gray area denotes the sum of statistical and systematic uncertainties on the background prediction.

Table 4.9: Number of expected and observed events in the channels $0\ell + 4\tau_h$ and $1\ell + 3\tau_h$ [2].

Process	$0\ell + 4\tau_h$	$1\ell + 3\tau_h$
SM HH $\to$ WW*WW* ($\times 30$)	0.2 ± 0.0	0.3 ± 0.0
SM HH $\to$ WW$^*\tau\tau$ ($\times 30$)	0.6 ± 0.1	0.1 ± 0.0
SM HH $\to \tau\tau\tau\tau$ ($\times 30$)	2.6 ± 0.4	1.3 ± 0.2
WZ	< 0.1	< 0.1
ZZ	1.9 ± 0.2	0.7 ± 0.1
Misidentified ℓ and τ_h[22]	2.2 ± 2.1	2.2 ± 1.6
Conversion electrons	< 0.1	< 0.1
Single Higgs boson	0.8 ± 0.4	0.4 ± 0.3
Other background	0.1 ± 0.1	< 0.1
Total expected background	5.0 ± 2.2	3.4 ± 1.6
Data	6	1

A potential HH signal event from the $1\ell + 3\tau_h$ channel is shown in Fig. 4.41. This event was observed in 2018 and has the highest BDT score (0.755) of the 6 selected events. That event features two reconstructed hadronic taus, one anti-tau and an electron with a considerable amount of missing transverse energy (MET). Different pairings of the final state objects hints that it is either a single-Higgs (VH) or a di-Higgs event.

The only candidate signal event from the $0\ell + 4\tau_h$ category is shown in Fig. 4.42. With the BDT score of 0.125, this event is most likely not a HH signal event.

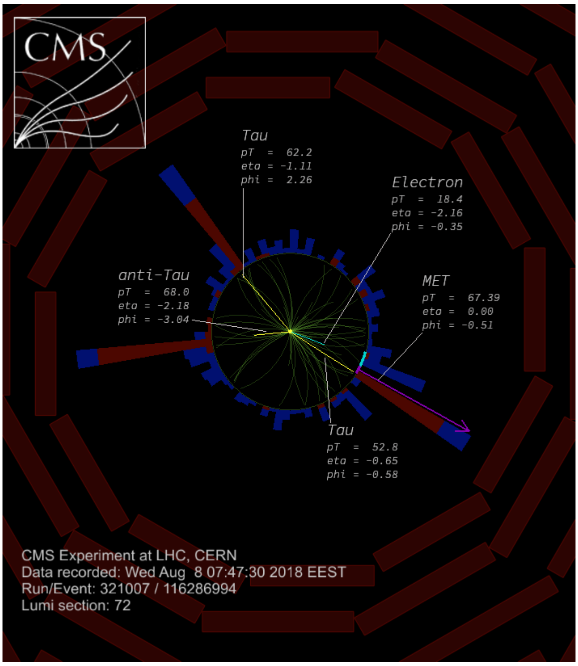


Figure 4.41: Event with the highest BDT score of 0.755 by the BDT trained on the JHEP04BM7 among the 6 selected HH signal candidates in the $1\ell + 3\tau_{\mathrm{h}}$. This event is potentially either actual HH signal event or a VH background event.

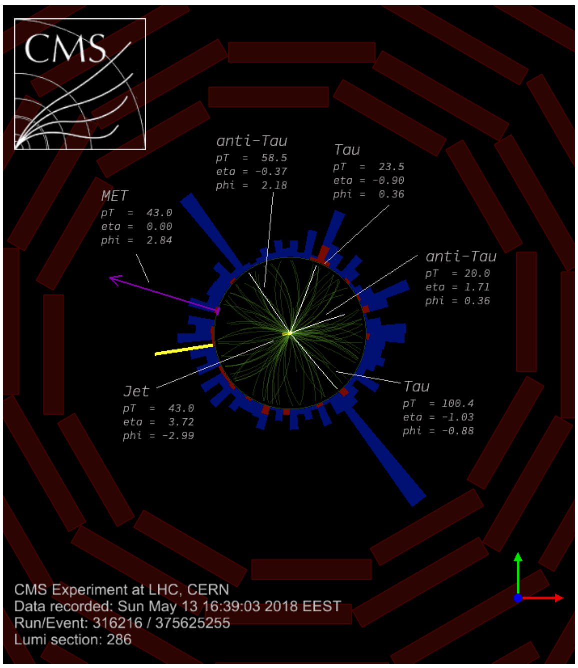


Figure 4.42: The only selected HH signal candidate found in the $0\ell + 4\tau_{\mathrm{h}}$ channel in 2018. With a score of 0.125 given by the BDT trained on the JHEP04BM7, this event is most probably not signal.

4.4.1 Systematic uncertainties

Signal extraction is affected by various experimental and theoretical effects that arise from imprecisely known or simulated effects and are treated as nuisance parameters. These effects, known as systematic uncertainties, are related to various kinematic properties, modifying the shape of the distributions in the discriminating observables as well as to the overall yield of the HH signal and the background processes. Most experimental effects correspond to data to MC corrections and data driven background estimations. In this section an overview of the systematic uncertainties affecting the $\mathrm{HH} \to \mathrm{multilepton}$ analysis are listed, while a more detailed description of the systematic uncertainties affecting to the $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$ channels is given.

As both $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$ have such low event yields, one of the biggest contibutor to the systematic uncertainties for these channels are related to the triggers. Namely, the di-τ trigger (lepton + τ_h cross trigger) used in the $0\ell + 4\tau_\mathrm{h}$ ($1\ell + 3\tau_\mathrm{h}$) channel is a source for the uncertainties arising from the τ_h legs of the trigger. However, the uncertainty in the efficiency of the lepton leg is neglected for the lepton + τ_h cross-trigger used in the $1\ell + 3\tau_\mathrm{h}$ channel. The size of the uncertainties is given by the TauPOG and depends on the kinematics (p_T, η, ϕ) as well as the decay mode of the τ. These trigger uncertainties are considered uncorrelated among all the analysis channels and are presumed to affect only the shape of the distribution of the discriminating variable.

Next, as fakes constitute a major part of the background for both channels, also the τ_h identification efficiency affects both channels considerably. Systematic uncertainty related to the τ_h identification efficiency is provided by the TauPOG and depends on the p_T and the decay mode of the τ_h. It is considered as a shape uncertainty and presumed to be uncorrelated for each year. However, it is dominated by statistical effects.

Another major contributor to the overall systematic uncertainty for the two channels is the uncertainty corresponding to the energy scale of the τ_h and is treated as a uncorrelated shape systematic for all the years.

Additionally, other analysis channels are also affected by jet energy scale, lepton identification and isolation, lepton trigger, jet energy resolution, b-tagging efficiency and mistag rate, signal and background rate, luminosity, pileup related uncertainties as well as by the L1 ECAL prefiring for the years 2016 and 2017.

4.4.2 Signal extraction

Signal extraction is done by performing a binned maximum likelihood fit using the BDT output distributions of the seven search channels as well as the two control regions. This is done by determining the signal strength $r = \frac{\sigma_{HH}^{fit}}{\sigma_{HH}^{theory}}$ for each physics scenario with a profile likelihood test statistic [115] that quantifies the probability of seeing the recorded data given the background model with and without the contribution from signal. Here we define the likelihood $\mathscr{L}$ in terms of the nuisance parameters θ that represent the systematic uncertainties introduced in the previous section.

Once the best fit signal strength $\hat{r}$ has been found, the significance Σ of it is defined as the ratio of it relative to the background only hypothesis, where the signal strength is set to $r = 0$:

$$\Sigma = -2\ln\left(\frac{\mathscr{L}(data|r=0,\hat{\theta})}{\mathscr{L}(data|r=\hat{r},\hat{\theta}_0)}\right). \tag{4.1}$$

Here the $\hat{\theta}$ and $\hat{\theta}_0$ denote the nuicance parameters for the best fit signal strength and the background-only hypothesis respectively. Using this definition above, we can now place upper limits on the signal strength by the means of asymptotic approximation [115] to define a 95% confidence interval for the signal strength.

However, if we would make decisions only based on the empirical data, we might introduce severe biases that could lead to a false discovery or rejection. For this reason we make use of Asimov dataset that we construct by fixing the observed yields to be equal to the predicted yields. The results obtained using the Asimov dataset are referred to as *expected*, whereas the ones obtained from the real measurements are referred to as *observed*.

A more detailed discussion regarding signal extraction can be found in Ref. [52].

4.4.3 Upper limits

In this section, 95% CL upper limit scans on the HH production cross section for the coupling strength modifiers for Higgs boson self-coupling κ_λ, Higgs boson coupling to top quark κ_t, and effective couplings of top quark to two Higgs bosons (c_2) and two Higgs bosons to two vector bosons (c_{2V}) for the channels $1\ell + 3\tau_\mathrm{h}$, $0\ell + 4\tau_\mathrm{h}$ and their combination are presented.

Limit scans for the coupling parameters are performed using the physics model described in Ref. [116]. When setting the limits on the cross section, the uncertainties for the theoretical cross section of HH production are frozen. For all limits one uses the non-resonant BDT using the JHEP04BM7 point as it focuses on currently important regions in κ_λ in the coupling scans .

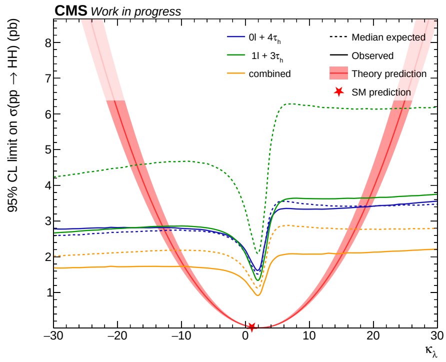
On Fig. 4.43a we see the scan over possible values of the Higgs boson self coupling strength modifier κ_λ. For the scan, all other Higgs boson couplings are set to the SM values. The intersection of the theory prediction and the observed (expected) gives the lower and upper limits of [-9.6, 15.25] ([-11.06, 17.21]) for the combination of these two channels, while the limits for the full $\mathrm{HH} \to \mathrm{multilepton}$ analysis are [-6.9, 11.1] ([-6.9, 11.7]).

Scan over the Higgs boson - top quark coupling strength modifier κ_t is shown in Fig. 4.43b. The observed (expected) lower and upper 95% CL limit for the combination of the two channels is [-2.64, 2.92] ([-2.82, 3.09]), while for the full $\mathrm{HH} \to \mathrm{multilepton}$ analysis they are [-1.96, 2.39] ([-2.00, 2.54]).

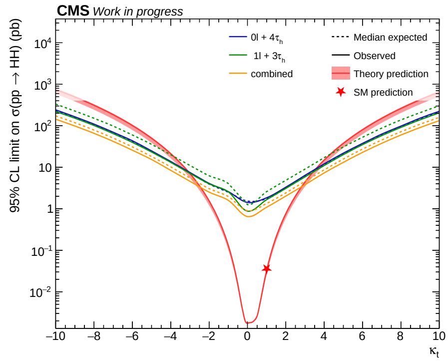
The limit scan for the EFT coupling parameter c_2 (shown on Fig. 4.43c) results in the observed (expected) lower and upper 95% CL limit for the combination of the two channels in being [-1.42, 1.86] ([-1.56, 1.98]) and for the full analysis [-1.05, 1.48] ([-0.96, 1.37]).

The 95% CL limits on the coupling strength modifier for non-resonant qqHH production c_{2v} is shown on Fig. 4.43d. The observed (expected) lower and upper 95% CL limit for the combination of the two channels in being [-4.69, 6.85] ([-4.75, 6.89]) and for the full analysis [-3.42, 5.56] ([-2.73, 4.83]).

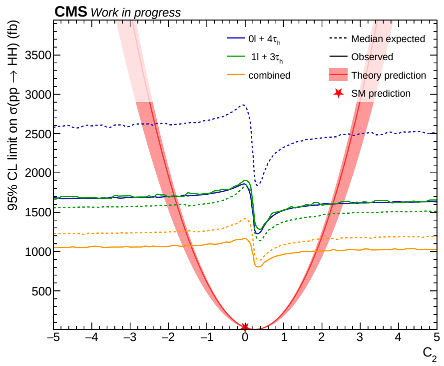
The contribution from the $0\ell + 4\tau_\mathrm{h}$ channel to the analysis is bigger for the expected values of κ_λ, κ_t and c_{2V}, while for the c_2 the stronger expected limit is coming from the $1\ell + 3\tau_\mathrm{h}$ channel.



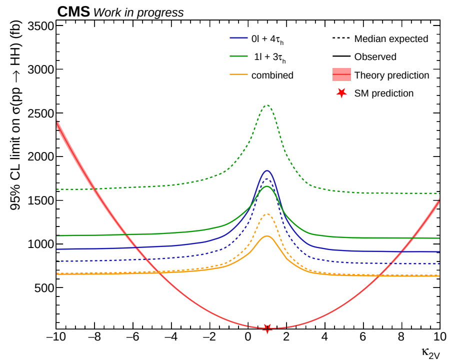
(a) κ_λ



(b) κ_t



(c) C_2



(d) C_{2V}

Figure 4.43: Upper limits for the parameters κ_λ, κ_t, C_2, C_{2V}.

The comparison of the expected and observed 95% CL upper limits of the $1\ell+3\tau_\mathrm{h}$ and $0\ell+4\tau_\mathrm{h}$ analysis channels, the combination of these these and for the full $\mathrm{HH}\to\mathrm{multilepton}$ analysis is shown in Table. 4.10. As both of these channels have very few events that pass the selection criteria, the contribution from these to the whole analysis is rather moderate.

Table 4.10: Lower and upper limits for $0\ell + 4\tau_h$, $1\ell + 3\tau_h$, combination of these two channels and for the full HH $\rightarrow$ multilepton analysis. There upper limits for the parameters κ_λ and C_2 for the full HH $\rightarrow$ multilepton analysis are taken from Ref. [2] and for κ_t from Ref. [22].

Parameter		$0\ell + 4\tau_h$	$1\ell + 3\tau_h$	**combined**	**HH $\rightarrow$ multilepton**
κ_λ	obs	[-12.67, 18.84]	[-12.96, 19.39]	[-9.60, 15.25]	[-6.9, 11.1]
	exp	[-12.85, 18.79]	[-17.00, 24.52]	[-11.06, 17.21]	[-6.9, 11.7]
κ_t	obs	[-3.14, 3.71]	[-3.02, 3.57]	[-2.64, 2.92]	[-1.96, 2.39]
	exp	[-3.07, 3.61]	[-3.91, 4.45]	[-2.82, 3.09]	[-2.00, 2.54]
C_2	obs	[-1.85, 2.29]	[-1.85, 2.29]	[-1.42, 1.86]	[-1.05, 1.48]
	exp	[-2.35, 2.78]	[-1.78, 2.20]	[-1.56, 1.98]	[-0.96, 1.37]
C_{2V}	obs	[-5.86, 7.99]	[-6.42, 8.62]	[-4.69, 6.85]	[-3.42, 5.56]
	exp	[-5.35, 7.50]	[-8.04, 10.00]	[-4.75, 6.89]	[-2.73, 4.83]

The observed and expected 95% CL upper limits on the SM HH production cross section for the two channels and their combination are shown on Fig. 4.44. The SM signal strength measurement is performed using the output of the BDT classifier that has been trained for SM non-resonant HH production. The observed (expected) upper limit for the combination of the two channels is 31 (38) times the SM cross section value, while for the full analysis the corresponding numbers are 21 and 19 for the observed and expected respectively.

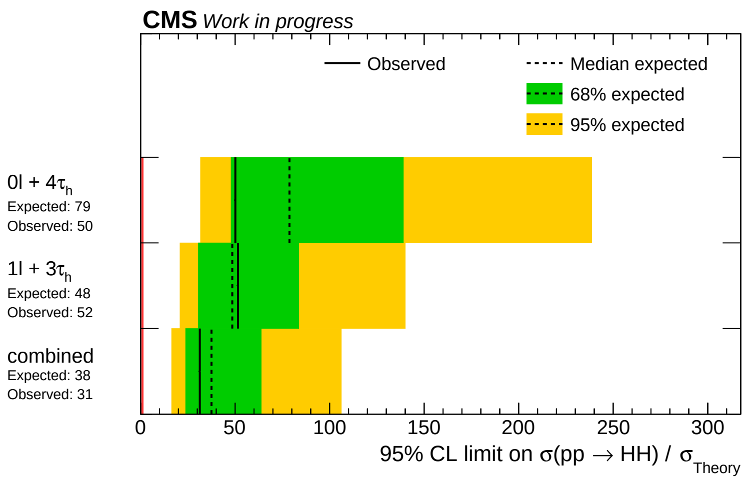


Figure 4.44: Observed and expected 95% CL upper limits on the SM HH production cross section. The observed (expected) upper limit for the combination of the two channels is 31 (38), while for the full HH $\rightarrow$ multilepton analysis it is 21 (19) (corresponding to 651 (592) fb). The numbers for the full analysis are taken from Ref. [2].

4.4.4 EFT Benchmarks

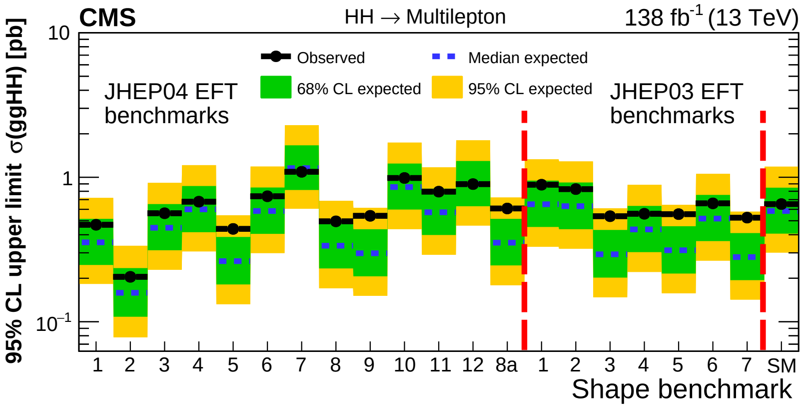
When setting the limits on the 20 benchmark scenarios, the one-hot-encoded value of the JHEP04 BM scenario or the kinematically closest JHEP03 BM scenario to a given JHEP04

BM point will be given as the input to the BDT. This means, we look at the the differential NLO HH cross section in m_{HH} and calculate the difference between the combinations of JHEP03 and JHEP04 BM scenarios. Then we assign for each JHEP03 scenario the JHEP04 that has the smallest difference when minimizing the metric $diff(s_1, s_2)$:

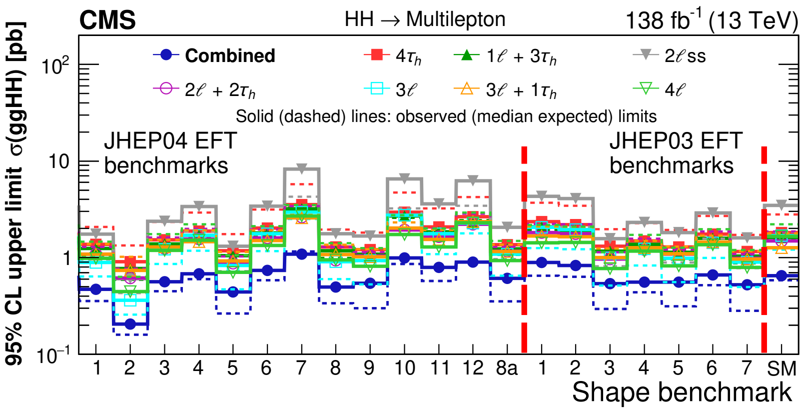
$$diff(s_1, s_2) = \sum_{i=0}^{bins} \left| \frac{N^{s1}_{events}(i) - N^{s2}_{events}(i)}{N^{s1}_{events}(i) + N^{s2}_{events}(i)} \right| , \tag{4.2}$$

where s_1 and s_2 denote the two scenarios for which the kinematic difference is calculated, and $N^{s1}_{events}(i)$ and $N^{s2}_{events}(i)$ the number of events in a given bin in the m_{HH} distribution.

Figs. 4.45 and 4.46 show the twenty benchmark scenarios that span a range of values in the κ_λ, κ_t, c_g, c_{2g} values, each of which corresponds to a different kinematic distribution, allowing us to measure the HH cross section at each point separately.

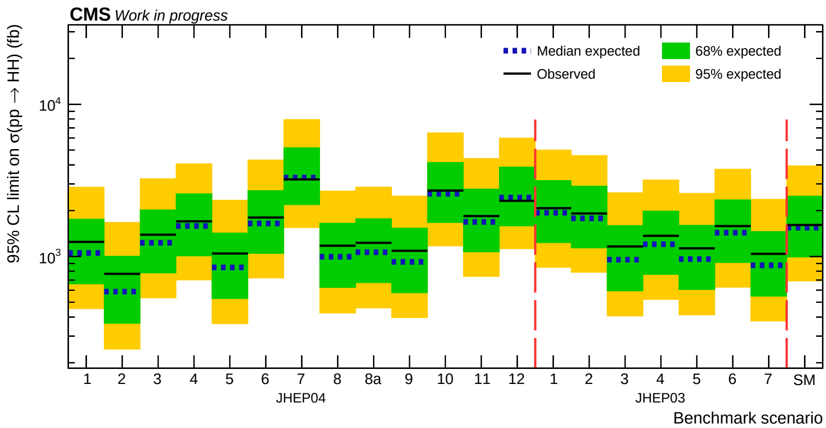


(a)

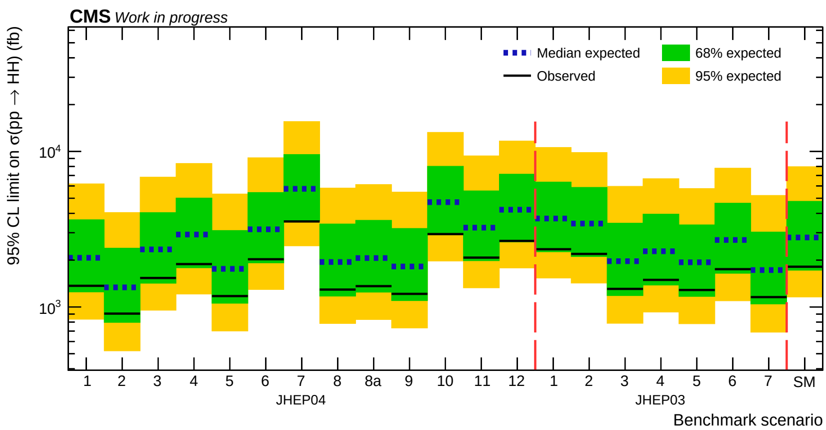


(b)

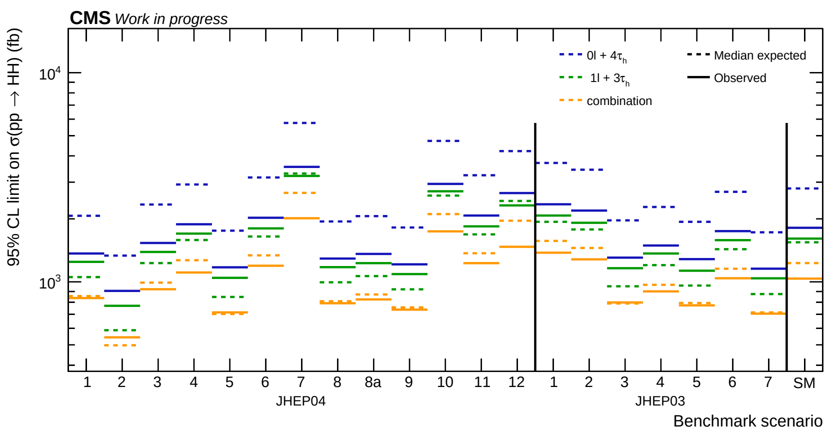
Figure 4.45: Upper limits at 95% CL on the production cross section for the SM, 13 (JHEP04) and 7 (JHEP03) EFT BSM benchmark scenarios for the full $\mathrm{HH} \to \mathrm{multilepton}$ *analysis (top) and for each analysis channel separately (bottom) [2].*



(a) $1\ell + 3\tau_h$



(b) $0\ell + 4\tau_h$



(c) Combination of $1\ell + 3\tau_h$ and $0\ell + 4\tau_h$.

Figure 4.46: Upper limits at 95% CL on the production cross section for the SM, 13 (JHEP04) and 7 (JHEP03) EFT BSM benchmark scenarios for the $1\ell + 3\tau_h$ (top), $0\ell + 4\tau_h$ (middle) and the combination of the two channels (bottom).

Comparing the 95% CL upper limit of the full $\mathrm{HH} \to \mathrm{multilepton}$ analysis shown in Fig. 4.45a and the combination of the two analysis channels focused on in this thesis we can see that the 95% CL upper limit of the full analysis given $\sim 2-3$ times stronger limits.

4.4.5 Resonant

The observed and expected limits on the resonant HH production cross section as a function of m_X for the full $\mathrm{HH} \to \mathrm{multilepton}$ analysis is shown on Figs. 4.47. The HH production cross section is evaluated at the mass points listed in Sec. 4.2, where the resonance mass m_X is given as an input to the resonant BDT.

As it is easier to distinguish signal from background for higher m_X values due to the increased acceptance, the limits are expected to be also more stricter at higher resonant masses m_X.

Also, as the signal efficiency and the BDT output shapes differ for different masses and spin scenarios, we performed a separate measurement for each mass and spin hypothesis. Depending on the spin and the mass hypothesis, the observed (expected) 95% CL upper limit for the resonant HH production cross section ranges from 0.18 to 0.90 (0.08 to 1.06) pb [2].

4.5 Conclusions and outlook

The $\mathrm{HH} \to \mathrm{multilepton}$ analysis [2] presented in this thesis was included in the combination of the Run2 CMS Higgs analyses in Ref. [29], which showed the existence of the quartic coupling VVHH ($\kappa_{2V} \neq 0$) with a significance of 6.6 standard deviations.

Furthermore, the exclusion region for κ_λ was tightened, with the lower bound now approaching 0, meaning the evidence for trilinear Higgs self-coupling is not in a too distant future. As the main strength of multilepton analysis is in constraining HH production at low m_{HH} region (corresponding to high κ_λ values), impact on the HH effort will be considerable.

Still, there are several aspects that can be improved for the ongoing LHC Run3[23] and the planned HL-LHC[24] $\mathrm{HH} \to \mathrm{multilepton}$ analyses.

The $\mathrm{HH} \to \mathrm{multilepton}$ analysis covered in this thesis is limited mostly by the available data and simulations. In addition to the fact that having more recorded and simulated events reduces the statistical uncertainties, then one would expect an increase in performance for background modeling and ML classifier training with more simulated events, thereby allowing a more refined analysis.

The aspects in the analysis strategy that can be improved are the following: firstly, as shown in Ref. [117] considering VBF as a separate signal category improves the the limits on c_{2V} by a factor 4. While the improvement might not be as pronounced in other analysis channels and/or for other Higgs parameters, the addition of the VBF category offers potentially a noticeable improvement to the analysis.

Secondly, more sophisticated ML methods could be used. For example in the channels that are not as statistically constrained, one could make use of neural networks. This then makes it possible to engineer new and meaningful features using LBN as described in Sec. 3.4.2.

[23] 2022-2025. The total integrated luminosity is expected to be $\sim 300\mathrm{fb}^{-1}$, which is more than twice the recorded luminosity in Run2

[24] Planned start is in the beginning of 2029. The total integrated luminosity is expected to be $\sim 3000\mathrm{fb}^{-1}$

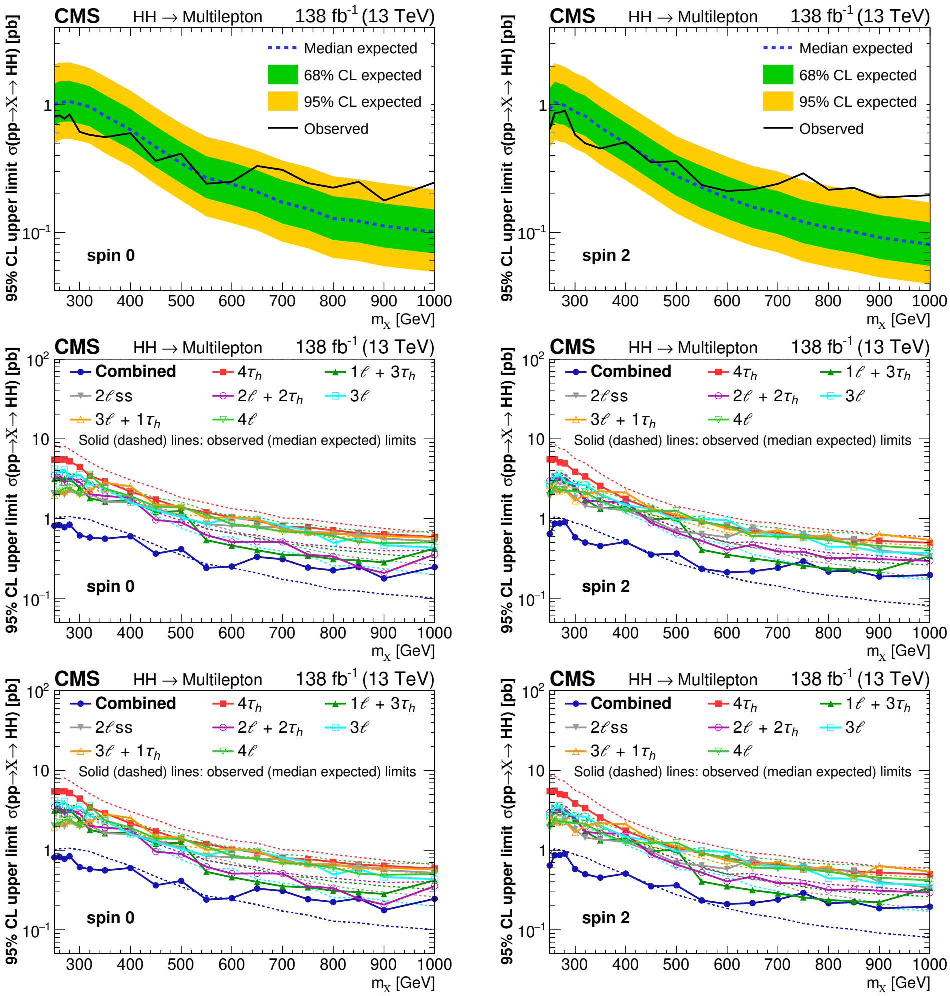


Figure 4.47: The observed and expected 95% CL upper limit on the production of new particles of spin-0 (left) and spin-2 (right) decaying into a pair of Higgs bosons in mass range of 250-1000 GeV. Plots on the top show the result of the combination of the seven channels, while the ones on the middle and bottom show the limits for each channel separately [2].

Next, in the current version of the analysis, the problem of imbalanced dataset is overcome by the means of oversampling. However, more sophisticated and performant sampling methods, as described Sec. 3.4.1, such as SMOTE or ADASYN could be employed. Alternatively, during training one could make use of focal loss [118], which has been shown to perform well with imbalanced datasets.

Furthermore, the currently employed method of calculating feature importances based on the number of times a features was used to split the data would not be suitable for neural networks. Also, this score might not be the best to describe the importance of features. For these purposes one could make use of metric such as SHAP, or even revisit the whole strategy of dropping variables (see Ref. [78]).

What's more, further subcategorizing channels based on the object multiplicity and/or flavor could be considered. Namely, splitting the $2\ell_{ss} + 0/1\tau_{\mathrm{h}}$ could be separate channels taking to account the lepton flavor flavors or splitting the channel based on tau multiplicity: $2\ell(ss) + 0/1\tau_{\mathrm{h}} \to 2\ell_{ss} + 0\tau_{\mathrm{h}}$ and $2\ell_{ss} + 1\tau_{\mathrm{h}}$. Having a separate channel that considers the two oppositely charged leptons $2\ell(os)$ is another possibility, which however most likely is greatly affected by large backgrounds. Either way the feasibility of these additions has to be studied further in a detailed manner.

Finally, as of yet there doesn't exist a multilepton analysis investigating the $\mathtt{X} \to \mathtt{YH}$[25]. As multilepton analysis is particularly sensitive to soft signatures, then the study of light mass Y will benefit from this a lot.

[25] X and Y being some new massive scalar particles

Acknowledgments

I would like to express my deepest gratitude to my supervisors, Mario Kadastik and Christian Veelken, whose guidance and mentorship were instrumental in shaping the direction of my research.

Heartfelt thanks are extended to Karl Ehatäht and Torben Lange for engaging in various discussions, offering valuable insights, and providing constructive feedback on the draft of my thesis, which contributed to refining the clarity of my work. Additionally, I would like to acknowledge Joosep Pata for his continuous guidance and thoughtful advice, as well as to Norman and Sven for the camaraderie and good laughs that lightened the mood during my doctoral studies. Gratitude is also extended to my other esteemed colleagues - Christine, Ram, Sandeep, Olin and Saswati.

I extend my heartfelt appreciation to the KBFI for providing the necessary resources and facilities that facilitated the completion of this thesis, with a special mention going to Lauri for keeping the compute cluster alive and well.

Big thanks to Arnaud Ferrari and Veronika Zadin for kindly agreeing to fulfil the opposing role in my PhD defence.

Moreover, I am profoundly grateful for the unwavering support from my family throughout my academic journey. A special mention is reserved for my wife, Sale-Iti, and my loyal canine companion, Drogomir, whose simple yet profound reminders to take breaks, enjoy a walk, and breathe in the crisp city air served as essential self-care amidst the academic rigor.

This work was financed by the Estonian Research Council grants PRG445 and MOBTT86.

Has anyone really been far even as decided to
use even go want to do look more like?

- Anonymous

Abstract
Measurement of Higgs Boson Properties in Leptonic Final States using ML-methods

The final elementary particle predicted by the Standard Model (SM) was experimentally discovered in 2012: the Higgs boson. During the ten years since it's discovery, a few properties of the Higgs boson still remain unmeasured. These parameters include for example the Higgs self-coupling λ and the Yukawa coupling to the top quark, y_t.

The subject of this thesis is the study of Higgs boson pair (HH) production, which allows the measurement of the Higgs boson self-coupling. The production of HH is studied in the decay channels to four vector bosons (VVVV), two vector bosons and two τ leptons (VV$\tau\tau$), and to four τ leptons ($\tau\tau\tau\tau$). The final states considered in the analysis feature multiple electrons (e), muons (μ), and hadronically decaying τ leptons (τ_h). Although the branching fractions for these decay channels are rather small, they have the advantage of being almost free of backgrounds. In order to increase the efficiency for the rare HH signal, the study includes seven complementary analysis channels defined by the multiplicity of e, μ, τ_h and their charge. The analyzed data has been recorded by the Compact Muon Solenoid (CMS) experiment located at the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN). The data used in this analysis was collected in 2016-2018 at a center-of-mass energy of 13 TeV and corresponds to an integrated luminosity of 138 fb^{-1}.

The main focus of this thesis is on two of the seven analysis channels - $0\ell + 4\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$. Both of these channels target the Higgs boson pair decays to four τ leptons. As a significant fraction of the background in these two channels stems from misidentified jets and electrons, the channels sensitivity can be increased by studying various τ reconstruction and identification algorithms.

The signal extraction in this analysis is done by the means of boosted desision tree (BDT) discriminator, namely XGBoost. In addition to the choice of the algorithm, a multitude of algorithm specific parameters, referred to as hyperparameters, needed to be specified prior to training. As the choice of the hyperparameters influences the performance of the algorithm in a major way, an in depth study of various hyperparameter optimization algorithms and their use cases was made.

The data recorded by CMS so far did not allow to observe a HH signal. Instead, an upper limit was set on the signal cross section, which amounts to 21 times the SM prediction. These limits were used to constrain new physics contributions in the context of an effective field theory (EFT). Upper limits were also set on the contribution of new heavy particles decaying to Higgs boson pairs, which are predicted by varios theories beyond the SM: spin-0 and spin-2 particles in the mass range of 250 GeV to 1 TeV. No statistically significant excess over the SM expectation was observed.

Kokkuvõte
Higgsi bosoni omaduste mõõtmine leptoneid sisaldavates kanalites kasutades masinõppe meetodeid

2012. aastal avastatud Higgsi boson oli viimane eksperimentaalselt tõestatud Standardmudeli (SM) poolt ennustatud elementaarosake. Kuigi sellest avastusest on möödas üle kümne aasta, on endiselt osa Higgsi bosoni parameetreid mõõtmata. Nende parameetrite hulka kuuluvad näiteks Higgsi eneseinteraktsiooni seoseparameeter λ ning Yukawa seoseparameeter, mis määratleb Higgsi bosoni interaktsiooni tugevuse top-kvargiga y_t.

Antud doktoritöö uurib Higgsi bosoni paaride (HH) teket, mille otsemõõtmiste abil saab piiritleda Higgsi eneseinteraktsiooni seoseparameetrit vastava tekkeprotsessi ristlõikest. HH protsessi uuritakse siin nelja vektor-bosoni (VVVV), kahe vektor-bosoni ja kahe τ leptoniga (VV$\tau\tau$), ning nelja τ leptoniga ($\tau\tau\tau\tau$) lagukanalites, kus lõppolekutes esineb mitu elektroni (e), müüonit (μ) ja hadrooniliselt lagunenud tau leptonit (τ_h). Sellised lõppolekud on harvad, kuid on see eest peaaegu vabad taustaprotsessidest. Signaali osatähtsuse suuredamiseks uuritakse mitmeleptoniliseid lõppolekuid seitsmes ortogonaalses analüüsikanalis, mis on defineeritud leptonite arvukuse ja elektrilaengu järgi. Mõõtmiste andmed pärinevad Compact Muon Solenoid (CMS) mis asub Suurel Hadroni Põrgutil (LHC) Euroopa Tuumauuringute Keskuses (CERN). Andmed koguti aastatel 2016-2018 massikeskme energiaga 13 TeV ning mis vastab integreeritud luminositeedile 138 fb^{-1}.

Antud doktoritöö fookus on kahel analüüsi kanalil, $0\ell+4\tau_\mathrm{h}$ ja $1\ell+3\tau_\mathrm{h}$. Mõlemad kanalid uurivad peamiselt HH lagunemist neljaks τ leptoniks. Kuna suur osa taustaprotsessidest on tingitud valesti identifitseeritud elektronidest ja osakeste jugadest, saab kanali tundlikkust suurendada uurides erinevaid τ leptoni rekonstrueerimise ja identifitseerimise algoritme.

Signaali ekstraheerimiseks kasutati XGBoost algoritmi, mis on üks võimendatud otsustuspuude (BDT) alaliikidest. Peale sobiva algoritmi valikut on vaja täpsustada enne treeningut suur hulk algoritmi spetsiifilisi parameetreid, samuti tuntud ka kui hüperparameetrid. Kuna hüperparameetrite valik mõjutab analüüsi tulemusi tugevalt, uuriti sügavuti hüperparameetrite optimiseerimise algoritmide sooritusvõimet ning kasutusjuhtumeid.

CMS eksperimendi poolt salvestatud andmete põhjal vajalikku tundlikkust HH signaali nägemiseks ei saavutatud. Seetõttu seati kasutatud andmete põhjal ülempiir maksimaalsele signaali tugevusele, mis on võrdne 21 kordsele Standardmudeli poolt ennustatud tekkeristlõikele. Neid ülempiire kasutati piiritlemaks uue füüsika panust efektiivse väljateooria (EFT) kontekstis. Ülempiirid seati ka uue massiivse osakese lagunemisele Higgsi bosoni paarideks, mida ennustavad erinevad Standardmudeli laiendused — spin-0 ja spin-2 osakestele massivahemikus 250 GeV kuni 1 TeV. Antud analüüs ei leitud statistiliselt märkimisväärseid ülejääke, jäädes seega vastavusse Standardmudeli ennustusega 95% usaldusnivool.

Bibliography

[1] L. Tani, D. Rand, C. Veelken, and M. Kadastik, “Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics”, *Eur. Phys. J. C*, vol. 81, no. 2, p. 170, 2021. DOI: 10.1140/epjc/s10052-021-08950-y. arXiv: 2011.04434 [hep-ex] (cit. on pp. 7, 54, 55, 83, 113, 215).

[2] CMS collaboration, “Search for Higgs boson pairs decaying to $\mathrm{WW}^*\mathrm{WW}^*$, $\mathrm{WW}^*\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s}$ = 13 TeV”, *JHEP*, vol. 07, p. 095, 2023. DOI: 10.1007/JHEP07(2023)095. arXiv: 2206.10268 [hep-ex] (cit. on pp. 7, 24, 25, 59, 61, 64, 65, 85, 90, 91, 93, 94, 125, 215).

[3] L. Tani and C. Veelken, “Comparison of Bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning applications in high energy physics”, *Comput. Phys. Commun.*, vol. 294, p. 108 955, 2024. DOI: 10.1016/j.cpc.2023.108955. arXiv: 2201.06809 [physics.data-an] (cit. on pp. 7, 54, 56, 193, 215).

[4] T. Lange, S. Nandan, J. Pata, L. Tani, and C. Veelken, “Tau lepton identification and reconstruction: A new frontier for jet-tagging ML algorithms”, *Comput. Phys. Commun.*, vol. 298, p. 109 095, 2024. DOI: 10.1016/j.cpc.2024.109095. arXiv: 2307.07747 [hep-ex] (cit. on pp. 7, 41, 201, 215).

[5] D. Rand, “Genetic algorithms in machine learning for parameter space optimization”, Ph.D. dissertation, Tallinn University of Technology, School of Science, 2019 (cit. on p. 8).

[6] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST*, vol. 3, S08004, 2008. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on pp. 13, 26, 31, 34, 35).

[7] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*. ATLAS-TDR-014, 1999 (cit. on pp. 13, 26).

[8] L. Evans, “The Large Hadron Collider (LHC)”, *New J. Phys.*, vol. 9, J. Hewett, J. Jaros, T. Kamae, and C. Prescott, Eds., p. 335, 2007. DOI: 10.1088/1367-2630/9/9/335 (cit. on pp. 13, 26–28).

[9] A. Robertson, R. Massey, and V. Eke, “What does the Bullet Cluster tell us about self-interacting dark matter?”, *Mon. Not. Roy. Astron. Soc.*, vol. 465, no. 1, pp. 569–587, 2017. DOI: 10.1093/mnras/stw2670. arXiv: 1605.04307 [astro-ph.CO] (cit. on p. 13).

[10] M. Persic, P. Salucci, and F. Stel, “The universal rotation curve of spiral galaxies—i. the dark matter connection”, *Monthly Notices of the Royal Astronomical Society*, vol. 281, no. 1, pp. 27–47, 1996 (cit. on p. 13).

[11] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, “A Theory of Dark Matter”, *Phys. Rev. D*, vol. 79, p. 015 014, 2009. DOI: 10.1103/PhysRevD.79.015014. arXiv: 0810.0713 [hep-ph] (cit. on p. 13).

[12] N. Craig, J. Galloway, and S. Thomas, “Searching for Signs of the Second Higgs Doublet”, May 2013. arXiv: 1305.2424 [hep-ph] (cit. on pp. 13, 24).

[13] C. Englert, T. Plehn, D. Zerwas, and P. M. Zerwas, "Exploring the Higgs portal", *Phys. Lett. B*, vol. 703, pp. 298–305, 2011. DOI: 10.1016/j.physletb.2011.08.002. arXiv: 1106.3097 [hep-ph] (cit. on pp. 13, 24).

[14] R. Grober and M. Muhlleitner, "Composite Higgs Boson Pair Production at the LHC", *JHEP*, vol. 06, p. 020, 2011. DOI: 10.1007/JHEP06(2011)020. arXiv: 1012.1562 [hep-ph] (cit. on pp. 13, 24).

[15] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.*, vol. 83, pp. 3370–3373, 1999. DOI: 10.1103/PhysRevLett.83.3370. arXiv: hep-ph/9905221 (cit. on pp. 13, 24).

[16] K. G. Wilson, "Confinement of quarks", *Physical review D*, vol. 10, no. 8, p. 2445, 1974 (cit. on p. 16).

[17] S. M. Bilenky, C. Giunti, and W. Grimus, "Phenomenology of neutrino oscillations", *Progress in Particle and Nuclear Physics*, vol. 43, pp. 1–86, 1999 (cit. on p. 16).

[18] G. Bertone and D. Hooper, "History of dark matter", *Reviews of Modern Physics*, vol. 90, no. 4, p. 045 002, 2018 (cit. on p. 16).

[19] L. Roszkowski, E. M. Sessolo, and S. Trojanowski, "WIMP dark matter candidates and searches—current status and future prospects", *Rept. Prog. Phys.*, vol. 81, no. 6, p. 066 201, 2018. DOI: 10.1088/1361-6633/aab913. arXiv: 1707.06277 [hep-ph] (cit. on p. 16).

[20] ATLAS Collaboration, "Search for heavy resonances decaying into a Z or W boson and a Higgs boson in final states with leptons and b-jets in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP*, vol. 06, p. 016, 2023. DOI: 10.1007/JHEP06(2023)016. arXiv: 2207.00230 [hep-ex] (cit. on p. 16).

[21] W. Buchmüller and M. Plümacher, "Matter–antimatter asymmetry and neutrino properties", *Physics reports*, vol. 320, no. 1-6, pp. 329–339, 1999 (cit. on p. 16).

[22] T. Lange, "Search for rare Higgs boson decays at a center of mass energy of $\sqrt{s}$ = 13 TeV with the CMS Experiment at the LHC", Ph.D. dissertation, Universität Hamburg, U. Hamburg, Dept. Phys., 2022 (cit. on pp. 17, 60, 66, 67, 90).

[23] M. Thomson, *Modern particle physics*. Cambridge University Press, 2013 (cit. on p. 17).

[24] D. Griffiths, *Introduction to elementary particles*. John Wiley & Sons, 2020 (cit. on p. 17).

[25] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, "A Historical Profile of the Higgs Boson", in *The standard theory of particle physics: Essays to celebrate CERN's 60th anniversary*, L. Maiani and L. Rolandi, Eds. 2016, pp. 255–274. DOI: 10.1142/9789814733519_0014. arXiv: 1504.07217 [hep-ph] (cit. on p. 18).

[26] C. Grojean, "Higgs physics", 2017. 1708.00794: 1802.03888 (cit. on p. 19).

[27] G. Aad *et al.*, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", *Phys. Rev. Lett.*, vol. 114, p. 191 803, 2015. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex] (cit. on p. 20).

[28] D. de Florian *et al.*, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", vol. 2/2017, Oct. 2016. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph] (cit. on pp. 20, 21).

[29] CMS Collaboration, “A portrait of the Higgs boson by the CMS experiment ten years after the discovery.”, *Nature*, vol. 607, no. 7917, pp. 60–68, 2022. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043 [hep-ex] (cit. on pp. 20, 59, 93).

[30] ATLAS Collaboration, “Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ decay channel using 139 fb^{-1} of $\sqrt{s}$ = 13 TeV pp collisions recorded by the ATLAS detector at the LHC”, *Phys. Lett. B*, vol. 843, p. 137 880, 2023. DOI: 10.1016/j.physletb.2023.137880. arXiv: 2207.00320 [hep-ex] (cit. on p. 20).

[31] E. Rossi, “Measurement of Higgs-boson self-coupling with single-Higgs and double-Higgs production channels”, Oct. 2020. arXiv: 2010.05252 [hep-ex] (cit. on p. 20).

[32] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca, and M. Spira, “$gg \to HH$: Combined uncertainties”, *Phys. Rev. D*, vol. 103, no. 5, p. 056 002, 2021. DOI: 10.1103/PhysRevD.103.056002. arXiv: 2008.11626 [hep-ph] (cit. on p. 20).

[33] B. Di Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, “Higgs boson potential at colliders: Status and perspectives”, *Reviews in Physics*, vol. 5, p. 100 045, 2020 (cit. on p. 22).

[34] F. Bishara, R. Contino, and J. Rojo, “Higgs pair production in vector-boson fusion at the LHC and beyond”, *Eur. Phys. J. C*, vol. 77, no. 7, p. 481, 2017. DOI: 10.1140/epjc/s10052-017-5037-9. arXiv: 1611.03860 [hep-ph] (cit. on p. 21).

[35] A. Carvalho, M. Dall’Osso, T. Dorigo, F. Goertz, C. A. Gottardo, and M. Tosi, “Higgs Pair Production: Choosing Benchmarks With Cluster Analysis”, *JHEP*, vol. 04, p. 126, 2016. DOI: 10.1007/JHEP04(2016)126. arXiv: 1507.02245 [hep-ph] (cit. on pp. 23, 24, 64).

[36] R. Frederix *et al.*, “Higgs pair production at the LHC with NLO and parton-shower effects”, *Physics Letters B*, vol. 732, pp. 142–149, 2014 (cit. on p. 23).

[37] J. Alison *et al.*, “Higgs boson potential at colliders: Status and perspectives”, *Rev. Phys.*, vol. 5, B. Di Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, Eds., p. 100 045, 2020. DOI: 10.1016/j.revip.2020.100045. arXiv: 1910.00012 [hep-ph] (cit. on p. 24).

[38] M. Capozi and G. Heinrich, “Exploring anomalous couplings in Higgs boson pair production through shape analysis”, *JHEP*, vol. 03, p. 091, 2020. DOI: 10.1007/JHEP03(2020)091. arXiv: 1908.08923 [hep-ph] (cit. on pp. 24, 64).

[39] O. Brüning, H. Burkhardt, and S. Myers, “The large hadron collider”, *Progress in Particle and Nuclear Physics*, vol. 67, no. 3, pp. 705–734, 2012 (cit. on pp. 26, 27).

[40] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST*, vol. 12, no. 10, P10003, 2017. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det] (cit. on pp. 26, 36, 39).

[41] S. Myers and E. Picasso, “The design, construction and commissioning of the CERN large Electron–Positron collider”, *Contemporary Physics*, vol. 31, no. 6, pp. 387–403, 1990 (cit. on p. 26).

[42] R. D. Ball *et al.*, “Parton distributions from high-precision collider data”, *Eur. Phys. J. C*, vol. 77, no. 10, p. 663, 2017. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph] (cit. on p. 29).

[43] Izaak Neutelings, *CMS coordinate system*, https://tikz.net/axis3d_cms, Accessed: 28.08.2023 (cit. on p. 30).

[44] Tracker Group of the CMS Collaboration, "The CMS Phase-1 pixel detector upgrade", *Journal of instrumentation*, vol. 16, no. 02, P02027, 2021 (cit. on pp. 31, 32).

[45] CMS HCAL/ECAL Collaboration, "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c", *Eur. Phys. J. C*, vol. 60, pp. 359–373, 2009, [Erratum: Eur.Phys.J.C 61, 353–356 (2009)]. DOI: 10.1140/epjc/s10052-009-0959-5 (cit. on p. 33).

[46] M. Tosi, "The CMS trigger in Run 2", Tech. Rep., 2017 (cit. on p. 34).

[47] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, "Jet Flavour Classification Using DeepJet", *JINST*, vol. 15, no. 12, P12012, 2020. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex] (cit. on pp. 39, 61).

[48] CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV", *JINST*, vol. 13, no. 10, P10005, 2018. DOI: 10.1088/1748-0221/13/10/P10005. arXiv: 1809.02816 [hep-ex] (cit. on p. 41).

[49] CMS Collaboration, "Identification of hadronic tau lepton decays using a deep neural network", *JINST*, vol. 17, P07023, 2022. DOI: 10.1088/1748-0221/17/07/P07023. arXiv: 2201.08458 [hep-ex] (cit. on pp. 41, 42).

[50] CMS Collaboration, "GNN-based end-to-end reconstruction in the CMS Phase 2 High-Granularity Calorimeter", *J. Phys. Conf. Ser.*, vol. 2438, no. 1, p. 012 090, 2023. DOI: 10.1088/1742-6596/2438/1/012090. arXiv: 2203.01189 [physics.ins-det] (cit. on p. 41).

[51] H. Qu, C. Li, and S. Qian, "Particle Transformer for Jet Tagging", Feb. 2022. arXiv: 2202.03772 [hep-ph] (cit. on p. 41).

[52] K. Ehatäht, "Measurement of higgs boson couplings in final states featuring multiple leptons and hadronic τ decay products", Ph.D. dissertation, Tallinn University of Technology, 2023 (cit. on pp. 42, 67, 83, 88).

[53] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc., 2022 (cit. on p. 43).

[54] B. Neal *et al.*, "A modern take on the bias-variance tradeoff in neural networks", 2018. arXiv: 1810.08591 [cs.LG] (cit. on p. 44).

[55] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *et al.*, *Learning internal representations by error propagation*, 1985 (cit. on p. 45).

[56] S. Ruder, "An overview of gradient descent optimization algorithms", 2016. arXiv: 1609.04747 [cs.LG] (cit. on p. 45).

[57] Z. Zhang, "Improved adam optimizer for deep neural networks", in *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, IEEE, 2018, pp. 1–2 (cit. on p. 45).

[58] P. Moritz, R. Nishihara, and M. Jordan, "A linearly-convergent stochastic L-BFGS algorithm", in *Artificial Intelligence and Statistics*, PMLR, 2016, pp. 249–258 (cit. on p. 45).

[59] T. G. Dietterich, "Ensemble methods in machine learning", in *International workshop on multiple classifier systems*, Springer, 2000, pp. 1–15 (cit. on p. 47).

[60] L. Breiman, “Bagging predictors”, *Machine learning*, vol. 24, pp. 123–140, 1996 (cit. on p. 47).

[61] R. E. Schapire *et al.*, “A brief introduction to boosting”, in *Ijcai*, Citeseer, vol. 99, 1999, pp. 1401–1406 (cit. on p. 47).

[62] B. Pavlyshenko, “Using stacking approaches for machine learning models”, in *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, IEEE, 2018, pp. 255–258 (cit. on p. 47).

[63] R. E. Schapire, “Explaining AdaBoost”, in *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, Springer, 2013, pp. 37–52 (cit. on p. 47).

[64] J. H. Friedman, “Greedy function approximation: A gradient boosting machine”, *Annals of statistics*, pp. 1189–1232, 2001 (cit. on p. 47).

[65] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system”, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794 (cit. on pp. 47, 67, 84).

[66] J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, “Parton shower uncertainties in jet substructure analyses with deep neural networks”, *Physical Review D*, vol. 95, no. 1, p. 014 018, 2017 (cit. on p. 48).

[67] B. W. Yap, K. A. Rani, H. A. A. Rahman, S. Fong, Z. Khairudin, and N. N. Abdullah, “An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets”, in *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*, Springer, 2014, pp. 13–22 (cit. on pp. 48, 49).

[68] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, “Generative adversarial networks: An overview”, *IEEE signal processing magazine*, vol. 35, no. 1, pp. 53–65, 2018 (cit. on p. 48).

[69] C. Doersch, “Tutorial on variational autoencoders”, 2016. arXiv: 1606.05908 [stat.ML] (cit. on p. 49).

[70] J. Apostolakis *et al.*, “HEP Software Foundation Community White Paper Working Group - Detector Simulation”, V. Elvira and J. Harvey, Eds., Mar. 2018. arXiv: 1803.04165 [physics.comp-ph] (cit. on p. 49).

[71] D. Salamani, A. Zaborowska, and W. Pokorski, “MetaHEP: Meta learning for fast shower simulation of high energy physics experiments”, *Phys. Lett. B*, vol. 844, p. 138 079, 2023. DOI: 10.1016/j.physletb.2023.138079 (cit. on p. 49).

[72] D. Freer and G.-Z. Yang, “Data augmentation for self-paced motor imagery classification with C-LSTM”, *Journal of neural engineering*, vol. 17, no. 1, p. 016 041, 2020 (cit. on p. 49).

[73] *Automating Data Augmentation: Practice, Theory and New Direction*, https://ai.stanford.edu/blog/data-augmentation/, Accessed: 2023-09-26 (cit. on p. 49).

[74] T. Au, M.-L. I. Chin, and G. Ma, “Mining rare events data by sampling and boosting: A case study”, in *Information Systems, Technology and Management: 4th International Conference, ICISTM 2010, Bangkok, Thailand, March 11-13, 2010. Proceedings 4*, Springer, 2010, pp. 373–379 (cit. on p. 49).

[75] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique”, *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002 (cit. on p. 49).

[76] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning", in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, IEEE, 2008, pp. 1322–1328 (cit. on pp. 49, 50).

[77] M. Erdmann, E. Geiser, Y. Rath, and M. Rieger, "Lorentz Boost Networks: Autonomous Physics-Inspired Feature Engineering", *JINST*, vol. 14, no. 06, P06006, 2019. DOI: 10.1088/1748-0221/14/06/P06006. arXiv: 1812.09722 [hep-ex] (cit. on pp. 50, 51).

[78] B. Gregorutti, B. Michel, and P. Saint-Pierre, "Correlation and variable importance in random forests", *Statistics and Computing*, vol. 27, pp. 659–678, 2017 (cit. on pp. 51, 52, 95).

[79] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using shapley additive explanation and application for real data in hospital", *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106 584, 2022 (cit. on p. 52).

[80] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", *Advances in neural information processing systems*, vol. 30, 2017 (cit. on p. 52).

[81] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles", 2018. arXiv: 1802.03888 [cs.LG] (cit. on p. 52).

[82] J. Kennedy and R. Eberhart, "Particle swarm optimization", in *Proceedings of ICNN'95-International Conference on Neural Networks*, 1995, p. 1942 (cit. on p. 54).

[83] J. H. Holland, "Genetic algorithms", *Scientific american*, vol. 267, no. 1, p. 66, 1992 (cit. on pp. 54, 55).

[84] H. J. Kushner, "A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise", *Journal of Basic Engineering*, vol. 86, no. 1, p. 97, Mar. 1964, ISSN: 0021-9223. DOI: 10.1115/1.3653121. eprint: https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/86/1/97/5763745/97_1.pdf. [Online]. Available: https://doi.org/10.1115/1.3653121 (cit. on p. 54).

[85] A. G. Zhilinskas, "Single-step Bayesian search method for an extremum of functions of a single variable", *Cybernetics*, vol. 11, no. 1, p. 160, 1975. DOI: 10.1007/BF01069961. [Online]. Available: https://doi.org/10.1007/BF01069961 (cit. on p. 54).

[86] J. Močkus, V. Tiesis, and A. Zilinskas, "The application of Bayesian methods for seeking the extremum", *Towards Global Optimization*, vol. 2, p. 117, Sep. 2014 (cit. on p. 54).

[87] J. Močkus, "On Bayesian methods for seeking the extremum", in *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*, G. I. Marchuk, Ed., Berlin, Heidelberg: Springer, 1975, p. 400, ISBN: 978-3-540-37497-8 (cit. on p. 54).

[88] J. Močkus, *Bayesian approach to global optimization*. Springer, 1989. DOI: 10.1007/978-94-009-0909-0 (cit. on p. 54).

[89] L. Li *et al.*, "Massively parallel hyperparameter tuning", 2018 (cit. on p. 54).

[90] F. Van Den Bergh and A. P. Engelbrecht, "A study of particle swarm optimization particle trajectories", *Information sciences*, vol. 176, no. 8, p. 937, 2006 (cit. on p. 54).

[91] J. Yang and C. K. Soh, "Structural optimization by genetic algorithms with tournament selection", *Journal of Computing in Civil Engineering*, vol. 11, no. 3, p. 195, 1997 (cit. on p. 55).

[92] N. M. Razali, J. Geraghty, *et al.*, "Genetic algorithm performance with different selection strategies in solving TSP", in *Proceedings, world congress on engineering*, 2011, p. 1 (cit. on p. 55).

[93] B. L. Miller, D. E. Goldberg, *et al.*, "Genetic algorithms, tournament selection, and the effects of noise", *Complex systems*, vol. 9, no. 3, pp. 193–212, 1995 (cit. on p. 55).

[94] A. Lipowski and D. Lipowska, "Roulette-wheel selection via stochastic acceptance", *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 6, pp. 2193–2196, 2012 (cit. on p. 55).

[95] W. M. Spears and K. A. De Jong, "An analysis of multi-point crossover", in *Foundations of genetic algorithms*, vol. 1, 1991, p. 301 (cit. on p. 55).

[96] K. A. De Jong and W. M. Spears, "A formal analysis of the role of multi-point crossover in genetic algorithms", *Annals of mathematics and Artificial intelligence*, vol. 5, no. 1, p. 1, 1992 (cit. on p. 55).

[97] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions", *Journal of Global Optimization*, vol. 13, no. 4, p. 455, 1998. DOI: 10.1023/A:1008306431147. [Online]. Available: https://doi.org/10.1023/A:1008306431147 (cit. on p. 56).

[98] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms", *Advances in neural information processing systems*, vol. 25, 2012 (cit. on p. 56).

[99] Y.-W. Shang and Y.-H. Qiu, "A note on the extended Rosenbrock function", *Evolutionary Computation*, vol. 14, no. 1, p. 119, 2006 (cit. on p. 56).

[100] ATLAS collaboration, "Dataset from the ATLAS Higgs boson machine learning challenge", in *CERN Open Data Portal*, 2014 (cit. on p. 56).

[101] H. White, "A reality check for data snooping", *Econometrica*, vol. 68, no. 5, pp. 1097–1126, 2000 (cit. on p. 58).

[102] T. Fushiki, "Estimation of prediction error by using k-fold cross-validation", *Statistics and Computing*, vol. 21, pp. 137–146, 2011 (cit. on p. 58).

[103] CMS Collaboration, "Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s}$ = 13TeV", *Phys. Lett. B*, vol. 778, pp. 101–127, 2018. DOI: 10.1016/j.physletb.2018.01.001. arXiv: 1707.02909 [hep-ex] (cit. on p. 59).

[104] CMS Collaboration, "Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s}$ =13 TeV", *Phys. Lett. B*, vol. 842, p. 137 531, 2023. DOI: 10.1016/j.physletb.2022.137531. arXiv: 2206.09401 [hep-ex] (cit. on p. 59).

[105] CMS Collaboration, "Search for nonresonant higgs boson pair production in the $b\bar{b}b\bar{b}$ final state at $\sqrt{s}$ = 13 tev", *Journal of High Energy Physics*, vol. 2019, no. 4, p. 112, 2019 (cit. on p. 59).

[106] CMS Collaboration, "Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at s=13 TeV", *Phys. Rev. Lett.*, vol. 129, no. 8, p. 081 802, 2022. DOI: 10.1103/PhysRevLett.129.081802. arXiv: 2202.09617 [hep-ex] (cit. on p. 59).

[107] CMS Collaboration, "Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s}$ = 13 TeV", *JHEP*, vol. 03, p. 257, 2021. DOI: 10.1007/JHEP03(2021)257. arXiv: 2011.12373 [hep-ex] (cit. on p. 59).

[108] CMS Collaboration, "Search for a new resonance decaying into two spin-0 bosons in a final state with two photons and two bottom quarks in proton-proton collisions at $\sqrt{s}$ = 13 TeV", Oct. 2023. arXiv: 2310.01643 [hep-ex] (cit. on p. 59).

[109] CMS Collaboration, "Search for nonresonant Higgs boson pair production in the four leptons plus twob jets final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV", *JHEP*, vol. 06, p. 130, 2023. DOI: 10.1007/JHEP06(2023)130. arXiv: 2206.10657 [hep-ex] (cit. on p. 59).

[110] CMS Collaboration, "Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C*, vol. 81, no. 4, p. 378, 2021. DOI: 10.1140/epjc/s10052-021-09014-x. arXiv: 2011.03652 [hep-ex] (cit. on p. 59).

[111] S. Agostinelli *et al.*, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 63).

[112] G. Buchalla, M. Capozi, A. Celis, G. Heinrich, and L. Scyboz, "Higgs boson pair production in non-linear Effective Field Theory with full m_t-dependence at NLO QCD", *JHEP*, vol. 09, p. 057, 2018. DOI: 10.1007/JHEP09(2018)057. arXiv: 1806.05162 [hep-ph] (cit. on pp. 64, 65).

[113] A. Carvalho, F. Goertz, K. Mimasu, M. Gouzevitch, and A. Aggarwal, "On the reinterpretation of non-resonant searches for Higgs boson pairs", *JHEP*, vol. 02, p. 049, 2021. DOI: 10.1007/JHEP02(2021)049. arXiv: 1710.08261 [hep-ph] (cit. on p. 65).

[114] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, "Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques", *J. Phys. Conf. Ser.*, vol. 513, D. L. Groep and D. Bonacorsi, Eds., p. 022 035, 2014. DOI: 10.1088/1742-6596/513/2/022035 (cit. on p. 68).

[115] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C*, vol. 71, p. 1554, 2011, [Erratum: Eur.Phys.J.C 73, 2501 (2013)]. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an] (cit. on pp. 87, 88).

[116] *Custom HH physics model implementing gluon gluon fusion (ggf / gghh), vector boson fusion (vbf / qqhh) and V boson associated production (vhh) modes*. https://gitlab.cern.ch/hh/tools/inference/-/blob/master/dhi/models/hh_model.py, Accessed: 2023-10-17, commit-ID: c79f9714 (cit. on p. 88).

[117] *Prospects for Di-Higgs boson searches with multilepton final states in the qqHH production mode*, https://indico.cern.ch/event/1147717/contributions/5082189/attachments/2525715/4344013/CBC22.pdf, Accessed: 2023-10-17 (cit. on p. 93).

[118] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection”, Aug. 2017. arXiv: 1708.02002 [cs.CV] (cit. on p. 95).

List of Figures

List of Tables

Appendix 1

I

L. Tani *et al.*, “Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics”, *Eur. Phys. J. C*, vol. 81, no. 2, p. 170, 2021. DOI: 10.1140/epjc/s10052-021-08950-y. arXiv: 2011.04434 [hep-ex]

Eur. Phys. J. C (2021) 81:170
https://doi.org/10.1140/epjc/s10052-021-08950-y

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics

Laurits Tani[a], **Diana Rand**[b], **Christian Veelken**[c], **Mario Kadastik**[d]

National Institute of Chemical Physics and Biophysics (NICPB), Akadeemia tee 23, 12618 Tallinn, Estonia

Received: 10 November 2020 / Accepted: 2 February 2021 / Published online: 19 February 2021

Abstract The analysis of vast amounts of data constitutes a major challenge in modern high energy physics experiments. Machine learning (ML) methods, typically trained on simulated data, are often employed to facilitate this task. Several choices need to be made by the user when training the ML algorithm. In addition to deciding which ML algorithm to use and choosing suitable observables as inputs, users typically need to choose among a plethora of algorithm-specific parameters. We refer to parameters that need to be chosen by the user as hyperparameters. These are to be distinguished from parameters that the ML algorithm learns autonomously during the training, without intervention by the user. The choice of hyperparameters is conventionally done manually by the user and often has a significant impact on the performance of the ML algorithm. In this paper, we explore two evolutionary algorithms: particle swarm optimization and genetic algorithm, for the purposes of performing the choice of optimal hyperparameter values in an autonomous manner. Both of these algorithms will be tested on different datasets and compared to alternative methods.

1 Introduction

Owing to the large amount of data recorded by contemporary high energy physics (HEP) experiments, the analysis of data relies on powerful computing facilities. Machine learning (ML) methods are used extensively to aid the data analysis [1,2]. Boosted decision trees (BDTs) [3] and artificial neural networks (ANNs) [4] are commonly used in HEP experiments. Even though these methods may aid the data analysis task significantly, their usage in practical HEP applications is not trivial. This is because, in order to achieve optimal results, a set of parameters, referred to as hyperparameters in the literature [5], need to be chosen by the user, depending on the given task and data.

The subject of this paper is to describe two evolutionary algorithms [6], which allow to find a set of optimal hyperparameters in an autonomous manner. The evolutionary algorithms studied in this paper are particle swarm optimization (PSO) [7] and genetic algorithm (GA) [8].

The task of finding optimal hyperparameter values can be recast as function maximization. One considers a mapping from a point h in hyperparameter space $\mathcal{H}$ to a "score" value $s(h)$, which quantifies the performance of the ML algorithm for a given task. Using a suitable encoding for hyperparameters of non-floating-point type, the hyperparameter space $\mathcal{H}$ can be taken to be the Euclidean space $\mathbb{R}^{\mathrm{N}}$, with N denoting the number of hyperparameters. Formally, the optimal hyperparameters, denoted by the symbol $\hat{h}$, are those that satisfy the condition:

$$\hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmax}}\; s(h)\,, \tag{1}$$

where $s : \mathcal{H} \mapsto \mathbb{R}$ refers to the objective function that maps a point h in $\mathcal{H}$ to a score $s(h)$. Recasting the hyperparameter optimization task as a function maximization problem allows to evaluate the performance of the PSO and GA on reference problems on function maximization from literature, as well as to compare their performance with alternative methods.

The paper is organized as follows: in Sects. 2 and 3, we describe the PSO and GA, respectively. In Sect. 4, we apply both evolutionary algorithms to a well-known function minimization problem from the literature, based on the Rosenbrock function [9], as well as to a typical data analysis task from the domain of HEP, the "ATLAS Higgs boson machine learning challenge" [10]. We conclude the paper with a summary in Sect. 5.

[a] e-mail: laurits.tani@cern.ch (corresponding author)
[b] e-mail: diana.rand@cern.ch
[c] e-mail: christian.veelken@cern.ch
[d] e-mail: mario.kadastik@cern.ch

2 Particle swarm optimization

Particle swarm optimization (PSO) [7] represents a computational method for optimizing continuous nonlinear functions. The method is effective for optimizing a wide range of functions. In common with other evolutionary algorithms, such as the GA, the PSO method is inspired by nature.

As the name of the method implies, the maximization of the objective function by the PSO is performed by a *swarm of particles*. The particles traverse the hyperparameter space $\mathcal{H}$, with the position of each particle representing one set of hyperparameters h. Having a swarm of particles allows the exploration of multiple points in the space $\mathcal{H}$ in parallel, thereby allowing for a highly parallel implementation of the PSO algorithm on a computer. The evolution of the particle swarm proceeds in iterations denoted by the letter k. In each iteration a new position $\mathbf{x}_i^{k+1}$ is computed for each particle i according to the relation:

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + w \cdot \mathbf{p}_i^k + \mathbf{F}_i^k \tag{2}$$

where $\mathbf{x}_i^k$ denotes the current position of the particle and $\mathbf{p}_i^k$ its *momentum*. The momentum term $w \cdot \mathbf{p}_i^k$ represents the inertia for particles to change their direction when traversing the space $\mathcal{H}$. The symbol $\mathbf{F}_i^k$ represents an attractive *force*, which has the effect for particles to move towards previously discovered extrema of the objective function. The momentum term causes a tendency for the particles to continue moving in their current direction, past the previously found extrema. This behaviour increases the exploration of the hyperparameter space $\mathcal{H}$ and is found to improve performance [7]. The coefficient w is referred to as *inertial weight* in the literature [11], though the term *damping weight* might be actually more descriptive as suggested in reference [12].

Our implementation of the PSO algorithm distinguishes between the personal best location $\hat{x}_i^k = \{x \in \mathcal{H} \wedge \hat{x}_i^k = x_i^{k'}$ for $k' \leq k \wedge s(x_i^k) \leq s(\hat{x}_i^k) \forall k' \leq k\}$ and the best known global extremum $\hat{\hat{x}}^k = \mathrm{argmax}\{\hat{x}_i^k\}$:

$$\mathbf{F}_i^k = c_1 \cdot r_1 \cdot (\hat{\mathbf{x}}_i^k - \mathbf{x}_i^k) + c_2 \cdot r_2 \cdot (\hat{\hat{\mathbf{x}}}^k - \mathbf{x}_i^k). \tag{3}$$

The coefficients c_1 and c_2 are referred to as the *cognitive* and the *social* weights in the literature [13], and the symbols r_1 and r_2 represent random numbers, which are drawn from an uniform distribution in the interval [0, 1]. The known global extremum $\hat{\hat{x}}^k$ for each particle is updated at each iteration by propagating the personal best location of a subset of the population, referred to as *info*. The number of particles in this subset is denoted by N_{info}. Restricting the computation of $\hat{\hat{x}}^k$ to a subset of particles helps to avoid premature convergence of the swarm to a local minimum.

We choose the coefficients c_1 and c_2 to be equal to 2, such that the particles move past their previously found target about half of the time if the inertial weight w would be negligible [7].

After each iteration the momenta are updated according to the rule:

$$\mathbf{p}_i^{k+1} = \mathbf{x}_i^{k+1} - \mathbf{x}_i^k \tag{4}$$

The positions $\mathbf{x}_i^0$ of all particles i are initialized randomly within the hyperparameter space $\mathcal{H}$, while all momenta $\mathbf{p}_i^0$ are randomly initialized within one quarter of the range of each hyperparameter.

The relation between the inertial weight w and the size of the coefficients c_1 and c_2 controls the influence of global (wide-ranging) versus local (nearby) exploration abilities of the particles. A larger inertial weight w allows the particles to move into unexplored regions of the hyperparameter space $\mathcal{H}$, whereas a small value of w causes the particle to hone in on local and global extrema found previously [11].

A suitable selection of w can provide a balance between global and local exploration abilities and thus require fewer iterations on average to find the optimum [11]. As discussed in Ref. [12], one may expect the performance of the PSO algorithm to be improved if one sets the inertial weight w to a large value for the first iterations of the PSO algorithm and gradually reduces w as the swarm evolves. Doing this allows the particles to explore the hyperparameter space $\mathcal{H}$ as fast as possible. By gradually reducing the value of w during subsequent iterations, when the approximate location of the extremum has been established, one switches smoothly from the global exploration to a local exploration, thus improving on the accuracy of the found extrema. The idea is analogous to the gradual reduction of the temperature parameter in simulated annealing [12].

Each time the position of a particle would move outside the bounds of the hyperparameter space $\mathcal{H}$, the position of the particle is set to the boundary value and its momentum is set to zero, thereby reducing the probability that the same particle moves against the boundary again in the next iteration.

3 Genetic algorithm

The second evolutionary algorithm considered in this paper, the genetic algorithm (GA), is motivated by the concept of natural selection [8]. The GA maintains a population of possible solutions to the optimization problem, which evolve through multiple generations in order to produce the best solution.

Each possible solution is referred to as a *chromosome*. Each chromosome (see Fig. 1) represents one point in the

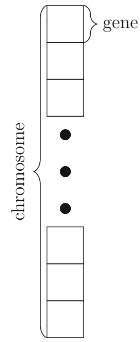


Fig. 1 A chromosome consisting of genes

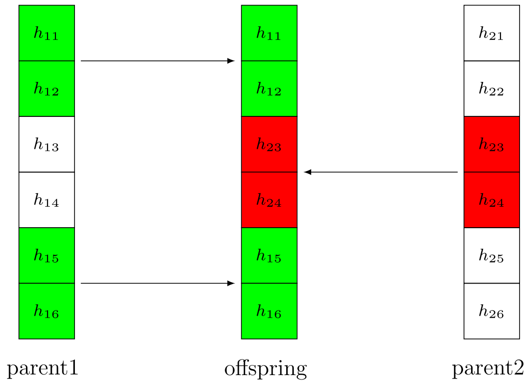


Fig. 2 Possible outcome of a 2-point crossover of two parents in case of 6-dimensional hyperparameter space $\mathcal{H}$, where h_{11} denotes the first hyperparameter of the first parent, h_{12} the second hyperparameter of the first parent, etc

hyperparameter space $\mathcal{H}$. Having multiple chromosomes allows the GA to explore multiple solutions in parallel.

The number of genes in a chromosome matches the dimension of the hyperparameter space $\mathcal{H}$.

The evolution towards the best solution is iterative. Each iteration corresponds to one generation in the evolution of all chromosomes and consists of 3 distinct stages: the selection of parents, the crossover of the genes, and the mutation.

The selection of parents is performed using the tournament method [14,15]. In each tournament a certain number of chromosomes compete to be selected as a parent for the next generation. The number of chromosomes participating in each tournament is denoted by the symbol N_{tour}. The participants are drawn from the population of chromosomes at random and are ranked in order of decreasing score $s(h)$. The participant with the highest score is selected as a parent with the probability P_{tour}. In case the chromosome with the highest score is not selected, the chromosome with the second highest score gets selected, again with the probability P_{tour}, and so on. The tournament ends when two chromosomes are selected in this way to be the parents.

A larger value of N_{tour} has the effect that the chromosome with a low score $s(h)$ has a smaller chance to be selected as the parent for the next generation, because there is a high probability that a chromosome with a better score participates in the same tournament. A smaller value of N_{tour} has the opposite effect. New tournaments are started until a sufficient number of pairs of parents are selected to produce the chromosomes for the next generation.

The chromosomes of two parents produce one new chromosome for the next generation by means of crossover [16,17]. We use *k-point crossover* in which the chromosomes of both parents are cut at k points (N_{cross} refers to the number of points, to avoid using the same symbol as for the number of iterations) and the chromosomes of the offspring are produced by randomly choosing chromosome segments from either parent (see Fig. 2).

The chromosomes of the offspring that are obtained by the crossover operation are subject to *mutation* [8], which aims to increase the diversity of the population, thereby allowing to explore domains in the hyperparameter space $\mathcal{H}$ not populated by chromosomes from the parent generation. Mutation also helps to avoid the population to get stuck in local minima.

In our implementation of the GA, the mutation of chromosomes is performed by adding a random number, drawn from a normal distribution with a mean of zero and a given width, to each gene. A high mutation rate has the effect of turning the GA into a random search. We avoid this effect by linearly decreasing the width of the normal distribution each iteration, with the initial width corresponding to a quarter of the maximum range of a given hyperparameter and the final width corresponding to zero.

Our implementation of the GA uses the concept of *elitism* [18]. Elitism means that the algorithm preserves a certain number of the best performing chromosomes within the population and passing the parent chromosomes on to the next generation together with their offspring. Elitism is found to improve the convergence toward an optimal solution. The number of parent chromosomes preserved in this manner is denoted by the symbol N_{elite}.

The convergence is further enhanced by *culling* [19], which means that we discard a certain number of chromosomes with the lowest score among the population before selecting the parents for the next generation. The number of parent chromosomes discarded in this way is referred to using the symbol N_{cull}. For each chromosome discarded by culling, we create a new chromosome with randomly initialized hyperparameter values to replace the one discarded.

Our implementation of the GA further allows to evolve groups of chromosomes in *subpopulations* [20]. The number of subpopulations is denoted by the symbol N_{subpop}. The selection of parents is restricted to the chromosomes from

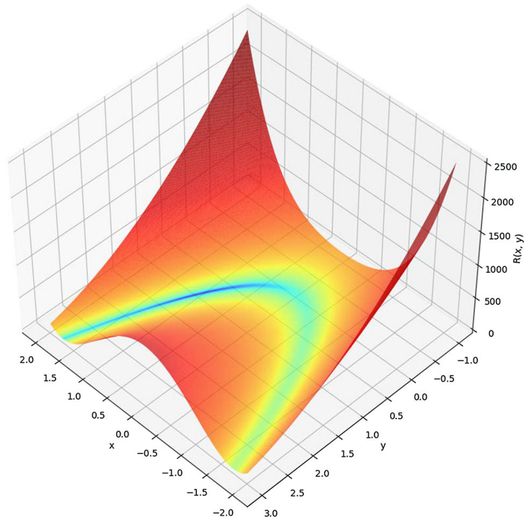


Fig. 3 The Rosenbrock function in a region around its global minimum, located at the position $(x, y) = (1, 1)$

the same subpopulation for the first $N_{subpop}^{generations}$ iterations of the algorithm. For the remaining iterations, the chromosomes from different subpopulations are allowed to mix freely.

4 Performance

The performance of both evolutionary algorithms, PSO and GA, is evaluated on two tasks: on the Rosenbrock function, which provides an example for a difficult function minimization problem, and on the ATLAS Higgs boson machine learning (ML) challenge, as a typical application of ML methods in HEP.

4.1 Rosenbrock function

The Rosenbrock function [9,21] represents a well-known trial function for evaluating the performance of function minimization algorithms. The function is defined as:

$$R(x, y) = (a - x)^2 + b(y - x^2)^2, \tag{5}$$

where the a and b are constants.

The Rosenbrock function has a global minimum at $(x, y) = (a, a^2)$. We chose to study the Rosenbrock function for the case $a = 1$ and $b = 100$. For the chosen values of a and b, the global minimum is located at the position $(x, y) = (1, 1)$, and the function value at the minimum is $R(1, 1) = 0$.

The challenge in finding the global minimum of the Rosenbrock function is that the function varies slowly along a curved valley, while rising steeply in direction orthogonal to the valley. Function minimization algorithms hence need to closely track the location of the valley. Figure 3 illustrates the Rosenbrock function in the region around the global minimum.

For the purpose of evaluating the performance of the PSO and GA, we treat the minimization of $R(x, y)$ as function of x and y as a two dimensional hyperparameter optimization problem, identifying x and y with the first and second hyperparameter respectively. The position of particles in case of the PSO algorithm and the value of the chromosomes in the case of GA are initialized within the range $[-500, +500] \times [-500, +500]$ and are enforced to stay within this range during the evolution of both algorithms.

4.1.1 Stopping criteria

In order to limit the computing time, we define a criterion when to stop the training of the PSO and GA. We use two criteria for this purpose and terminate the evolution when either criterion is fulfilled. The first criterion is an upper limit on the number of iterations, denoted by the symbol N_{iter}^{max}. Additionally, we terminate the evolution once the algorithm has found a point (x, y) for which $R(x, y) < 10^{-3}$.

4.1.2 Optimization methods

We compare the performance of the PSO and GA for finding the minimum of the Rosenbrock function with three alternative methods, the gradient descent algorithm [22], and two naive methods for choosing the hyperparameters, to which we refer to as "grid search" and "random guessing". The latter two serve as a cross-check. One would expect of course that evolutionary algorithms such as the PSO and GA outperform the naive methods.

(Modified) gradient descent We have modified the gradient descent (GD) algorithm in order to improve its performance on the Rosenbrock function. The issue is that the unmodified GD algorithm often 'zig-zags' from one side of the valley to the other, causing the algorithm to progress very slowly in the direction along the valley, towards the global minimum [23]. To prevent this 'zig-zag' behaviour and improve the convergence of the algorithm, we have modified the GD algorithm in the following way: at each iteration, the algorithm determines the direction of the steepest descent by numerical evaluation of the gradient at a given point h^k. The new position h^{k+1} is computed according to:

$$h^{k+1} = h^k + \delta \cdot \frac{\nabla h^k}{|\nabla h^k|}, \tag{6}$$

where the term $\frac{\nabla h^k}{|\nabla h^k|}$ represents the direction of the steepest descent and the step size δ represents the parameter of the algorithm.

Rather than moving immediately to the new position h^{k+1}, the modified GD algorithm computes the value of the objective function s at the new position $s(h^{k+1})$.

It then compares the actual decrease of the objective function, $s(h^{k+1}) - s(h^k)$, with the expected decrease, given the expression $\delta \cdot \frac{\nabla h^k}{|\nabla h^k|} \cdot \nabla h^k$. In case $s(h^{k+1}) - s(h^k) > 2 \cdot \delta \cdot \frac{\nabla h^k}{|\nabla h^k|} \cdot \nabla h^k$, we conclude that the step size δ is too large and needs to be reduced in order to avoid this 'zig-zag' behaviour.

In our implementation, we successively reduce the step size by a factor of two until the condition is satisfied. The algorithm then moves to the new position, the initial step size is restored, and the algorithm recomputes the gradient at the new position for the next iteration.

We choose the number of iterations for the GD algorithm to be 10^6 and the initial step size δ to be 10^{-2}.

Grid search This is a widely used hyperparameter optimization method available for example in the package scikit-learn [24]. This method is based on choosing N^d grid points in each dimension d of the hyperparameter space $\mathcal{H}$, evaluating the objective function s for all $\prod_{d=1}^{N} N^d_{grid}$ combinations of grid points, and selecting the best performing combination. The same number of evaluations of the objective function, $\prod_{d=1}^{N} N^d_{grid}$, is chosen to be the same as for the other algorithms, in order to compare all algorithms for the same time usage of computing time. Here we assume that the evaluation of the objective function consumes the majority of the computing time and the computations internal to the PSO and GA are negligible in comparison. We believe this assumption represents a very good approximation for practical approximations of these methods in HEP, discussed in the introduction, where one evaluation of s corresponds to one training of a ML algorithm. For the Rosenbrock function minimization task, we choose $N^1_{grid} = N^2_{grid} = 10^3$ grid points for each of the two dimensions, equidistantly within the interval [-500, +500] in each dimension.

Random Guessing In the random guessing (RNG) method, we draw a total of $N_p = 10^6$ points in the hyperparameter space $\mathcal{H}$ at random, sampling from a uniform distribution within the range $[-500, +500] \times [-500, +500]$. The point corresponding to the minimum of the objective function s over the set of these points is selected as the best-performing point of the RNG method. The number of points N_p is chosen such that the function is evaluated the same number of times for the RNG method as for the PSO, GA, GD and GS methods.

Table 1 Parameters of the GA used for the Rosenbrock function minimization task

Parameter	Value
N_{tour}	5
P_{tour}	0.4
N_{cross}	1
P_{mutate}	0.2
$N^{generations}_{subpop}$	90
N_{subpop}	5
N_{cull}	50
N_{elite}	25

Particle swarm optimization The same maximal number of 10^6 evaluations of the objective function s were used for the PSO, by setting the number of particles in the swarm to 100 and the maximum number of iterations to 10^4. The evaluation of the PSO was terminated before reaching the maximum number of iterations in case the global minimum $s(\hat{\hat{x}}^k)$ found by the PSO differed from the global minimum of the Rosenbrock function by less than 10^{-3}. The coefficients c_1 and c_2 were chosen to be 2 and the inertial weight w was chosen to linearly decrease from 0.8 to 0.4 as a function of iteration k. The number of informants N_{info} was set to 7.

Genetic Algorithm The same maximum number of 10^6 evaluations of the objective function were used for the GA, for which the number of chromosomes was chosen to be 10^4 and the maximum number of iterations to be 100. The same threshold for early termination of 10^{-3} was chosen for the GA, as for the PSO. The early termination triggers once $s(h) < 10^{-3}$ for the hyperparameter values h represented by any chromosome . The values of other parameters of the GA, used for the Rosenbrock function minimization task, are given in Table 1. During the first iterations of the algorithm, when subpopulations are used, the parameters N_{cull} and N_{elites} amount to 10 and 5 respectively.

4.1.3 Procedure for comparing different methods

Owing to the fact that the minima found by the GD, GS and RNG, PSO, GA methods depends on the values of random numbers that are used to initialize and/or evolve each algorithm, the performance of each method needs to be evaluated for a set of different 'trials', each trial using a different seed to produce a different sequence of random numbers.

4.1.4 Results

The distribution in $\hat{\hat{R}} = R(\hat{\hat{h}})$ at the minima $\hat{\hat{h}}$ found in 100 different trials is shown in Fig. 4. Numerical values of the

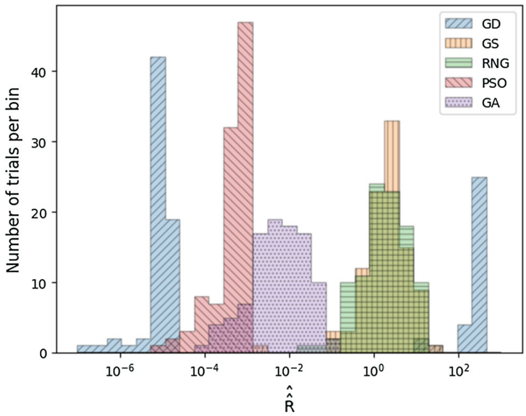


Fig. 4 Distribution in $\hat{\hat{R}} = R(\hat{\hat{h}})$ of the Rosenbrock function at the minimum $\hat{\hat{h}}$ found in 100 different trials for the GD, GS, RNG, PSO and GA

Table 2 Average value $\bar{R}$ and standard deviation σ_R achieved by the GD, GS, RNG, PSO, and GA methods in the Rosenbrock function minimization task

Method	$\bar{R}$	σ_R
GD	85.85	143.7
GS	3.29	3.89
RNG	3.11	3.44
PSO	0.00057	0.00030
GA	0.0014	0.0021

average $\bar{R}$ and of the width of the distribution, quantified by the standard deviation $\sigma_R = \sqrt{\frac{1}{99}\sum s}$, are shown in Table 2.
Discussion One can see in Fig. 4 that the GD method performs extraordinarily well in about half of the trials, while in the other half it fails to get close to the minimum of the Rosenbrock function at all. The poor performance of the GD method in the latter trials is due to the cases where the particle moves so slowly along the valley of the Rosenbrock function that the maximum number of 10^6 iterations is reached before the algorithm reaches the global minimum of the position $(x, y) = (1, 1)$.

The PSO algorithm achieves the lowest value $\bar{R}$, outperforming all other methods on the Rosenbrock function minimization task, followed by the GA. The PSO and GA also exhibit the lowest standard deviation σ_R, which means that their performance is robust against variations in the random choice of starting positions across different trials.

We remark that the early termination limited the average number of evaluations of the objective function to $\sim 7 \cdot 10^3$ for the PSO, while the early termination had little effect for the GA (as well as for the GD, GS, and RNG methods), which makes the performance of the PSO even more impressive. As expected, both evolutionary algorithms outperform all other methods.

4.2 The ATLAS Higgs boson machine learning challenge

The ATLAS Higgs boson machine learning challenge (HBC) [10] represents a typical application of ML algorithms to the field of HEP. The task of the HBC is to obtain an optimal separation of the Standard Model (SM) Higgs boson $\rightarrow \tau\tau$ signal from the large SM background. The background consist of Drell–Yan production of Z bosons, the production of W bosons in association with jets, and top quark pair production. Samples of signal and background events are generated by Monte Carlo (MC) simulation. Events are selected in the $\tau\tau \rightarrow \mu\bar{\nu}_\mu\nu_\tau\tau_h\nu$ final state, where we use the symbol τ_h to denote the hadronic decay of a τ lepton. Background contributions arising from multijet production without associated production of bosons or top quark are neglected.

In total 550,000 signal plus background events are provided by the organizers of the HBC, of which we use 80% for training the ML algorithm and 20% for testing the performance of the trained ML algorithm. We refer to the former as the train samples and to the latter as the test sample.

We utilize a BDT to perform the separation of the Higgs boson signal from backgrounds. For the BDT implementation, we chose the XGBoost package [25].

The objective function s for the hyperparameter optimization represents an approximation for the sensitivity to discover the Higgs boson signal in a physics analysis at the Large Hadron Collider (LHC). The function s was given by the organizers of the HBC and is referred to as the 'approximate mean significance' (AMS), which is defined by:

$$AMS(\theta_{cut}) = \sqrt{2 \cdot (s + b + b_r) \cdot ln\left[1 + \frac{s}{b + b_r}\right] - s}\ , \tag{7}$$

where b denotes the amount of background and s the amount of signal that passes a cut on the BDT output. The term b_r is introduced as a regularization in order to reduce the effect of statistical fluctuations of b and s, resulting from limited MC statistics (as discussed in Ref. [10]). The value of b_r was given by the organizers of the HBC and amounts to $b_r = 10$. The function $AMS(\theta_{cut})$, for $\theta_{cut} = 0.15$, is used as objective function for the BDT training.

Even with the addition of the b_r term, statistical fluctuations of the number of signal and background events passing the cut on the BDT output still causes a sizable difference between the AMS scores computed on the test and on the training sample. We find that the difference between test and

Table 3 Parameters of the GA used for the ATLAS Higgs boson machine learning challenge

Parameter	Value
N_{tour}	5
P_{tour}	0.4
N_{cross}	1
P_{mutate}	0.2
$N_{subpop}^{generations}$	90
N_{subpop}	5
N_{cull}	40
N_{elite}	7

Table 4 Default values of hyperparameters in the XGBoost package [26]

Parameter	Default value
Num-boost-round	10
Learning-rate	0.3
Max-depth	6
Gamma	0
Min-child-weight	1.0
Subsample	1.0
Colsample-bytree	1.0

training performance can be reduced and a higher AMS score on the test sample can be achieved if we use a modified version of Eq. (7) as the objective function for the BDT training. We refer to the modified version of Eq. (7) as d-AMS. The idea is to add a penalty term for the difference between the AMS scores on the test compared to the training sample, so that the BDT training (and the hyperparameter optimization) reduces this difference:

$$d\text{-}AMS = AMS_{test} - \kappa \cdot \max(0, [AMS_{test} - AMS_{train}]) \tag{8}$$

where the coefficient κ controls the strength of the penalty term. We find the choice $\kappa = 1.5$ to work well for a wide range of different ML applications that we tried. After the BDT training with a fixed $\theta_{cut} = 0.15$ has finished, the threshold θ_{cut} is optimized such that d-AMS attains its maximal value on the training sample.

The PSO was evolved for a maximum of 7000 evaluations of the objective function, using a swarm of 70 particles and a maximum number of 100 iterations. The coefficients c_1 and c_2 were both chosen to be equal to 2 and the inertial weight w was chosen to linearly decrease from 0.8 to 0.4 during the evolution of the PSO algorithm.

For the GA, we used 70 chromosomes and a maximum number of 100 iterations. The values of the other parameters are given in Table 3.

The XGBoost hyperparameters chosen and the default values for these parameters are given in Table 4. The parameter *num-boost-round* specifies the number of boosting iterations, corresponding to the number of trees in the BDT. The *learning-rate* parameter controls the effect that trees added at a later stage of the boosted iterations have on the output of the BDT relative to the effect of trees added at an earlier stage. Small values of the *learning-rate* parameter decrease the effect of trees added during the boosting iterations, thereby reducing the effect of boosting on the BDT output. The parameter *max-depth* specifies the maximum depth of a tree. The parameter *gamma* represents a regularization parameter, which aims to reduce overfitting. Large values of this parameter prevent the splitting of leaf nodes before the maximum depth of a tree is reached. The parameter *min-child-weight* specifies the minimum number of events that is required in each leaf node. The parameter *subsample* limits the number of training events that are used to grow each tree to a fraction of the full training sample. A value of this parameter smaller than one decreases overfitting. The parameter *colsample-bytree* specifies the number of different features that are used in a tree. A value of one means that all features are considered for splitting leaf nodes, while a value smaller than one restricts the number of features that are used in a tree to a subset of all features. The purpose of this restriction is to reduce overfitting. The number of features considered for each tree are drawn at random, independently for each boosting iteration.

The choice of all of these parameters typically represents a trade-off. Large values of the parameters *num-boost-round*, *learning-rate*, *max-depth*, *subsample*, and *colsample-bytree* increase the complexity of the BDT, while large values of the parameters *gamma*, and *min-child-weight* have a regularizing effect. BDTs with a higher complexity in general perform better in separating signal from background on the training sample, but typically are also more susceptible to overfitting.

The performance of the PSO and GA is assessed by comparing the AMS scores achieved on the test sample by a BDT trained with the default hyperparameters and with hyperparameters obtained by the RNG method compared to the AMS scores of BDTs trained with optimized hyperparameter values found by the PSO and GA.

Two criteria are used to stop the evolution of the PSO and GA. The first criterion is the number of iterations N_{iter}. Additionally, we terminate the evolution once the variance between the positions of the particles in the PSO or between chromosomes in the GA is below a certain threshold. The variance is quantified by the *compactness* (also known as the mean coefficient of variance), which is defined as:

Table 5 Hyperparameter values obtained by the RNG, PSO and the GA for the ATLAS Higgs boson machine learning challenge

Parameter	RNG	PSO	GA
Num-boost-round	295	153	451
Learning-rate	0.062	0.300	0.085
Max-depth	5	4	5
Gamma	0.98	3.86	2.99
Min-child-weight	173	323.6	442.2
Subsample	0.83	0.830	0.907
Colsample-bytree	0.7	1.0	0.3

Table 6 Performance of BDTs trained using the optimal values of the hyperparameters obtained by the PSO and by the GA compared to a BDT trained using the default values of hyperparameters in the XGBoost package [25] and with hyperparameters obtained by RNG, for the ATLAS Higgs boson machine learning challenge

Method	θ_{cut}	AMS score Public leaderboard	AMS score Private leaderboard
Default	0.175	3.170	3.200
RNG	0.152	3.620	3.608
PSO	0.134	3.628	3.655
GA	0.152	3.619	3.683

$$\text{compactness} = \frac{1}{N}\sum_{j=1}^{N}\frac{\sigma^j}{\bar{x}^j} \tag{9}$$

with

$$\sigma^j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i^j - \bar{x}^j)^2},$$

where N denotes the number of hyperparameters, n the number of particles or chromosomes, and $\bar{x}^j$ the mean value of the j-th hyperparameter over the population of particles or genes, respectively.

A low value of the compactness means that the hyperparameters of different particles or genes are very similar, indicating that the PSO or GA has converged to a single point in the hyperparameter space $\mathcal{H}$.

4.3 Results

The optimal values of the hyperparameters obtained with the RNG method, the PSO and the GA are given in Table 5. In Table 6 we compare the AMS scores obtained for these hyperparameter values to the AMS scores obtained with the default values of hyperparameters defined in the XGBoost package. The performance is evaluated for two samples of events, referred to as the *public* and *private* leaderboard samples. Both samples are provided by the organizers of the HBC and contain signal and background events that overlap with neither the test nor the train sample.

The performance achieved by the PSO and GA are very similar and about 12–13% higher than the performance obtained using the default values of hyperparameters. The results of the BDT trained using the hyperparameters obtained by the RNG method are similar to those obtained by the PSO and GA. Comparing the PSO and GA optimized hyperparameters, we find that all except *num-boost-round* and *learning-rate* parameters have similar values. The value of the *num-boost-round* parameter optimized by the GA is higher by about a factor of 2.9, while the value of the *learning-rate* parameter is lower by a factor of 3.5. The fact that the *learning-rate* parameter decreases by a factor that is similar to the increase of the *num-boost-round* parameter is not surprising: using a large number of trees and a lower learning rate has about the same effect as using a lower number of trees and a higher learning rate. The product of the *num-boost-round* and *learning-rate* parameters is more similar, differing only by a factor of 1.2 between the PSO and GA. The situation is different for the *colsample-bytree* parameter. It has a small effect on the d-AMS and AMS scores and is hence only loosely constrained during the hyperparameter optimization.

The parameters obtained by the RNG method are more different - only the *max-depth* parameter is very similar to those found by PSO and GA. Having *min-child-weight* roughly two times smaller than it was found for PSO and GA means making the model more prone to overfitting. However, this effect is overcome by having the product of the anti-correlated pair, *num-boost-round* and *learning-rate*, two times smaller from the ones obtained by PSO and GA, thus the model less susceptible to overfitting. Furthermore having three to four times smaller *gamma* helps the model generalize even further. Again the effect of *colsample-bytree* had negligible effect on the d-AMS and AMS scores.

5 Summary

Two evolutionary algorithms, the particle swarm optimization (PSO) and the genetic algorithm (GA), for choosing an optimal set of hyperparameters in applications of machine learning (ML) methods to data analyses in high energy physics (HEP) have been presented. The performance of both methods have been studied for a difficult function minimization task (Rosenbrock function) and for a typical data analysis test in the field of HEP [Higgs boson machine learning challenge (HBC)]. In the latter case, a boosted decision tree (BDT) has been used as ML algorithm. The PSO as well as the GA demonstrate their ability to find the optimal parameter value in the function minimization task. Compared to using the default values of hyperparameters, the optimization of the hyperparameter values improves the sensitivity of the data analysis, as quantified by the AMS score, by 12–13%. This improvement demonstrates that the optimization of hyperparameters is a worthwhile task for data analyses in the field of HEP. Randomly probing different hyperparameter sets and subsequently picking the best performing one showed similar performance to both PSO and GA. This can be attributed to the highly fluctuating hyperparameter space of this particular example. For a highly structured hyperparameter space, the gain of using a more sophisticated method, like PSO or GA, will be much higher, as was shown by the Rosenbrock minimization problem. The optimization of the hyperparameters by the PSO and GA is fully automated and relieve the user from manual tuning of the hyperparameters.

Acknowledgements This work was supported by the Estonian Research Council grants PRG445 and PRG780. Also, we would like to thank Joosep Pata for constructive criticism of the manuscript.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: Data used for the Higgs Boson Machine Learning Challenge is available made available by the organizers of that competition at http://opendata.cern.ch/record/328. The competition itself can be found at https://www.kaggle.com/c/higgs-boson.]

Funded by SCOAP^3.

References

1. K. Albertsson, P. Altoe, D. Anderson, J. Anderson, M. Andrews, J.P.A. Espinosa, A. Aurisano, L. Basara, A. Bevan, W. Bhimji, et al., arXiv:1807.02876 (2018)
2. S. Whiteson, D. Whiteson, Eng. Appl. Artif. Intell. **22**(8), 1203 (2009)
3. B.P. Roe, H.J. Yang, J. Zhu, Y. Liu, I. Stancu, G. McGregor, Nucl. Instrum. Methods A **543**, 577 (2005)
4. A.K. Jain, J. Mao, K.M. Mohiuddin, Computer **29**(3), 31 (1996)
5. M. Claesen, B. De Moor, arXiv preprint arXiv:1502.02127 (2015)
6. E. Zitzler, K. Deb, L. Thiele, Evol. Comput. **8**(2), 173 (2000)
7. J. Kennedy, R. Eberhart, in *Proceedings of ICNN'95-International Conference on Neural Networks* (1995), p. 1942
8. J.H. Holland, Sci. Am. **267**(1), 66 (1992)
9. Y.W. Shang, Y.H. Qiu, Evol. Comput. **14**(1), 119 (2006)
10. ATLAS collaboration, in *CERN Open Data Portal* (2014)
11. Y. Shi, R.C. Eberhart, in *International Conference on Evolutionary Programming* (Springer, 1998), p. 591
12. R.C. Eberhart, Y. Shi, in *International Conference on Evolutionary Programming* (Springer, 1998), p. 611
13. F. Van Den Bergh, A.P. Engelbrecht, Inf. Sci. **176**(8), 937 (2006)
14. J. Yang, C.K. Soh, J. Comput. Civ. Eng. **11**(3), 195 (1997)
15. N.M. Razali, J. Geraghty, et al., in *Proceedings, World Congress on Engineering* (2011), p. 1
16. W.M. Spears, K.A. De Jong, in *Foundations of Genetic Algorithms*, vol. 1 (Elsevier, 1991), p. 301
17. K.A. De Jong, W.M. Spears, Ann. Math. Artif. Intell. **5**(1), 1 (1992)
18. D. Bhandari, C. Murthy, S.K. Pal, Int. J. Pattern Recognit. Artif. Intell. **10**(06), 731 (1996)
19. E.B. Baum, D. Boneh, C. Garrett, Evol. Comput. **9**(1), 93 (2001)
20. R. Tanese, Distributed genetic algorithms for function optimization. Ph.D. thesis, University of Michigan (1989)
21. S. Kok, C. Sandrock, Evol. Comput. **17**(3), 437 (2009)
22. S. Ruder, arXiv:1609.04747 (2016)
23. M. Andrychowicz, M. Denil, S. Gomez, M.W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, N. De Freitas, Adv. Neural Inf. Process. Syst. 3981–3989 (2016)
24. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011)
25. T. Chen, C. Guestrin, in *Proceedings, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), p. 785
26. XGBoost Parameters. https://xgboost.readthedocs.io/en/latest/parameter.html. Online; Accessed 08 July 2020

Appendix 2

II

CMS collaboration, “Search for Higgs boson pairs decaying to WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s}$ = 13 TeV”, *JHEP*, vol. 07, p. 095, 2023. DOI: 10.1007/JHEP07(2023)095. arXiv: 2206.10268 [hep-ex]

Published for SISSA by Springer

Received: *June 21, 2022*
Accepted: *November 27, 2022*
Published: *July 12, 2023*

Search for Higgs boson pairs decaying to WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS collaboration

E-mail: cms-publication-committee-chair@cern.ch

Abstract: The results of a search for Higgs boson pair (HH) production in the WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ decay modes are presented. The search uses 138 fb^{-1} of proton-proton collision data recorded by the CMS experiment at the LHC at a center-of-mass energy of 13 TeV from 2016 to 2018. Analyzed events contain two, three, or four reconstructed leptons, including electrons, muons, and hadronically decaying tau leptons. No evidence for a signal is found in the data. Upper limits are set on the cross section for nonresonant HH production, as well as resonant production in which a new heavy particle decays to a pair of Higgs bosons. For nonresonant production, the observed (expected) upper limit on the cross section at 95% confidence level (CL) is 21.3 (19.4) times the standard model (SM) prediction. The observed (expected) ratio of the trilinear Higgs boson self-coupling to its value in the SM is constrained to be within the interval -6.9 to 11.1 (-6.9 to 11.7) at 95% CL, and limits are set on a variety of new-physics models using an effective field theory approach. The observed (expected) limits on the cross section for resonant HH production range from 0.18 to 0.90 (0.08 to 1.06) pb at 95% CL for new heavy-particle masses in the range 250–1000 GeV.

Keywords: Hadron-Hadron Scattering, Higgs Physics

ArXiv ePrint: 2206.10268

https://doi.org/10.1007/JHEP07(2023)095

JHEP07(2023)095

Contents

1 Introduction

Since the discovery of the Higgs (H) boson [1–3], many of its properties have already been measured with high precision [4–6]. One important property that remains largely unknown is the H boson self-coupling. A precise measurement of this coupling is necessary to determine the shape of the Higgs potential, and thus verify that the mechanism breaking the electroweak gauge symmetry is indeed the Higgs mechanism [7–12] of the standard model (SM) [13–15]. The SM predicts the existence of both trilinear and quartic H boson self-couplings. Due to the very low predicted cross section for triple H boson production, the SM quartic self-coupling will not be experimentally accessible at the CERN LHC, even with the full integrated luminosity of $3000\,\mathrm{fb}^{-1}$ scheduled to be delivered after the high-luminosity LHC upgrade [16, 17]. The strength of the trilinear self-coupling, however, can be determined using measurements of H boson pair (HH) production.

In the SM, most HH pairs are produced in two types of processes. The Feynman diagrams for the dominant "gluon fusion" (ggHH) process at leading order (LO) in perturbative quantum chromodynamics (QCD) are shown in figure 1. The left "triangle" diagram

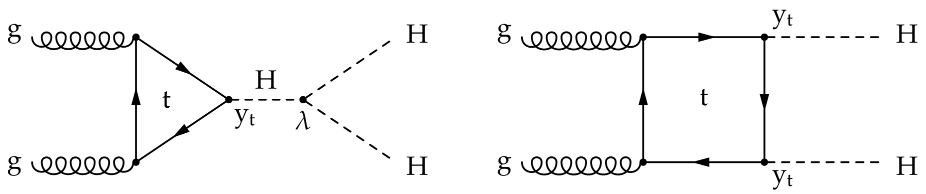


Figure 1. Leading order Feynman diagrams for SM nonresonant HH production via gluon fusion, including the "triangle" diagram (left) and the "box" diagram (right).

amplitude varies proportionally to the H boson self-coupling (λ) and the Yukawa coupling of the top quark (y_t), while the right "box" diagram amplitude is insensitive to λ and varies as y_t^2. The triangle and box diagrams interfere destructively, so the ggHH cross section exhibits a strong dependence on both λ and y_t. The ggHH cross section in the SM has been computed to be $31.1^{+2.1}_{-7.2}$ fb at next-to-next-to-LO (NNLO) accuracy in QCD using the $\mathrm{FT}_{\mathrm{approx}}$ scheme, in which the true top quark mass is used for the real radiation matrix elements, while the virtual part is computed using an infinite top quark mass [18]. The predicted SM cross section for the subdominant "vector boson fusion" (qqHH) process is 1.73 ± 0.04 fb at next-to-NNLO accuracy in QCD [19].

Deviations of the coupling strength modifiers $\kappa_\lambda = \lambda/\lambda^{\mathrm{SM}}$ and $\kappa_t = y_t/y_t^{\mathrm{SM}}$ from unity would affect both the rate of HH production and kinematic distributions of the HH signal. The HH invariant mass (m_{HH}) is particularly sensitive to changes in κ_λ and κ_t, as these couplings affect the triangle and box diagram amplitudes differently. Because SM ggHH and qqHH production do not include a heavy resonant particle, and typically result in a broad m_{HH} distribution, they are referred to as "nonresonant". Changes in κ_λ and κ_t also influence the rate of single Higgs boson production and the Higgs boson decay branching fractions [20, 21].

The presence of undiscovered particles or interactions, predicted by a variety of theoretical models beyond the SM, may alter the HH production rate as well as observable kinematic distributions. Such particles could give rise to loop diagrams similar to the one shown on the left of figure 1. These diagrams may significantly enhance the HH production rate, as they occur at the same loop level as HH production in the SM. Since no particles beyond those predicted by the SM have been observed so far, their mass may be at the TeV scale or higher, well above the scale of electroweak symmetry breaking. Loop contributions of such heavy particles can be approximated as contact interactions with the H boson using an effective field theory (EFT) approach [22, 23]. Following ref. [24], the contact interactions relevant for HH production are parametrized by the couplings c_g, c_{2g}, and c_2, referring to the interactions between two gluons and one H boson, two gluons and two H bosons, and two top quarks and two H bosons, respectively. The corresponding Feynman diagrams for ggHH production are shown in figure 2. The LO diagrams for qqHH production contain no gluons or top quarks, so the impacts of c_g, c_{2g}, and c_2 are only considered in the ggHH signal in this publication.

JHEP07(2023)095

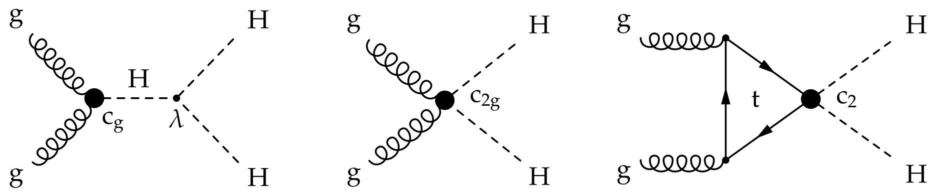


Figure 2. Leading order Feynman diagrams for nonresonant HH production via gluon fusion in an EFT approach, where loop-mediated contact interactions between (left) two gluons and one H boson, (middle) two gluons and two H bosons, and (right) two top quarks and two H bosons are parametrized by three effective couplings: c_g, c_{2g}, and c_2.

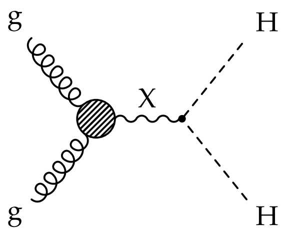


Figure 3. Leading order Feynman diagram for resonant HH production.

An excess of HH signal events may also result from decays of new heavy particles, denoted as X, into pairs of H bosons. Various theoretical models of new physics postulate such decays, in particular two-Higgs-doublet models [25, 26], composite-Higgs models [27, 28], Higgs portal models [29, 30], and models inspired by warped extra dimensions [31]. In the last class of models, the new heavy particles may have spin 0 ("radions") or spin 2 ("gravitons") [32]. In this paper, the resulting "resonant" HH production is sought for mass values of X from 250 to 1000 GeV, and the width of X is assumed to be negligible compared to the experimental resolution in m_{HH}. This would create a peak in the reconstructed m_{HH} distribution around the mass m_{X} of the resonance. The Feynman diagram for this process is shown in figure 3. For resonance masses above 1 TeV the strongest constraints are given by searches for HH production targeting H boson decays to bottom quarks [33–35], as the selection and reconstruction efficiency for hadronic decays increases, in particular in the trigger, and relevant backgrounds decrease with energy. For leptonic decay modes, the selection and reconstruction efficiency in general is high and as such do not increase notably for high masses above 1 TeV.

Phenomenological studies of the prospects for discovering HH signal in the $\mathrm{WW^*WW^*}$ decay mode are documented in refs. [36–40], where the symbol $*$ denotes virtual particles. The ATLAS Collaboration published results of a search for nonresonant and resonant HH pairs decaying to $\mathrm{WW^*WW^*}$ based on 36 fb^{-1} of proton-proton (pp) collision data recorded at $\sqrt{s} = 13$ TeV [41], placing an upper limit of 160 times the SM predicted cross section for nonresonant HH production at 95% confidence level (CL). Searches for HH production in pp collisions at $\sqrt{s} = 7$, 8, and 13 TeV have previously been performed by

JHEP07(2023)095

the CMS and ATLAS Collaborations in the decay modes $\mathrm{bb}\gamma\gamma$ [42, 43], bbbb [33, 44–47], $\mathrm{bb}\tau\tau$ [35, 48, 49], bbWW^* [34, 50–52], and $\mathrm{WW}^*\gamma\gamma$ [53]. Limits on HH production obtained from a combination of some of these analyses have been published by the CMS and ATLAS Collaborations [54, 55].

Searches targeting the $\mathrm{bb}\tau\tau$ [48], bbbb [45, 46], and $\mathrm{bb}\gamma\gamma$ [42] final states in CMS, and $\mathrm{bb}\tau\tau$ [35] and $\mathrm{bb}\gamma\gamma$ [43] in ATLAS, provide the strongest constraints on nonresonant HH production to date, with observed (expected) 95% CL upper limits ranging from 3.3 to 9.9 (3.9 to 7.8) times the SM predicted cross section. The corresponding lower bounds on κ_λ vary from -1.5 to -3.3 (-2.4 to -5.0 expected), with upper bounds between 6.7 and 9.4 (7.7 to 12.0 expected). The ATLAS $\mathrm{bb}\gamma\gamma$ analysis places a 95% CL upper limit of 0.64 pb on resonant HH production with a mass around 250 GeV (where 0.39 pb was expected) [43], while the ATLAS resonant bbbb search constrains higher mass hypotheses most strongly, with observed and expected limits around 0.01 pb at 1 TeV [33]. The ATLAS $\mathrm{bb}\tau\tau$ performs best for many mass points in between [35]. The only published HH search using an EFT approach comes from CMS in the $\mathrm{bb}\gamma\gamma$ final state, with 95% CL upper limits on the HH production cross section ranging from 0.1 to 0.6 pb, depending on the EFT scenario [42].

This paper presents the first search for H boson pairs decaying to $\mathrm{WW}^*\mathrm{WW}^*$, $\mathrm{WW}^*\tau\tau$, and $\tau\tau\tau\tau$. Both nonresonant and resonant HH production in final states with multiple reconstructed leptons, i.e., electrons (e), muons (μ), or hadronically decaying tau leptons (τ_h) are covered. The search is based on LHC pp collision data recorded by the CMS experiment at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}. Signal candidate events are subdivided into seven mutually exclusive "search categories" based on ℓ (e, μ) and τ_h multiplicity: two same-sign ℓ with fewer than two τ_h (2ℓss), three ℓ with no τ_h (3ℓ), four ℓ (4ℓ), three ℓ with one additional τ_h ($3\ell+1\tau_\mathrm{h}$), two ℓ with two τ_h ($2\ell+2\tau_\mathrm{h}$), one ℓ with three τ_h ($1\ell+3\tau_\mathrm{h}$), or four τ_h with no ℓ ($4\tau_\mathrm{h}$). In final states with a total of four ℓ and τ_h, the charge sum of all ℓ and τ_h candidates is required to be zero. The seven search categories target HH signal events in which the H boson pair decays into $\mathrm{WW}^*\mathrm{WW}^*$, $\mathrm{WW}^*\tau\tau$, or $\tau\tau\tau\tau$. Multivariate analysis (MVA) methods are used to distinguish the HH signal from backgrounds.

The paper is structured as follows. A brief overview of the CMS detector is given in section 2. Section 3 lists the data sets and simulation samples used. The reconstruction of e, μ, τ_h, and jets, along with various kinematic observables, is detailed in section 4. This is followed by a description, in section 5, of the event selection criteria defining the seven search categories. The multivariate methods used to distinguish the HH signal from backgrounds are detailed in section 6. The estimation of these backgrounds is described in section 7, followed by an outline of the relevant systematic uncertainties in section 8. The statistical procedure used to extract limits on the HH production rate in the SM, as well as constraints on SM coupling strengths, EFT benchmark scenarios, and resonant HH production rates are presented in section 9. The paper concludes with a summary in section 10.

JHEP07(2023)095

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of one barrel and two endcap sections. The silicon tracker measures charged particles within the pseudorapidity range $|\eta| < 2.5$ for data recorded in 2016, and within the range $|\eta| < 3.0$ for data recorded in 2017 and 2018. The ECAL is a fine-grained hermetic calorimeter with quasi-projective geometry, and is divided into a barrel region covering $|\eta| < 1.5$, and two endcaps that extend to $|\eta| = 3.0$. The HCAL barrel and endcaps similarly cover the region $|\eta| < 3.0$. Forward calorimeters extend beyond these endcaps to $|\eta| = 5.0$. Muons are detected within the range $|\eta| < 2.4$ by gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. Collision events of interest are selected using a two-tiered trigger system. The level-1 trigger, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select less than 100 kHz of events from a 40 MHz base event rate, within a fixed latency of 4 μs [56]. The second tier, known as the high-level trigger, is a processor farm which runs a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to around 1 kHz before data storage [57]. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the most relevant kinematic variables, can be found in ref. [58].

3 Data samples and Monte Carlo simulation

The analyzed pp collision data correspond to an integrated luminosity of 138 fb^{-1}, collected by the CMS detector over three years: 36 fb^{-1} in 2016, 42 fb^{-1} in 2017, and 60 fb^{-1} in 2018 [59–61]. This analysis uses triggers requiring one or more reconstructed e, μ, or τ_h candidates to be associated with the same collision vertex. The exact triggers and their thresholds varied slightly from year to year because of changes in luminosity and detector conditions, as well as improvements to the trigger algorithms. The transverse momentum (p_T) thresholds imposed by the trigger on the "leading" (highest p_T), "subleading" (second-highest p_T), and third e, μ, or τ_h, and the corresponding η requirements for each year are shown in table 1. All triggers include identification and isolation requirements on the e, μ, and τ_h candidates [57]. When combined, the triggers achieve an efficiency of 95–100% for simulated SM HH signal events in each of the seven search categories.

Monte Carlo (MC) simulated samples are used to model HH signal events and a wide range of SM background processes that produce final states with e, μ, or τ_h. Background MC samples include processes producing a single W or Z boson, two bosons (WW, WZ, ZZ, Wγ, and Zγ), three bosons (WWW, WWZ, WZZ, ZZZ, and WZγ), a single H boson (via gluon fusion, vector boson fusion, or associated production with a W or Z boson), a single top quark, a top quark-antiquark pair ($\mathrm{t\bar{t}}$), and top quarks associated with one or more bosons ($\mathrm{t\bar{t}W}$, $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}H}$, tHq, and tHW). All MC samples were generated using either

Trigger	Selection requirements for reconstructed e, μ, and τ_h objects
Single e	$p_T(e) > 27$–35 GeV
Single μ	$p_T(\mu) > 22$–27 GeV
Double e	$p_T(e) > 23,\ 12$ GeV
e + μ	$p_T(e) > 23$ GeV, $p_T(\mu) > 8$ GeV
μ + e	$p_T(\mu) > 23$ GeV, $p_T(e) > 8$–12 GeV
Double μ	$p_T(\mu) > 17,\ 8$ GeV
e + τ_h	$p_T(e) > 24$ GeV, $p_T(\tau_h) > 20$–30 GeV, $\lvert\eta(e, \tau_h)\rvert < 2.1$
μ + τ_h	$p_T(\mu) > 19$–20 GeV, $p_T(\tau_h) > 20$–27 GeV, $\lvert\eta(\mu, \tau_h)\rvert < 2.1$
Double τ_h	$p_T(\tau_h) > 35$–40 GeV, $\lvert\eta(\tau_h)\rvert < 2.1$
Triple e	$p_T(e) > 16,\ 12,\ 8$ GeV
Two e + μ	$p_T(e) > 12,\ 12$ GeV, $p_T(\mu) > 8$ GeV
Two μ + e	$p_T(\mu) > 9,\ 9$ GeV, $p_T(e) > 9$ GeV
Triple μ	$p_T(\mu) > 12,\ 10,\ 5$ GeV

Table 1. Selection requirements on p_T and η of reconstructed electrons (e), muons (μ), and hadronically decaying tau leptons (τ_h) applied by the triggers used in this analysis. The trigger p_T thresholds for leading, subleading, and third e, μ, or τ_h are separated by commas. For trigger thresholds that varied over time, the range of variation is indicated.

MadGraph5_aMC@NLO v2 [62, 63], POWHEG v2 [64–66], MCFM v7 [67–69], or PYTHIA v8.2 [70]. All samples that include a H boson were produced for a H boson mass of 125 GeV. Specific details of the simulated processes are summarized in table 2.

The parton distribution functions (PDFs) of the proton are modeled using the NNPDF3.0 and NNPDF3.1 PDF sets [85–89]. Parton shower, hadronization processes, and τ decays are modeled by PYTHIA, using the tunes CP5, CUETP8M1, CUETP8M2, or CUETP8M2T4 [90–92], depending on the process and the data-taking period that is being modeled. The matching of matrix elements to parton showers is performed using the MLM scheme [93] for the LO samples and the FxFx scheme [94] for the NLO samples. The interactions of particles with the CMS detector material was simulated in detail using Geant4 [95]. Simulated events were reconstructed using the same procedure as in data. The response of the trigger is included in the simulation. Additional pp interactions (pileup) were generated with PYTHIA and overlaid on all MC events, with event weights used to match the collision multiplicity to the distribution inferred from data. Residual differences between data and simulation are rectified by applying corrections to simulated events.

A variety of HH signal samples were generated at LO and NLO accuracy in QCD to simulate nonresonant HH production, covering the ggHH and qqHH production processes, with the H bosons decaying to either WW^*, ZZ^*, or ττ. The NLO samples are used to extract the rate of the HH signal from the data, while LO samples with a larger number of simulated events are used to train machine learning algorithms. Separate ggHH samples are produced for SM HH production and for a total of twelve EFT benchmark (BM) scenarios in the Higgs Effective Field Theory (HEFT) approach [24]. These benchmarks, along with

JHEP07(2023)095

Process	MC generator (order)	Cross section order
ggHH	MadGraph5_aMC@NLO v2 (LO) [71, 72]	NNLO $\mathrm{FT}_{\mathrm{approx}}$
	POWHEG v2 (NLO) [73–75]	
qqHH	MadGraph5_aMC@NLO v2 (LO)	N3LO
Single H boson production		
(via gluon fusion)	POWHEG v2 (NLO) [76]	N3LO QCD, NLO EW
(via vector boson fusion)	POWHEG v2 (NLO) [77]	NNLO QCD, NLO EW
(with a W or a Z boson)	POWHEG v2 (NLO) [78]	NNLO QCD, NLO EW
(with a pair of top quarks)	MadGraph5_aMC@NLO v2 (NLO)	NLO QCD, NLO EW
(with a single top quark)	MadGraph5_aMC@NLO v2 (LO)	NLO
W	MadGraph5_aMC@NLO v2 (LO)	NNLO
Z	MadGraph5_aMC@NLO v2 (LO)	NNLO QCD, NLO EW
WW (double-parton interaction)	PYTHIA v8.2 (LO)	LO
(same-sign pair)	MadGraph5_aMC@NLO v2 (LO)	LO
(opposite-sign pair)	POWHEG v2 (NLO) [79, 80]	NNLO
WZ	MadGraph5_aMC@NLO v2 (NLO)	NNLO
ZZ (quark-initiated)	POWHEG v2 (NLO) [79, 80]	NNLO
(gluon-initiated)	MCFM v7 (LO) [81]	NLO
$\mathrm{W}\gamma$, $\mathrm{Z}\gamma$, $\mathrm{WZ}\gamma$, $\mathrm{t}\gamma$, $\mathrm{t\bar{t}}\gamma$	MadGraph5_aMC@NLO v2 (NLO)	NLO
WWW, WWZ, WZZ, ZZZ	MadGraph5_aMC@NLO v2 (NLO)	NLO
Single top	POWHEG v2 (NLO) [82]	NLO
(with a W boson)	POWHEG v2 (NLO) [83]	NLO
(with a Z boson)	MadGraph5_aMC@NLO v2 (NLO)	NLO
$\mathrm{t\bar{t}}$	POWHEG v2 (NLO) [84]	NNLO
$\mathrm{t\bar{t}t\bar{t}}$	MadGraph5_aMC@NLO v2 (NLO)	NLO
$\mathrm{t\bar{t}W}$	MadGraph5_aMC@NLO v2	NLO QCD, NLO EW
	(NLO QCD, NLO EW)	
$\mathrm{t\bar{t}Z}$	MadGraph5_aMC@NLO v2 (NLO)	NLO QCD, NLO EW
$\mathrm{t\bar{t}WW}$, $\mathrm{t\bar{t}WZ}$, $\mathrm{t\bar{t}ZZ}$	MadGraph5_aMC@NLO v2 (LO)	LO

Table 2. The MC generators that are used to simulate HH signal and background processes. The order of MC simulation and cross section calculation both refer to the perturbative expansion in QCD. Additional higher order electroweak (EW) corrections, if present, are indicated separately.

JHEP07(2023)095

the seven benchmarks from ref. [96], represent different combinations of κ_λ, κ_t, $\mathrm{c_g}$, $\mathrm{c_{2g}}$, and $\mathrm{c_2}$ HEFT parameter values, and are chosen to probe distinct classes of HH kinematic configurations. These benchmarks are referred to as JHEP04 BM1-12, and JHEP03 BM1-7, respectively. The benchmark JHEP04 BM8 is complemented by a modified version of this benchmark, published in ref. [97], denoted as JHEP04 BM8a. The parameter values of these twenty BM scenarios are shown in table 3. The values of the $\mathrm{c_g}$ and $\mathrm{c_{2g}}$ couplings published in ref. [96] have been scaled by factors of 1.5 and −3, respectively, to convert them to the convention introduced for these couplings in ref. [24]. In order to increase the number of simulated events and to model kinematic configurations not explicitly generated, such as JHEP03 BM1-7, the ggHH samples are merged and the events in the merged samples are reweighted, using the procedure documented in ref. [98], to match the distributions in m_HH and $|\cos\theta^*|$ computed at NLO accuracy and published in ref. [97]. This procedure is applied to the LO and NLO ggHH samples separately. The symbol $\cos\theta^*$ denotes the cosine of the polar angle of one H with respect to the beam axis in the HH rest frame. The qqHH samples are produced only for SM HH production.

Benchmark	κ_λ	κ_t	c_2	c_g	c_{2g}
JHEP04 BM1	7.5	1.0	−1.0	0.0	0.0
JHEP04 BM2	1.0	1.0	0.5	−0.8	0.6
JHEP04 BM3	1.0	1.0	−1.5	0.0	−0.8
JHEP04 BM4	−3.5	1.5	−3.0	0.0	0.0
JHEP04 BM5	1.0	1.0	0.0	0.8	−1.0
JHEP04 BM6	2.4	1.0	0.0	0.2	−0.2
JHEP04 BM7	5.0	1.0	0.0	0.2	−0.2
JHEP04 BM8	15.0	1.0	0.0	−1.0	1.0
JHEP04 BM8a	1.0	1.0	0.5	4/15	0.0
JHEP04 BM9	1.0	1.0	1.0	−0.6	0.6
JHEP04 BM10	10.0	1.5	−1.0	0.0	0.0
JHEP04 BM11	2.4	1.0	0.0	1.0	−1.0
JHEP04 BM12	15.0	1.0	1.0	0.0	0.0
JHEP03 BM1	3.94	0.94	−1/3	0.75	−1
JHEP03 BM2	6.84	0.61	1/3	0	1
JHEP03 BM3	2.21	1.05	−1/3	0.75	−1.5
JHEP03 BM4	2.79	0.61	1/3	−0.75	−0.5
JHEP03 BM5	3.95	1.17	−1/3	0.25	1.5
JHEP03 BM6	5.68	0.83	1/3	−0.75	−1
JHEP03 BM7	−0.10	0.94	1	0.25	0.5
SM	1.0	1.0	0.0	0.0	0.0

Table 3. Parameter values for κ_λ, κ_t, c_2, c_g, and c_{2g} in MC samples modeling twenty benchmark scenarios in the EFT approach, plus SM HH production.

Resonant HH production was simulated at LO for both spin-0 (radion) and spin-2 (graviton) scenarios with m_X values of 250, 260, 270, 280, 300, 320, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, and 1000 GeV.

4 Event reconstruction

The CMS particle-flow (PF) algorithm [99] aims to reconstruct and identify each individual particle in an event, using an optimized combination of information from the various elements of the CMS detector. The particles are subsequently classified into five mutually exclusive types: electrons, muons, photons, and charged and neutral hadrons. These particles are then combined to reconstruct hadronic τ decays, jets, and the missing transverse momentum in the event.

The candidate vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex. The physics objects used for this determination are the jets, clustered using the infrared and collinear safe anti-k_T algorithm [100, 101], with the tracks assigned to candidate vertices as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the p_T of those jets.

Electrons are reconstructed within the geometric acceptance of the tracking detectors ($|\eta| < 2.5$) by combining information from the tracker and the ECAL [102]. They are initially identified using an MVA classifier which distinguishes real electrons from hadrons, along with requirements that the track be associated with the collision vertex, and limits on hadronic energy deposits separated by $\Delta R < 0.4$ from the electrons (their "isolation"). The angular separation between two particles is defined as $\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}$, where the symbol ϕ refers to the azimuthal angle of the particle. Electrons passing this initial selection are referred to as "loose". In this analysis, events with electrons originating from hadron decays ("nonprompt"), or with hadrons misidentified as electrons, constitute the largest source of background. This motivates the use of an additional MVA classifier, which is trained to select "prompt" electrons from W, Z, and τ lepton decays, and to reject nonprompt or misidentified electrons. This MVA classifier was previously used for measurements of $\mathrm{t\bar{t}H}$ production in events with multiple leptons [103]. It combines observables comparing measurements of the electron energy and direction in the tracker and the ECAL, the compactness of the electron cluster, the bremsstrahlung emitted along the electron trajectory, and the electron isolation. Two levels of thresholds on the output of this MVA classifier are used in the analysis, referred to as the "tight" and "medium" electron selections for the more and less restrictive thresholds, respectively. The tight selection has an average efficiency of 60% for electrons from SM HH decays. Only the electrons passing the tight selections are used to reconstruct signal candidate events, while data events with electrons passing the medium selections and failing the tight selections are used to estimate the contribution of misidentified- and nonprompt-electron backgrounds in each search category. Compared to ref. [103], this analysis uses lower thresholds on the MVA classifier output for the medium and tight electron selections, in order to increase the efficiency in particular for low-p_{T} electrons, which frequently appear in the HH signal events studied in this analysis. Electrons from photon conversions in the tracker are suppressed by requiring that the track is missing no hits in the innermost layers of the silicon tracker, and is not matched to a reconstructed conversion vertex. In the 2ℓss category, further electron selection criteria are applied, which require agreement among three independent measurements of the electron charge, including the Gaussian sum filter and Kalman filter track curvatures, as well as the ECAL supercluster position [104]. The remaining charge misidentification rate is measured to be less than 0.1% for $|\eta| < 1.479$, and under 0.4% for $|\eta| > 1.479$. The charge quality requirement reduces the electron identification efficiency by about 4%.

Muons are reconstructed by extrapolating tracks in the silicon tracker to hits in the gas-ionization muon detectors embedded in the steel flux-return yoke outside the solenoid [105]. To pass the initial loose identification requirement for this analysis, muons must satisfy criteria related to isolation and track proximity to the primary interaction vertex, as well as track quality observables and matching between the tracker and muon chambers. Additional requirements on the prompt vs. nonprompt muon identification MVA classifier from ref. [103] serve to select muons passing a tight selection for signal candidate events, and a medium selection for nonprompt background estimation. Inputs to this MVA classifier include energy deposits close to the muon in the ECAL and HCAL, the hits and track segments reconstructed in the muon detectors located outside the solenoid, the quality of

JHEP07(2023)095

the spatial matching between the track segments reconstructed in the silicon tracker and in the muon detectors, and the isolation of the muon with respect to other particles. Again, lower selection thresholds on the MVA classifier output compared to ref. [103] bring higher efficiency for the HH signal, amounting to 80% per muon in simulated SM HH events for the tight selection. In the 2ℓss channel, the uncertainty in the curvature of the muon track is required to be less than 20% to ensure a high-quality charge measurement [103]. This requirement reduces the muon identification efficiency by about 2%.

Hadronic decays of tau leptons are identified using the "hadrons-plus-strips" algorithm [106]. This algorithm classifies individual hadronic decay modes of the τ by combining charged hadrons from the PF reconstruction with neutral pions. The latter are reconstructed by clustering electrons and photons into rectangular strips, which are narrow in η but wide in the ϕ direction. The spread in ϕ accounts for photons originating from neutral pion decays that convert into electron-positron (e^-e^+) pairs while traversing the silicon tracker. The e^- and e^+ are bent in opposite directions in ϕ by the magnetic field, and may further emit bremsstrahlung photons before reaching the ECAL. The decay modes considered in this analysis produce one charged pion or kaon plus up to two neutral pions (collectively referred to as "one-prong" τ_h), or three charged pions or kaons plus zero or one neutral pion (referred to as "three-prong" τ_h). The DeepTau algorithm [107] distinguishes true τ_h objects from quark and gluon jets, electrons, and muons using a convolutional artificial neural network (NN) [108] with 42 high-level observables as input, together with low-level information obtained from the silicon tracker, ECAL, HCAL, and the muon detectors. The former include the p_T, η, ϕ, and mass of the τ_h candidate, the reconstructed τ_h decay mode, its isolation with respect to charged and neutral particles, and the estimated distance that the τ lepton traverses between its production and decay. For three-prong τ_h candidates, this distance is determined by reconstructing the decay vertex, while for one-prong τ_h candidates, the transverse impact parameter of the charged pion track with respect to the primary pp interaction vertex is used as an estimate of the distance. The low-level information quantifies the particle activity within two $\eta \times \phi$ grids, centered on the direction of the τ_h candidate: an inner grid of size 0.2×0.2, filled with 0.02×0.02 cells, and an outer grid of size 0.5×0.5 (partially overlapping with the inner grid), with 0.05×0.05 cells. Selected τ_h candidates in this analysis must have $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$, and are subjected to two levels of thresholds on the NN output that separates τ_h from quark and gluon jets, referred to as the tight and medium τ_h selections, respectively.

Hadronic jets (j) are reconstructed with the anti-k_T algorithm using the particles reconstructed with the PF algorithm as input, and serve to identify $H \to WW^* \to jj\ell\nu$ decays in this analysis. Jets reconstructed with size parameters of 0.4 ("small-radius jets") and 0.8 ("large-radius jets") are both used: two small-radius jets to reconstruct the two quarks from low-p_T W boson decays, or a single large-radius jet to reconstruct high-p_T W boson decays, where the quarks are collimated. Overlap between small-radius jets and electrons, muons, and τ_h is resolved by discarding those small-radius jets that contain one or more PF particles matched to an electron, a muon, or a constituent of a τ_h passing the medium selection criteria. In case of large-radius jets, electrons and muons passing the loose selection are removed from the collection of PF particles used as input to the jet reconstruction, so that leptons produced in $\mathrm{H} \to \mathrm{WW}^* \to \mathrm{jj}\ell\nu$ decays of Lorentz-boosted H bosons are not clustered into those jets.

The effect of pileup on the reconstruction of large-radius jets is mitigated by applying the pileup per particle identification algorithm (PUPPI) [109, 110] to the collection of particles used as input to the jet reconstruction. For small-radius jets, the effect of pileup is reduced by removing charged particles identified with pileup vertices from the jet reconstruction, and applying corrections to the jet energy to account for neutral particles from pileup.

After calibration, the jet energy resolution at the central rapidities amounts to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [111]. This analysis only considers jets reconstructed in the region $|\eta| < 2.4$. Small-radius jets must have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, while large-radius jets must have $p_{\mathrm{T}} > 170\,\mathrm{GeV}$. Additional criteria requiring that each large-radius jet contain exactly two identifiable, energetic subjets are applied to specifically select those from boosted hadronic W boson decays [112].

Events containing small-radius jets identified with the hadronization of bottom quarks (b jets) are vetoed in this analysis. The DeepJet algorithm [113] exploits observables related to the long lifetime of b hadrons and the higher particle multiplicity and mass of b jets compared to light quark and gluon jets. Both “loose” and “medium” b jet selections on the DeepJet output are employed in this analysis, corresponding to b jet selection efficiencies of 84 and 70%, while the misidentification rates for light-quark or gluon jets are 11 and 1.1%, respectively.

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is computed as the negative vector p_{T} sum of all the particles reconstructed by the PF algorithm in an event, and its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$ [114]. The $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is modified to account for corrections to the energy scale of the reconstructed jets in the event. A linear discriminant, denoted as $p_{\mathrm{T}}^{\mathrm{miss,LD}}$, is employed to remove background events in which the reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ arises from resolution effects. The discriminant is defined by the relation $p_{\mathrm{T}}^{\mathrm{miss,LD}} = 0.6 p_{\mathrm{T}}^{\mathrm{miss}} + 0.4 H_{\mathrm{T}}^{\mathrm{miss}}$, where $H_{\mathrm{T}}^{\mathrm{miss}}$ corresponds to the magnitude of the vector p_{T} sum of e, μ, and τ_{h} passing the medium selection criteria, and small-radius jets satisfying the criteria detailed above [115].

5 Event selection

Events are selected with the aim of maximizing the acceptance for HH decays to $\mathrm{WW}^*\mathrm{WW}^*$, $\mathrm{WW}^*\tau\tau$, and $\tau\tau\tau\tau$, while simultaneously rejecting the large backgrounds from multijet production, single and pair production of W and Z bosons, and $\mathrm{t\bar{t}}$ production. To achieve this, each event must contain multiple reconstructed ℓ or τ_{h} associated with the primary interaction vertex. The ℓ and τ_{h} may originate from the decay of a W boson or a τ lepton. Seven mutually exclusive search categories, distinguished by the number of reconstructed ℓ and τ_{h} candidates, are included in the analysis: 2ℓss, 3ℓ, 4ℓ, $3\ell + 1\tau_{\mathrm{h}}$, $2\ell + 2\tau_{\mathrm{h}}$, $1\ell + 3\tau_{\mathrm{h}}$, and $4\tau_{\mathrm{h}}$. Here “ss” indicates a same-sign $\ell\ell$ pair, with two leptons of identical electric charge. The ℓ and τ_{h} candidates selected in any of the seven search categories must pass the tight selection criteria described in section 4. In addition,

JHEP07(2023)095

they are required to pass category-specific p_T thresholds motivated by the trigger selection. Further requirements are placed on the sum of ℓ and τ_h charges, and, in two categories, on the discriminant $p_\mathrm{T}^\mathrm{miss,LD}$ and the multiplicity of jets.

The leading and subleading leptons in the 2ℓss category must pass p_T selection thresholds of 25 and 15 GeV, respectively. Events in this category are required to contain two or more small-radius jets, or at least one large-radius jet, targeting hadronic W boson decays. Dielectron events must have $p_\mathrm{T}^\mathrm{miss,LD} > 30\,\mathrm{GeV}$ and $m(\ell\ell) < 81\,\mathrm{GeV}$ or $m(\ell\ell) > 101\,\mathrm{GeV}$, in order to suppress charge-misidentified $\mathrm{Z} \to \mathrm{ee}$ background. If the event contains a τ_h, the charge of the τ_h must be opposite to the charge of the leptons. After this selection, the main backgrounds in the 2ℓss category arise from WZ production, from $\mathrm{W}\gamma$ events in which the photon converts into an $\mathrm{e}^-\mathrm{e}^+$ pair and either the e^- or the e^+ is not reconstructed, and from events in which one or both reconstructed leptons are due to a nonprompt ℓ or a misidentified hadron, as shown in table 6. The "other" background given in the table is dominated by same-sign W boson pairs and WWW production. The $\mathrm{WW}^*\mathrm{WW}^*$ decay mode accounts for roughly 70% of SM HH signal events selected in the 2ℓss category, with $\mathrm{WW}^*\tau\tau$ events accounting for the other 30%.

In the 3ℓ category, the leading, subleading, and third ℓ are required to have p_T values greater than 25, 15, and 10 GeV, respectively, and the sum of their charges must be either $+1$ or -1. At least one small- or large-radius jet must be present, and the $p_\mathrm{T}^\mathrm{miss,LD}$ quantity must be greater than 30 GeV, or 45 GeV if there is at least one same-flavor opposite-sign (SFOS) $\ell\ell$ pair in the event. Again, backgrounds are dominated by WZ production and events with misidentified ℓ. Notable contributions to the "other" background arise from WWW and WWZ production. The signal composition is similar to the 2ℓss category.

The 4ℓ category has identical lepton selection criteria to the 3ℓ category, except that the third ℓ must have $p_\mathrm{T} > 15\,\mathrm{GeV}$, and a fourth ℓ with $p_\mathrm{T} > 10\,\mathrm{GeV}$ is required, and the sum of the four lepton charges is required to be equal to zero. In this category and all the remaining categories, there are no selection requirements on jets or $p_\mathrm{T}^\mathrm{miss,LD}$. Almost 70% of signal events come from the $\mathrm{WW}^*\mathrm{WW}^*$ decay mode, and about 30% from $\mathrm{WW}^*\tau\tau$, while ZZ production accounts for 85% of the background.

Events in the $3\ell + 1\tau_\mathrm{h}$ category are required to satisfy the 3ℓ criteria on the ℓ objects, except that an additional τ_h with $p_\mathrm{T} > 20\,\mathrm{GeV}$ and charge opposite to the sum of the ℓ charges is required. Background events in which the reconstructed τ_h fails a loose selection on the NN output of the DeepTau algorithm that separates τ_h from electrons, or falls near the ECAL barrel-endcap transition region in $1.460 < |\eta| < 1.558$ are removed. About 70% of signal events come from the $\mathrm{WW}^*\tau\tau$ decay mode, while ZZ production and events with at least one misidentified ℓ or τ_h dominate the background.

In the $2\ell+2\tau_\mathrm{h}$ category, the leading and subleading ℓ are required to pass p_T thresholds of 25 and 15 GeV, while the two τ_h must have $p_\mathrm{T} > 20\,\mathrm{GeV}$. The sum of ℓ plus τ_h charges is required to be zero. Signal contributions are mostly from the $\mathrm{WW}^*\tau\tau$ (60%) and $\tau\tau\tau\tau$ (40%) decay modes, while background contributions arise from ZZ production and events with a misidentified ℓ or τ_h candidate.

In the $1\ell + 3\tau_\mathrm{h}$ category, the ℓ is required to satisfy the conditions $|\eta| < 2.1$ and $p_\mathrm{T} > 20$ (15) GeV if it is an electron (muon). The leading, subleading, and third τ_h must have $p_T > 40$, 30, and 20 GeV, respectively, and the sum of τ_h and ℓ charges is required to be zero. Background events containing a $Z \to ee$ decay where one electron is misidentified as a τ_h are vetoed by discarding events containing an e-τ_h pair of opposite charge and mass $71 < m(e\tau_h) < 101$ GeV, and in which the τ_h either fails a loose selection on the discriminant that separates τ_h from electrons, or falls into the region $1.460 < |\eta| < 1.558$. Around 80% of HH signal events selected in the $1\ell + 3\tau_h$ category are from $\tau\tau\tau\tau$ and 20% from the $WW^*\tau\tau$ decay mode, while the majority of background events stem from ZZ production or contain a misidentified ℓ or τ_h.

The $4\tau_h$ category requires the leading and subleading τ_h to pass p_T thresholds of 40 and 30 GeV, respectively, and the third and fourth τ_h to have $p_T > 20$ GeV. Given the extremely low backgrounds in this category, no charge sum criterion or $Z \to ee$ veto is applied. Almost all signal events come from the $\tau\tau\tau\tau$ decay mode, while 55% of the background events contain at least one misidentified τ_h candidate, and the remainder arises from ZZ (30%) and single Higgs boson (15%) production.

In all seven search categories, the background contamination from processes with top quarks is reduced by discarding events with at least one selected small-radius jet passing the medium b jet identification, or at least two passing the loose b jet identification. Leptons originating from low-mass Drell–Yan production, decays of J/ψ and Υ mesons, cascade decays of bottom quarks, and photon conversions are removed by vetoing events containing any pair of loose ℓ with mass $m(\ell\ell) < 12$ GeV. To eliminate overlap with events selected in the ongoing search for HH production in the $b\bar{b}ZZ$, $ZZ \to 4\ell$ decay mode, no event in the 2ℓss, 3ℓ, and 4ℓ categories may contain two SFOS loose $\ell\ell$ pairs with a mass of the four-ℓ system of less than 140 GeV. In addition, to reduce the $Z \to \ell\ell$ background, these three categories along with $2\ell + 2\tau_h$ and $3\ell + 1\tau_h$ exclude events where any SFOS loose $\ell\ell$ pair has an invariant mass of 81–101 GeV (Z boson veto).

A summary of the event selection criteria applied in the different categories is given in table 4. Criteria that are common to all seven search categories are given in table 5.

Two control regions (CRs) are used to validate the modeling of the WZ and ZZ backgrounds. These CRs match the signal regions of the 3ℓ and 4ℓ categories, but with the Z boson veto inverted, and are referred to as the "3ℓ WZ" CR and "4ℓ ZZ" CR, respectively.

The number of events selected in the signal regions of each of the seven search categories and in the 3ℓ WZ and 4ℓ ZZ CRs are given in table 6. The contribution expected from nonresonant HH production with event kinematics as predicted by the SM, but 30 times the SM cross section, is given separately for HH decays into WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ in the upper three rows of each table. The event yields given in the rows labeled WW^*WW^* include a small contribution from HH decays into WW^*ZZ^* and ZZ^*ZZ^*, and, similarly, the numbers quoted in the rows labeled $WW^*\tau\tau$ include a small contribution from HH decays into $ZZ^*\tau\tau$.

6 Analysis strategy

The rate of the HH signal is extracted through a binned maximum likelihood (ML) fit to the distributions in the output of boosted decision tree (BDT) classifiers [116], which are

Category	2ℓss	3ℓ	4ℓ
Targeted HH decays	WW^*WW^*	WW^*WW^*	WW^*WW^*
Trigger	Single- and double-lepton	Single-, double- and triple-lepton	Single-, double- and triple-lepton
Lepton p_T	>25 / 15 GeV	>25 / 15 / 10 GeV	>25 / 15 / 15 / 10 GeV
Lepton charge sum	±2, with charge quality requirements applied	±1	0
Dilepton invariant mass	$\lvert m_{\ell\ell} - m_Z \rvert > 10\,\text{GeV}$ †	$\lvert m_{\ell\ell} - m_Z \rvert > 10\,\text{GeV}$ ‡	$\lvert m_{\ell\ell} - m_Z \rvert > 10\,\text{GeV}$ ‡
Jets	≥2 small-radius jets or ≥1 large-radius jet	≥1 small-radius jet or ≥1 large-radius jet	—
Missing p_T	$p_T^{\text{miss,LD}} > 30\,\text{GeV}$ §	$p_T^{\text{miss,LD}} > 30\,\text{GeV}$ $^{\parallel}$	—

Category	$3\ell + 1\tau_h$	$2\ell + 2\tau_h$
Targeted HH decays	$WW^*\tau\tau$	$WW^*\tau\tau$, $\tau\tau\tau\tau$
Trigger	Single-, double-, and triple-lepton	Single- and double-lepton
Lepton p_T	>25 / 15 / 10 GeV	>25 / 15 GeV
τ_h p_T	>20 GeV	>20 GeV
Lepton and τ_h charge	ℓ and τ_h charges sum to 0	ℓ and τ_h charges sum to 0
Dilepton invariant mass	$\lvert m_{\ell\ell} - m_Z \rvert > 10\,\text{GeV}$ ‡	$\lvert m_{\ell\ell} - m_Z \rvert > 10\,\text{GeV}$ ‡

Category	$1\ell + 3\tau_h$	$4\tau_h$
Targeted HH decays	$\tau\tau\tau\tau$	$\tau\tau\tau\tau$
Trigger	Single-lepton, lepton+τ_h and double-τ_h	Double-τ_h
Lepton η	$\lvert\eta\rvert < 2.1$	—
Lepton p_T	>20 GeV (e) or >15 GeV (μ)	—
τ_h p_T	>40 / 30 / 20 GeV	>40 / 30 / 20 / 20 GeV
Lepton and τ_h charge	ℓ and τ_h charges sum to 0	τ_h charges sum to 0
$Z \to ee$ veto	$\lvert m_{e\tau_h} - 86\,\text{GeV} \rvert > 15\,\text{GeV}$ ¶	—

† Applied to all SFOS $\ell\ell$ pairs and electron pairs with the same charge.
‡ Applied to all SFOS $\ell\ell$ pairs.
§ Only applied to events containing two electrons.
$^{\parallel}$ Tightened to $p_T^{\text{miss,LD}} > 45\,\text{GeV}$ if event contains a SFOS $\ell\ell$ pair.
¶ For τ_h classified as electrons by the DeepTau algorithm or with $1.460 < \lvert\eta\rvert < 1.558$.

Table 4. Event selection criteria applied in the seven search categories. The p_T thresholds for ℓ and τ_h with the highest, second-, third-, and fourth-highest p_T are separated by slashes. The symbol "—" indicates that no requirement is applied.

Object and event properties	Selection criteria
Lepton and τ_h pseudorapidity	$\lvert\eta\rvert < 2.5$ for e, $\lvert\eta\rvert < 2.4$ for μ, $\lvert\eta\rvert < 2.3$ for τ_h
Dilepton invariant mass	$m_{\ell\ell} > 12\,\text{GeV}$ (all $\ell\ell$ pairs)
Four-lepton invariant mass	$m_{4\ell} > 140\,\text{GeV}$ (any two SFOS $\ell\ell$ pairs)
b jet veto	0 medium and ≤ 1 loose b-tagged small-radius jet

Table 5. Reconstructed object and event selection requirements in all seven search categories. Electrons or muons in the $\ell\ell$ pairs include any leptons passing the loose selection criteria.

trained to discriminate the HH signal from backgrounds, along with kinematic distributions from the two CRs above. The data from each of the three years are fit separately. Three classifiers are trained for each of the seven search categories using a mix of MC simulation

JHEP07(2023)095

Process	$2\ell ss$	3ℓ	4ℓ
SM HH $\rightarrow$ WW*WW* ($\times$ 30)	73 $\pm$ 6	33 $\pm$ 3	2.2 $\pm$ 0.2
SM HH $\rightarrow$ WW$^*\tau\tau$ ($\times$ 30)	31 $\pm$ 3	12 $\pm$ 1	0.9 $\pm$ 0.1
SM HH $\rightarrow \tau\tau\tau\tau$ ($\times$ 30)	3 $\pm$ 0	1 $\pm$ 0	0.1 $\pm$ 0.0
WZ	1999 $\pm$ 122	1318 $\pm$ 78	0.4 $\pm$ 0.1
ZZ	121 $\pm$ 3	109 $\pm$ 3	53.9 $\pm$ 3.1
Misidentified ℓ	4842 $\pm$ 1327	510 $\pm$ 94	2.2 $\pm$ 1.1
Conversion electrons	804 $\pm$ 174	117 $\pm$ 24	0.7 $\pm$ 0.3
Electron charge misid.	394 $\pm$ 61	—	—
Single Higgs boson	214 $\pm$ 6	61 $\pm$ 1	2.4 $\pm$ 0.3
Other backgrounds	2740 $\pm$ 338	289 $\pm$ 29	4.0 $\pm$ 0.5
Total expected background	11 114 $\pm$ 1387	2404 $\pm$ 128	63.7 $\pm$ 3.3
Data	10 344	2621	62

Process	$3\ell + 1\tau_h$	$2\ell + 2\tau_h$	$1\ell + 3\tau_h$	$4\tau_h$
SM HH $\rightarrow$ WW*WW* ($\times$ 30)	0.9 $\pm$ 0.1	0.2 $\pm$ 0.0	0.2 $\pm$ 0.0	0.3 $\pm$ 0.0
SM HH $\rightarrow$ WW$^*\tau\tau$ ($\times$ 30)	4.1 $\pm$ 0.3	3.9 $\pm$ 0.4	0.6 $\pm$ 0.1	0.1 $\pm$ 0.0
SM HH $\rightarrow \tau\tau\tau\tau$ ($\times$ 30)	0.9 $\pm$ 0.1	2.3 $\pm$ 0.3	2.6 $\pm$ 0.4	1.3 $\pm$ 0.2
WZ	0.2 $\pm$ 0.0	<0.1	<0.1	<0.1
ZZ	24.1 $\pm$ 1.4	18.4 $\pm$ 1.3	1.9 $\pm$ 0.2	0.7 $\pm$ 0.1
Misidentified ℓ and τ_h	23.9 $\pm$ 6.6	31.9 $\pm$ 10.1	2.2 $\pm$ 2.1	2.2 $\pm$ 1.6
Conversion electrons	0.1 $\pm$ 0.0	0.1 $\pm$ 0.1	<0.1	<0.1
Single Higgs boson	3.8 $\pm$ 0.4	2.8 $\pm$ 0.7	0.8 $\pm$ 0.4	0.4 $\pm$ 0.3
Other backgrounds	2.8 $\pm$ 0.4	2.2 $\pm$ 0.8	0.1 $\pm$ 0.1	<0.1
Total expected background	54.9 $\pm$ 6.8	55.4 $\pm$ 10.3	5.0 $\pm$ 2.2	3.4 $\pm$ 1.6
Data	55	55	6	1

Process	3ℓ WZ CR	4ℓ ZZ CR
WZ	12 565 $\pm$ 705	<1
ZZ	765 $\pm$ 47	2000 $\pm$ 108
Misidentified ℓ	804 $\pm$ 211	13 $\pm$ 4
Conversion electrons	106 $\pm$ 21	2 $\pm$ 0
Other backgrounds	625 $\pm$ 76	60 $\pm$ 8
Total expected background	14 866 $\pm$ 742	2074 $\pm$ 108
Data	14 994	2096

Table 6. The number of expected and observed events in each of the seven search categories, and in two CRs, which validate the modeling of the WZ and ZZ backgrounds. The symbol "—" indicates that the background is not relevant for the category. The HH signal represents the sum of the ggHH and qqHH production processes and is normalized to 30 times the event yield expected in the SM, corresponding to a cross section of about 1 pb. The event yields are obtained by performing the event selection and applying appropriate corrections to the simulated events. Quoted uncertainties represent the sum of statistical and systematic components. Uncertainties that are smaller than half the value of the least significant digit have been rounded to zero.

from all three years, targeting nonresonant HH production and resonant HH production from the decay of heavy particles of spin 0 and of spin 2. The binning is chosen with the objective of maximizing the sensitivity for the HH signal, while maintaining sufficient background events in each bin to keep the statistical uncertainty in the background prediction under control. In the two categories with high event yields (2ℓss and 3ℓ) the BDT output binning is chosen such that each bin contains a similar number of expected HH signal events. The four categories containing events with τ_h ($3\ell + 1\tau_h$, $2\ell + 2\tau_h$, $1\ell + 3\tau_h$, and $4\tau_h$) have low event yields and sizable background contributions arising from the misidentification of ℓ and τ_h candidates, which are determined from data and statistically limited. For these categories, we choose the binning for each BDT output distribution such that a similar number of expected background events is contained in each bin. In the 4ℓ category, the fact that the background is dominated by ZZ production, which is modeled by the MC simulation with low statistical uncertainties, allows one to choose the binning in the same way as for the 2ℓss and 3ℓ categories. The number of bins is determined by the condition that the relative statistical uncertainty in the background prediction in each bin does not exceed 15%. Higher bin numbers correspond to a higher BDT output value, and feature a higher signal-to-background ratio. For the SM HH signal, the bins with the highest BDT output values feature a signal-to-background ratio up to 10 times higher than the inclusive ratio in each category.

The inputs to the BDT classifiers differ by search category and include the p_T and η of reconstructed ℓ and τ_h; the angular separation ΔR and invariant mass of $\ell\ell$, $\ell\tau_h$, and $\tau_h\tau_h$ pairs; the ΔR and invariant mass between an ℓ or τ_h candidate and the nearest jet(s); the number of jets in the event; the discriminant $p_T^{\text{miss,LD}}$; the scalar p_T sum of all reconstructed e, μ, τ_h, and jets; the "visible" mass of the Higgs boson pair, given by the mass of the system of reconstructed e, μ, τ_h, and jets; and where applicable, the "full" mass of the HH system, including neutrinos, reconstructed using the algorithm from ref. [117] designed for reconstructing Higgs pair decays into τ leptons. This algorithm targets HH signal events decaying to $\tau\tau\tau\tau$ and thus works best in the $4\tau_h$ and $1\ell + 3\tau_h$ search categories. Distributions in some of the observables used as inputs to the BDT classifiers in the 2ℓss and 3ℓ categories are shown in figure 4.

These observables are complemented by further inputs, which parametrize the BDT as a function of the model parameters: the Higgs boson couplings λ, y_t, c_g, c_{2g}, and c_2 for nonresonant HH production, and the mass of the heavy particle X in resonant HH production. When training the BDT that targets nonresonant HH production, the values for the couplings are chosen according to the twelve EFT benchmark scenarios given in ref. [24] and the SM, indicated by thirteen binary inputs to the BDT. The BDT classifiers used for the analysis of resonant HH production are trained separately for spin-0 and spin-2 on the full set of resonance masses listed in section 3, and the resonance mass is used as an input to the BDT. Each simulated background event is replicated multiple times in the training sample, with different values assigned to the Higgs boson couplings and the mass of the heavy particle X.

The training is performed using simulated samples of signal and background events. The signal events used in the training consist of ggHH events in the HH decay modes

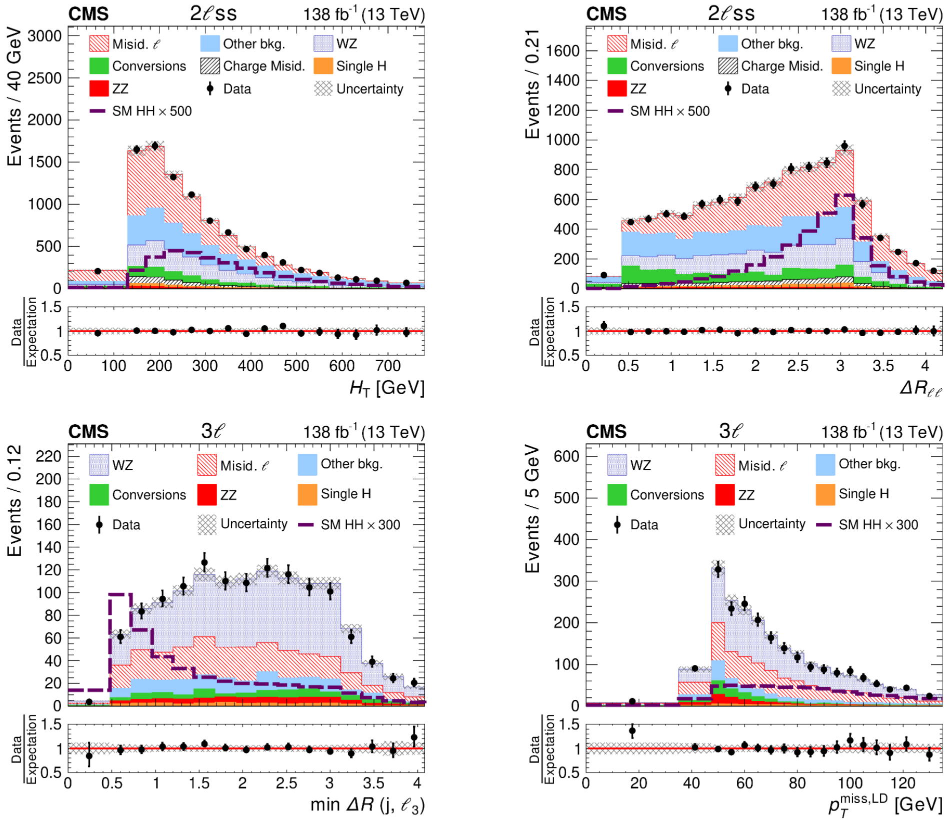


Figure 4. Distributions in a few observables used as inputs to the BDT classifiers in the 2ℓss and 3ℓ categories: the scalar p_{T} sum, denoted as H_{T}, of the two reconstructed ℓ and all small-radius jets in the 2ℓss category (upper left); the angular separation ΔR between the two ℓ in the 2ℓss category (upper right); the angular separation between ℓ_3 and the nearest small-radius jet in the 3ℓ category (lower left); and $p_{\mathrm{T}}^{\text{miss,LD}}$ in the 3ℓ category (lower right). The ℓ_3 in the 3ℓ category is defined as the ℓ that is not part of the opposite-sign $\ell\ell$ pair of lowest mass. The normalization and shape of the distributions expected for the different background processes are shown for the values of nuisance parameters obtained from an ML fit in which the HH signal is constrained to be zero. The gray shaded area indicates the sum of statistical and systematic uncertainties on the background prediction obtained from this ML fit.

$\mathrm{WW^*WW^*}$, $\mathrm{WW^*}\tau\tau$, and $\tau\tau\tau\tau$. Background contributions arising from the misidentification of ℓ and τ_{h} candidates and from the mismeasurement of the electron charge are included in the simulation. The signal and background events used in the training are required to pass the event selection criteria for the respective search category, described in section 5. The number of training events is increased by applying the medium ℓ and τ_{h} identification criteria instead of the tight ones. Weights are applied to background events arising from different sources, such that the relative fractions of different types of back-

JHEP07(2023)095

grounds in the training match the fractions expected in the signal region of the analysis, i.e. when the tight ℓ and τ_h identification criteria are applied. The MC samples used for the BDT training overlap with the samples used to model signal and background contributions in the analysis. To avoid potential biases, the training samples are split into two samples of equal size, based on even and odd event numbers. The BDTs trained on even events are evaluated on odd events, and vice versa, thereby ensuring that BDTs are not trained and evaluated on the same events. The training is performed using the XGBoost algorithm [118], interfaced to the Scikit-learn machine learning library [119]. The parameters of the BDT training (so-called "hyperparameters") are optimized using the particle swarm optimization algorithm described in ref. [120].

7 Background estimation

Background contributions are classified as either "reducible" or "irreducible". In this analysis, three types of reducible backgrounds are considered, arising from misidentified ℓ or τ_h, electron charge misidentification, and electrons from photon conversions. Background events in which all selected ℓ and τ_h come from W, Z, or H boson decays, and are reconstructed with the correct charge, are considered "irreducible". The ℓ/τ_h misidentification and electron charge misidentification backgrounds are both determined from data, while electron conversions and irreducible backgrounds are modeled using MC simulation.

The ℓ/τ_h misidentification background (which includes nonprompt leptons) is the largest reducible background in all search categories. Nonprompt ℓ are either electrons or muons produced in bottom and charm quark decays, or muons that originate from pion and kaon decays. Hadronic jets may also be misidentified as electrons or τ_h. The ℓ/τ_h misidentification background estimate is detailed in section 7.1. The electron charge misidentification background is only relevant for the 2ℓss search category, and is described in section 7.2. The modeling of photon conversion events by the MC simulation has been validated in data as described in refs. [103, 121].

The main contribution to the irreducible background arises from WZ production in the 2ℓss and 3ℓ categories, and ZZ production in the remaining five categories. The production of pairs of bosons (γ, W, Z, or H) other than WZ, ZZ and HH, and production of bosons with top quarks, including Wγ, Zγ, WH, ZH, tH, $\mathrm{t\bar{t}H}$, tW, $\mathrm{t\bar{t}W}$, tZ, $\mathrm{t\bar{t}Z}$, tγ, and $\mathrm{t\bar{t}}\gamma$, constitute subdominant additional backgrounds. The tZ and $\mathrm{t\bar{t}Z}$ backgrounds also include contributions from off-shell $\mathrm{t\bar{t}}\gamma^*$ and tγ^* production. Background processes which include at least one top quark are suppressed by the b jet veto described in section 5, but are still sizable compared to the expected HH signal. All irreducible backgrounds are modeled using the MC simulation.

The modeling of the dominant irreducible WZ and ZZ backgrounds is validated using the "3ℓ WZ" and "4ℓ ZZ" CRs introduced in section 5. Distributions in kinematic observables from these CRs (shown in figure 5) are included in the ML fit that is used to extract the HH signal, described in section 9. This provides in-situ constraints on the WZ and ZZ backgrounds and on systematic uncertainties related to lepton identification and trigger efficiency. The transverse mass, $m_\mathrm{T} = \sqrt{2\, p_\mathrm{T}^{\ell}\, p_\mathrm{T}^{\text{miss}}\, (1 - \cos\Delta\phi)}$, in the 3ℓ WZ CR is computed using the ℓ that is not identified as originating from the Z boson decay. The

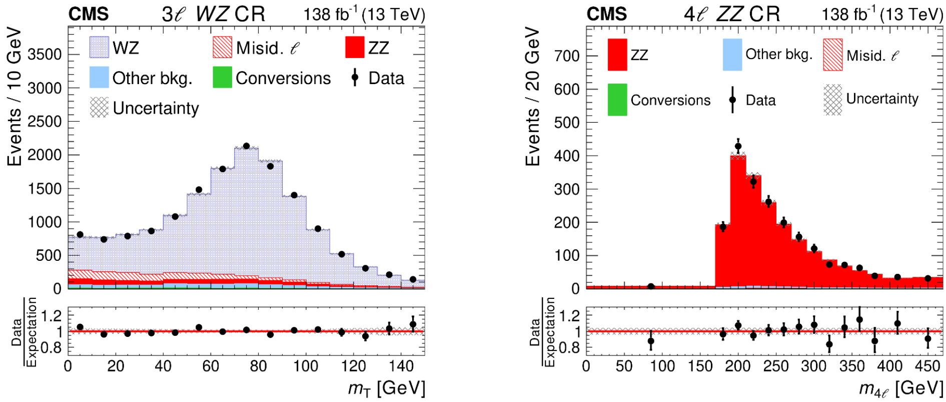


Figure 5. Distributions in m_T in the 3ℓ WZ CR (left) and in $m_{4\ell}$ in the 4ℓ ZZ CR (right). The normalization and shape of the distributions expected for WZ, ZZ, and other background processes are shown for the values of nuisance parameters obtained from the ML fit described in section 9. The gray shaded area indicates the sum of statistical and systematic uncertainties on the background prediction obtained from the ML fit.

symbol $\Delta\phi$ refers to the angle in the transverse plane between the ℓ momentum and the $\vec{p}_\mathrm{T}^\mathrm{miss}$. The observable $m_{4\ell}$ refers to the mass of the 4ℓ system in the 4ℓ ZZ CR.

The modeling of the reducible ℓ/τ_h misidentification background is validated in two further CRs, the "2ℓss CR" and the "$2\ell + 2\tau_\mathrm{h}$ CR". They are based on the signal regions (SRs) of the 2ℓss and $2\ell + 2\tau_\mathrm{h}$ categories. In the 2ℓss CR, no b jet veto is applied, and at least one small-radius jet passing the medium b jet identification is required. The $2\ell + 2\tau_\mathrm{h}$ CR differs from the SR of the $2\ell + 2\tau_\mathrm{h}$ category in that the sum of ℓ plus τ_h charges is required to be non-zero, and no Z boson veto is applied. The 2ℓss CR is dominated by events with misidentified ℓ, while the $2\ell+2\tau_\mathrm{h}$ CR is dominated by events with misidentified τ_h. Distributions in the transverse mass m_T in the 2ℓss CR and in the mass of the HH candidate in the $2\ell + 2\tau_\mathrm{h}$ CR, reconstructed by the algorithm described in ref. [117], are shown in figure 6. The transverse mass in the 2ℓss CR is computed using the leading ℓ. The data agree well with the background prediction in both CRs.

Simulated events are only considered as irreducible background if every selected e, μ, and τ_h candidate matches a prompt MC generator-level counterpart. Events with at least one selected electron from a photon conversion, and the remaining ℓ and τ_h candidates matched to prompt leptons in MC simulation, are classified as conversion background. Electrons that are misidentified as τ_h, and τ_h that are misidentified as e are also modeled using the MC simulation. All other simulated events are discarded, as the ℓ/τ_h misidentification and charge misidentification backgrounds are estimated from data, as described below.

7.1 Lepton and τ_h misidentification background

The background from events with misidentified ℓ and τ_h candidates is estimated using the "fake factor" or "FF" method from ref. [115]. An estimate of this background's contribution

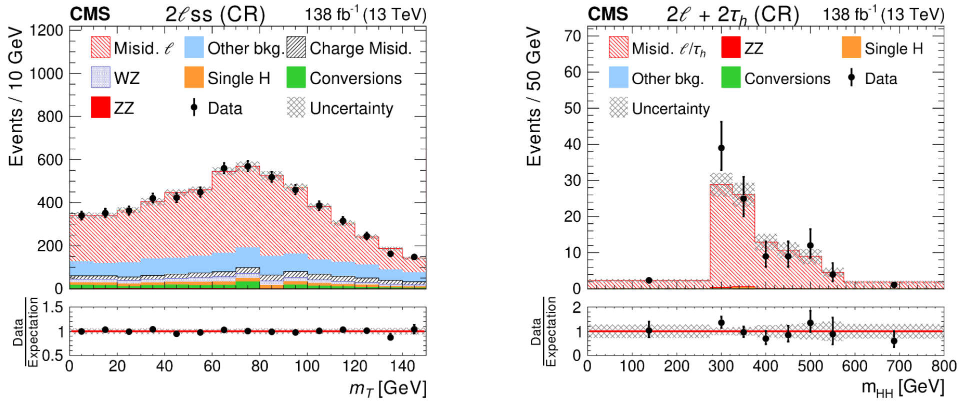


Figure 6. Distributions in m_{T} in the 2ℓss CR (left) and in the mass of the HH candidate in the $2\ell{+}2\tau_{\mathrm{h}}$ CR (right). The normalization and shape of the distributions expected for the misidentified ℓ/τ_{h} background and other background processes are shown for the values of nuisance parameters obtained from an ML fit in which the HH signal is constrained to be zero. The gray shaded area indicates the sum of statistical and systematic uncertainties on the background prediction obtained from this ML fit.

to the SR of each search category is obtained by selecting a sample of events that satisfy all selection criteria of the SR for the respective search category, except that the e, μ, and τ_{h} are required to pass the medium selections instead of the tight ones. The sample of events thus obtained is referred to as the application region (AR) of the FF method. Events in which every ℓ and τ_{h} satisfies the tight selections are excluded from the AR.

The prediction for misidentification backgrounds in the SR is obtained by applying suitably chosen weights w to the events selected in the AR, where w is given by the expression

$$w = (-1)^{n+1} \prod_{i=1}^{n} \frac{f_i(p_{\mathrm{T}}, \eta)}{1 - f_i(p_{\mathrm{T}}, \eta)}. \tag{7.1}$$

The product extends over all e, μ, and τ_{h} that pass the medium, but fail the tight selection criteria, and n refers to the total number of such ℓ and τ_{h}. The symbol $f_i(p_{\mathrm{T}}, \eta)$ corresponds to the probability for a single e, μ, or τ_{h} that passes the medium selection to also pass the tight selection. These probabilities are measured separately for e, μ, and τ_{h} candidates, parametrized as a function of p_{T} and η, and vary between 5 and 30%. The contributions of irreducible backgrounds to the AR are subtracted based on the MC expectation of such processes. The alternating sign in eq. (7.1) is necessary to avoid double-counting arising from events with more than one misidentified ℓ or τ_{h} [115].

The probabilities $f_i(p_{\mathrm{T}}, \eta)$ for electrons and muons are measured in multijet events, as described in ref. [103]. The $f_i(p_{\mathrm{T}}, \eta)$ for τ_{h} are measured using $\mathrm{Z} \to \mu\mu$+jets events, where the misidentified τ_{h} candidates arise from quark or gluon jets. These events are selected by requiring a muon pair passing the tight selection, with opposite charge and invariant mass

JHEP07(2023)095

$60 < m_{\mu\mu} < 120\,\text{GeV}$, plus at least one τ_h candidate that passes the medium τ_h selection. The leading and subleading muons must have $p_T > 25$ and $15\,\text{GeV}$, respectively. Events must also pass the b jet veto described in section 5 to remove $t\bar{t}$ background.

7.2 Charge misidentification background

The electron charge misidentification background in the 2ℓss category is estimated using the method described in ref. [103]. A sample of dielectron events passing all selection criteria of the SR of the 2ℓss category, except that both electrons are required to have opposite- instead of same-sign charge, is selected and assigned appropriately chosen weights. The weights are computed by summing the probabilities for the charge of either electron to be mismeasured. The probability for the mismeasurement of the electron charge is determined using $Z \to ee$ events, and ranges from under 0.1% in the barrel up to 0.4% in the endcap. The probability for mismeasuring the charge of muons is negligible [103].

8 Systematic uncertainties

Multiple sources of systematic uncertainty affect the predicted event yields, the distributions in the output of the BDT classifiers, or both. These uncertainties may be theoretical, affecting the predicted cross section or decay kinematics of the collision process, or experimental, accounting for differences in object reconstruction and calibration between data and the MC simulation, or for uncertainties on the estimates of the ℓ/τ_h misidentification and electron charge misidentification background obtained from data. The systematic uncertainties may be correlated or uncorrelated across the three data-taking years, and among the various signal and background processes considered in the analysis.

The SM prediction for the ggHH production cross section at $\sqrt{s} = 13\,\text{TeV}$ has a relative uncertainty of $+6.7\%/-23.2\%$ [122], while the qqHH cross section uncertainty is $\pm 2.1\%$ [19]. The predicted H boson decay branching fractions to WW^*, $\tau\tau$, and ZZ^* have relative uncertainties of 1.54%, 1.65%, and 1.54%, respectively [123]. Correlations between these uncertainties have a negligible effect. Alternate HH predictions are generated with the renormalization and factorization scales varied up and down by a factor of 2. Variations that increase the factorization scale and decrease the renormalization scale (and vice versa) are excluded, following the recommendation of ref. [123]. All theoretical uncertainties in the HH signal model are correlated across all three data-taking years and among the seven search categories. The uncertainties in the H boson decay branching fractions and the effect of renormalization and factorization scale uncertainties in the signal acceptance impact the measurement of cross sections for both nonresonant and resonant HH production. Conversely, the uncertainties in the SM prediction for the ggHH and qqHH cross sections only affect the measurement of the HH production rate as a ratio to the SM prediction.

Theoretical uncertainties also affect the irreducible background prediction. The relative uncertainties in the cross sections of the dominant WZ and ZZ backgrounds are 2.1 and 6.3%, respectively [124–126]. The uncertainties in the cross sections for the subdominant single H boson backgrounds range from 2 to 9% for ggHH, qqHH, WH, and ZH. The cross sections for the production of W, Z, or H bosons with one or two top quarks are known with uncertainties of 8–15%. The event yields of extremely rare backgrounds not mentioned above (e.g., triple boson or four top quark production) are given a conservative uncertainty of 50%, since the analysis has little sensitivity to these processes. Following ref. [103], background contributions arising from photon conversions are assigned a 30% yield uncertainty. The theoretical uncertainties affecting background cross sections are partially correlated among different processes. Here, contributions arising from uncertainties in the proton PDFs are correlated among processes with a similar initial state. Processes involving single H boson production are an exception. These uncertainties are uncorrelated from other background processes but correlated among each other depending on the initial state. Uncertainties arising from the choice of the renormalization and factorization scales are correlated for processes with similar production modes, for example among all processes involving diboson production (WW, WZ, ZZ, Wγ, and Zγ). Uncertainties in α_s are correlated among all background processes. The theoretical cross section uncertainties for signal processes are uncorrelated with those of background processes, but otherwise follow the same uncertainty scheme for proton PDF, scale, and α_s contributions. All theoretical cross section uncertainties are treated as correlated across the different data-taking years and among all seven search categories.

The rate of the misidentified ℓ/τ_h background is assigned a 30% uncertainty in all search categories, to account for variations in the misidentification rates between the ARs of the FF method and the multijet (Z $\rightarrow \mu\mu$+jets) event samples used to measure the $f_i(p_\mathrm{T}, \eta)$ for e and μ (τ_h). In the $3\ell + 1\tau_\mathrm{h}$ and $1\ell + 3\tau_\mathrm{h}$ categories, an additional uncertainty of 30% (uncorrelated with the other 30% uncertainty) is assigned to the rate of the misidentified ℓ/τ_h background, to account for the extra uncertainty arising from the modified τ_h selection criteria that suppress the misidentification of electrons as τ_h. The effect of statistical uncertainties in the probabilities $f_i(p_\mathrm{T}, \eta)$ for electrons and muons is evaluated by varying these probabilities in bins of p_T and η and determining the resulting change in the shape of the BDT classifier output distribution obtained for the misidentified ℓ/τ_h background. For τ_h, the effect of statistical uncertainties in $f_i(p_\mathrm{T}, \eta)$ is evaluated by fitting the probabilities in bins of η with functions that are linear in p_T, varying the slope of these functions up and down within the uncertainties obtained from the fit, and determining the resulting change in the shape of the BDT classifier output distribution.

An additional uncertainty in the BDT output shape in each category is evaluated for events with a nonprompt or misidentified ℓ or τ_h as follows: Simulated events passing all signal selection criteria are compared to those with at least one ℓ or τ_h candidate failing the tight identification criteria, scaled according to the FF method described in section 7, but with the probabilities $f_i(p_\mathrm{T}, \eta)$ taken from the MC simulation instead of from the data. The ratio of these two shapes is fitted with a linear function, which is convoluted with the misidentified ℓ/τ_h background prediction from the data to serve as an uncertainty in the BDT output shape for these events in the SR. The systematic uncertainties associated with the misidentified ℓ/τ_h background prediction and the uncertainty associated with the electron charge misidentification rate are treated as uncorrelated among the different data-taking years.

The rate of the electron charge misidentification background in the 2ℓss category is assigned a 30% uncertainty. It covers the uncertainty on the electron charge misidentification rates measured in $Z \to ee$ events, including the effect of background contamination in these samples, and accounts for differences observed in the following "closure" test: Simulated events are required to pass all signal selection criteria of the 2ℓss category, except that the two leptons are required to have opposite-sign electric charges. The selected events are scaled according to the electron charge misidentification probability in simulated events, determined by applying the procedure detailed in section 7.2 to MC simulation. The resulting background estimate is compared to the one obtained by applying the nominal signal selection criteria of the 2ℓss category to simulated events.

Uncertainties in the modeling of the trigger and object reconstruction efficiency affect all signal and background processes that are estimated using MC simulation. Trigger efficiencies for events with at least two ℓ are compared between data and MC simulation in control regions enriched in the $t\bar{t}$, WZ, and ZZ background processes, as a function of lepton flavor, p_T, and η. This results in a small p_T-dependent uncertainty correlated between the 2ℓss and $2\ell + 2\tau_h$ categories, and a 1% normalization uncertainty, which is correlated among the 3ℓ, $3\ell + 1\tau_h$, and 4ℓ categories. The data-to-simulation agreement in trigger efficiency for the $4\tau_h$ and $1\ell + 3\tau_h$ categories is computed using an independent set of data, as a function of the p_T and η of the ℓ and all τ_h, and the reconstructed decay modes of all τ_h. The trigger uncertainties for these two categories are treated as uncorrelated. All systematic uncertainties related to trigger modeling are correlated across different physics processes, but uncorrelated among the three data-taking years.

The uncertainties in the reconstruction and identification efficiencies for e, μ, and τ_h candidates have been measured in Z boson enriched regions in data for each level of identification criteria (tight, medium, and loose), and are applied to each event as a function of p_T and η for leptons and of p_T and the reconstructed hadronic decay mode for τ_h. The reconstructed τ_h energy has an uncertainty of around 1%, depending on the data-taking year and reconstructed τ_h decay mode. These uncertainties affect the predicted rate and BDT output shape for signal and background, and are correlated among the different physics processes, but uncorrelated across different data-taking years.

The jet energy scale and resolution are determined using dijet control regions [111, 127]. The jet energy scale is evaluated using 11 separate components, accounting for partial correlations between the data recorded in different years. The jet energy resolution uncertainty is uncorrelated among the three data-taking years. Jet energy uncertainties are also propagated to the p_T^{miss} calculation. An additional uncertainty in $\vec{p}_T^{miss}$ comes from uncertainty in the energy of "unclustered" PF hadrons (PF hadrons not clustered into either small- or large-radius jets), which is uncorrelated across different years. The probability for true b jets to fail the multivariate b jet identification criteria, or for jets from gluons or light flavored quarks to be misidentified as b jets, is compared in data and MC simulation in event regions that are enriched in light-flavor quark or gluon, or heavy-flavor jets. The resulting uncertainty in the data-to-simulation agreement affects the yields and BDT output shapes of multiple physics processes. The statistical component of this uncertainty is treated as uncorrelated across different data-taking years, while other experimental sources are correlated.

The integrated luminosities for data collected in 2016, 2017, and 2018 have 1.2–2.5% individual uncertainties [59–61], while the overall uncertainty for the 2016–2018 period is

1.6%. The uncertainty in the measured cross section for inelastic pp collisions, amounting to 5% [128], is taken into account by varying the number of pileup interactions in MC simulation, which impacts the jet reconstruction and the isolation of ℓ and τ_h.

The sources of systematic uncertainty which create the largest uncertainties in the measured ratio of the HH production cross section to its SM prediction are the theoretical uncertainties in the HH production cross section and decay branching fractions (25%), the uncertainties in the rate and shape of backgrounds from misidentified ℓ or τ_h (22%), and in the rates of backgrounds modeled using MC simulation (13%). These uncertainties in the signal measurement are determined by removing uncertainties that correspond to a given systematic source from an ML fit to pseudodata, as described in section 9, and subtracting the obtained uncertainty in the signal measurement in quadrature from the total uncertainty. The impacts of systematic uncertainties are small compared to the effect of the statistical uncertainty in the data (79%), and are comparable to the statistical uncertainties in the distributions in the BDT classifier output for background processes (33%). The latter includes the effect of statistical uncertainties in the MC simulation and in the ℓ/τ_h misidentification and electron charge misidentification backgrounds obtained from data. All other sources of uncertainty have an impact of 5% or less.

9 Results

The data selected in the seven search categories are tested against multiple HH production hypotheses: the SM prediction; variations of the SM coupling strength modifiers κ_λ, κ_t, κ_V, and κ_{2V}; the effective couplings c_g, c_{2g}, and c_2 in the EFT approach; and resonant production of H boson pairs originating from the decay of heavy particles with spins of 0 or 2 and masses m_X ranging from 250 to 1000 GeV. In each case, the data observed in the seven search categories is fit simultaneously to a model composed of the background prediction (with uncertainties) and the HH signal hypothesis under consideration. The distributions in m_T in the 3ℓ WZ CR and in $m_{4\ell}$ in the 4ℓ ZZ CR shown in figure 5 are included in these fits, in order to obtain in-situ constraints on the systematic uncertainties described in section 8. This in turn reduces the uncertainties in the signal and background predictions.

The SM "signal strength" parameter μ is defined as the ratio of the measured HH production cross section to its predicted value in the SM. This parameter modifies the expected signal yield by the same proportion in each category. By contrast, variations in the κ modifiers may affect the signal yields in each category differently, and also change the BDT classifier output shape for HH events. The twenty benchmark scenarios spanning combinations of κ_λ, κ_t, c_g, c_{2g}, and c_2 values in the coupling parameter space each correspond to different kinematic distributions, so the HH production cross section for each point is measured separately. Similarly, signal efficiency and BDT classifier output shapes vary dramatically for different resonant masses, and thus a separate measurement is performed for each mass and spin hypothesis. The SM signal strength measurement is performed using the output of the BDT classifier that has been trained for SM nonresonant HH production, while the κ_λ measurement uses the BDT trained for benchmark scenario

JHEP04 BM7. In the scenario JHEP04 BM7, the m_{HH} value tends to be close to the lower limit of 250 GeV, which matches the event kinematics for nonresonant HH production in the κ_λ range of the expected limit. When setting limits on the twenty different benchmark scenarios, the binary BDT inputs correspond to the given scenario, or in case of the benchmarks from ref. [96] the kinematically closest scenario. In case of resonant HH production, the BDT input for the resonance mass is set to the m_{X} value for which the limit is computed.

The SM signal strength is measured using a profile likelihood test statistic [129], with systematic uncertainties treated as nuisance parameters θ in a frequentist approach [130]. The effect of variations in θ on the shape of the BDT classifier output distribution for the HH signal and for background processes is incorporated into the ML fit using the technique described in ref. [131]. Statistical uncertainties in these distributions are also taken into account using the approach detailed in ref. [131]. The likelihood ratio q_μ for a fixed "test" signal strength value μ is

$$q_\mu = -2\Delta \ln \mathcal{L} = -2 \ln \frac{\mathcal{L}(\mathrm{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\mathrm{data}|\hat{\mu}, \hat{\theta})},$$

where $\hat{\mu}$ and $\hat{\theta}$ are the signal strength and nuisance parameter values that give the maximum value of the likelihood function $\mathcal{L}$ for the given set of data (requiring $\hat{\mu} \geq 0$), and $\hat{\theta}_\mu$ is the set of θ values which maximize $\mathcal{L}$ for the fixed μ. The 95% confidence level (CL) upper limit for μ is obtained using the $\mathrm{CL_s}$ criterion [132, 133], with q_μ set to 0 when $\mu < \hat{\mu}$. The probabilities to observe a given value of the likelihood ratio q_μ under the signal-plus-background and background-only hypotheses are computed using the asymptotic approximation from ref. [129]. The limits on μ obtained using the asymptotic approximation, match the limits obtained with toy MC experiments [130] within 10%. The SM coupling strength modifiers and the cross sections for the various HH production hypotheses are measured by scanning the likelihood ratio q_μ as a function of μ. Theoretical and experimental uncertainties affecting the signal and background yields or the shape of the BDT classifier output distributions may be correlated or uncorrelated across different years, search categories, and BDT output bins, as described in section 8.

For the case of nonresonant HH production with event kinematics as predicted by the SM, the best-fit value of the HH production rate, obtained from the simultaneous fit of all seven search categories, amounts to $\hat{\mu} = 2 \pm 8\,(\mathrm{stat.}) \pm 6\,(\mathrm{syst.})$ times the SM expectation. The measured value of the signal strength refers to the sum of ggHH and qqHH production and is compatible with both the SM and background-only hypotheses, within statistical and systematic uncertainties. Distributions in the output of the BDT classifier for SM nonresonant HH production in the seven search categories are shown in figures 7 and 8, and the corresponding expected event yields are given in table 7. The data excess in the rightmost bin of the BDT classifier output distribution for the 3ℓ category is not statistically significant: 11 events are observed in this bin, while $5.2 \pm 0.7\,(\mathrm{stat.}) \pm 0.2\,(\mathrm{syst.})$ are expected from background processes, amounting to a local significance of about 1.7 standard deviations. The observed (expected) 95% CL upper limit on the cross section for

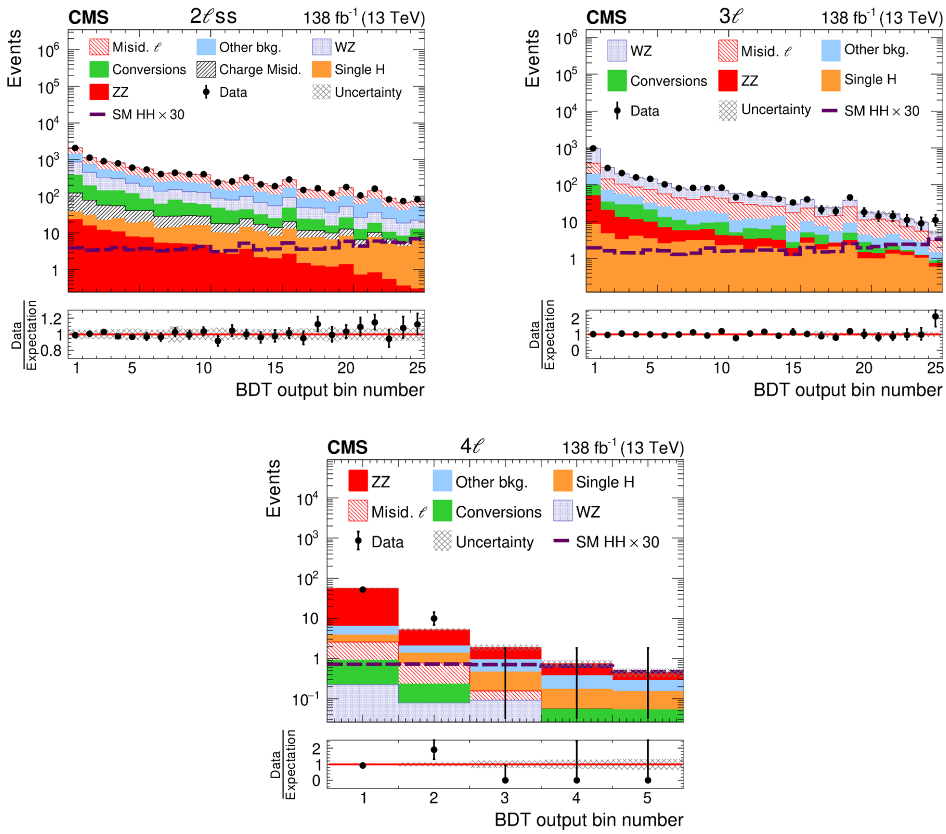


Figure 7. Distribution in the output of the BDT trained for nonresonant HH production and evaluated for the benchmark scenario JHEP04 BM7 for the 2ℓss (upper left), 3ℓ (upper right), and 4ℓ (lower) categories. The SM HH signal is shown for a cross section amounting to 30 times the value predicted in the SM. The normalization and shape of the distributions expected for the background processes are shown for the values of nuisance parameters obtained from the ML fit of the signal+background hypothesis to the data. The gray shaded area indicates the sum of statistical and systematic uncertainties on the background prediction obtained from the ML fit. No data events are observed in the three rightmost bins of the BDT output distribution in the 4ℓ category.

nonresonant HH production is 651 (592) fb. Taking into account the theoretical uncertainties in the SM HH production cross section, this corresponds to an observed (expected) limit on the nonresonant HH production rate of 21.3 (19.4) times the SM expectation. These limits are shown in figure 9 for individual categories and for the combination of all seven search categories, which is referred to as the "HH $\to$ multilepton" result. The 3ℓ and $1\ell + 3\tau_h$ categories are the most sensitive to SM HH production, followed closely by the other categories.

JHEP07(2023)095

Process	2ℓss	3ℓ	4ℓ
SM HH $\to$ WW*WW* ($\times$ 30)	73 $\pm$ 6	33 $\pm$ 3	2.2 $\pm$ 0.2
SM HH $\to$ WW$^*\tau\tau$ ($\times$ 30)	31 $\pm$ 3	12 $\pm$ 1	0.9 $\pm$ 0.1
SM HH $\to$ $\tau\tau\tau\tau$ ($\times$ 30)	3 $\pm$ 0	1 $\pm$ 0	0.1 $\pm$ 0.0
WZ	2003 $\pm$ 58	1321 $\pm$ 27	0.4 $\pm$ 0.1
ZZ	121 $\pm$ 2	109 $\pm$ 2	54.7 $\pm$ 1.8
Misidentified ℓ	3939 $\pm$ 267	670 $\pm$ 55	2.3 $\pm$ 1.0
Conversion electrons	1009 $\pm$ 170	146 $\pm$ 24	0.9 $\pm$ 0.4
Electron charge misid.	366 $\pm$ 52	—	—
Single Higgs boson	216 $\pm$ 4	62 $\pm$ 1	2.4 $\pm$ 0.3
Other backgrounds	2690 $\pm$ 224	293 $\pm$ 20	4.1 $\pm$ 0.4
Total expected background	10 346 $\pm$ 396	2601 $\pm$ 68	64.8 $\pm$ 2.1
Data	10 344	2621	62

Process	$3\ell + 1\tau_h$	$2\ell + 2\tau_h$	$1\ell + 3\tau_h$	$4\tau_h$
SM HH $\to$ WW*WW* ($\times$ 30)	0.9 $\pm$ 0.1	0.2 $\pm$ 0.0	0.2 $\pm$ 0.0	0.3 $\pm$ 0.0
SM HH $\to$ WW$^*\tau\tau$ ($\times$ 30)	4.1 $\pm$ 0.3	3.9 $\pm$ 0.4	0.6 $\pm$ 0.1	0.1 $\pm$ 0.0
SM HH $\to$ $\tau\tau\tau\tau$ ($\times$ 30)	0.9 $\pm$ 0.1	2.3 $\pm$ 0.3	2.6 $\pm$ 0.4	1.3 $\pm$ 0.2
WZ	0.2 $\pm$ 0.0	<0.1	<0.1	<0.1
ZZ	24.3 $\pm$ 0.8	18.5 $\pm$ 1.0	1.9 $\pm$ 0.2	0.7 $\pm$ 0.1
Misidentified ℓ and τ_h	25.1 $\pm$ 4.4	33.5 $\pm$ 4.6	2.1 $\pm$ 1.7	1.5 $\pm$ 0.9
Conversion electrons	0.1 $\pm$ 0.0	0.1 $\pm$ 0.1	<0.1	<0.1
Single Higgs boson	3.8 $\pm$ 0.2	2.9 $\pm$ 0.5	0.8 $\pm$ 0.4	0.4 $\pm$ 0.1
Other backgrounds	2.7 $\pm$ 0.3	2.1 $\pm$ 0.4	0.1 $\pm$ 0.0	<0.1
Total expected background	56.2 $\pm$ 4.5	57.0 $\pm$ 4.8	4.9 $\pm$ 1.7	2.6 $\pm$ 0.9
Data	55	55	6	1

Process	3ℓ WZ CR	4ℓ ZZ CR
WZ	12 546 $\pm$ 148	<1
ZZ	799 $\pm$ 24	2032 $\pm$ 60
Misidentified ℓ	908 $\pm$ 122	13 $\pm$ 4
Conversion electrons	134 $\pm$ 22	3 $\pm$ 0
Other backgrounds	620 $\pm$ 54	59 $\pm$ 6
Total expected background	15 006 $\pm$ 202	2108 $\pm$ 60
Data	14 994	2096

Table 7. The number of expected and observed events in each of the seven search categories, and in two CRs, which validate the modeling of the WZ and ZZ backgrounds. The ℓ/τ_h misidentification and electron charge misidentification backgrounds are determined from data, as described in section 7, while the HH signal and all other backgrounds are modeled using MC simulation. The symbol "—" indicates that the background is not relevant for the category. The HH signal represents the sum of the ggHH and qqHH production processes and is normalized to 30 times the event yield expected in the SM, corresponding to a cross section of about 1 pb. The expected event yields are computed for the values of nuisance parameters obtained from the ML fit described in section 9. Quoted uncertainties represent the sum of statistical and systematic components. Uncertainties that are smaller than half the value of the least significant digit have been rounded to zero.

JHEP07(2023)095

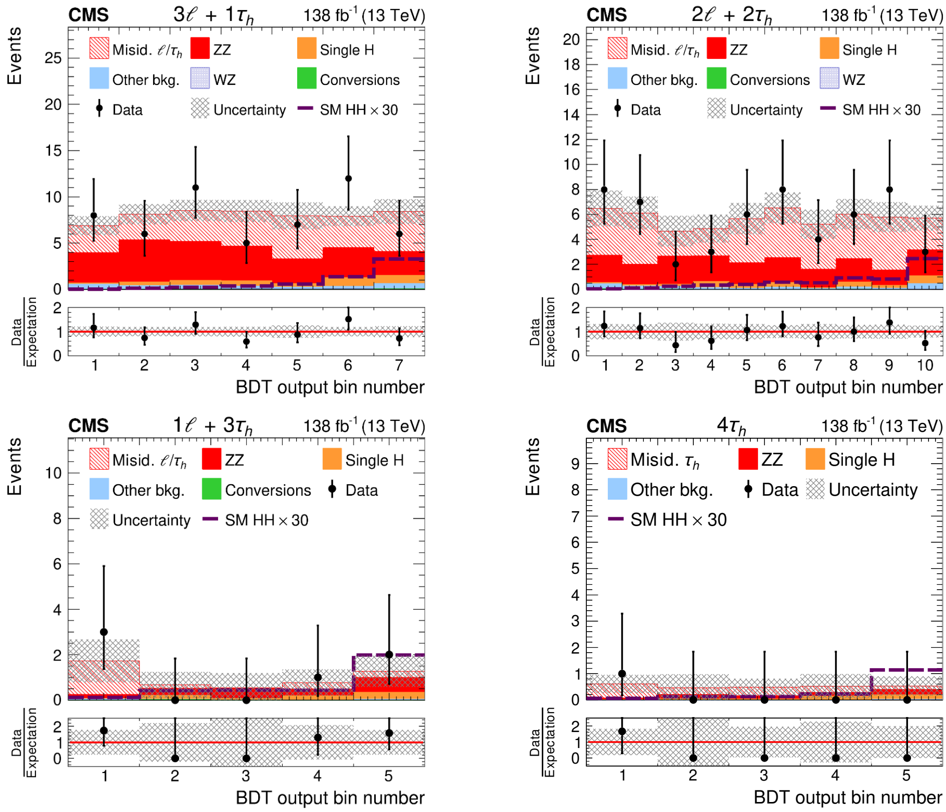


Figure 8. Distribution in the output of the BDT trained for nonresonant HH production and evaluated for the benchmark scenario JHEP04 BM7 for the $3\ell + 1\tau_h$ (upper left), $2\ell + 2\tau_h$ (upper right), $1\ell + 3\tau_h$ (lower left), and $4\tau_h$ (lower right) categories. The SM HH signal is shown for a cross section amounting to 30 times the value predicted in the SM. The normalization and shape of the distributions expected for the background processes are shown for the values of nuisance parameters obtained from the ML fit of the signal+background hypothesis to the data. The gray shaded area indicates the sum of statistical and systematic uncertainties on the background prediction obtained from the ML fit.

The observed (expected) 95% CL interval for the H boson trilinear self-coupling strength modifier is measured to be $-6.9 < \kappa_\lambda < 11.1$ ($-6.9 < \kappa_\lambda < 11.7$). The upper limit on κ_λ is one of the strongest constraints on this fundamental SM parameter to date, with only HH searches in the $bb\gamma\gamma$ [42, 43] and bbbb [45] decay modes providing tighter bounds. The observed and expected upper limits on the HH production cross section as a function of κ_λ, obtained from the simultaneous fit of all seven search categories, are shown in figure 10, along with the limits obtained for each category individually.

The observed and expected limits on the ggHH production cross section for the twenty benchmark scenarios are shown in figure 11 and summarized in table 8. Signal contribu-

JHEP07(2023)095

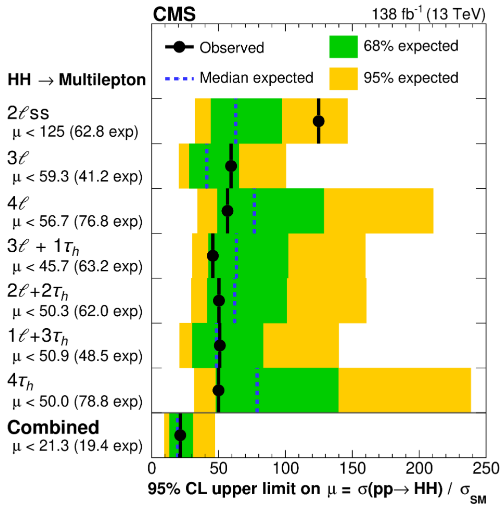


Figure 9. Observed and expected 95% CL upper limits on the SM HH production cross section, obtained for both individual search categories and from a simultaneous fit of all seven categories combined.

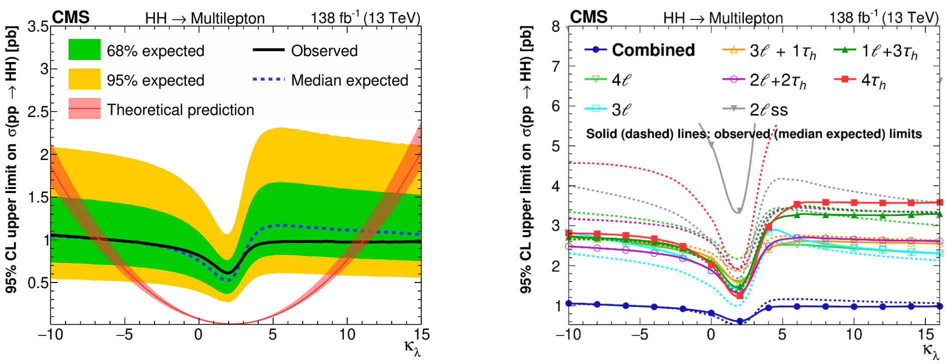


Figure 10. Observed and expected 95% CL upper limits on the HH production cross section as a function of the H boson self-coupling strength modifier κ_λ. All H boson couplings other than λ are assumed to have the values predicted in the SM. The left plot shows the result obtained by combining all seven search categories, while the right plot shows the limits obtained for each category separately. The red curve in the left plot represents the SM prediction for the HH production cross section as a function of κ_λ, and the red shaded band the theoretical uncertainty in this prediction.

JHEP07(2023)095

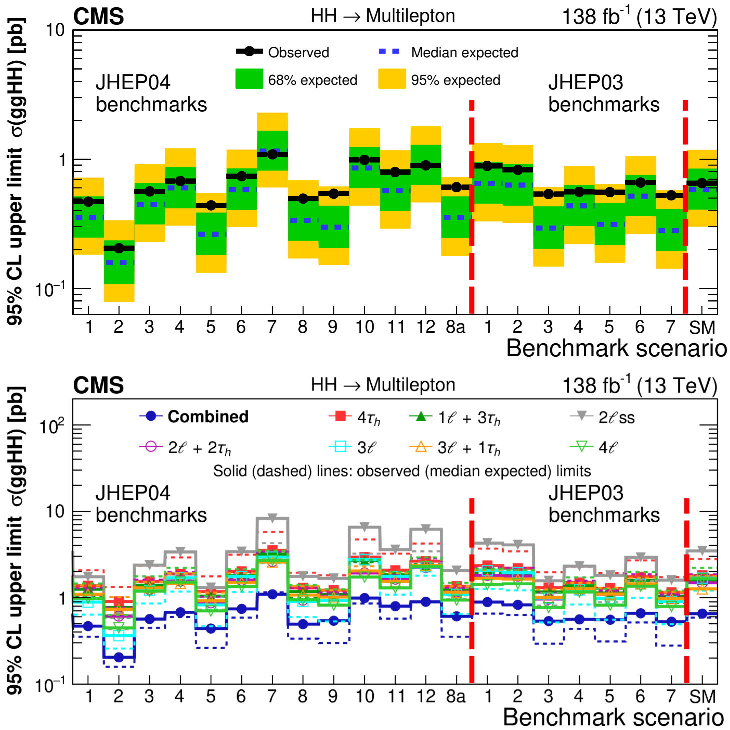


Figure 11. Observed and expected 95% CL upper limits on the HH production cross section for the twelve benchmark scenarios from ref. [24], the additional benchmark scenario 8a from ref. [97], the seven benchmark scenarios from ref. [96], and for the SM. The upper plot shows the result obtained by combining all seven search categories, while the lower plot shows the limits obtained for each category separately, and the combined limit.

tions from the qqHH process, at the rate expected in the SM, are about two orders of magnitude lower than the limits that we set on the rate of the ggHH signal in these measurements and are therefore neglected. The observed (expected) limits on nonresonant HH production in the different benchmark scenarios range from 0.21 to 1.09 (0.16 to 1.16) pb, depending on the scenario. These limits are a factor of 2–3 higher than those obtained by the CMS measurement in the bb$\gamma\gamma$ final state [42]. The variation in expected limits reflects differences in the m_{HH} distribution among the benchmark scenarios, which in turn affect the p_{T} and angles between the particles produced in the H boson decays. As a consequence, the signal acceptance can change, along with the separation of the HH signal from backgrounds through the BDT classifiers described in section 6. The most and least stringent limits on the cross section are expected for the benchmark scenarios JHEP04 BM2 and BM7, respectively. The former has a pronounced tail of the m_{HH} distribution extending to high values, while the latter is characterized by low m_{HH} values, as seen in figure 5 of ref. [24].

JHEP07(2023)095

JHEP04 benchmark	Observed (expected) limit [fb]
BM1	469 (354)
BM2	205 (159)
BM3	563 (447)
BM4	677 (600)
BM5	439 (263)
BM6	739 (584)
BM7	1090 (1156)
BM8	495 (336)
BM9	541 (298)
BM10	988 (855)
BM11	795 (572)
BM12	897 (898)
BM8a	608 (353)
JHEP03 benchmark	**Observed (expected) limit [fb]**
BM1	888 (650)
BM2	828 (632)
BM3	538 (293)
BM4	559 (436)
BM5	556 (313)
BM6	660 (518)
BM7	525 (280)

Table 8. Observed (expected) 95% CL upper limits on the ggHH production cross section for the twelve benchmark scenarios from ref. [24], the additional benchmark scenario 8a from ref. [97] and the seven benchmark scenarios from ref. [96]. The corresponding observed (expected) upper limit for the SM is 652 (583) fb. The limits correspond to the combination of all seven search categories.

Figure 12 shows the observed and expected upper limits on the HH production cross section as a function of the coupling c_2, and the region excluded in the κ_t–c_2 plane. The effects of variations in κ_λ and κ_t on the rate of the SM single H boson background [21] and on the H boson decay branching fractions [20] are taken into account when computing these limits and those shown in figure 10. The magnitude of these effects is typically 5 to 10% within the scanned range of κ_λ and κ_t. Assuming κ_t and κ_λ are both equal to 1, the coupling c_2 is observed (expected) to be constrained to the interval $-1.05 < c_2 < 1.48$ ($-0.96 < c_2 < 1.37$) at 95% CL.

Similar to the right part of figure 12, Figure 13 shows the observed and expected regions excluded in the κ_t–κ_λ and κ_λ–c_2 planes.

Figure 14 shows the observed and expected limits on the resonant HH production cross section as a function of m_X for a spin-0 or spin-2 particle X decaying to HH. The mass points probed are listed in the fourth paragraph of section 3. The limits are expected to become more stringent as m_X increases, as the acceptance for the HH signal increases and the signal can be more easily distinguished from backgrounds. The observed (expected) 95% CL upper limits on the resonant HH production cross section range from 0.18 to 0.90 (0.08 to 1.06) pb, depending on the mass and spin. Tabulated results are provided

JHEP07(2023)095

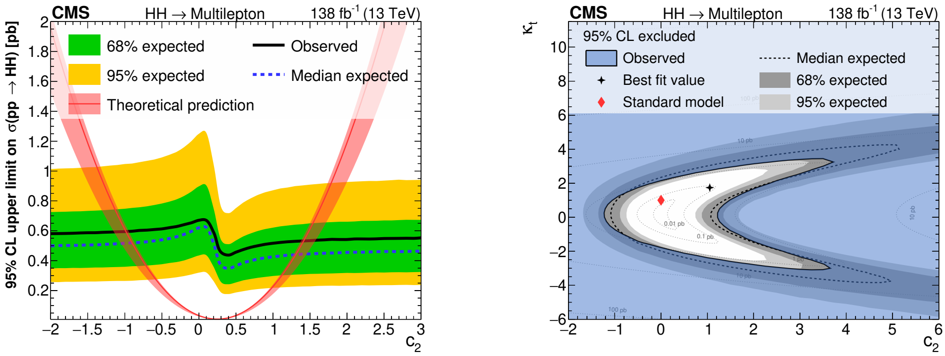


Figure 12. Observed and expected limits on the HH production cross section as a function of the effective coupling c_2 (left), and the region excluded in the κ_t–c_2 plane (right). All limits are computed at 95% CL. H boson couplings other than the ones shown in the plots (c_2 in the left plot and c_2 and κ_t in the right plot) are assumed to have the values predicted by the SM.

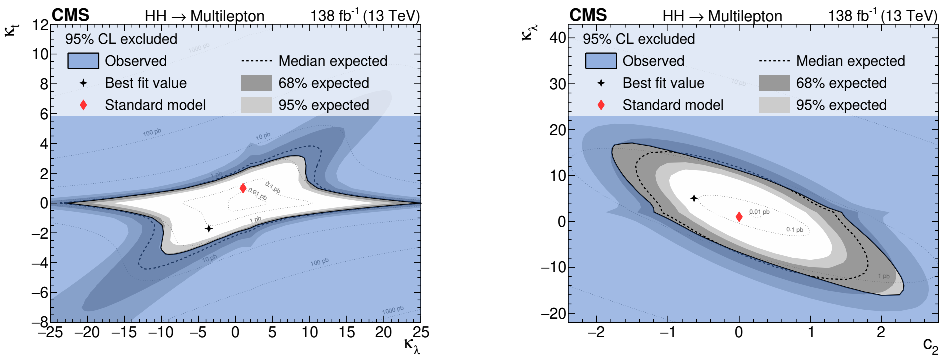


Figure 13. Observed and expected regions excluded in the κ_t–κ_λ (left) and κ_λ–c_2 (right) planes. H boson couplings other than the ones shown in the plots (κ_λ and κ_t in the left plot, and c_2 and κ_λ in the right plot) are assumed to have the values predicted by the SM.

in the HEPData record for this analysis [134]. Only the ATLAS search in the $\mathrm{bb}\gamma\gamma$ final state achieves more stringent limits at low masses (close to 250 GeV) [43], while the low-mass limits from ATLAS in the $\mathrm{bb}\tau\tau$ decay mode are roughly the same [35]. Both these analyses, along with the ATLAS search for bbbb decays [33], set much more stringent limits at higher masses.

For $m_\mathrm{X} \gtrsim 600$ GeV, the observed limit is less stringent than the expected limit, due to a small excess of events in the data that is concentrated near $m_\mathrm{X} = 750$ GeV in the 2ℓss and 3ℓ categories. The distributions in the output of the BDT classifier targeting resonances with spin 2 and mass 750 GeV in the 2ℓss and 3ℓ categories are shown in figure 15. A small excess of events can be seen in the rightmost bin of both distributions. In the 2ℓss (3ℓ) category, 42 (17) events are observed in this bin in the data, while $27.3 \pm 2.8\,(\text{stat.}) \pm 0.7\,(\text{syst.})$ ($8.0 \pm 0.8\,(\text{stat.}) \pm 0.5\,(\text{syst.})$) are expected from background processes, amounting to a

JHEP07(2023)095

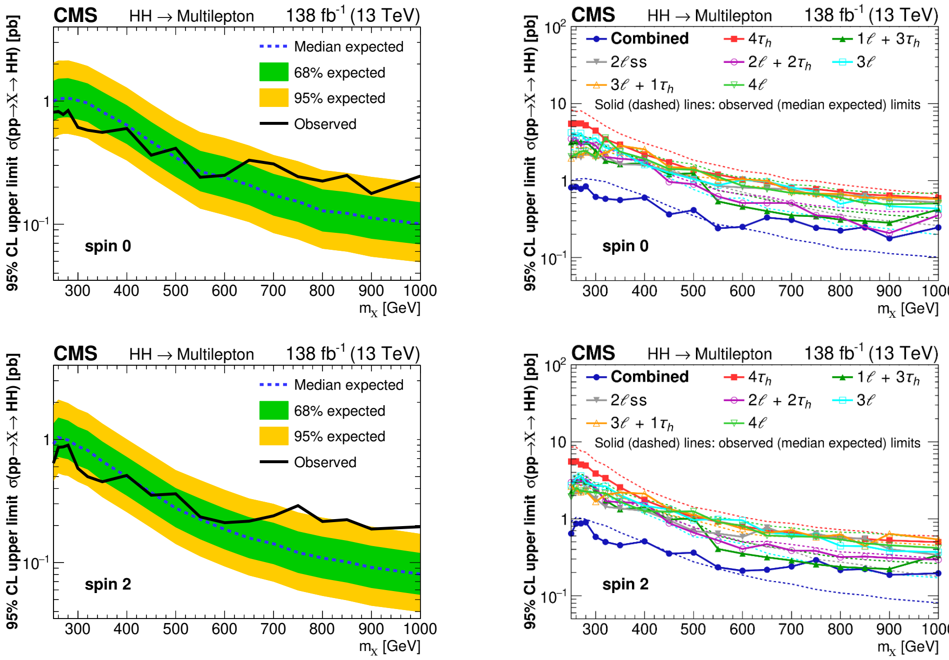


Figure 14. Observed and expected 95% CL upper limits on the production of new particles X of spin 0 (upper) and spin 2 (lower) and mass m_X in the range 250–1000 GeV, which decay to H boson pairs. The plot on the left shows the result obtained by combining all seven search categories, while the plot on the right shows the limits obtained for each category separately, and the combined limit.

local significance of about 2.1 (2.1) standard deviations. The excess affects the observed limits in a broad mass range from 600 to 1000 GeV. No measurement is made for masses above 1000 GeV, as limits on HH decays producing at least one bottom quark pair are much more stringent in this phase space [33, 34]. The presence of multiple neutrinos in HH signal events in these categories, coming from W boson or τ lepton decays, limits the experimental resolution on m_X and causes the BDT classifier output distributions to be highly correlated for resonances of similar mass. No significant excess is observed in any of the other five search categories. The significance for the combination of all seven search categories at 750 GeV amounts to 1.9 standard deviations, without accounting for the “look elsewhere effect” [135].

10 Summary

The results of a search for nonresonant and resonant Higgs boson pair (HH) production in final states with multiple reconstructed leptons, including electrons and muons (ℓ) and hadronically decaying tau leptons (τ_h), has been presented. The search targets the HH decay modes WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$, using proton-proton collision data recorded by the CMS experiment at a center-of-mass energy of 13 TeV and corresponding to an

JHEP07(2023)095

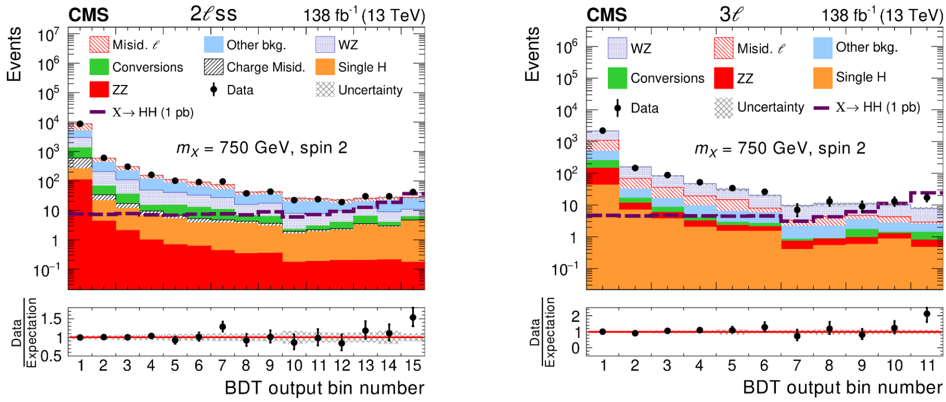


Figure 15. Distribution in BDT classifier output for resonances of spin 2 and mass 750 GeV in the 2ℓss (left) and 3ℓ (right) categories. The resonant HH signal is shown for a cross section amounting to 1 pb. The distributions expected for the background processes are shown for the values of nuisance parameters obtained from the ML fit of the signal+background hypothesis to the data.

integrated luminosity of 138 fb^{-1}. Seven search categories, distinguished by ℓ and τ_h multiplicity, are included in the analysis: 2ℓss, 3ℓ, 4ℓ, $3\ell + 1\tau_h$, $2\ell + 2\tau_h$, $1\ell + 3\tau_h$, and $4\tau_h$, where "ss" indicates an $\ell\ell$ pair with the same charge. No evidence for a signal is found in the data. Upper limits on the cross sections for both nonresonant and resonant HH production are set. The observed (expected) limits on the nonresonant HH production cross section in twenty EFT benchmark scenarios range from 0.21 to 1.09 (0.16 to 1.16) pb at 95% confidence level (CL), depending on the scenario. For nonresonant HH production with event kinematics as predicted by the standard model (SM), the observed (expected) 95% CL upper limit on the HH production rate is 21.3 (19.4) times the rate expected in the SM. The results of the search for nonresonant HH production are used to exclude regions in the plane of the H boson coupling to the top quark, y_t, and of the trilinear Higgs boson self-coupling, λ. Assuming y_t has the value expected in the SM, the observed (expected) 95% CL interval for λ is between -6.9 and 11.1 (-6.9 and 11.7) times the value expected in the SM. The resonant production of H boson pairs, resulting from decays of new heavy particles X with mass m_X, is probed within the mass range 250–1000 GeV. The corresponding observed (expected) 95% CL upper limits on the cross section for resonant HH production range from 0.18 to 0.90 (0.08 to 1.06) pb, depending on the mass and spin of the resonance.

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide

JHEP07(2023)095

LHC Computing Grid and other centers for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC, the CMS detector, and the supporting computing infrastructure provided by the following funding agencies: BMBWF and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, FAPERGS, and FAPESP (Brazil); MES and BNSF (Bulgaria); CERN; CAS, MoST, and NSFC (China); MINCIENCIAS (Colombia); MSES and CSF (Croatia); RIF (Cyprus); SENESCYT (Ecuador); MoER, ERC PUT and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRI (Greece); NKFIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIP and NRF (Republic of Korea); MES (Latvia); LAS (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MOS (Montenegro); MBIE (New Zealand); PAEC (Pakistan); MES and NSC (Poland); FCT (Portugal); MESTD (Serbia); MCIN/AEI and PCTI (Spain); MOSTR (Sri Lanka); Swiss Funding Agencies (Switzerland); MST (Taipei); MHESI and NSTDA (Thailand); TUBITAK and TENMAK (Turkey); NASU (Ukraine); STFC (United Kingdom); DOE and NSF (USA).

Individuals have received support from the Marie-Curie program and the European Research Council and Horizon 2020 Grant, contract Nos. 675440, 724704, 752730, 758316, 765710, 824093, 884104, and COST Action CA16108 (European Union); the Leventis Foundation; the Alfred P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the F.R.S.-FNRS and FWO (Belgium) under the "Excellence of Science — EOS" — be.h project n. 30820817; the Beijing Municipal Science & Technology Commission, No. Z191100007219010; the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Hellenic Foundation for Research and Innovation (HFRI), Project Number 2288 (Greece); the Deutsche Forschungsgemeinschaft (DFG), under Germany's Excellence Strategy — EXC 2121 "Quantum Universe" — 390833306, and under project number 400140256 — GRK2497; the Hungarian Academy of Sciences, the New National Excellence Program — ÚNKP, the NKFIH research grants K 124845, K 124850, K 128713, K 128786, K 129058, K 131991, K 133046, K 138136, K 143460, K 143477, 2020-2.2.1-ED-2021-00181, and TKP2021-NKTA-64 (Hungary); the Council of Science and Industrial Research, India; the Latvian Council of Science; the Ministry of Education and Science, project no. 2022/WK/14, and the National Science Center, contracts Opus 2021/41/B/ST2/01369 and 2021/43/B/ST2/01552 (Poland); the Fundação para a Ciência e a Tecnologia, grant CEECIND/01334/2018 (Portugal); the National Priorities Research Program by Qatar National Research Fund; MCIN/AEI/10.13039/501100011033, ERDF "a way of making Europe", and the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia María de Maeztu, grant MDM-2017-0765 and Programa Severo Ochoa del Principado de Asturias (Spain); the Chulalongkorn Academic into Its 2nd Century Project Advancement Project, and the National Science, Research and Innovation Fund via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation, grant B05F650021

(Thailand); the Kavli Foundation; the Nvidia Corporation; the SuperMicro Corporation; the Welch Foundation, contract C-1845; and the Weston Havens Foundation (USA).

References

[1] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1 [arXiv:1207.7214] [INSPIRE].

[2] CMS collaboration, *Observation of a new boson at a mass of* 125 *GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30 [arXiv:1207.7235] [INSPIRE].

[3] CMS collaboration, *Observation of a new boson with mass near* 125 *GeV in pp collisions at* $\sqrt{s} = 7$ *and* 8 *TeV*, JHEP **06** (2013) 081 [arXiv:1303.4571] [INSPIRE].

[4] CMS collaboration, *Combined measurements of Higgs boson couplings in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 421 [arXiv:1809.10733] [INSPIRE].

[5] ATLAS collaboration, *Combined measurements of Higgs boson production and decay using up to* 80 fb^{-1} *of proton-proton collision data at* $\sqrt{s} = 13$ *TeV collected with the ATLAS experiment*, Phys. Rev. D **101** (2020) 012002 [arXiv:1909.02845] [INSPIRE].

[6] CMS collaboration, *Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production*, Nature Phys. **18** (2022) 1329 [arXiv:2202.06923] [INSPIRE].

[7] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321 [INSPIRE].

[8] P.W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132 [INSPIRE].

[9] P.W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508 [INSPIRE].

[10] G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585 [INSPIRE].

[11] P.W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. **145** (1966) 1156 [INSPIRE].

[12] T.W.B. Kibble, *Symmetry breaking in non-Abelian gauge theories*, Phys. Rev. **155** (1967) 1554 [INSPIRE].

[13] S.L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579 [INSPIRE].

[14] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264 [INSPIRE].

[15] A. Salam, *Weak and electromagnetic interactions*, Conf. Proc. C **680519** (1968) 367 [INSPIRE].

[16] D. de Florian, I. Fabre and J. Mazzitelli, *Triple Higgs production at hadron colliders at NNLO in QCD*, JHEP **03** (2020) 155 [arXiv:1912.02760] [INSPIRE].

[17] O. Aberle et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN-2017-007-M (2017) [DOI:10.23731/CYRM-2017-004] [INSPIRE].

[18] M. Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*, JHEP **05** (2018) 059 [arXiv:1803.02463] [INSPIRE].

[19] F.A. Dreyer and A. Karlberg, *Vector-boson fusion Higgs pair production at N^3LO*, Phys. Rev. D **98** (2018) 114016 [arXiv:1811.07906] [INSPIRE].

[20] G. Degrassi, P.P. Giardino, F. Maltoni and D. Pagani, *Probing the Higgs self coupling via single Higgs production at the LHC*, JHEP **12** (2016) 080 [arXiv:1607.04251] [INSPIRE].

[21] F. Maltoni, D. Pagani, A. Shivaji and X. Zhao, *Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC*, Eur. Phys. J. C **77** (2017) 887 [arXiv:1709.08649] [INSPIRE].

[22] W. Buchmuller and D. Wyler, *Effective Lagrangian analysis of new interactions and flavor conservation*, Nucl. Phys. B **268** (1986) 621 [INSPIRE].

[23] B. Grzadkowski, M. Iskrzynski, M. Misiak and J. Rosiek, *Dimension-six terms in the Standard Model Lagrangian*, JHEP **10** (2010) 085 [arXiv:1008.4884] [INSPIRE].

[24] A. Carvalho et al., *Higgs pair production: choosing benchmarks with cluster analysis*, JHEP **04** (2016) 126 [arXiv:1507.02245] [INSPIRE].

[25] N. Craig, J. Galloway and S. Thomas, *Searching for signs of the second Higgs doublet*, arXiv:1305.2424 [INSPIRE].

[26] D.T. Nhung, M. Mühlleitner, J. Streicher and K. Walz, *Higher order corrections to the trilinear Higgs self-couplings in the real NMSSM*, JHEP **11** (2013) 181 [arXiv:1306.3926] [INSPIRE].

[27] R. Grober and M. Mühlleitner, *Composite Higgs boson pair production at the LHC*, JHEP **06** (2011) 020 [arXiv:1012.1562] [INSPIRE].

[28] R. Contino et al., *Strong double Higgs production at the LHC*, JHEP **05** (2010) 089 [arXiv:1002.1011] [INSPIRE].

[29] C. Englert, T. Plehn, D. Zerwas and P.M. Zerwas, *Exploring the Higgs portal*, Phys. Lett. B **703** (2011) 298 [arXiv:1106.3097] [INSPIRE].

[30] J.M. No and M. Ramsey-Musolf, *Probing the Higgs portal at the LHC through resonant di-Higgs production*, Phys. Rev. D **89** (2014) 095031 [arXiv:1310.6035] [INSPIRE].

[31] L. Randall and R. Sundrum, *A large mass hierarchy from a small extra dimension*, Phys. Rev. Lett. **83** (1999) 3370 [hep-ph/9905221] [INSPIRE].

[32] K.-M. Cheung, *Phenomenology of radion in Randall-Sundrum scenario*, Phys. Rev. D **63** (2001) 056007 [hep-ph/0009232] [INSPIRE].

[33] ATLAS collaboration, *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **105** (2022) 092002 [arXiv:2202.07288] [INSPIRE].

[34] CMS collaboration, *Search for heavy resonances decaying to a pair of Lorentz-boosted Higgs bosons in final states with leptons and a bottom quark pair at $\sqrt{s} = 13$ TeV*, JHEP **05** (2022) 005 [arXiv:2112.03161] [INSPIRE].

[35] ATLAS collaboration, *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector*, arXiv:2209.10910 [INSPIRE].

[36] U. Baur, T. Plehn and D.L. Rainwater, *Measuring the Higgs boson self coupling at the LHC and finite top mass matrix elements*, Phys. Rev. Lett. **89** (2002) 151801 [hep-ph/0206024] [INSPIRE].

[37] U. Baur, T. Plehn and D.L. Rainwater, *Determining the Higgs boson selfcoupling at hadron colliders*, Phys. Rev. D **67** (2003) 033003 [hep-ph/0211224] [INSPIRE].

[38] Q. Li, Z. Li, Q.-S. Yan and X. Zhao, *Probe Higgs boson pair production via the $3\ell 2j+\not{E}$ mode*, Phys. Rev. D **92** (2015) 014015 [arXiv:1503.07611] [INSPIRE].

[39] A. Adhikary et al., *Revisiting the non-resonant Higgs pair production at the HL-LHC*, JHEP **07** (2018) 116 [arXiv:1712.05346] [INSPIRE].

[40] J. Ren et al., *LHC search of new Higgs boson via resonant di-Higgs production with decays into 4W*, JHEP **06** (2018) 090 [arXiv:1706.05980] [INSPIRE].

[41] ATLAS collaboration, *Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$ decay channel using ATLAS data recorded at $\sqrt{s}=13$ TeV*, JHEP **05** (2019) 124 [arXiv:1811.11028] [INSPIRE].

[42] CMS collaboration, *Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s}=13$ TeV*, JHEP **03** (2021) 257 [arXiv:2011.12373] [INSPIRE].

[43] ATLAS collaboration, *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector*, Phys. Rev. D **106** (2022) 052001 [arXiv:2112.11876] [INSPIRE].

[44] ATLAS collaboration, *Search for the $HH \to b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s}=13$ TeV with the ATLAS detector*, JHEP **07** (2020) 108 [*Erratum ibid.* **01** (2021) 145] [arXiv:2001.05178] [INSPIRE].

[45] CMS collaboration, *Search for Higgs boson pair production in the four b quark final state in proton-proton collisions at $\sqrt{s}=13$ TeV*, Phys. Rev. Lett. **129** (2022) 081802 [arXiv:2202.09617] [INSPIRE].

[46] CMS collaboration, *Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks*, arXiv:2205.06667 [INSPIRE].

[47] CMS collaboration, *Search for resonant pair production of Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at 13 TeV*, JHEP **08** (2018) 152 [arXiv:1806.03548] [INSPIRE].

[48] CMS collaboration, *Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s}=13$ TeV*, Phys. Lett. B **842** (2023) 137531 [arXiv:2206.09401] [INSPIRE].

[49] ATLAS collaboration, *Reconstruction and identification of boosted di-τ systems in a search for Higgs boson pairs using 13 TeV proton-proton collision data in ATLAS*, JHEP **11** (2020) 163 [arXiv:2007.14811] [INSPIRE].

[50] CMS collaboration, *Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state in proton-proton collisions at $\sqrt{s}=13$ TeV*, JHEP **01** (2018) 054 [arXiv:1708.04188] [INSPIRE].

JHEP07(2023)095

[51] ATLAS collaboration, *Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **801** (2020) 135145 [arXiv:1908.06765] [INSPIRE].

[52] ATLAS collaboration, *Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **04** (2019) 092 [arXiv:1811.04671] [INSPIRE].

[53] ATLAS collaboration, *Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 1007 [arXiv:1807.08567] [INSPIRE].

[54] CMS collaboration, *Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **122** (2019) 121803 [arXiv:1811.09689] [INSPIRE].

[55] ATLAS collaboration, *Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **800** (2020) 135103 [arXiv:1906.02025] [INSPIRE].

[56] CMS collaboration, *Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV*, 2020 JINST **15** P10017 [arXiv:2006.10165] [INSPIRE].

[57] CMS collaboration, *The CMS trigger system*, 2017 JINST **12** P01020 [arXiv:1609.02366] [INSPIRE].

[58] CMS collaboration, *The CMS Experiment at the CERN LHC*, 2008 JINST **3** S08004 [INSPIRE].

[59] CMS collaboration, *Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS*, Eur. Phys. J. C **81** (2021) 800 [arXiv:2104.01927] [INSPIRE].

[60] CMS collaboration, *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*, CMS-PAS-LUM-17-004, CERN, Geneva (2018).

[61] CMS collaboration, *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*, CMS-PAS-LUM-18-002, CERN, Geneva (2019).

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

[63] A. Kalogeropoulos and J. Alwall, *The SysCalc code: a tool to derive theoretical systematic uncertainties*, arXiv:1801.08401 [INSPIRE].

[64] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040 [hep-ph/0409146] [INSPIRE].

[65] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070 [arXiv:0709.2092] [INSPIRE].

[66] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043 [arXiv:1002.2581] [INSPIRE].

[67] J.M. Campbell and R.K. Ellis, *An Update on vector boson pair production at hadron colliders*, Phys. Rev. D **60** (1999) 113006 [hep-ph/9905386] [INSPIRE].

[68] J.M. Campbell, R.K. Ellis and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **07** (2011) 018 [arXiv:1105.0020] [INSPIRE].

[69] J.M. Campbell, R.K. Ellis and W.T. Giele, *A multi-threaded version of MCFM*, *Eur. Phys. J. C* **75** (2015) 246 [arXiv:1503.06182] [INSPIRE].

[70] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[71] B. Hespel, D. Lopez-Val and E. Vryonidou, *Higgs pair production via gluon fusion in the Two-Higgs-Doublet Model*, *JHEP* **09** (2014) 124 [arXiv:1407.0281] [INSPIRE].

[72] A. Carvalho, *Gravity particles from warped extra dimensions, predictions for LHC*, arXiv:1404.0102 [INSPIRE].

[73] G. Heinrich et al., *NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers*, *JHEP* **08** (2017) 088 [arXiv:1703.09252] [INSPIRE].

[74] G. Heinrich et al., *Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects*, *JHEP* **06** (2019) 066 [arXiv:1903.08137] [INSPIRE].

[75] G. Heinrich, S.P. Jones, M. Kerner and L. Scyboz, *A non-linear EFT description of $gg \to HH$ at NLO interfaced to POWHEG*, *JHEP* **10** (2020) 021 [arXiv:2006.16877] [INSPIRE].

[76] E. Bagnaschi, G. Degrassi, P. Slavich and A. Vicini, *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM*, *JHEP* **02** (2012) 088 [arXiv:1111.2854] [INSPIRE].

[77] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, *JHEP* **02** (2010) 037 [arXiv:0911.5299] [INSPIRE].

[78] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^{\pm}/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, *JHEP* **10** (2013) 083 [arXiv:1306.2542] [INSPIRE].

[79] T. Melia, P. Nason, R. Rontsch and G. Zanderighi, *W^+W^-, WZ and ZZ production in the POWHEG BOX*, *JHEP* **11** (2011) 078 [arXiv:1107.5051] [INSPIRE].

[80] P. Nason and G. Zanderighi, *W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2*, *Eur. Phys. J. C* **74** (2014) 2702 [arXiv:1311.1365] [INSPIRE].

[81] J.M. Campbell, R.K. Ellis and C. Williams, *Bounding the Higgs width at the LHC using full analytic results for $gg \to e^-e^+\mu^-\mu^+$*, *JHEP* **04** (2014) 060 [arXiv:1311.3589] [INSPIRE].

[82] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, *JHEP* **09** (2012) 130 [arXiv:1207.5391] [INSPIRE].

[83] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547 [arXiv:1009.2450] [INSPIRE].

[84] S. Frixione, P. Nason and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126 [arXiv:0707.3088] [INSPIRE].

[85] NNPDF collaboration, *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [arXiv:1410.8849] [INSPIRE].

JHEP07(2023)095

[86] NNPDF collaboration, *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663 [arXiv:1706.00428] [INSPIRE].

[87] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43** (2016) 023001 [arXiv:1510.03865] [INSPIRE].

[88] J. Rojo et al., *The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II*, *J. Phys. G* **42** (2015) 103103 [arXiv:1507.00556] [INSPIRE].

[89] A. Accardi et al., *A critical appraisal and evaluation of modern PDFs*, *Eur. Phys. J. C* **76** (2016) 471 [arXiv:1603.08906] [INSPIRE].

[90] CMS collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements*, *Eur. Phys. J. C* **76** (2016) 155 [arXiv:1512.00815] [INSPIRE].

[91] CMS collaboration, *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*, *Eur. Phys. J. C* **80** (2020) 4 [arXiv:1903.12179] [INSPIRE].

[92] P. Skands, S. Carrazza and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 tune*, *Eur. Phys. J. C* **74** (2014) 3024 [arXiv:1404.5630] [INSPIRE].

[93] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J. C* **53** (2008) 473 [arXiv:0706.2569] [INSPIRE].

[94] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061 [arXiv:1209.6215] [INSPIRE].

[95] GEANT4 collaboration, *GEANT4 — a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 [INSPIRE].

[96] M. Capozi and G. Heinrich, *Exploring anomalous couplings in Higgs boson pair production through shape analysis*, *JHEP* **03** (2020) 091 [arXiv:1908.08923] [INSPIRE].

[97] G. Buchalla et al., *Higgs boson pair production in non-linear Effective Field Theory with full m_t-dependence at NLO QCD*, *JHEP* **09** (2018) 057 [arXiv:1806.05162] [INSPIRE].

[98] A. Carvalho et al., *On the reinterpretation of non-resonant searches for Higgs boson pairs*, *JHEP* **02** (2021) 049 [arXiv:1710.08261] [INSPIRE].

[99] CMS collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, 2017 *JINST* **12** P10003 [arXiv:1706.04965] [INSPIRE].

[100] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[101] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[102] CMS collaboration, *Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC*, 2021 *JINST* **16** P05014 [arXiv:2012.06888] [INSPIRE].

[103] CMS collaboration, *Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **81** (2021) 378 [arXiv:2011.03652] [INSPIRE].

[104] CMS collaboration, *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV*, 2015 *JINST* **10** P06005 [arXiv:1502.02701] [INSPIRE].

[105] CMS collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*, 2018 *JINST* **13** P06015 [arXiv:1804.04528] [INSPIRE].

[106] CMS collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV*, 2018 *JINST* **13** P10005 [arXiv:1809.02816] [INSPIRE].

[107] CMS collaboration, *Identification of hadronic tau lepton decays using a deep neural network*, 2022 *JINST* **17** P07023 [arXiv:2201.08458] [INSPIRE].

[108] Y. LeCun, *Generalization and network design strategies*, in *Connectionism in perspective*, R. Pfeife et al. eds., Zurich, Switzerland (1989).

[109] CMS collaboration, *Pileup mitigation at CMS in 13 TeV data*, 2020 *JINST* **15** P09018 [arXiv:2003.00503] [INSPIRE].

[110] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup per particle identification*, *JHEP* **10** (2014) 059 [arXiv:1407.6013] [INSPIRE].

[111] CMS collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at* 8 *TeV*, 2017 *JINST* **12** P02014 [arXiv:1607.03663] [INSPIRE].

[112] CMS collaboration, *Search for resonances decaying to a pair of Higgs bosons in the $b\bar{b}q\bar{q}'\ell\nu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **10** (2019) 125 [arXiv:1904.04193] [INSPIRE].

[113] E. Bols et al., *Jet flavour classification using DeepJet*, 2020 *JINST* **15** P12012 [arXiv:2008.10519] [INSPIRE].

[114] CMS collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector*, 2019 *JINST* **14** P07004 [arXiv:1903.06078] [INSPIRE].

[115] CMS collaboration, *Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV*, *JHEP* **08** (2018) 066 [arXiv:1803.05485] [INSPIRE].

[116] L. Breiman, J. Friedman, R.A. Olshen and C.J. Stone, *Classification and regression trees*, Chapman and Hall/CRC (1984) [INSPIRE].

[117] K. Ehatäht, L. Marzola and C. Veelken, *Reconstruction of the mass of Higgs boson pairs in events with Higgs boson pairs decaying into four τ leptons*, arXiv:1809.06140 [INSPIRE].

[118] T. Chen and C. Guestrin, *XGBoost: a scalable tree boosting system*, arXiv:1603.02754 [DOI:10.1145/2939672.2939785] [INSPIRE].

[119] F. Pedregosa et al., *Scikit-learn: machine learning in Python*, *J. Machine Learning Res.* **12** (2011) 2825 [arXiv:1201.0490] [INSPIRE].

[120] L. Tani, D. Rand, C. Veelken and M. Kadastik, *Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics*, *Eur. Phys. J. C* **81** (2021) 170 [arXiv:2011.04434] [INSPIRE].

[121] CMS collaboration, *Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **03** (2018) 166 [arXiv:1709.05406] [INSPIRE].

JHEP07(2023)095

[122] J. Baglio et al., *$gg \to HH$: combined uncertainties*, *Phys. Rev. D* **103** (2021) 056002 [arXiv:2008.11626] [INSPIRE].

[123] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the nature of the Higgs sector*, arXiv:1610.07922 [DOI:10.23731/CYRM-2017-002] [INSPIRE].

[124] M. Grazzini, S. Kallweit, D. Rathlev and M. Wiesemann, *$W^{\pm}Z$ production at hadron colliders in NNLO QCD*, *Phys. Lett. B* **761** (2016) 179 [arXiv:1604.08576] [INSPIRE].

[125] F. Cascioli et al., *ZZ production at hadron colliders in NNLO QCD*, *Phys. Lett. B* **735** (2014) 311 [arXiv:1405.2219] [INSPIRE].

[126] F. Caola, K. Melnikov, R. Röntsch and L. Tancredi, *QCD corrections to ZZ production in gluon fusion at the LHC*, *Phys. Rev. D* **92** (2015) 094028 [arXiv:1509.06734] [INSPIRE].

[127] CMS collaboration, *Jet algorithms performance in 13 TeV data*, CMS-PAS-JME-16-003, CERN, Geneva (2017).

[128] CMS collaboration, *Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV*, *JHEP* **07** (2018) 161 [arXiv:1802.02613] [INSPIRE].

[129] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [*Erratum ibid.* **73** (2013) 2501] [arXiv:1007.1727] [INSPIRE].

[130] ATLAS, CMS collaboration and L.H.C.H.C. Group, *Procedure for the LHC Higgs boson search combination in summer 2011*, ATL-PHYS-PUB-2011-011 (2011).

[131] J.S. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, in the proceedings of the *PHYSTAT 2011*, (2011), p. 115–120 [DOI:10.5170/CERN-2011-006.115] [arXiv:1103.0354] [INSPIRE].

[132] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth. A* **434** (1999) 435 [hep-ex/9902006] [INSPIRE].

[133] A.L. Read, *Presentation of search results: The CL_s technique*, *J. Phys. G* **28** (2002) 2693 [INSPIRE].

[134] CMS collaboration, *Search for Higgs boson pairs decaying to WWWW, $WW\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV*, https://www.hepdata.net/record/ins2098277 [DOI:10.17182/HEPDATA.130795].

[135] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, *Eur. Phys. J. C* **70** (2010) 525 [arXiv:1005.1891] [INSPIRE].

JHEP07(2023)095

The CMS collaboration

Yerevan Physics Institute, Yerevan, Armenia

A. Tumasyan[1]

Institut für Hochenergiephysik, Vienna, Austria

W. Adam, J.W. Andrejkovic, T. Bergauer, S. Chatterjee, K. Damanakis, M. Dragicevic, A. Escalante Del Valle, M. Jeitler[2], N. Krammer, L. Lechner, D. Liko, I. Mikulec, P. Paulitsch, F.M. Pitters, J. Schieck[2], R. Schöfbeck, D. Schwarz, S. Templ, W. Waltenberger, C.-E. Wulz[2]

Universiteit Antwerpen, Antwerpen, Belgium

M.R. Darwish[3], E.A. De Wolf, T. Janssen, T. Kello[4], A. Lelek, H. Rejeb Sfar, P. Van Mechelen, N. Van Remortel

Vrije Universiteit Brussel, Brussel, Belgium

E.S. Bols, J. D'Hondt, A. De Moor, M. Delcourt, H. El Faham, S. Lowette, S. Moortgat, A. Morton, D. Müller, A.R. Sahasransu, S. Tavernier, W. Van Doninck, D. Vannerom

Université Libre de Bruxelles, Bruxelles, Belgium

D. Beghin, B. Clerbaux, G. De Lentdecker, L. Favart, J. Jaramillo, K. Lee, M. Mahdavikhorrami, I. Makarenko, A. Malara, S. Paredes, L. Pétré, A. Popov, N. Postiau, E. Starling, L. Thomas, M. Vanden Bemden, C. Vander Velde, P. Vanlaer

Ghent University, Ghent, Belgium

D. Dobur, J. Knolle, L. Lambrecht, G. Mestdach, M. Niedziela, C. Rendón, C. Roskas, A. Samalan, K. Skovpen, M. Tytgat, N. Van Den Bossche, B. Vermassen, L. Wezenbeek

Université Catholique de Louvain, Louvain-la-Neuve, Belgium

A. Benecke, A. Bethani, G. Bruno, F. Bury, C. Caputo, P. David, C. Delaere, I.S. Donertas, A. Giammanco, K. Jaffel, Sa. Jain, V. Lemaitre, K. Mondal, J. Prisciandaro, A. Taliercio, T.T. Tran, P. Vischia, S. Wertz

Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil

G.A. Alves, E. Coelho, C. Hensel, A. Moraes, P. Rebello Teles

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

W.L. Aldá Júnior, M. Alves Gallo Pereira, M. Barroso Ferreira Filho, H. Brandao Malbouisson, W. Carvalho, J. Chinellato[5], E.M. Da Costa, G.G. Da Silveira[6], D. De Jesus Damiao, V. Dos Santos Sousa, S. Fonseca De Souza, J. Martins[7], C. Mora Herrera, K. Mota Amarilo, L. Mundim, H. Nogima, A. Santoro, S.M. Silva Do Amaral, A. Sznajder, M. Thiel, F. Torres Da Silva De Araujo[8], A. Vilela Pereira

JHEP07(2023)095

Universidade Estadual Paulista, Universidade Federal do ABC, São Paulo, Brazil

C.A. Bernardes[6], L. Calligaris, T.R. Fernandez Perez Tomei, E.M. Gregores, P.G. Mercadante, S.F. Novaes, Sandra S. Padula

Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria

A. Aleksandrov, G. Antchev, R. Hadjiiska, P. Iaydjiev, M. Misheva, M. Rodozov, M. Shopova, G. Sultanov

University of Sofia, Sofia, Bulgaria

A. Dimitrov, T. Ivanov, L. Litov, B. Pavlov, P. Petkov, A. Petrov

Beihang University, Beijing, China

T. Cheng, T. Javaid[9], M. Mittal, L. Yuan

Department of Physics, Tsinghua University, Beijing, China

M. Ahmad, G. Bauer[10], Z. Hu, K. Yi[10,11]

Institute of High Energy Physics, Beijing, China

G.M. Chen[9], H.S. Chen[9], M. Chen[9], F. Iemmi, C.H. Jiang, A. Kapoor, H. Liao, Z.-A. Liu[12], V. Milosevic, F. Monti, R. Sharma, J. Tao, J. Thomas-Wilsker, J. Wang, H. Zhang, J. Zhao

State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China

A. Agapitos, Y. An, Y. Ban, C. Chen, A. Levin, Q. Li, X. Lyu, Y. Mao, S.J. Qian, X. Sun, D. Wang, J. Xiao, H. Yang

Sun Yat-Sen University, Guangzhou, China

M. Lu, Z. You

Institute of Modern Physics and Key Laboratory of Nuclear Physics and Ion-beam Application (MOE) — Fudan University, Shanghai, China

X. Gao[4], D. Leggat, H. Okawa, Y. Zhang

Zhejiang University, Hangzhou, Zhejiang, China

Z. Lin, M. Xiao

Universidad de Los Andes, Bogota, Colombia

C. Avila, A. Cabrera, C. Florez, J. Fraga

Universidad de Antioquia, Medellin, Colombia

J. Mejia Guisao, F. Ramirez, J.D. Ruiz Alvarez

University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia

D. Giljanovic, N. Godinovic, D. Lelas, I. Puljak

JHEP07(2023)095

University of Split, Faculty of Science, Split, Croatia
Z. Antunovic, M. Kovac, T. Sculac

Institute Rudjer Boskovic, Zagreb, Croatia
V. Brigljevic, B.K. Chitroda, D. Ferencek, D. Majumder, M. Roguljic, A. Starodumov[13], T. Susa

University of Cyprus, Nicosia, Cyprus
A. Attikis, K. Christoforou, G. Kole, M. Kolosova, S. Konstantinou, J. Mousa, C. Nicolaou, F. Ptochos, P.A. Razis, H. Rykaczewski, H. Saka

Charles University, Prague, Czech Republic
M. Finger[13], M. Finger Jr.[13], A. Kveton

Escuela Politecnica Nacional, Quito, Ecuador
E. Ayala

Universidad San Francisco de Quito, Quito, Ecuador
E. Carrera Jarrin

Academy of Scientific Research and Technology of the Arab Republic of Egypt, Egyptian Network of High Energy Physics, Cairo, Egypt
H. Abdalla[14], Y. Assran[15,16]

Center for High Energy Physics (CHEP-FU), Fayoum University, El-Fayoum, Egypt
M.A. Mahmoud, Y. Mohammed

National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
S. Bhowmik, R.K. Dewanjee, K. Ehataht, M. Kadastik, S. Nandan, C. Nielsen, J. Pata, M. Raidal, L. Tani, C. Veelken

Department of Physics, University of Helsinki, Helsinki, Finland
P. Eerola, H. Kirschenmann, K. Osterberg, M. Voutilainen

Helsinki Institute of Physics, Helsinki, Finland
S. Bharthuar, E. Brücken, F. Garcia, J. Havukainen, M.S. Kim, R. Kinnunen, T. Lampén, K. Lassila-Perini, S. Lehti, T. Lindén, M. Lotti, L. Martikainen, M. Myllymäki, J. Ott, M.m. Rantanen, H. Siikonen, E. Tuominen, J. Tuominiemi

Lappeenranta-Lahti University of Technology, Lappeenranta, Finland
P. Luukka, H. Petrow, T. Tuuva

IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
C. Amendola, M. Besancon, F. Couderc, M. Dejardin, D. Denegri, J.L. Faure, F. Ferri, S. Ganjour, P. Gras, G. Hamel de Monchenault, P. Jarry, V. Lohezic, J. Malcles, J. Rander, A. Rosowsky, M.Ö. Sahin, A. Savoy-Navarro[17], P. Simkina, M. Titov

JHEP07(2023)095

Laboratoire Leprince-Ringuet, CNRS/IN2P3, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France

S. Ahuja, C. Baldenegro Barrera, F. Beaudette, M. Bonanomi, A. Buchot Perraguin, P. Busson, A. Cappati, C. Charlot, O. Davignon, B. Diab, G. Falmagne, B.A. Fontana Santos Alves, S. Ghosh, R. Granier de Cassagnac, A. Hakimi, B. Harikrishnan, J. Motta, M. Nguyen, C. Ochando, L. Portales, J. Rembser, R. Salerno, U. Sarkar, J.B. Sauvan, Y. Sirois, A. Tarabini, E. Vernazza, A. Zabi, A. Zghiche

Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

J.-L. Agram[18], J. Andrea, D. Apparu, D. Bloch, G. Bourgatte, J.-M. Brom, E.C. Chabert, C. Collard, D. Darej, U. Goerlach, C. Grimault, A.-C. Le Bihan, E. Nibigira, P. Van Hove

Institut de Physique des 2 Infinis de Lyon (IP2I), Villeurbanne, France

S. Beauceron, C. Bernet, G. Boudoul, C. Camen, A. Carle, N. Chanon, D. Contardo, P. Depasse, H. El Mamouni, J. Fay, S. Gascon, M. Gouzevitch, G. Grenier, B. Ille, I.B. Laktineh, M. Lethuillier, L. Mirabito, S. Perries, K. Shchablo, V. Sordini, L. Torterotot, M. Vander Donckt, S. Viret

Georgian Technical University, Tbilisi, Georgia

I. Bagaturia[19], I. Lomidze, Z. Tsamalaidze[13]

RWTH Aachen University, I. Physikalisches Institut, Aachen, Germany

V. Botta, L. Feld, K. Klein, M. Lipinski, D. Meuser, A. Pauls, N. Röwert, M. Teroerde

RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany

A. Dodonova, N. Eich, D. Eliseev, M. Erdmann, P. Fackeldey, B. Fischer, T. Hebbeker, K. Hoepfner, F. Ivone, M.y. Lee, L. Mastrolorenzo, M. Merschmeyer, A. Meyer, S. Mondal, S. Mukherjee, D. Noll, A. Novak, A. Pozdnyakov, Y. Rath, H. Reithler, A. Schmidt, S.C. Schuler, A. Sharma, L. Vigilante, S. Wiedenbeck, S. Zaleski

RWTH Aachen University, III. Physikalisches Institut B, Aachen, Germany

C. Dziwok, G. Flügge, W. Haj Ahmad[20], O. Hlushchenko, T. Kress, A. Nowack, O. Pooth, A. Stahl[21], T. Ziemons, A. Zotz

Deutsches Elektronen-Synchrotron, Hamburg, Germany

H. Aarup Petersen, M. Aldaya Martin, P. Asmuss, S. Baxter, M. Bayatmakou, O. Behnke, A. Bermúdez Martínez, S. Bhattacharya, A.A. Bin Anuar, F. Blekman[22], K. Borras[23], D. Brunner, A. Campbell, A. Cardini, C. Cheng, F. Colombina, S. Consuegra Rodríguez, G. Correia Silva, M. De Silva, L. Didukh, G. Eckerlin, D. Eckstein, L.I. Estevez Banos, O. Filatov, E. Gallo[22], A. Geiser, A. Giraldi, G. Greau, A. Grohsjean, V. Guglielmi, M. Guthoff, A. Jafari[24], N.Z. Jomhari, B. Kaech, A. Kasem[23], M. Kasemann, H. Kaveh, C. Kleinwort, R. Kogler,

D. Krücker, W. Lange, K. Lipka, W. Lohmann[25], R. Mankel, I.-A. Melzer-Pellmann, M. Mendizabal Morentin, J. Metwally, A.B. Meyer, G. Milella, M. Mormile, A. Mussgiller, A. Nürnberg, Y. Otarid, D. Pérez Adán, A. Raspereza, B. Ribeiro Lopes, J. Rübenach, A. Saggio, A. Saibel, M. Savitskyi, M. Scham[26,23], V. Scheurer, S. Schnake[23], P. Schütze, C. Schwanenberger[22], M. Shchedrolosiev, R.E. Sosa Ricardo, D. Stafford, N. Tonon[†], M. Van De Klundert, F. Vazzoler, R. Walsh, D. Walter, Q. Wang, Y. Wen, K. Wichmann, L. Wiens[23], C. Wissing, S. Wuchterl, A. Zimermmane Castro Santos

University of Hamburg, Hamburg, Germany
R. Aggleton, S. Albrecht, S. Bein, L. Benato, P. Connor, K. De Leo, M. Eich, K. El Morabit, F. Feindt, A. Fröhlich, C. Garbers, E. Garutti, M. Hajheidari, J. Haller, A. Hinzmann, G. Kasieczka, R. Klanner, W. Korcari, T. Kramer, V. Kutzner, J. Lange, T. Lange, A. Lobanov, C. Matthies, A. Mehta, L. Moureaux, M. Mrowietz, A. Nigamova, K.J. Pena Rodriguez, M. Rieger, O. Rieger, P. Schleper, M. Schröder, J. Schwandt, H. Stadie, G. Steinbrück, A. Tews, M. Wolf

Karlsruher Institut fuer Technologie, Karlsruhe, Germany
J. Bechtel, S. Brommer, M. Burkart, E. Butz, R. Caspart, T. Chwalek, A. Dierlamm, A. Droll, N. Faltermann, M. Giffels, J.O. Gosewisch, A. Gottmann, F. Hartmann[21], C. Heidecker, M. Horzela, U. Husemann, P. Keicher, R. Koppenhöfer, S. Maier, S. Mitra, Th. Müller, M. Neukum, G. Quast, K. Rabbertz, J. Rauser, D. Savoiu, M. Schnepf, D. Seith, I. Shvetsov, H.J. Simonis, R. Ulrich, J. van der Linden, R.F. Von Cube, M. Wassmer, M. Weber, S. Wieland, R. Wolf, S. Wozniewski, S. Wunsch

Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece
G. Anagnostou, P. Assiouras, G. Daskalakis, A. Kyriakis, A. Stakia

National and Kapodistrian University of Athens, Athens, Greece
M. Diamantopoulou, D. Karasavvas, P. Kontaxakis, A. Manousakis-Katsikakis, A. Panagiotou, I. Papavergou, N. Saoulidou, K. Theofilatos, E. Tziaferi, K. Vellidis, E. Vourliotis

National Technical University of Athens, Athens, Greece
G. Bakas, K. Kousouris, I. Papakrivopoulos, G. Tsipolitis, A. Zacharopoulou

University of Ioánnina, Ioánnina, Greece
K. Adamidis, I. Bestintzanos, I. Evangelou, C. Foudas, P. Gianneios, P. Katsoulis, P. Kokkas, N. Manthos, I. Papadopoulos, J. Strologas

MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
M. Csanád, K. Farkas, M.M.A. Gadallah[27], S. Lökös[28], P. Major, K. Mandal, G. Pásztor, A.J. Rádl, O. Surányi, G.I. Veres

JHEP07(2023)095

Wigner Research Centre for Physics, Budapest, Hungary
M. Bartók[29], G. Bencze, C. Hajdu, D. Horvath[30,31], F. Sikler, V. Veszpremi

Institute of Nuclear Research ATOMKI, Debrecen, Hungary
N. Beni, S. Czellar, D. Fasanella, J. Karancsi[29], J. Molnar, Z. Szillasi, D. Teyssier

Institute of Physics, University of Debrecen, Debrecen, Hungary
P. Raics, B. Ujvari[32], G. Zilizi

Karoly Robert Campus, MATE Institute of Technology, Gyongyos, Hungary
T. Csorgo[33], F. Nemes[33], T. Novak

Panjab University, Chandigarh, India
J. Babbar, S. Bansal, S.B. Beri, V. Bhatnagar, G. Chaudhary, S. Chauhan, N. Dhingra[34], R. Gupta, A. Kaur, H. Kaur, M. Kaur, P. Kumari, M. Meena, K. Sandeep, J.B. Singh[35], A. K. Virdi

University of Delhi, Delhi, India
A. Ahmed, A. Bhardwaj, B.C. Choudhary, M. Gola, S. Keshri, A. Kumar, M. Naimuddin, P. Priyanka, K. Ranjan, S. Saumya, A. Shah

Saha Institute of Nuclear Physics, HBNI, Kolkata, India
R. Bhattacharya, S. Bhattacharya, D. Bhowmik, S. Dutta, S. Dutta, B. Gomber[36], M. Maity[37], P. Palit, P.K. Rout, G. Saha, B. Sahu, S. Sarkar

Indian Institute of Technology Madras, Madras, India
P.K. Behera, S.C. Behera, P. Kalbhor, J.R. Komaragiri[38], D. Kumar[38], A. Muhammad, L. Panwar[38], R. Pradhan, P.R. Pujahari, A. Sharma, A.K. Sikdar, P.C. Tiwari[38]

Bhabha Atomic Research Centre, Mumbai, India
K. Naskar[39]

Tata Institute of Fundamental Research-A, Mumbai, India
T. Aziz, S. Dugad, M. Kumar, G.B. Mohanty, P. Suryadevara

Tata Institute of Fundamental Research-B, Mumbai, India
S. Banerjee, R. Chudasama, M. Guchait, S. Karmakar, S. Kumar, G. Majumder, K. Mazumdar, S. Mukherjee

National Institute of Science Education and Research, An OCC of Homi Bhabha National Institute, Bhubaneswar, Odisha, India
S. Bahinipati[40], C. Kar, P. Mal, T. Mishra, V.K. Muraleedharan Nair Bindhu[41], A. Nayak[41], P. Saha, N. Sur, S.K. Swain, D. Vats[41]

Indian Institute of Science Education and Research (IISER), Pune, India
A. Alpana, S. Dube, B. Kansal, A. Laha, S. Pandey, A. Rastogi, S. Sharma

JHEP07(2023)095

Isfahan University of Technology, Isfahan, Iran
H. Bakhshiansohi[42], E. Khazaie, M. Zeinali[43]

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
S. Chenarani[44], S.M. Etesami, M. Khakzad, M. Mohammadi Najafabadi

University College Dublin, Dublin, Ireland
M. Grunewald

INFN Sezione di Bari[a], Università di Bari[b], Politecnico di Bari[c], Bari, Italy
M. Abbrescia[a,b], R. Aly[a,c,45], C. Aruta[a,b], A. Colaleo[a], D. Creanza[a,c], N. De Filippis[a,c], M. De Palma[a,b], A. Di Florio[a,b], W. Elmetenawee[a,b], F. Errico[a,b], L. Fiore[a], G. Iaselli[a,c], M. Ince[a,b], S. Lezki[a,b], G. Maggi[a,c], M. Maggi[a], I. Margjeka[a,b], V. Mastrapasqua[a,b], S. My[a,b], S. Nuzzo[a,b], A. Pellecchia[a,b], A. Pompili[a,b], G. Pugliese[a,c], R. Radogna[a], D. Ramos[a], A. Ranieri[a], G. Selvaggi[a,b], L. Silvestris[a], F.M. Simone[a,b], Ü. Sözbilir[a], R. Venditti[a], P. Verwilligen[a]

INFN Sezione di Bologna[a], Università di Bologna[b], Bologna, Italy
G. Abbiendi[a], C. Battilana[a,b], D. Bonacorsi[a,b], L. Borgonovi[a], L. Brigliadori[a], R. Campanini[a,b], P. Capiluppi[a,b], A. Castro[a,b], F.R. Cavallo[a], M. Cuffiani[a,b], G.M. Dallavalle[a], T. Diotalevi[a,b], F. Fabbri[a], A. Fanfani[a,b], P. Giacomelli[a], L. Giommi[a,b], C. Grandi[a], L. Guiducci[a,b], S. Lo Meo[a,46], L. Lunerti[a,b], S. Marcellini[a], G. Masetti[a], F.L. Navarria[a,b], A. Perrotta[a], F. Primavera[a,b], A.M. Rossi[a,b], T. Rovelli[a,b], G.P. Siroli[a,b]

INFN Sezione di Catania[a], Università di Catania[b], Catania, Italy
S. Costa[a,b,47], A. Di Mattia[a], R. Potenza[a,b], A. Tricomi[a,b,47], C. Tuve[a,b]

INFN Sezione di Firenze[a], Università di Firenze[b], Firenze, Italy
G. Barbagli[a], B. Camaiani[a,b], A. Cassese[a], R. Ceccarelli[a,b], V. Ciulli[a,b], C. Civinini[a], R. D'Alessandro[a,b], E. Focardi[a,b], G. Latino[a,b], P. Lenzi[a,b], M. Lizzo[a,b], M. Meschini[a], S. Paoletti[a], R. Seidita[a,b], G. Sguazzoni[a], L. Viliani[a]

INFN Laboratori Nazionali di Frascati, Frascati, Italy
L. Benussi, S. Bianco, D. Piccolo

INFN Sezione di Genova[a], Università di Genova[b], Genova, Italy
M. Bozzo[a,b], F. Ferro[a], R. Mulargia[a], E. Robutti[a], S. Tosi[a,b]

INFN Sezione di Milano-Bicocca[a], Università di Milano-Bicocca[b], Milano, Italy
A. Benaglia[a], G. Boldrini[a], F. Brivio[a,b], F. Cetorelli[a,b], F. De Guio[a,b], M.E. Dinardo[a,b], P. Dini[a], S. Gennai[a], A. Ghezzi[a,b], P. Govoni[a,b], L. Guzzi[a,b], M.T. Lucchini[a,b], M. Malberti[a], S. Malvezzi[a], A. Massironi[a], D. Menasce[a],

JHEP07(2023)095

L. Moroni[a], M. Paganoni[a,b], D. Pedrini[a], B.S. Pinolini[a], S. Ragazzi[a,b], N. Redaelli[a], T. Tabarelli de Fatis[a,b], D. Zuolo[a,b]

INFN Sezione di Napoli[a], Università di Napoli 'Federico II'[b], Napoli, Italy; Università della Basilicata[c], Potenza, Italy; Università G. Marconi[d], Roma, Italy

S. Buontempo[a], F. Carnevali[a,b], N. Cavallo[a,c], A. De Iorio[a,b], F. Fabozzi[a,c], A.O.M. Iorio[a,b], L. Lista[a,b,48], S. Meola[a,d,21], P. Paolucci[a,21], B. Rossi[a], C. Sciacca[a,b]

INFN Sezione di Padova[a], Università di Padova[b], Padova, Italy; Università di Trento[c], Trento, Italy

P. Azzi[a], N. Bacchetta[a], D. Bisello[a,b], P. Bortignon[a], A. Bragagnolo[a,b], R. Carlin[a,b], P. Checchia[a], T. Dorigo[a], F. Gasparini[a,b], U. Gasparini[a,b], G. Grosso[a], L. Layer[a,49], E. Lusiani[a], M. Margoni[a,b], F. Marini[a], A.T. Meneguzzo[a,b], J. Pazzini[a,b], P. Ronchese[a,b], R. Rossin[a,b], F. Simonetto[a,b], G. Strong[a], M. Tosi[a,b], H. Yarar[a,b], M. Zanetti[a,b], P. Zotto[a,b], A. Zucchetta[a,b], G. Zumerle[a,b]

INFN Sezione di Pavia[a], Università di Pavia[b], Pavia, Italy

C. Aimè[a,b], A. Braghieri[a], S. Calzaferri[a,b], D. Fiorina[a,b], P. Montagna[a,b], V. Re[a], C. Riccardi[a,b], P. Salvini[a], I. Vai[a], P. Vitulo[a,b]

INFN Sezione di Perugia[a], Università di Perugia[b], Perugia, Italy

P. Asenov[a,50], G.M. Bilei[a], D. Ciangottini[a,b], L. Fanò[a,b], M. Magherini[a,b], G. Mantovani[a,b], V. Mariani[a,b], M. Menichelli[a], F. Moscatelli[a,50], A. Piccinelli[a,b], M. Presilla[a,b], A. Rossi[a,b], A. Santocchia[a,b], D. Spiga[a], T. Tedeschi[a,b]

INFN Sezione di Pisa[a], Università di Pisa[b], Scuola Normale Superiore di Pisa[c], Pisa, Italy; Università di Siena[d], Siena, Italy

P. Azzurri[a], G. Bagliesi[a], V. Bertacchi[a,c], L. Bianchini[a,b], T. Boccali[a], E. Bossini[a,b], D. Bruschini[a,c], R. Castaldi[a], M.A. Ciocci[a,b], V. D'Amante[a,d], R. Dell'Orso[a], M.R. Di Domenico[a,d], S. Donato[a], A. Giassi[a], F. Ligabue[a,c], E. Manca[a,c], G. Mandorli[a,c], D. Matos Figueiredo[a], A. Messineo[a,b], M. Musich[a,b], F. Palla[a], S. Parolia[a,b], G. Ramirez-Sanchez[a,c], A. Rizzi[a,b], G. Rolandi[a,c], S. Roy Chowdhury[a,c], A. Scribano[a], N. Shafiei[a,b], P. Spagnolo[a], R. Tenchini[a], G. Tonelli[a,b], N. Turini[a,d], A. Venturi[a], P.G. Verdini[a]

INFN Sezione di Roma[a], Sapienza Università di Roma[b], Roma, Italy

P. Barria[a], M. Campana[a,b], F. Cavallari[a], D. Del Re[a,b], E. Di Marco[a], M. Diemoz[a], E. Longo[a,b], P. Meridiani[a], G. Organtini[a,b], F. Pandolfi[a], R. Paramatti[a,b], C. Quaranta[a,b], S. Rahatlou[a,b], C. Rovelli[a], F. Santanastasio[a,b], L. Soffi[a], R. Tramontano[a,b]

INFN Sezione di Torino[a], Università di Torino[b], Torino, Italy; Università del Piemonte Orientale[c], Novara, Italy

N. Amapane[a,b], R. Arcidiacono[a,c], S. Argiro[a,b], M. Arneodo[a,c], N. Bartosik[a], R. Bellan[a,b], A. Bellora[a,b], J. Berenguer Antequera[a,b], C. Biino[a], N. Cartiglia[a],

JHEP07(2023)095

M. Costa[a,b], R. Covarelli[a,b], N. Demaria[a], M. Grippo[a,b], B. Kiani[a,b], F. Legger[a], C. Mariotti[a], S. Maselli[a], A. Mecca[a,b], E. Migliore[a,b], E. Monteil[a,b], M. Monteno[a], M.M. Obertino[a,b], G. Ortona[a], L. Pacher[a,b], N. Pastrone[a], M. Pelliccioni[a], M. Ruspa[a,c], K. Shchelina[a], F. Siviero[a,b], V. Sola[a], A. Solano[a,b], D. Soldi[a,b], A. Staiano[a], M. Tornago[a,b], D. Trocino[a], G. Umoret[a,b], A. Vagnerini[a,b]

INFN Sezione di Trieste[a], Università di Trieste[b], Trieste, Italy
S. Belforte[a], V. Candelise[a,b], M. Casarsa[a], F. Cossutti[a], A. Da Rold[a,b], G. Della Ricca[a,b], G. Sorrentino[a,b]

Kyungpook National University, Daegu, Korea
S. Dogra, C. Huh, B. Kim, D.H. Kim, G.N. Kim, J. Kim, J. Lee, S.W. Lee, C.S. Moon, Y.D. Oh, S.I. Pak, S. Sekmen, Y.C. Yang

Chonnam National University, Institute for Universe and Elementary Particles, Kwangju, Korea
H. Kim, D.H. Moon

Hanyang University, Seoul, Korea
E. Asilar, J. Choi, T.J. Kim, J. Park

Korea University, Seoul, Korea
S. Cho, S. Choi, B. Hong, K. Lee, K.S. Lee, J. Lim, J. Park, S.K. Park, J. Yoo

Kyung Hee University, Department of Physics, Seoul, Korea
J. Goh

Sejong University, Seoul, Korea
H. S. Kim, Y. Kim

Seoul National University, Seoul, Korea
J. Almond, J.H. Bhyun, J. Choi, S. Jeon, J. Kim, J. Kim, J.S. Kim, S. Ko, H. Kwon, H. Lee, S. Lee, B.H. Oh, M. Oh, S.B. Oh, H. Seo, U.K. Yang, I. Yoon

University of Seoul, Seoul, Korea
W. Jang, D.Y. Kang, Y. Kang, D. Kim, S. Kim, B. Ko, J.S.H. Lee, Y. Lee, J.A. Merlin, I.C. Park, Y. Roh, M.S. Ryu, D. Song, Watson, I.J., S. Yang

Yonsei University, Department of Physics, Seoul, Korea
S. Ha, H.D. Yoo

Sungkyunkwan University, Suwon, Korea
M. Choi, H. Lee, Y. Lee, I. Yu

College of Engineering and Technology, American University of the Middle East (AUM), Dasman, Kuwait
T. Beyrouthy, Y. Maghrbi

JHEP07(2023)095

Riga Technical University, Riga, Latvia
K. Dreimanis, V. Veckalns

Vilnius University, Vilnius, Lithuania
M. Ambrozas, A. Carvalho Antunes De Oliveira, A. Juodagalvis, A. Rinkevicius, G. Tamulaitis

National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia
N. Bin Norjoharuddeen, S.Y. Hoh[51], I. Yusuff[51], Z. Zolkapli

Universidad de Sonora (UNISON), Hermosillo, Mexico
J.F. Benitez, A. Castaneda Hernandez, H.A. Encinas Acosta, L.G. Gallegos Maríñez, M. León Coello, J.A. Murillo Quijada, A. Sehrawat, L. Valencia Palomo

Centro de Investigacion y de Estudios Avanzados del IPN, Mexico City, Mexico
G. Ayala, H. Castilla-Valdez, E. De La Cruz-Burelo, I. Heredia-De La Cruz[52], R. Lopez-Fernandez, C.A. Mondragon Herrera, D.A. Perez Navarro, A. Sánchez Hernández

Universidad Iberoamericana, Mexico City, Mexico
C. Oropeza Barrera, F. Vazquez Valencia

Benemerita Universidad Autonoma de Puebla, Puebla, Mexico
I. Pedraza, H.A. Salazar Ibarguen, C. Uribe Estrada

University of Montenegro, Podgorica, Montenegro
I. Bubanja, J. Mijuskovic[53], N. Raicevic

National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan
A. Ahmad, M.I. Asghar, A. Awais, M.I.M. Awan, H.R. Hoorani, W.A. Khan, M. Shoaib, M. Waqas

AGH University of Science and Technology Faculty of Computer Science, Electronics and Telecommunications, Krakow, Poland
V. Avati, L. Grzanka, M. Malawski

National Centre for Nuclear Research, Swierk, Poland
H. Bialkowska, M. Bluj, B. Boimska, M. Górski, M. Kazana, M. Szleper, P. Zalewski

Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland
K. Bunkowski, K. Doroba, A. Kalinowski, M. Konecki, J. Krolikowski

JHEP07(2023)095

Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal

M. Araujo, P. Bargassa, D. Bastos, A. Boletti, P. Faccioli, M. Gallinaro, J. Hollar, N. Leonardo, T. Niknejad, M. Pisano, J. Seixas, O. Toldaiev, J. Varela

VINCA Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

P. Adzic[54], M. Dordevic, P. Milenovic, J. Milosevic

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

M. Aguilar-Benitez, J. Alcaraz Maestre, A. Álvarez Fernández, M. Barrio Luna, Cristina F. Bedoya, C.A. Carrillo Montoya, M. Cepeda, M. Cerrada, N. Colino, B. De La Cruz, A. Delgado Peris, J.P. Fernández Ramos, J. Flix, M.C. Fouz, O. Gonzalez Lopez, S. Goy Lopez, J.M. Hernandez, M.I. Josa, J. León Holgado, D. Moran, C. Perez Dengra, A. Pérez-Calero Yzquierdo, J. Puerta Pelayo, I. Redondo, D.D. Redondo Ferrero, L. Romero, S. Sánchez Navas, J. Sastre, L. Urda Gómez, J. Vazquez Escobar, C. Willmott

Universidad Autónoma de Madrid, Madrid, Spain

J.F. de Trocóniz

Universidad de Oviedo, Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), Oviedo, Spain

B. Alvarez Gonzalez, J. Cuevas, J. Fernandez Menendez, S. Folgueras, I. Gonzalez Caballero, J.R. González Fernández, E. Palencia Cortezon, C. Ramón Álvarez, V. Rodríguez Bouza, A. Soto Rodríguez, A. Trapote, N. Trevisani, C. Vico Villalba

Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain

J.A. Brochero Cifuentes, I.J. Cabrillo, A. Calderon, J. Duarte Campderros, M. Fernandez, C. Fernandez Madrazo, P.J. Fernández Manteca, A. García Alonso, G. Gomez, C. Lasaosa García, C. Martinez Rivero, P. Martinez Ruiz del Arbol, F. Matorras, P. Matorras Cuevas, J. Piedra Gomez, C. Prieels, A. Ruiz-Jimeno, L. Scodellaro, I. Vila, J.M. Vizan Garcia

University of Colombo, Colombo, Sri Lanka

M.K. Jayananda, B. Kailasapathy[55], D.U.J. Sonnadara, D.D.C. Wickramarathna

University of Ruhuna, Department of Physics, Matara, Sri Lanka

W.G.D. Dharmaratna, K. Liyanage, N. Perera, N. Wickramage

JHEP07(2023)095

CERN, European Organization for Nuclear Research, Geneva, Switzerland
T.K. Aarrestad, D. Abbaneo, J. Alimena, E. Auffray, G. Auzinger, J. Baechler, P. Baillon†, D. Barney, J. Bendavid, M. Bianco, B. Bilin, A. Bocci, E. Brondolin, C. Caillol, T. Camporesi, G. Cerminara, N. Chernyavskaya, S.S. Chhibra, S. Choudhury, M. Cipriani, L. Cristella, D. d'Enterria, A. Dabrowski, A. David, A. De Roeck, M.M. Defranchis, M. Deile, M. Dobson, M. Dünser, A. Elliott-Peisert, F. Fallavollita[56], A. Florent, L. Forthomme, W. Funk, S. Ghosh, S. Giani, D. Gigi, K. Gill, F. Glege, L. Gouskos, E. Govorkova, M. Haranko, J. Hegeman, V. Innocente, T. James, P. Janot, J. Kaspar, J. Kieseler, M. Komm, N. Kratochwil, S. Laurila, P. Lecoq, A. Lintuluoto, C. Lourenço, B. Maier, L. Malgeri, M. Mannelli, A.C. Marini, F. Meijers, S. Mersi, E. Meschi, F. Moortgat, M. Mulders, S. Orfanelli, L. Orsini, F. Pantaleo, E. Perez, M. Peruzzi, A. Petrilli, G. Petrucciani, A. Pfeiffer, M. Pierini, D. Piparo, M. Pitt, H. Qu, T. Quast, D. Rabady, A. Racz, G. Reales Gutiérrez, M. Rovere, H. Sakulin, S. Scarfi, M. Selvaggi, A. Sharma, P. Silva, W. Snoeys, P. Sphicas[57], A.G. Stahl Leiton, S. Summers, K. Tatar, V.R. Tavolaro, D. Treille, P. Tropea, A. Tsirou, J. Wanczyk[58], K.A. Wozniak, W.D. Zeuner

Paul Scherrer Institut, Villigen, Switzerland
L. Caminada[59], A. Ebrahimi, W. Erdmann, R. Horisberger, Q. Ingram, H.C. Kaestli, D. Kotlinski, C. Lange, M. Missiroli[59], L. Noehte[59], T. Rohe

ETH Zurich — Institute for Particle Physics and Astrophysics (IPA), Zurich, Switzerland
K. Androsov[58], M. Backhaus, P. Berger, A. Calandri, A. De Cosa, G. Dissertori, M. Dittmar, M. Donegà, C. Dorfer, F. Eble, K. Gedia, F. Glessgen, T.A. Gómez Espinosa, C. Grab, D. Hits, W. Lustermann, A.-M. Lyon, R.A. Manzoni, L. Marchese, C. Martin Perez, A. Mascellani[58], M.T. Meinhard, F. Nessi-Tedaldi, J. Niedziela, F. Pauss, V. Perovic, S. Pigazzini, M.G. Ratti, M. Reichmann, C. Reissel, T. Reitenspiess, B. Ristic, D. Ruini, D.A. Sanz Becerra, J. Steggemann[58], D. Valsecchi[21], R. Wallny

Universität Zürich, Zurich, Switzerland
C. Amsler[60], P. Bärtschi, C. Botta, D. Brzhechko, M.F. Canelli, K. Cormier, A. De Wit, R. Del Burgo, J.K. Heikkilä, M. Huwiler, W. Jin, A. Jofrehei, B. Kilminster, S. Leontsinis, S.P. Liechti, A. Macchiolo, P. Meiring, V.M. Mikuni, U. Molinatti, I. Neutelings, A. Reimers, P. Robmann, S. Sanchez Cruz, K. Schweiger, M. Senger, Y. Takahashi

National Central University, Chung-Li, Taiwan
C. Adloff[61], C.M. Kuo, W. Lin, S.S. Yu

National Taiwan University (NTU), Taipei, Taiwan
L. Ceard, Y. Chao, K.F. Chen, P.H. Chen, P.s. Chen, H. Cheng, W.-S. Hou, Y.y. Li, R.-S. Lu, E. Paganis, A. Psallidas, A. Steen, H.y. Wu, E. Yazgan, P.r. Yu

JHEP07(2023)095

Chulalongkorn University, Faculty of Science, Department of Physics, Bangkok, Thailand

C. Asawatangtrakuldee, N. Srimanobhas

Çukurova University, Physics Department, Science and Art Faculty, Adana, Turkey

D. Agyel, F. Boran, Z.S. Demiroglu, F. Dolek, I. Dumanoglu[62], E. Eskut, Y. Guler[63], E. Gurpinar Guler[63], C. Isik, O. Kara, A. Kayis Topaksu, U. Kiminsu, G. Onengut, K. Ozdemir[64], A. Polatoz, A.E. Simsek, B. Tali[65], U.G. Tok, S. Turkcapar, E. Uslan, I.S. Zorbakir

Middle East Technical University, Physics Department, Ankara, Turkey

G. Karapinar, K. Ocalan[66], M. Yalvac[67]

Bogazici University, Istanbul, Turkey

B. Akgun, I.O. Atakisi, E. Gülmez, M. Kaya[68], O. Kaya[69], Ö. Özçelik, S. Tekten[70]

Istanbul Technical University, Istanbul, Turkey

A. Cakir, K. Cankocak[62], Y. Komurcu, S. Sen[71]

Istanbul University, Istanbul, Turkey

O. Aydilek, S. Cerci[65], B. Hacisahinoglu, I. Hos[72], B. Isildak[73], B. Kaynak, S. Ozkorucuklu, C. Simsek, D. Sunar Cerci[65]

Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkiv, Ukraine

B. Grynyov

National Science Centre, Kharkiv Institute of Physics and Technology, Kharkiv, Ukraine

L. Levchuk

University of Bristol, Bristol, United Kingdom

D. Anthony, E. Bhal, J.J. Brooke, A. Bundock, E. Clement, D. Cussans, H. Flacher, M. Glowacki, J. Goldstein, G.P. Heath, H.F. Heath, L. Kreczko, B. Krikler, S. Paramesvaran, S. Seif El Nasr-Storey, V.J. Smith, N. Stylianou[74], K. Walkingshaw Pass, R. White

Rutherford Appleton Laboratory, Didcot, United Kingdom

A.H. Ball, K.W. Bell, A. Belyaev[75], C. Brew, R.M. Brown, D.J.A. Cockerill, C. Cooke, K.V. Ellis, K. Harder, S. Harper, M.-L. Holmberg[76], J. Linacre, K. Manolopoulos, D.M. Newbold, E. Olaiya, D. Petyt, T. Reis, T. Schuh, C.H. Shepherd-Themistocleous, I.R. Tomalin, T. Williams

JHEP07(2023)095

Imperial College, London, United Kingdom

R. Bainbridge, P. Bloch, S. Bonomally, J. Borg, S. Breeze, O. Buchmuller, V. Cepaitis, G.S. Chahal[77], D. Colling, P. Dauncey, G. Davies, M. Della Negra, S. Fayer, G. Fedi, G. Hall, M.H. Hassanshahi, G. Iles, J. Langford, L. Lyons, A.-M. Magnan, S. Malik, D.G. Monk, J. Nash[78], M. Pesaresi, B.C. Radburn-Smith, D.M. Raymond, A. Richards, A. Rose, E. Scott, C. Seez, A. Shtipliyski, R. Shukla, A. Tapper, K. Uchida, T. Virdee[21], M. Vojinovic, N. Wardle, D. Winterbottom

Brunel University, Uxbridge, United Kingdom

K. Coldham, J.E. Cole, A. Khan, P. Kyberd, I.D. Reid, L. Teodorescu, S. Zahid

Baylor University, Waco, Texas, USA

S. Abdullin, A. Brinkerhoff, B. Caraway, J. Dittmann, K. Hatakeyama, A.R. Kanuganti, B. McMaster, M. Saunders, S. Sawant, C. Sutantawibul, J. Wilson

Catholic University of America, Washington, DC, USA

R. Bartek, A. Dominguez, R. Uniyal, A.M. Vargas Hernandez

The University of Alabama, Tuscaloosa, Alabama, USA

A. Buccilli, S.I. Cooper, D. Di Croce, S.V. Gleyzer, C. Henderson, C.U. Perez, P. Rumerio[79], C. West

Boston University, Boston, Massachusetts, USA

A. Akpinar, A. Albert, D. Arcaro, C. Cosby, Z. Demiragli, C. Erice, E. Fontanesi, D. Gastler, S. May, J. Rohlf, K. Salyer, D. Sperka, D. Spitzbart, I. Suarez, A. Tsatsos, S. Yuan

Brown University, Providence, Rhode Island, USA

G. Benelli, B. Burkle, X. Coubez[23], D. Cutts, M. Hadley, U. Heintz, J.M. Hogan[80], T. Kwon, G. Landsberg, K.T. Lau, D. Li, M. Lukasik, J. Luo, M. Narain, N. Pervan, S. Sagir[81], F. Simpson, E. Usai, W.Y. Wong, X. Yan, D. Yu, W. Zhang

University of California, Davis, Davis, California, USA

J. Bonilla, C. Brainerd, R. Breedon, M. Calderon De La Barca Sanchez, M. Chertok, J. Conway, P.T. Cox, R. Erbacher, G. Haza, F. Jensen, O. Kukral, G. Mocellin, M. Mulhearn, D. Pellett, B. Regnery, D. Taylor, Y. Yao, F. Zhang

University of California, Los Angeles, California, USA

M. Bachtis, R. Cousins, A. Datta, D. Hamilton, J. Hauser, M. Ignatenko, M.A. Iqbal, T. Lam, W.A. Nash, S. Regnard, D. Saltzberg, B. Stone, V. Valuev

University of California, Riverside, Riverside, California, USA

Y. Chen, R. Clare, J.W. Gary, M. Gordon, G. Hanson, G. Karapostoli, O.R. Long, N. Manganelli, W. Si, S. Wimpenny, Y. Zhang

JHEP07(2023)095

University of California, San Diego, La Jolla, California, USA

J.G. Branson, P. Chang, S. Cittolin, S. Cooperstein, D. Diaz, J. Duarte, R. Gerosa, L. Giannini, J. Guiang, R. Kansal, V. Krutelyov, R. Lee, J. Letts, M. Masciovecchio, F. Mokhtar, M. Pieri, B.V. Sathia Narayanan, V. Sharma, M. Tadel, F. Würthwein, Y. Xiang, A. Yagil

University of California, Santa Barbara — Department of Physics, Santa Barbara, California, USA

N. Amin, C. Campagnari, M. Citron, G. Collura, A. Dorsett, V. Dutta, J. Incandela, M. Kilpatrick, J. Kim, B. Marsh, P. Masterson, H. Mei, M. Oshiro, M. Quinnan, J. Richman, U. Sarica, F. Setti, J. Sheplock, P. Siddireddy, D. Stuart, S. Wang

California Institute of Technology, Pasadena, California, USA

A. Bornheim, O. Cerri, I. Dutta, J.M. Lawhorn, N. Lu, J. Mao, H.B. Newman, T. Q. Nguyen, M. Spiropulu, J.R. Vlimant, C. Wang, S. Xie, Z. Zhang, R.Y. Zhu

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

J. Alison, S. An, M.B. Andrews, P. Bryant, T. Ferguson, A. Harilal, C. Liu, T. Mudholkar, S. Murthy, M. Paulini, A. Roberts, A. Sanchez, W. Terrill

University of Colorado Boulder, Boulder, Colorado, USA

J.P. Cumalat, W.T. Ford, A. Hassani, G. Karathanasis, E. MacDonald, R. Patel, A. Perloff, C. Savard, N. Schonbeck, K. Stenson, K.A. Ulmer, S.R. Wagner, N. Zipper

Cornell University, Ithaca, New York, USA

J. Alexander, S. Bright-Thonney, X. Chen, Y. Cheng, D.J. Cranshaw, J. Fan, X. Fan, D. Gadkari, S. Hogan, J. Monroy, J.R. Patterson, D. Quach, J. Reichert, M. Reid, A. Ryd, J. Thom, P. Wittich, R. Zou

Fermi National Accelerator Laboratory, Batavia, Illinois, USA

M. Albrow, M. Alyari, G. Apollinari, A. Apresyan, L.A.T. Bauerdick, D. Berry, J. Berryhill, P.C. Bhat, K. Burkett, J.N. Butler, A. Canepa, G.B. Cerati, H.W.K. Cheung, F. Chlebana, K.F. Di Petrillo, J. Dickinson, V.D. Elvira, Y. Feng, J. Freeman, A. Gandrakota, Z. Gecse, L. Gray, D. Green, S. Grünendahl, O. Gutsche, R.M. Harris, R. Heller, T.C. Herwig, J. Hirschauer, L. Horyn, B. Jayatilaka, S. Jindariani, M. Johnson, U. Joshi, T. Klijnsma, B. Klima, K.H.M. Kwok, S. Lammel, D. Lincoln, R. Lipton, T. Liu, C. Madrid, K. Maeshima, C. Mantilla, D. Mason, P. McBride, P. Merkel, S. Mrenna, S. Nahn, J. Ngadiuba, V. Papadimitriou, K. Pedro, C. Pena[82], F. Ravera, A. Reinsvold Hall[83], L. Ristori, E. Sexton-Kennedy, N. Smith, A. Soha, L. Spiegel, J. Strait, L. Taylor, S. Tkaczyk, N.V. Tran, L. Uplegger, E.W. Vaandering, H.A. Weber, I. Zoi

University of Florida, Gainesville, Florida, USA
P. Avery, D. Bourilkov, L. Cadamuro, V. Cherepanov, R.D. Field, D. Guerrero, M. Kim, E. Koenig, J. Konigsberg, A. Korytov, K.H. Lo, K. Matchev, N. Menendez, G. Mitselmakher, A. Muthirakalayil Madhu, N. Rawal, D. Rosenzweig, S. Rosenzweig, K. Shi, J. Wang, Z. Wu

Florida State University, Tallahassee, Florida, USA
T. Adams, A. Askew, R. Habibullah, V. Hagopian, R. Khurana, T. Kolberg, G. Martinez, H. Prosper, C. Schiber, O. Viazlo, R. Yohay, J. Zhang

Florida Institute of Technology, Melbourne, Florida, USA
M.M. Baarmand, S. Butalla, T. Elkafrawy[84], M. Hohlmann, R. Kumar Verma, D. Noonan, M. Rahmani, F. Yumiceva

University of Illinois at Chicago (UIC), Chicago, Illinois, USA
M.R. Adams, H. Becerril Gonzalez, R. Cavanaugh, S. Dittmer, O. Evdokimov, C.E. Gerber, D.J. Hofman, D. S. Lemos, A.H. Merrit, C. Mills, G. Oh, T. Roy, S. Rudrabhatla, M.B. Tonjes, N. Varelas, J. Viinikainen, X. Wang, Z. Ye

The University of Iowa, Iowa City, Iowa, USA
M. Alhusseini, K. Dilsiz[85], L. Emediato, R.P. Gandrajula, O.K. Köseyan, J.-P. Merlo, A. Mestvirishvili[86], J. Nachtman, H. Ogul[87], Y. Onel, A. Penzo, C. Snyder, E. Tiras[88]

Johns Hopkins University, Baltimore, Maryland, USA
O. Amram, B. Blumenfeld, L. Corcodilos, J. Davis, A.V. Gritsan, S. Kyriacou, P. Maksimovic, J. Roskes, M. Swartz, T.Á. Vámi

The University of Kansas, Lawrence, Kansas, USA
A. Abreu, J. Anguiano, P. Baringer, A. Bean, Z. Flowers, T. Isidori, S. Khalil, J. King, G. Krintiras, M. Lazarovits, C. Le Mahieu, C. Lindsey, J. Marquez, N. Minafra, M. Murray, M. Nickel, C. Rogan, C. Royon, R. Salvatico, S. Sanders, E. Schmitz, C. Smith, Q. Wang, Z. Warner, J. Williams, G. Wilson

Kansas State University, Manhattan, Kansas, USA
S. Duric, A. Ivanov, K. Kaadze, D. Kim, Y. Maravin, T. Mitchell, A. Modak, D. Roy

Lawrence Livermore National Laboratory, Livermore, California, USA
F. Rebassoo, D. Wright

University of Maryland, College Park, Maryland, USA
E. Adams, A. Baden, O. Baron, A. Belloni, S.C. Eno, N.J. Hadley, S. Jabeen, R.G. Kellogg, T. Koeth, Y. Lai, S. Lascio, A.C. Mignerey, S. Nabili, C. Palmer, C. Papageorgakis, M. Seidel, L. Wang, K. Wong

JHEP07(2023)095

Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

D. Abercrombie, G. Andreassi, R. Bi, W. Busza, I.A. Cali, Y. Chen, M. D'Alfonso, G. Gomez-Ceballos, M. Goncharov, P. Harris, M. Hu, M. Klute, D. Kovalskyi, J. Krupa, Y.-J. Lee, C. Mironov, C. Paus, D. Rankin, C. Roland, G. Roland, Z. Shi, G.S.F. Stephans, J. Wang, B. Wyslouch

University of Minnesota, Minneapolis, Minnesota, USA

R.M. Chatterjee, A. Evans, J. Hiltbrand, Sh. Jain, B.M. Joshi, M. Krohn, Y. Kubota, J. Mans, M. Revering, R. Rusack, R. Saradhy, N. Schroeder, N. Strobbe, M.A. Wadud

University of Mississippi, Oxford, Mississippi, USA

L.M. Cremaldi

University of Nebraska-Lincoln, Lincoln, Nebraska, USA

K. Bloom, M. Bryson, S. Chauhan, D.R. Claes, C. Fangmeier, L. Finco, F. Golf, C. Joo, I. Kravchenko, I. Reed, J.E. Siado, G.R. Snow[†], W. Tabb, A. Wightman, F. Yan, A.G. Zecchinelli

State University of New York at Buffalo, Buffalo, New York, USA

G. Agarwal, H. Bandyopadhyay, L. Hay, I. Iashvili, A. Kharchilava, C. McLean, D. Nguyen, J. Pekkanen, S. Rappoccio, A. Williams

Northeastern University, Boston, Massachusetts, USA

G. Alverson, E. Barberis, Y. Haddad, Y. Han, A. Hortiangtham, A. Krishna, J. Li, J. Lidrych, G. Madigan, B. Marzocchi, D.M. Morse, V. Nguyen, T. Orimoto, A. Parker, L. Skinnari, A. Tishelman-Charny, T. Wamorkar, B. Wang, A. Wisecarver, D. Wood

Northwestern University, Evanston, Illinois, USA

S. Bhattacharya, J. Bueghly, Z. Chen, A. Gilbert, T. Gunter, K.A. Hahn, Y. Liu, N. Odell, M.H. Schmitt, M. Velasco

University of Notre Dame, Notre Dame, Indiana, USA

R. Band, R. Bucci, M. Cremonesi, A. Das, R. Goldouzian, M. Hildreth, K. Hurtado Anampa, C. Jessop, K. Lannon, J. Lawrence, N. Loukas, L. Lutton, J. Mariano, N. Marinelli, I. Mcalister, T. McCauley, C. Mcgrady, K. Mohrman, C. Moore, Y. Musienko[13], R. Ruchti, A. Townsend, M. Wayne, H. Yockey, M. Zarucki, L. Zygala

The Ohio State University, Columbus, Ohio, USA

B. Bylsma, L.S. Durkin, B. Francis, C. Hill, A. Lesauvage, M. Nunez Ornelas, K. Wei, B.L. Winer, B.R. Yates

JHEP07(2023)095

Princeton University, Princeton, New Jersey, USA
F.M. Addesa, B. Bonham, G. Dezoort, P. Elmer, A. Frankenthal, B. Greenberg, N. Haubrich, S. Higginbotham, A. Kalogeropoulos, G. Kopp, S. Kwan, D. Lange, D. Marlow, K. Mei, I. Ojalvo, J. Olsen, D. Stickland, C. Tully

University of Puerto Rico, Mayaguez, Puerto Rico, USA
S. Malik, S. Norberg

Purdue University, West Lafayette, Indiana, USA
A.S. Bakshi, V.E. Barnes, R. Chawla, S. Das, L. Gutay, M. Jones, A.W. Jung, D. Kondratyev, A.M. Koshy, M. Liu, G. Negro, N. Neumeister, G. Paspalaki, S. Piperov, A. Purohit, J.F. Schulte, M. Stojanovic, J. Thieman, F. Wang, R. Xiao, W. Xie

Purdue University Northwest, Hammond, Indiana, USA
J. Dolen, N. Parashar

Rice University, Houston, Texas, USA
D. Acosta, A. Baty, T. Carnahan, M. Decaro, S. Dildick, K.M. Ecklund, S. Freed, P. Gardner, F.J.M. Geurts, A. Kumar, W. Li, B.P. Padley, R. Redjimi, J. Rotter, W. Shi, S. Yang, E. Yigitbasi, L. Zhang[89], Y. Zhang, X. Zuo

University of Rochester, Rochester, New York, USA
A. Bodek, P. de Barbaro, R. Demina, J.L. Dulemba, C. Fallon, T. Ferbel, M. Galanti, A. Garcia-Bellido, O. Hindrichs, A. Khukhunaishvili, E. Ranken, R. Taus, G.P. Van Onsem

The Rockefeller University, New York, New York, USA
K. Goulianos

Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA
B. Chiarito, J.P. Chou, Y. Gershtein, E. Halkiadakis, A. Hart, M. Heindl, O. Karacheban[25], I. Laflotte, A. Lath, R. Montalvo, K. Nash, M. Osherson, S. Salur, S. Schnetzer, S. Somalwar, R. Stone, S.A. Thayil, S. Thomas, H. Wang

University of Tennessee, Knoxville, Tennessee, USA
H. Acharya, A.G. Delannoy, S. Fiorendi, T. Holmes, S. Spanier

Texas A&M University, College Station, Texas, USA
O. Bouhali[90], M. Dalchenko, A. Delgado, R. Eusebi, J. Gilmore, T. Huang, T. Kamon[91], H. Kim, S. Luo, S. Malhotra, R. Mueller, D. Overton, D. Rathjens, A. Safonov

Texas Tech University, Lubbock, Texas, USA
N. Akchurin, J. Damgov, V. Hegde, K. Lamichhane, S.W. Lee, T. Mengke, S. Muthumuni, T. Peltola, I. Volobouev, Z. Wang, A. Whitbeck

JHEP07(2023)095

Vanderbilt University, Nashville, Tennessee, USA

E. Appelt, S. Greene, A. Gurrola, W. Johns, A. Melo, F. Romeo, P. Sheldon, S. Tuo, J. Velkovska

University of Virginia, Charlottesville, Virginia, USA

B. Cardwell, B. Cox, G. Cummings, J. Hakala, R. Hirosky, M. Joyce, A. Ledovskoy, A. Li, C. Neu, C.E. Perez Lara, B. Tannenwald

Wayne State University, Detroit, Michigan, USA

P.E. Karchin, N. Poudyal

University of Wisconsin — Madison, Madison, Wisconsin, USA

S. Banerjee, K. Black, T. Bose, S. Dasu, I. De Bruyn, P. Everaerts, C. Galloni, H. He, M. Herndon, A. Herve, C.K. Koraka, A. Lanaro, A. Loeliger, R. Loveless, J. Madhusudanan Sreekala, A. Mallampalli, A. Mohammadi, D. Pinna, A. Savin, V. Shang, V. Sharma, W.H. Smith, D. Teague, S. Trembath-Reichert, W. Vetens

Authors affiliated with an institute or an international laboratory covered by a cooperation agreement with CERN

S. Afanasiev, V. Andreev, Yu. Andreev, T. Aushev, M. Azarkin, A. Babaev, A. Belyaev, V. Blinov[92], E. Boos, V. Borshch, D. Budkouski, V. Bunichev, M. Chadeeva[92], V. Chekhovsky, A. Dermenev, T. Dimova[92], I. Dremin, M. Dubinin[82], L. Dudko, V. Epshteyn, A. Ershov, G. Gavrilov, V. Gavrilov, S. Gninenko, V. Golovtcov, N. Golubev, I. Golutvin, I. Gorbunov, V. Ivanchenko, Y. Ivanov, V. Kachanov, L. Kardapoltsev[92], V. Karjavine, A. Karneyeu, V. Kim[92], M. Kirakosyan, D. Kirpichnikov, M. Kirsanov, V. Klyukhin, O. Kodolova[93], D. Konstantinov, V. Korenkov, A. Kozyrev[92], N. Krasnikov, E. Kuznetsova[94], A. Lanev, P. Levchenko, A. Litomin, N. Lychkovskaya, V. Makarenko, A. Malakhov, V. Matveev[92], V. Murzin, A. Nikitenko[95], S. Obraztsov, V. Okhotnikov, A. Oskin, I. Ovtin[92], V. Palichik, P. Parygin, V. Perelygin, M. Perfilov, S. Petrushanko, G. Pivovarov, V. Popov, E. Popova, O. Radchenko[92], V. Rusinov, M. Savina, V. Savrin, D. Selivanova, V. Shalaev, S. Shmatov, S. Shulha, Y. Skovpen[92], S. Slabospitskii, V. Smirnov, D. Sosnov, A. Stepennov, V. Sulimov, E. Tcherniaev, A. Terkulov, O. Teryaev, I. Tlisova, M. Toms, A. Toropin, L. Uvarov, A. Uzunian, E. Vlasov, A. Vorobyev, N. Voytishin, B.S. Yuldashev[96], A. Zarubin, I. Zhizhin, A. Zhokin

† *Deceased*
[1] *Also at Yerevan State University, Yerevan, Armenia*
[2] *Also at TU Wien, Vienna, Austria*
[3] *Also at Institute of Basic and Applied Sciences, Faculty of Engineering, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt*
[4] *Also at Université Libre de Bruxelles, Bruxelles, Belgium*
[5] *Also at Universidade Estadual de Campinas, Campinas, Brazil*
[6] *Also at Federal University of Rio Grande do Sul, Porto Alegre, Brazil*
[7] *Also at UFMS, Nova Andradina, Brazil*

JHEP07(2023)095

[8] *Also at The University of the State of Amazonas, Manaus, Brazil*
[9] *Also at University of Chinese Academy of Sciences, Beijing, China*
[10] *Also at Nanjing Normal University Department of Physics, Nanjing, China*
[11] *Now at The University of Iowa, Iowa City, Iowa, USA*
[12] *Also at University of Chinese Academy of Sciences, Beijing, China*
[13] *Also at an institute or an international laboratory covered by a cooperation agreement with CERN*
[14] *Also at Cairo University, Cairo, Egypt*
[15] *Also at Suez University, Suez, Egypt*
[16] *Now at British University in Egypt, Cairo, Egypt*
[17] *Also at Purdue University, West Lafayette, Indiana, USA*
[18] *Also at Université de Haute Alsace, Mulhouse, France*
[19] *Also at Ilia State University, Tbilisi, Georgia*
[20] *Also at Erzincan Binali Yildirim University, Erzincan, Turkey*
[21] *Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland*
[22] *Also at University of Hamburg, Hamburg, Germany*
[23] *Also at RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany*
[24] *Also at Isfahan University of Technology, Isfahan, Iran*
[25] *Also at Brandenburg University of Technology, Cottbus, Germany*
[26] *Also at Forschungszentrum Jülich, Juelich, Germany*
[27] *Also at Physics Department, Faculty of Science, Assiut University, Assiut, Egypt*
[28] *Also at Karoly Robert Campus, MATE Institute of Technology, Gyongyos, Hungary*
[29] *Also at Institute of Physics, University of Debrecen, Debrecen, Hungary*
[30] *Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary*
[31] *Now at Universitatea Babes-Bolyai — Facultatea de Fizica, Cluj-Napoca, Romania*
[32] *Also at Faculty of Informatics, University of Debrecen, Debrecen, Hungary*
[33] *Also at Wigner Research Centre for Physics, Budapest, Hungary*
[34] *Also at Punjab Agricultural University, Ludhiana, India*
[35] *Also at UPES — University of Petroleum and Energy Studies, Dehradun, India*
[36] *Also at University of Hyderabad, Hyderabad, India*
[37] *Also at University of Visva-Bharati, Santiniketan, India*
[38] *Also at Indian Institute of Science (IISc), Bangalore, India*
[39] *Also at Indian Institute of Technology (IIT), Mumbai, India*
[40] *Also at IIT Bhubaneswar, Bhubaneswar, India*
[41] *Also at Institute of Physics, Bhubaneswar, India*
[42] *Also at Deutsches Elektronen-Synchrotron, Hamburg, Germany*
[43] *Also at Sharif University of Technology, Tehran, Iran*
[44] *Also at Department of Physics, University of Science and Technology of Mazandaran, Behshahr, Iran*
[45] *Also at Helwan University, Cairo, Egypt*
[46] *Also at Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Bologna, Italy*
[47] *Also at Centro Siciliano di Fisica Nucleare e di Struttura Della Materia, Catania, Italy*
[48] *Also at Scuola Superiore Meridionale, Università di Napoli 'Federico II', Napoli, Italy*
[49] *Also at Università di Napoli 'Federico II', Napoli, Italy*
[50] *Also at Consiglio Nazionale delle Ricerche — Istituto Officina dei Materiali, Perugia, Italy*
[51] *Also at Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia*
[52] *Also at Consejo Nacional de Ciencia y Tecnología, Mexico City, Mexico*
[53] *Also at IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France*
[54] *Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia*
[55] *Also at Trincomalee Campus, Eastern University, Sri Lanka, Nilaveli, Sri Lanka*
[56] *Also at INFN Sezione di Pavia, Università di Pavia, Pavia, Italy*
[57] *Also at National and Kapodistrian University of Athens, Athens, Greece*

[58] Also at Ecole Polytechnique Fédérale Lausanne, Lausanne, Switzerland
[59] Also at Universität Zürich, Zurich, Switzerland
[60] Also at Stefan Meyer Institute for Subatomic Physics, Vienna, Austria
[61] Also at Laboratoire d'Annecy-le-Vieux de Physique des Particules, IN2P3-CNRS, Annecy-le-Vieux, France
[62] Also at Near East University, Research Center of Experimental Health Science, Mersin, Turkey
[63] Also at Konya Technical University, Konya, Turkey
[64] Also at Izmir Bakircay University, Izmir, Turkey
[65] Also at Adiyaman University, Adiyaman, Turkey
[66] Also at Necmettin Erbakan University, Konya, Turkey
[67] Also at Bozok Universitetesi Rektörlügü, Yozgat, Turkey
[68] Also at Marmara University, Istanbul, Turkey
[69] Also at Milli Savunma University, Istanbul, Turkey
[70] Also at Kafkas University, Kars, Turkey
[71] Also at Hacettepe University, Ankara, Turkey
[72] Also at Istanbul University — Cerrahpasa, Faculty of Engineering, Istanbul, Turkey
[73] Also at Yildiz Technical University, Istanbul, Turkey
[74] Also at Vrije Universiteit Brussel, Brussel, Belgium
[75] Also at School of Physics and Astronomy, University of Southampton, Southampton, United Kingdom
[76] Also at University of Bristol, Bristol, United Kingdom
[77] Also at IPPP Durham University, Durham, United Kingdom
[78] Also at Monash University, Faculty of Science, Clayton, Australia
[79] Also at Università di Torino, Torino, Italy
[80] Also at Bethel University, St. Paul, Minnesota, USA
[81] Also at Karamanoğlu Mehmetbey University, Karaman, Turkey
[82] Also at California Institute of Technology, Pasadena, California, USA
[83] Also at United States Naval Academy, Annapolis, Maryland, USA
[84] Also at Ain Shams University, Cairo, Egypt
[85] Also at Bingol University, Bingol, Turkey
[86] Also at Georgian Technical University, Tbilisi, Georgia
[87] Also at Sinop University, Sinop, Turkey
[88] Also at Erciyes University, Kayseri, Turkey
[89] Also at Institute of Modern Physics and Key Laboratory of Nuclear Physics and Ion-beam Application (MOE) — Fudan University, Shanghai, China
[90] Also at Texas A&M University at Qatar, Doha, Qatar
[91] Also at Kyungpook National University, Daegu, Korea
[92] Also at another institute or international laboratory covered by a cooperation agreement with CERN
[93] Also at Yerevan Physics Institute, Yerevan, Armenia
[94] Now at University of Florida, Gainesville, Florida, USA
[95] Also at Imperial College, London, United Kingdom
[96] Also at Institute of Nuclear Physics of the Uzbekistan Academy of Sciences, Tashkent, Uzbekistan

Appendix 3

III

L. Tani and C. Veelken, “Comparison of Bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning applications in high energy physics”, *Comput. Phys. Commun.*, vol. 294, p. 108 955, 2024. DOI: `10.1016/j.cpc.2023.108955`. arXiv: `2201.06809 [physics.data-an]`

Computer Physics Communications 294 (2024) 108955

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Computational Physics

Comparison of Bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning applications in high energy physics ☆

Laurits Tani *, Christian Veelken

National Institute of Chemical Physics and Biophysics (NICPB), Akadeemia tee 23, 12618 Tallinn, Estonia

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.8171318

Keywords:
Hyperparameter optimization
High energy physics
Evolutionary algorithms
Machine learning

ABSTRACT

When using machine learning (ML) techniques, users typically need to choose a plethora of algorithm-specific parameters, referred to as hyperparameters. In this paper, we compare the performance of two algorithms, particle swarm optimisation (PSO) and Bayesian optimisation (BO), for the autonomous determination of these hyperparameters in applications to different ML tasks typical for the field of high energy physics (HEP). Our evaluation of the performance includes a comparison of the capability of the PSO and BO algorithms to make efficient use of the highly parallel computing resources that are characteristic of contemporary HEP experiments.

1. Introduction

Machine learning (ML) methods often aid the analysis of the vast amounts of data that are produced by contemporary high energy physics (HEP) experiments. The ML algorithms often feature tunable parameters, referred to as hyperparameters, which need to be chosen by the user and often have a significant effect on the algorithm's performance. In a previous publication [1] we presented two different algorithms, particle swarm optimisation (PSO) and the genetic algorithm, for the autonomous determination of these hyperparameters. In the present paper we compare the performance of the PSO algorithm, the more promising of the two algorithms studied in our previous publication, to the performance of the Bayesian optimisation (BO) [2–6] algorithm. The latter is widely used for the task of finding optimal hyperparameter values in the context of ML applications since the pioneering work of Refs. [7–9]. The "asynchronous successive halving algorithm" (ASHA) [10] is an alternative algorithm for optimising the values of hyperparameters, which is popular in the ML community outside the field of HEP.

As in our previous publication, we formulate the task of determining the set of optimal hyperparameter values as a function maximisation problem. More specifically, given an ML algorithm $\mathcal{A}$, we seek to find a point h in the space $\mathcal{H}$ of hyperparameters, such that the performance of the ML algorithm $\mathcal{A}$ attains its maximum at this point:

$$\hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmax}}\ s(h),$$

where the objective (or "score") function (OF) $s(h)$ quantifies the performance of the ML algorithm $\mathcal{A}$, and the point h at which $s(h)$ attains its maximum is denoted by the symbol $\hat{h}$. We compare the performance of the PSO and BO algorithms on two benchmark tasks, the task of finding the minimum of the Rosenbrock function [11], and on a typical data analysis task in the field of HEP, the "ATLAS Higgs boson machine learning challenge" [12].

An important aspect in applications of ML algorithms in the field of HEP is an algorithm's capability to make efficient use of massively parallel computing facilities. A single training of an ML algorithm on a single machine may take several hours, days, or, in extreme cases, even weeks. In the context of the hyperparameter optimisation task, such a single training corresponds to a single evaluation of the OF $s(h)$. In order for the hyperparameter optimisation task to finish within an acceptable time, different evaluations of the OF, *i.e.* different ML trainings, need to be executed in parallel. The computing facilities of contemporary HEP experiments typically allow users concurrent access to hundreds, sometimes even thousands, of machines. It is therefore of high practical relevance whether the PSO and BO algorithms can organise the hyperparameter optimisation task such that hundreds of ML trainings can be executed in parallel. We find that both algorithms fulfill this requirement, but do exhibit some differences in performance compared

☆ The review of this paper was arranged by Prof. Z. Was.

* Corresponding author.
E-mail address: laurits.tani@cern.ch (L. Tani).

https://doi.org/10.1016/j.cpc.2023.108955
Received 21 July 2023; Received in revised form 20 September 2023; Accepted 3 October 2023
Available online 6 October 2023

to the case that all ML trainings are executed sequentially on a single machine.

The manuscript is structured as follows: In Section 2, we present the main concepts of the BO algorithm. The PSO algorithm is described in Ref. [1]. In Section 3, we compare the performance of the PSO and BO algorithms on the two benchmark tasks. The study of the parallelisation capability of the two algorithms is presented in Section 4. We conclude the paper with a summary in Section 5.

2. Bayesian optimisation

The BO algorithm is designed to facilitate the numerical maximisation of objective functions $s(h)$ which are time-consuming to evaluate and for which the analytical form is in general not known. The BO algorithm further allows to solve the maximisation task without using derivative information on $s(h)$.

This is achieved by performing the maximisation not on $s(h)$ directly, but on an approximation of $s(h)$, which is referred to as the "surrogate" function (SF). The SF is chosen such that it is fast to evaluate and its analytic form, including its derivative, is known. We use a Gaussian process [13] with the Matérn kernel [14] for the SF in this paper. For each point $h \in \mathcal{H}$, the SF provides two values: an estimate for the value $s(h)$ of the OF and a confidence interval. The latter represents an estimate of the accuracy of the approximation of the OF by the SF at this point.

The numerical maximisation of the OF is performed by an iterative procedure. Each iteration consists of two steps: The first step consists of finding the next point h for which to perform the time-consuming evaluation of the OF. The task of finding this point is performed by an "acquisition function" (AF). The inputs to the AF consist of the estimate for the values $s(h)$ provided by the SF and the estimated accuracy of the approximation of the OF by the SF at this point. In the second step, the SF is updated with the information of the value of the OF at the point h, in order to improve the accuracy of the approximation. Each evaluation of the OF thus serves two purposes: first, to find points h where the OF attains a higher value $s(h)$ compared to previously found points and second, to improve the accuracy of the approximation of the OF by the SF. We use the expected improvement (EI) [15] for the AF in this paper. The time-consuming evaluation of the OF is performed at the point where the AF reaches its maximum. The search for the maximum of the AF is performed numerically, using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) [16–19] algorithm. The numeric search for the maximum takes advantage of the fact that both the SF and the AF are fast to evaluate, which allows multiple evaluations of the SF and AF to be made in order to find the best point for which to make the time-consuming next evaluation of the OF. The EI acquisition function has a parameter, denoted by the symbol ξ, that allows to regulate how much importance is given to finding points h where the OF attains a high value ("exploitation") versus finding points which improve the accuracy of the approximation ("exploration"). The former are typically located in the neighbourhood of previously found points, while the latter are typically located in previously unexplored regions of the space $\mathcal{H}$. Higher values of ξ result in more exploration and lower values in more exploitation. The update of the SF after each evaluation of the OF is the feature to which the Bayesian optimisation algorithm owes its name. The SF is referred to as "prior function" before evaluating the OF and as "posterior function" afterwards.

The BO algorithm is started by evaluating the OF at an initial set of points and fitting the SF to the values $s(h)$ of the OF at these points. This initial set of points is chosen with the objective of populating the space $\mathcal{H}$ uniformly. We use the Latin hypercube [20] algorithm to obtain this initial set of points. After fitting the SF to the values $s(h)$ of the OF at these points, the BO algorithm enters the iterative phase: Given the SF, the BFGS algorithm is used to find the location where the AF attains its maximum. The OF is then evaluated at this point and the SF is updated. These steps are repeated until either the algorithm has converged, indicated by changes in the value of the OF that fall below a given threshold, or the computing resources are exhausted, *i.e.* a maximum number of iterations or a computing-time limit is reached.

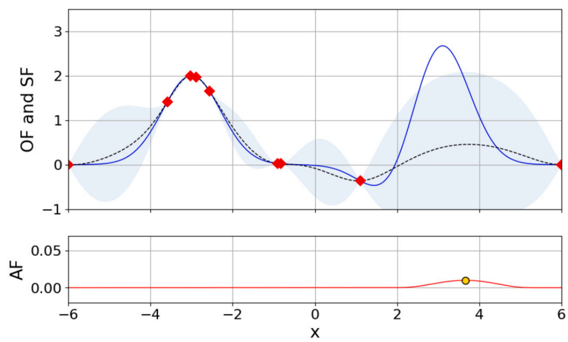


Fig. 1. Example for the operation of the BO algorithm: iteration 7. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

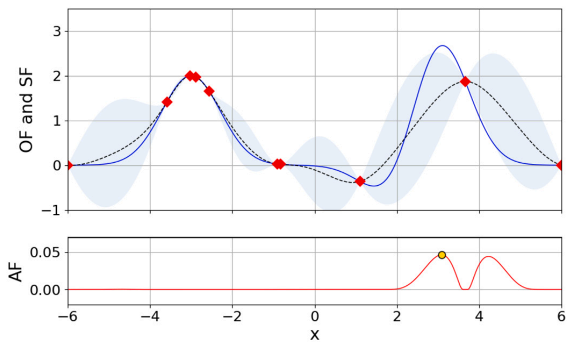


Fig. 2. Example for the operation of the BO algorithm: iteration 8.

The operation of the BO algorithm is illustrated by means of an example in Figs. 1 and 2, which visualise the seventh and eighth iteration of the BO algorithm, respectively. The OF is represented by the solid blue line and the SF by the dashed black line in the upper part of each figure. The red diamond-shaped markers indicate the points where the OF has been evaluated previously (including 2 initial points, which were chosen randomly in this example). The shaded blue area represents the confidence interval that quantifies the accuracy of the approximation of the OF by the SF. The AF is represented by the solid red line in the lower part of the figures. The yellow circle along the red line indicates the maximum of the AF, *i.e.* the point h where the next evaluation of the OF is performed.

The BO algorithm has been implemented as described above by the authors. We have validated our implementation by comparing its performance to the implementation of the BO algorithm included in the scikit-optimize [21] package.

The BO algorithm that we described above allows to evaluate the OF at one point h per iteration of the algorithm. This is well suited for executing the BO algorithm sequentially on a single machine. For the purpose of optimising the hyperparameters of ML algorithms the training of which is performed on massively parallel computing facilities, the iterative phase of the BO algorithm needs to be extended to provide multiple points h per iteration of the BO algorithm. This extension of the BO algorithm is non-trivial and still a field of active research [22]. In this paper, we follow the approach described in Ref. [23], referred to as "multi-points expected improvement" (q-EI), and use the implementation provided by Ref. [24] when running the hyperparameter parameter optimisation on multiple machines in parallel.

Table 1
Parameter settings for the BO algorithm. The parameter N_{init}^{points} refers to the number of points, obtained from the Latin hypercube algorithm, that are used to initialise the BO algorithm. The parameter ξ regulates the relative importance of exploitation versus exploration of the EI acquisition function.

Parameter	Value
N_{init}^{points}	100
ξ	0.01

Table 2
Parameter settings for the PSO algorithm. The parameters are described in Ref. [1].

Parameter	Value
N_{info}	10
c_1	1.62
c_2	1.62
w_{min}	0.4
w_{max}	0.8

3. Performance

In this section, the performance of the PSO and BO algorithms is compared on the two benchmark tasks: the finding of the minimum of the Rosenbrock function and the ATLAS Higgs boson machine learning challenge.

3.1. Rosenbrock function

The Rosenbrock function [11] is a widely used trial function for evaluating the performance of function minimisation algorithms. It is defined as:

$$R(x, y) = (a - x)^2 + b(y - x^2)^2 . \tag{1}$$

Its domain is the x–y plane and it depends on two parameters a and b. The global minimum of the Rosenbrock function is located at $(x, y) = (a, a^2)$. We choose the two parameters to be $a = 1$ and $b = 10$, resulting in the minimum to be located at $(x, y) = (1, 1)$ and the value of the Rosenbrock function at the minimum to be $R(1, 1) = 0$. See Fig. 3 in Ref. [1] for a visualisation of the Rosenbrock function around the minimum.

We treat the task of finding the minimum of the Rosenbrock function as a two dimensional hyperparameter optimisation problem. The optimal set of hyperparameters is scanned in the range of $[-500, +500] \times [-500, +500]$.

Both the BO and the PSO algorithms were executed for 30 iterations. During each iteration, 100 different hyperparameter sets were evaluated in parallel. The parameter settings for both algorithms are given in Tables 1 and 2. The minimisation of the Rosenbrock function was repeated for 1000 trials, using a different random number seed for each trial.

Performance We denote the location of the minimum found in each trial i by the symbol $\hat{\hat{h}}_i$ and the value of the Rosenbrock function at these points by the symbol $\hat{\hat{R}}_i = R(\hat{\hat{h}}_i)$. The average of these values over the 1000 trials, $\langle \hat{\hat{R}} \rangle = \frac{1}{1000} \cdot \sum_{i=1}^{1000} \hat{\hat{R}}_i$, is shown in Fig. 3. For the first 10 iterations, the BO algorithm converges faster to the minimum than the PSO algorithm, but improves less rapidly than the PSO algorithm for more than 10 iterations. This difference in the rate of convergence is expected, since the BO algorithm has been developed for applications where the number of evaluations of the OF is in the order of a few hundred to a thousand [22] (here it is 3000, given by the product of 30 iterations times 100 different hyperparameter sets that are evaluated in parallel per iteration). Neither the BO nor the PSO algorithm converges to the true minimum $R = 0$ within 30 iterations. We have shown in Ref. [1] that the PSO algorithm converges to the true minimum after 10^4 iterations. The shaded band in Fig. 3 represents the standard deviation of the $\hat{\hat{R}}_i$ over the 1000 trials.

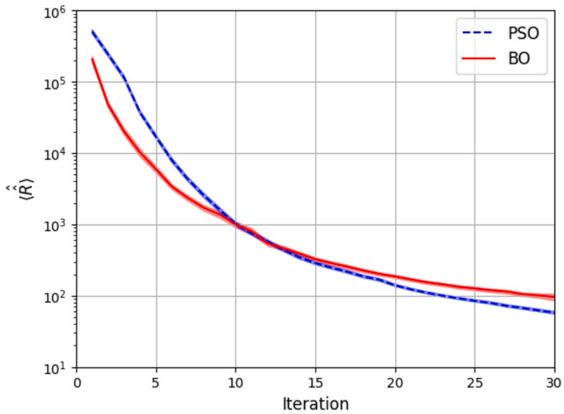


Fig. 3. Minimum of the Rosenbrock function found by the BO and PSO algorithms, as function of the number of iterations. The lines represent the average performance over 1000 trials and the shaded bands the variation (standard deviation) of the performance.

Table 3
Minimum and maximum values of the hyperparameters for the HBC. The hyperparameters are detailed in Ref. [12].

Hyperparameter	min	max
num-boost-round	1	500
learning-rate	10^{-5}	1
max-depth	1	6
gamma	0	5
min-child-weight	0	500
subsample	0.8	1
colsample-bytree	0.3	1

3.2. The ATLAS Higgs Boson machine learning challenge

The *ATLAS Higgs Boson machine learning challenge (HBC)* [12] constitutes the second benchmark task for evaluating the performance of the BO and PSO algorithms. As in our previous publication [1], we choose a boosted decision tree (BDT), implemented in the XGBoost [25] package, for the ML algorithm. The hyperparameter optimisation is performed with respect to the 7 hyperparameters given in Table 3 of Ref. [1]. During the optimisation, each of the 7 hyperparameters is restricted to be within the range given in Table 3.

The optimisation of the hyperparameters by the BO and PSO algorithms was run for 30 iterations. 70 BDTs, initialised with different random number seeds, were trained in parallel per iteration. The parameter settings used for the BO and PSO algorithm are the same as for the task of finding the minimum of the Rosenbrock function and are given in Tables 1 and 2, except for the parameter N_{info} of the PSO algorithm, which was set to 7 for the HBC task. The 550000 signal and background events provided by the organisers of the HBC are split into training and test sets as described in Ref. [1]. The "modified approximate mean significance" (d-AMS) defined by Eq. (8) of Ref. [1] was used as objective function for the BDT training. The d-AMS scores were computed setting the coefficient κ that controls the penalty term against overtraining to 0.3. The hyperparameter optimisation was repeated for 100 trials.

Table 4
Average values (μ) of the hyperparameters found by the BO and PSO algorithms and their standard deviation (σ) for the HBC task.

Hyperparameter	BO		PSO	
	μ	σ	μ	σ
num-boost-round	364.3	85.2	413.7	85.4
learning-rate	0.126	0.041	0.089	0.024
max-depth	3.9	0.7	4.5	0.7
gamma	1.78	2.00	2.81	1.47
min-child-weight	352.7	130.1	440.4	72.4
subsample	0.861	0.081	0.871	0.066
colsample-bytree	0.852	0.268	0.859	0.145

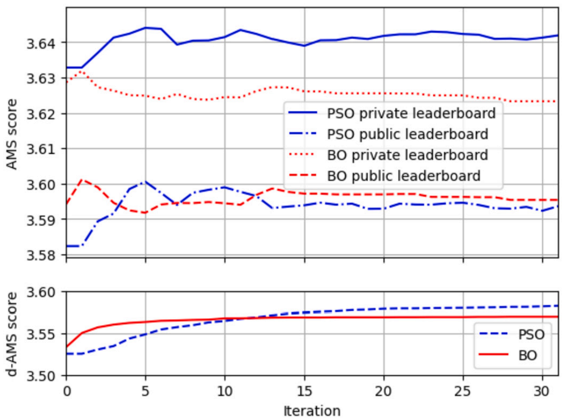


Fig. 4. Evolution of the AMS score as function of the number of iterations for the HBC task. The lines represent the average performance over 100 trials.

3.2.1. Results

The average hyperparameter values found by the BO and PSO algorithms over the 100 trials and the standard deviation of these values are reported in Table 4. The values found by both algorithms are similar. The *gamma* hyperparameter has little effect on the d-AMS score and is thus not well constrained by the optimisation.

The performance of BDTs is evaluated on two separate samples of events, referred to as the *public* and *private* leaderboard samples [1,12]. Following Ref. [1], the performance is quantified by the “approximate mean significance” (AMS) score. The latter is averaged over the 100 trials. The motivation for using the d-AMS score as objective function for the training, while the (final) performance is quantified using the AMS score is detailed in Ref. [1]. The evolution of the average performance as function of the number of iterations is shown in Fig. 4. The bottom part of the figure shows the evolution of the d-AMS score, the objective function used for the BDT training. While the d-AMS score keeps increasing monotonously with more iterations, the AMS scores on the public and private leaderboard samples reach a plateau already after 5–10 iterations. In subsequent iterations, the AMS scores start to fluctuate around the plateau. The magnitude of the fluctuations is of $\mathcal{O}(10^{-2})$. The differences in performance between the public and private leaderboard samples are compatible with a statistical fluctuation of the signal and background events contained in these samples [12]. The BO and PSO algorithms achieve a similar performance on the HBC task.

4. Parallelisation capability

The capability of the BO and PSO algorithms to make efficient use of parallel computing resources is studied on the task of finding the minimum of the Rosenbrock function. We denote by $N_{parallel}$ the number of points h in hyperparameter space that are evaluated in parallel per iteration of the BO and PSO algorithms. The value of $N_{parallel}$ represents the number of ML trainings that are executed in parallel on different machines. Ideally, the duration (“wall time”) of the hyperparameter optimisation task should decrease inversely proportional to $N_{parallel}$, except for some (small) overhead imposed by the BO and PSO algorithms.

Bayesian optimisation Based on the literature [22], one expects the BO algorithm to perform best when all ML trainings are performed sequentially on a single machine, with one point h in hyperparameter space evaluated per iteration. We have compared the performance, quantified by the values of $\langle\hat{\hat{R}}\rangle$, obtained when executing our implementation of the BO algorithm (described in Section 2) sequentially on a single machine ($N_{parallel} = 1$) with the performance obtained by the q-EI algorithm from Ref. [24], with the latter running either $N_{parallel} = 2$, 5, 10, 25, 50, 100, 250, 500 or 1000 ML trainings in parallel. We find that the values of $\langle\hat{\hat{R}}\rangle$ are very similar in all cases and mainly depend on the total number of evaluations of the Rosenbrock function, given by the product of the number of iterations times $N_{parallel}$. For 3000 evaluations, the value of $\langle\hat{\hat{R}}\rangle$ amounts to about 10^2 for all values of $N_{parallel}$ that we have tried.

The computing (CPU) time overhead imposed by the BO algorithm for updating the SF and finding the maximum of the AF amounts to about 50 CPU hours on a 2.30 GHz Intel® Xeon® E5-2695 v3 processor. The overhead does not vary much as function of $N_{parallel}$. The task of updating the SF and finding the maximum of the AF needs to run (as a “supervisor” task) on a single machine. The overhead is sizeable compared to the computing time spent on performing 3000 evaluations of the Rosenbrock function, which only takes 0.06 CPU seconds, but becomes less important for “real” ML training applications: The total computing time spent on the HPC task (with 30 iterations and 70 ML trainings running in parallel) amounts to about 500 CPU hours.

Particle swarm optimisation In contrast to the BO algorithm, the capability to make efficient use of parallel computing resources is an intrinsic feature of the PSO algorithm. We find that the performance of the PSO algorithm depends on the value of $N_{parallel}$, which corresponds to the number of particles in the swarm, in a non-trivial manner. Both too small and too large values of $N_{parallel}$ degrade the performance of the PSO algorithm compared to the optimal value. We find that the optimal value of $N_{parallel}$ amounts to about 2% times the total number (N_{tot}) of evaluations of the Rosenbrock function, *i.e.* the best performance of the PSO algorithm is achieved when using $N_{parallel} = 0.02 \cdot N_{tot}$ particles in the swarm and a fixed number of 50 iterations (up to 10000 evaluations of the Rosenbrock function that we have tried). We recommend to set the parameter N_{info} of the PSO algorithm to 10% times $N_{parallel}$.

We find that the CPU time overhead imposed by the PSO algorithm (spent on updating the positions and momenta of the particles in the swarm) is negligible for both benchmark tasks.

The CPU time overhead limits the “speedup” (reduction in wall time) that one can achieve by increasing the number of points in hyperparameter space that are evaluated in parallel. The effect is referred to as “Amdahl’s law” [26] in the literature and visualized in Fig. 5. An ideal algorithm with negligible overhead would achieve a speedup factor equal to $N_{parallel}$ in this figure. The speedup achieved by the PSO algorithm on both benchmark tasks is very close to the ideal case. The CPU time overhead limits the speedup factor achievable by the BO algorithm to about 3 for the task of finding of the minimum of the Rosenbrock function and to about 12 for the HBC task.

5. Summary

We have compared the performance of two autonomous algorithms for the optimisation of hyperparameters, Bayesian optimisation (BO)

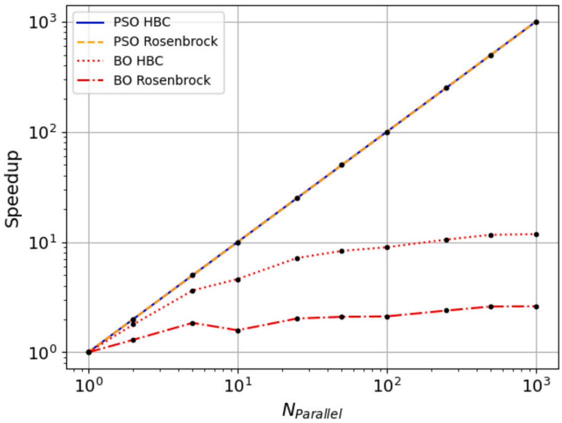


Fig. 5. Amdahl's law: Parallelisation properties of the BO and PSO algorithms on the two benchmark tasks. The two curves for the PSO algorithm are very close.

and particle swarm optimisation (PSO), on two benchmark tasks typical for ML applications in the field of high energy physics: the task of finding the minimum of the Rosenbrock function and the ATLAS Higgs boson machine learning challenge.

We find that the BO algorithm performs better than the PSO algorithm when the total number of evaluations of the Rosenbrock function (equivalent to the number of ML trainings) is of the order of a few hundred to a few thousand. If the number of evaluations (ML trainings) is large, the PSO algorithm outperforms the BO algorithm.

The capability of both algorithms to make efficient use of parallel computing resources is good. In particular, we find that the "multi-points expected improvement" of the BO algorithm provides similar performance when running on parallel computing resources compared to executing the BO algorithm sequentially on a single machine. In the case of the PSO algorithm, we found that the best performance is achieved by setting the number of particles in the swarm to 2% times the total number of function evaluations (ML trainings) and using a fixed number of 50 iterations.

We found that the BO algorithm may add a significant computational overhead to the task of finding the optimal hyperparameter values, while for the PSO algorithm the overhead is insignificant.

CRediT authorship contribution statement

Laurits Tani: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Christian Veelken:** Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data is made available by the organizers of the ATLAS Higgs Boson Machine Learning Challenge. Software used to produce the results is published on Zenodo: https://doi.org/10.5281/zenodo.8171318.

Acknowledgements

We would like to thank our colleague Joosep Pata for helpful discussions. This work has been supported by the Estonian Research Council grant PRG445.

References

[1] L. Tani, D. Rand, C. Veelken, M. Kadastik, Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics, Eur. Phys. J. C 81 (2) (2021) 1.
[2] H.J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, J. Basic Eng. 86 (1) (1964) 97, https://doi.org/10.1115/1.3653121.
[3] A.G. Zhilinskas, Single-step Bayesian search method for an extremum of functions of a single variable, Cybernetics 11 (1) (1975) 160, https://doi.org/10.1007/BF01069961.
[4] J. Močkus, V. Tiesis, A. Zilinskas, The application of Bayesian methods for seeking the extremum, in: Towards Global Optimization 2, 2014, p. 117.
[5] J. Močkus, On Bayesian methods for seeking the extremum, in: G.I. Marchuk (Ed.), Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974, Springer, Berlin, Heidelberg, 1975, p. 400.
[6] J. Močkus, Bayesian Approach to Global Optimization, Springer, 1989.
[7] J. Snoek, H. Larochelle, R.P. Adams, Practical Bayesian optimization of machine learning algorithms, in: Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 2, NIPS'12, Curran Associates Inc., Red Hook, NY, USA, 2012, p. 2951.
[8] J. Bergstra, D. Yamins, D.D. Cox, Hyperopt: a Python library for optimizing the hyperparameters of machine learning algorithms, in: Stéfan van der Walt, Jarrod Millman, Katy Huff (Eds.), Proceedings of the 12th Python in Science Conference, 2013, p. 13.
[9] C. Thornton, F. Hutter, H.H. Hoos, K. Leyton-Brown, Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 847.
[10] L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, M. Hardt, B. Recht, A. Talwalkar, A system for massively parallel hyperparameter tuning, arXiv:1810.05934, 2020.
[11] Y.-W. Shang, Y.-H. Qiu, A note on the extended Rosenbrock function, Evol. Comput. 14 (1) (2006) 119.
[12] ATLAS collaboration, Dataset from the ATLAS Higgs boson machine learning challenge, in: CERN Open Data Portal, 2014.
[13] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2005.
[14] M.G. Genton, Classes of kernels for machine learning: a statistics perspective, J. Mach. Learn. Res. 2 (2002) 299.
[15] D.R. Jones, M. Schonlau, W.J. Welch, Efficient global optimization of expensive black-box functions, J. Glob. Optim. 13 (4) (1998) 455, https://doi.org/10.1023/A:1008306431147.
[16] C.G. Broyden, The convergence of a class of double-rank minimization algorithms 1. General considerations, IMA J. Appl. Math. 6 (1) (1970) 76, https://doi.org/10.1093/imamat/6.1.76.
[17] R. Fletcher, A new approach to variable metric algorithms, Comput. J. 13 (3) (1970) 317, https://doi.org/10.1093/comjnl/13.3.317.
[18] D. Goldfarb, A family of variable-metric methods derived by variational means, Math. Comput. 24 (109) (1970) 23, http://www.jstor.org/stable/2004873.
[19] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization, Math. Comput. 24 (111) (1970) 647, http://www.jstor.org/stable/2004840.
[20] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (2) (1979) 239, http://www.jstor.org/stable/1268522.
[21] F. Pedregosa, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825, arXiv:1201.0490.
[22] P.I. Frazier, A tutorial on Bayesian optimization, arXiv:1807.02811, 2018.
[23] D. Ginsbourger, R.L. Riche, L. Carraro, A multi-points criterion for deterministic parallel global optimization based on Gaussian processes, Tech. rep., HAL open science, hal-00260579, 2008, https://hal.archives-ouvertes.fr/hal-00260579.
[24] J. Wang, S.C. Clark, E. Liu, P.I. Frazier, Parallel Bayesian global optimization of expensive functions, Oper. Res. 68 (2020) 1850.
[25] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, p. 785.
[26] G.M. Amdahl, Validity of the single processor approach to achieving large scale computing capabilities, in: Proceedings of the April 18–20, 1967, Spring Joint Computer Conference, AFIPS '67 (Spring), Association for Computing Machinery, New York, NY, USA, 1967, p. 483.

Appendix 4

IV

T. Lange *et al.*, “Tau lepton identification and reconstruction: A new frontier for jet-tagging ML algorithms”, *Comput. Phys. Commun.*, vol. 298, p. 109 095, 2024. DOI: 10.1016/j.cpc.2024.109095. arXiv: 2307.07747 [hep-ex]

Computer Physics Communications 298 (2024) 109095

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Computational Physics

Tau lepton identification and reconstruction: A new frontier for jet-tagging ML algorithms ☆

Torben Lange, Saswati Nandan, Joosep Pata, Laurits Tani *, Christian Veelken

National Institute of Chemical Physics and Biophysics (NICPB), Rävala pst 10, 10143 Tallinn, Estonia

ARTICLE INFO

Keywords:
Tau identification
Machine learning
ParticleTransformer
LorentzNet
DeepTau

ABSTRACT

Identifying and reconstructing hadronic τ decays (τ_h) is an important task at current and future high-energy physics experiments, as τ_h represent an important tool to analyze the production of Higgs and electroweak bosons as well as to search for physics beyond the Standard Model. The identification of τ_h can be viewed as a generalization and extension of jet-flavour tagging, which has in the recent years undergone significant progress due to the use of deep learning. Based on a granular simulation with realistic detector effects and a particle flow-based event reconstruction, we show in this paper that deep learning-based jet-flavour-tagging algorithms are powerful τ_h identifiers. Specifically, we show that jet-flavour-tagging algorithms such as LorentzNet and ParticleTransformer can be adapted in an end-to-end fashion for discriminating τ_h from quark and gluon jets. We find that the end-to-end transformer-based approach significantly outperforms contemporary state-of-the-art τ_h reconstruction and identification algorithms currently in use at the Large Hadron Collider.

1. Introduction

Jets constitute an important experimental signature at current and future high-energy physics experiments. The task of identifying or "tagging" the parton that originated the jet based on its constituent structure has undergone remarkable progress in recent years, driven by supervised machine learning on open datasets, see e.g. Refs. [1,2] for a review. In particular, the application of advanced deep-learning (DL) techniques, originally developed for image, natural-language and point-cloud processing, has enabled fine-grained and robust classification based on the rich information available in the particle constituents of a jet. Hadronic τ decays (τ_h) can be regarded as a special type of highly-collimated jets of low particle multiplicity. This motivated us to study the prospects for applying the same techniques to the task of identifying τ_h based on the particle content of the τ_h candidates. The results of this study are presented in this paper.

With a lifetime of 2.9×10^{-13} seconds, the τ lepton decays almost instantaneously. It can thus not be detected directly, but the particles produced in the τ decay can. The reconstruction and identification of τ leptons is thus based on the reconstruction of the τ decay products. In about one-third of the cases the τ decays to an electron or muon plus two neutrinos. In the remaining two-thirds of the cases, the τ decays into one neutrino plus a system of hadrons, consisting of typically either one or three charged pions ($\pi^{\pm}$) or kaons ($K^{\pm}$) and up to two neutral pions (π^0). The π^0 mesons decay nearly instantly and almost exclusively to a pair of photons. Decays of τ leptons into five charged mesons are rare [3]. In the energy range of interest, the $\pi^{\pm}$ and $K^{\pm}$ produced in the τ decays are difficult to distinguish experimentally. We collectively refer to them using the symbol $h^{\pm}$. We further introduce the symbol τ_h to refer to the system of all hadrons produced in the τ decay. The decays of τ^+ and τ^- are related by charge conjugation invariance.

The τ lepton is instrumental for Standard Model (SM) precision measurements as well as for searches for physics beyond the SM (BSM). Measurements of the τ lepton's properties, such as its lifetime, mass, and branching fractions allow to test the universality between lepton generations of the charged-current coupling, while measurements of its spin polarization allow to probe the neutral-current coupling of the τ lepton [4,5]. Measurements of hadronic τ decays allow to study perturbative and non-perturbative effects in quantum chromodynamics. A variety of models for BSM physics predict new particles that predominantly decay to τ leptons, such as models with extended gauge symmetries that manifest themselves through heavy charged and neutral gauge bosons [6–8], models of third generation lepto-quarks [9], supersymmetric models [10–18], and models with an extended Higgs

☆ The review of this paper was arranged by Prof. Z. Was.

* Corresponding author.
E-mail address: laurits.tani@cern.ch (L. Tani).

https://doi.org/10.1016/j.cpc.2024.109095
Received 6 December 2023; Accepted 14 January 2024
Available online 18 January 2024
0010-4655/

sector [19–23]. The τ lepton is also important for tests of lepton-flavour violation, which may reveal itself in τ decays [4,5] as well as in lepton-flavour violating decays of Z [24,25] and Higgs (H) [26,27] bosons to a τ lepton and an electron or muon. The sizable coupling of the τ lepton to the H boson has been used to probe the H boson coupling to fermions [28,29] and to study the Higgs potential [30,31].

In this paper, we focus on the identification of hadronic τ decays. The e and μ produced in τ decays can be reconstructed and identified using standard algorithms for electron and muon reconstruction.[1] Our study is performed in electron–positron (e^+e^-) collisions at the CLIC linear collider [34] at a centre-of-mass energy of $\sqrt{s} = 380$ GeV. The motivation for performing the study in e^+e^- collisions is twofold: first, because a detailed simulation of the CLICdet detector [35] and a performant event reconstruction based on the particle-flow (PF) approach [36–38] is publicly available. The PF approach combines information provided by tracking detectors with calorimeter information. Our experience at the Large Hadron Collider (LHC) is that the PF approach greatly benefits the identification of hadronic τ decays. Second, the literature on τ_h identification at future high-energy e^+e^- experiments is sparse and often based on simple algorithms similar to those used by the ATLAS and CMS collaborations during the start-up of the LHC [39,40]. Relevant references are [41–43]. Refs. [44–46,4] focus on distinguishing between individual hadronic decay modes of the τ lepton, which is important in particular for measurements of its branching fractions and for τ spin polarization measurements.

The main result of this paper is that the advancements in DL techniques that drove the progress in jet-flavour tagging significantly improve the identification of hadronic τ decays. We compare the performance of two recently published algorithms for jet-flavour tagging, LorentzNet [47] and ParticleTransformer [48], to state-of-the-art τ_h reconstruction and identification algorithms currently in use at the LHC, the "hadrons-plus-strips" (HPS) [49,50] and DeepTau [51] algorithms. The reconstruction and identification of τ_h by the latter algorithms proceeds in two steps: In the first step the τ_h are reconstructed by the HPS algorithm, and in the second step the reconstructed τ_h are identified by the DeepTau algorithm, where "identification" refers to discriminating the τ_h from quark and gluon jets. The HPS algorithm has been developed by domain experts and does not employ machine-learning techniques, while the DeepTau algorithm is based on a convolutional deep neural network (DNN). We have retrained the DeepTau algorithm for e^+e^- collisions, using the same event samples for its training as for the LorentzNet and ParticleTransformer algorithms. We use the combination of the HPS algorithm for τ_h reconstruction and DeepTau algorithm for τ_h identification as reference for state-of-the-art algorithms against which we compare the performance of the DL-based algorithms. The latter perform the tasks of τ_h reconstruction and identification in a single step, following an end-to-end approach. Our choice of the LorentzNet and ParticleTransformer algorithms is based on Ref. [47], which reported that the LorentzNet algorithm outperforms alternative DL-based jet-flavour tagging algorithms such as the ResNeXt-50 [52], ParticleFlowNetwork [53], and ParticleNet [54] algorithms on different jet tagging tasks. The ParticleTransformer algorithm has been developed by the same group of authors as the LorentzNet algorithm. It extends the latter by using additional input variables, which we expect may increase the τ_h identification performance. We believe this result to be applicable to future high-energy e^+e^- experiments such as CEPC [42], CLIC [34], FCC-ee [55], and ILC [56] as well as to proton–proton (pp) collisions at the LHC.

The paper is structured as follows: In Section 2, we detail the simulated samples of τ_h and of quark and gluon jets that we use to study the τ_h reconstruction and identification performance. The reconstruction of muons, electrons, photons, charged and neutral hadrons via the PF method, which are used as input for the τ_h identification, is described in the same section. In Section 3, we present the LorentzNet and ParticleTransformer algorithms. The performance of these algorithms is compared to the performance of the HPS and DeepTau algorithms in Section 4. The HPS and DeepTau algorithms are described in the appendix. Our motivation for documenting the HPS and DeepTau algorithms in the appendix is to concisely summarize the relevant information of Refs. [49–51] and to record the few adjustments that we have made to adapt the algorithms to e^+e^- collisions. We conclude the paper with a summary and an outlook in Section 5.

2. Monte Carlo samples and event reconstruction

The optimization and subsequent performance evaluation of the τ_h identification algorithms is based on a set of Monte Carlo (MC) event samples. The samples are generated for e^+e^- collisions at a centre-of-mass energy of $\sqrt{s} = 380$ GeV, using the program PYTHIA8 [57]. We generate 1 million "signal" events of $Z/\gamma^* \to \tau\tau$ and ZH, $H \to \tau\tau$ each and 2 million "background" events of $Z/\gamma^* \to q\bar{q}'$. The ZH, $H \to \tau\tau$ sample is used to train the LorentzNet, ParticleTransformer, and DeepTau algorithms and to evaluate their performance. The $Z/\gamma^* \to \tau\tau$ sample is used as a cross check when evaluating the algorithms' performance, to examine that the algorithms did not exploit differences in event kinematics between the ZH, $H \to \tau\tau$ signal and the $Z/\gamma^* \to q\bar{q}'$ background, as this would result in an overly optimistic assessment of the algorithms' performance. To this end, suitable chosen weights are applied to the samples of τ_h and jets used for the training of the DeepTau, LorentzNet, and ParticleTransformer algorithms. The weights are chosen such that the distributions in polar angle θ and transverse momentum p_T of the reconstructed jets become identical for the ZH, $H \to \tau\tau$ signal and the $Z/\gamma^* \to q\bar{q}'$ background. Since we expect the probability for a quark or gluon jet to be misidentified as τ_h to be in the order to 10^{-2} or below, it is particularly important to have sufficient background statistics. The τ decays are simulated using PYTHIA8, and the generator tune and other settings are based on Ref. [58]. No $\gamma\gamma \to$ hadrons overlay background is included in our simulation. The study of this overlay is left to future work.

The stable particles from PYTHIA8 are passed to a full GEANT4 [59]-based detector simulation and subsequent reconstruction based on the CLICdet detector [35] and the MARLIN reconstruction code [60], interfaced in the KEY4HEP [61] software package. We use the CLICdet detector (`CLIC_o3_v14`), as the detector design has been thoroughly studied, and a rather complete implementation of tracker, calorimeter and particle flow reconstruction is available. The CLICdet detector is optimized for precision physics and is based on a silicon pixel detector and tracker, a Si-W electromagnetic calorimeter and a scintillating hadronic calorimeter, encased in a 4T solenoid. More details about the expected physics performance, including track and jet energy reconstruction properties, is available in [62] and references therein. Moreover, the CLICdet detector is conceptually similar to the proposed CLD detector for the FCC-ee [63], and thus a relevant benchmark model for particle identification and reconstruction studies.

Based on the MARLIN reconstruction in KEY4HEP, the output of the simulation and reconstruction chain is a set of PANDORA [36–38] particle flow candidates for each event, described by a four-momentum, a charge and a particle identification label: electron (e), muon (μ), photon (γ), charged hadron ($h^{\pm}$), and neutral hadron (h^0). The PANDORA particle flow algorithm aggregates calorimeter hits to clusters and combines tracks and calorimeter clusters to reconstruct stable particle candidates.

[1] The small distance that a τ lepton typically travels between its production and decay, results in a finite impact parameter of the electron or muon track with respect to the primary event vertex, which provides a handle to distinguish e and μ produced in the primary e^+e^- collision from those resulting from τ decays. The neutrinos produced in leptonic τ decays provide another handle to this end. Their momenta can be inferred from energy and momentum conservation or computed, with improved resolution, by means of dedicated algorithms [32,33].

The ParticleTransformer and DeepTau algorithms use the transverse (d_{xy}) and longitudinal (d_{z}) impact parameters of tracks to improve the discrimination of τ_{h} and jets. In jets arising from the hadronization of light quarks and gluons, the d_{xy} and d_{z} of the tracks are expected to be compatible with zero within their respective uncertainties ($\sigma_{d\mathrm{xy}}$ and $\sigma_{d\mathrm{z}}$), while non-zero impact parameters are expected for the charged particles produced in τ decays, reflecting the small distance that τ leptons travel between their production and decay. The d_{xy} and d_{z} are not part of the KEY4HEP format used in PANDORA and thus need to be computed for the work presented in this paper. As the distances that τ leptons travel between their production and decay are typically small compared to the expected radius of curvature of the tracks originating from the τ decay, we simplify the task of computing the d_{xy} and d_{z} by using a linear approximation to the equations of motion for a charged particle in a magnetic field [64]. Details of this approximation are documented in the repository of the code [65].

The particle flow candidates are clustered to jets using the generalized k_{t} algorithm for $\mathrm{e}^+\mathrm{e}^-$ collisions (`ee_genkt`) [66] with $p = -1$ and $R = 0.4$. All jets considered in this paper are required to satisfy the conditions $10 < \theta < 170°$ and $p_{\mathrm{T}} > 20$ GeV, where the symbol θ refers to the polar angle, with the z-axis taken to be the beam axis, and p_{T} to the transverse momentum. The condition on θ selects jets within the geometric acceptance of the tracking detector. Due to the Lorentz boost in τ direction, the particles produced in τ decays become more collimated in the detector as the energy of the τ lepton increases. The selection $p_{\mathrm{T}} > 20$ GeV ensures that, in the signal samples, all particles produced in τ decays are within a narrow inner region of the jet.

The jets reconstructed in $\mathrm{Z}/\gamma^* \to \tau\tau$ and ZH, $\mathrm{H} \to \tau\tau$ signal events are required to be matched to generator-level τ_{h} within $\Delta R < 0.4$, while those reconstructed in $\mathrm{Z}/\gamma^* \to \mathrm{q}\bar{\mathrm{q}}'$ background events are required to be matched to either a quark or a gluon on generator level. The distance ΔR between generator-level and reconstructed particles is computed as:

$$\Delta R = \sqrt{(\theta_i - \theta_j)^2 + (\phi_i - \phi_j)^2}\,, \tag{1}$$

where the symbol ϕ denotes the azimuthal angle of the jet, the subscript i refers to the direction of the reconstructed jet, and the subscript j to that of the generator-level τ_{h}, quark, or gluon. Reconstructed jets that are close to generator-level electrons or muons are removed from the signal and background samples. This leaves us with approximately 1.2 and 0.9 million reconstructed jets in the $\mathrm{Z}/\gamma^* \to \tau\tau$ and ZH, $\mathrm{H} \to \tau\tau$ signal samples, respectively, and 3.5 million jets in the background sample.

For each jet, we store the associated particle-flow candidates as well as the jet four momentum. The samples are shuffled and divided to mutually exclusive train-validation-test samples with a 26 : 9 : 65% split. The training dataset is used to optimize the parameters of the LorentzNet, ParticleTransformer, and DeepTau algorithms, while the validation dataset is used to monitor the training. The final performance of the different algorithms is evaluated using the test dataset.

3. Tau identification algorithms

The discrimination between τ_{h} from jets takes advantage of the fact that τ_{h} are typically more collimated and contain fewer particles compared to quark and gluon jets. Distributions of the jet radius and of the number of particles in the jet are shown in Fig. 1 for illustration. The jet radius $\langle r \rangle$ is defined by:

$$\langle r \rangle = \frac{\sum_i p_{\mathrm{T}}^i \, \Delta R}{\sum_i p_{\mathrm{T}}^i}\,, \tag{2}$$

where the sum extends over all particles i within the jet and the distance ΔR between the direction of the jet j and particle i is given by Eq. (1). We also show the distribution in jet mass. Quark and gluon jets typically have a higher mass than τ_{h}, reflecting the higher particle multiplicity and wider angular spread, while the mass of τ_{h} is bounded from above by the τ lepton mass of 1.777 GeV [3].

Each τ_{h} identification algorithm considered in this paper is seeded by the collection of jets described in Section 2 and uses the constituent particles of these jets as input to the τ_{h} reconstruction. The output of each algorithm is a discriminant $\mathcal{D}_\tau$ in the range 0 to 1, where a value of 1 means that the jet is identified as τ_{h}, while a value of 0 means that the jet is identified as originating from the hadronization of either a quark or a gluon.

All constituents of the jet are considered in the τ_{h} reconstruction. No selection on θ and p_{T} of the jet constituents is applied, because we observed that not applying such selection improves (by a small amount) the τ_{h} identification performance. We remark that it may be necessary to apply a p_{T} threshold on the jet constituents in order to reduce the effect of the $\gamma\gamma \to \mathrm{hadrons}$ overlay background, which is not included in our simulation, or in case particles of low p_{T} are not well modelled by the MC simulation.

The performance of each algorithm is evaluated in terms of τ_{h} identification efficiency and of the misidentification rate for quark and gluon jets. We denote the former by the symbol ε_τ and the latter by P_{misid}. The τ_{h} identification efficiency corresponds to the probability for a genuine τ_{h} in the $\mathrm{Z}/\gamma^* \to \tau\tau$ and ZH, $\mathrm{H} \to \tau\tau$ signal samples to pass a selection on the discriminant $\mathcal{D}_\tau$, while the misidentification rate refers to the probability for jets that originate from the hadronization of a quark or gluon in the $\mathrm{Z}/\gamma^* \to \mathrm{q}\bar{\mathrm{q}}'$ background sample to pass this selection. The probability is defined by:

$$\mathcal{P} = \frac{p_{\mathrm{T}}^{\mathrm{rec}} > 20 \text{ GeV} \;\&\; 10 < \theta_{\mathrm{rec}} < 170° \;\&\; \mathcal{D}_\tau > \mathcal{T}}{p_{\mathrm{T}}^{\text{gen-X}} > 20 \text{ GeV} \;\&\; 10 < \theta_{\text{gen-X}} < 170°}\,, \tag{3}$$

where the symbols $\mathcal{P}$ and X refer to the efficiency ε_τ (to the misidentification rate P_{misid}) and to the system of charged hadrons and neutral pions produced in the τ decay (to the quark or gluon that initiated the jet) in case of signal (background). The symbols $p_{\mathrm{T}}^{\mathrm{rec}}$ and θ_{rec} refer to the transverse momentum and polar angle of the jet that seeds the τ_{h} reconstruction in case of the LorentzNet and ParticleTransformer algorithms and to the p_{T} and θ of the τ_{h} object reconstructed by the HPS algorithm in case of the DeepTau algorithm. The symbol "&" denotes conjunction and $\mathcal{T}$ refers to the threshold imposed on the discriminant $\mathcal{D}_\tau$. The acceptance criteria defined by the denominator are applied in the numerator also. The efficiency ε_τ as well as the misidentification rate P_{misid} depend on the threshold $\mathcal{T}$ and in general vary with p_{T} and θ.

3.1. LorentzNet

The LorentzNet algorithm [47] employs a DNN architecture based on the attention mechanism [67,68]. The algorithm uses as input the four-momentum, type, and charge of the N particles of highest p_{T} among the jet's particle constituents, plus the two beam particles (the colliding e^+ and e^-). In case the jet contains fewer than N particles, the "missing" particles are represented by zeros. The number N of input particles constitutes a parameter of the algorithm, which needs to be chosen by the user. Based on Fig. 1, we choose $N = 25$. The particle type (e, μ, γ, $\mathrm{h}^\pm$, h^0) is passed in one-hot encoded format [69]. The architecture of the DNN is designed such that the output of the algorithm is equivariant to proper orthochronous Lorentz transformations (translations, rotations, and boosts) of the particles' four-momenta and is invariant to permutations of any pair of particles. The latter property means that the LorentzNet algorithm is invariant to the ordering in which the jet's particle constituents are presented to the algorithm. A function $\phi : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{m}}$ is equivariant to actions of the Lorentz group G if the following relationship holds: $\phi(g(x)) = g(\phi(x))$, where $x \in \mathbb{R}^{\mathrm{n}}$, $\phi(x) \in \mathbb{R}^{\mathrm{m}}$, and $g \in G$. The Lorentz equivariance introduces an inductive bias into the algorithm, with the aim of improving the algorithm's

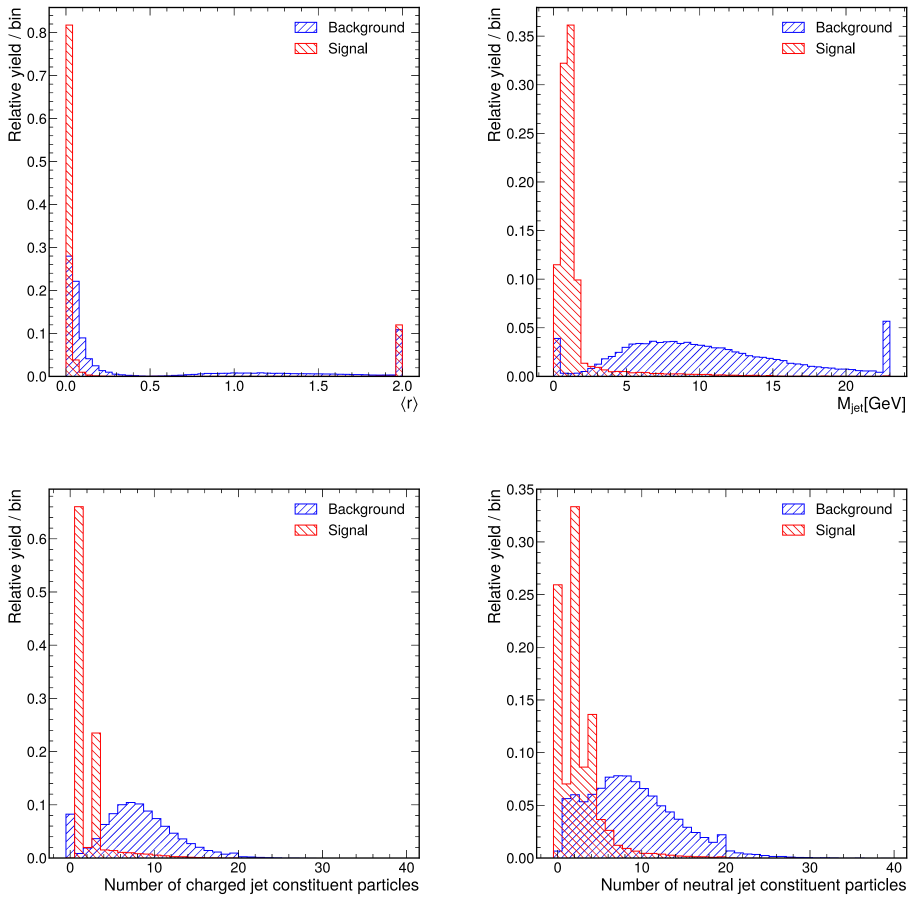


Fig. 1. Distributions in the jet radius $\langle r \rangle$ (upper left), the mass M_{jet} of the jet (upper right), and in the number N of charged (lower left) and neutral (lower right) particles in the jet. The distributions are plotted on generator level for jets originating from hadronic τ decays ("signal") compared to quark and gluon jets ("background"). The rightmost bins of each distribution represent overflow bins.

capability for generalization. It is demonstrated in Refs. [70,47] that this DNN architecture is flexible enough to approximate any Lorentz equivariant function. The algorithm is implemented in PyTorch [71]. All parameters of the LorentzNet algorithm, which need to be chosen by the user, are set to the values given in Section 3.3 of Ref. [47], except for the parameter N, as explained above, and the parameter c, which we set to $c = 0.005$.

The training is performed in batches of 128 jets and for a maximum of 100 epochs. The algorithm has $2.3 \cdot 10^5$ trainable parameters, which are updated after each epoch using the AdamW [72] optimizer, in order to minimize the loss on the training dataset. The learning rate is varied according to the one-cycle policy [73] during the training, with the maximum learning rate set to 10^{-3}. The focal loss from Ref. [74] with $\gamma = 2$ is used for the loss function. We found that this choice of loss function improves the separation of τ_{h} from quark and gluon jets in particular for intermediate values of the DNN output, compared to using binary cross-entropy loss. The loss on the validation dataset is computed after each training epoch. The final DNN parameters are chosen to be those that minimize the loss on the validation dataset.

Distributions in the discriminant $\mathcal{D}_\tau$ for the training and test datasets are shown in Fig. 2. The distributions on the test sample are represented by solid lines, while the dashed lines represent the distributions on the training sample. A moderate amount of overtraining can be seen in the figure. The main effect of the overtraining is that the tail of the background distribution in the region of high values of the discriminant $\mathcal{D}_\tau$ is more pronounced for the test sample than for the training sample.

3.2. Particle transformer

The ParticleTransformer algorithm [48] is also based on the attention mechanism [67,68]. Its architecture has originally been developed in the context of natural-language processing and is referred to as Transformer model in the literature [75,76]. The ParticleTransformer extends the LorentzNet algorithm by using additional input variables, notably the transverse and longitudinal impact parameters of charged particles, plus four observables, which represent properties of particle pairs. The algorithm is implemented in PyTorch [71]. For the per-particle features, we use the 17 observables given in Table 2 of Ref. [48]. The four pairwise features are the distance ΔR between the particles, given by Eq. (1), the mass of the particle pair, and the two observables:

$$k_{\text{t}} = \min\left(p_{\text{T}}^i, p_{\text{T}}^j\right) \Delta R$$
$$z = \min\left(p_{\text{T}}^i, p_{\text{T}}^j\right) / \left(p_{\text{T}}^i + p_{\text{T}}^j\right) , \quad (4)$$

where the superscripts i and j refer to the first and second particle of the pair, respectively. The 25 particles of highest p_{T} among the particle

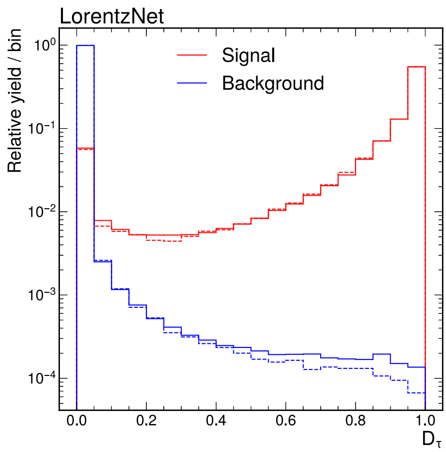


Fig. 2. Distribution in the discriminant $\mathcal{D}_\tau$ for the LorentzNet algorithm. The solid curves refer to the test dataset and the dashed curves to the training dataset. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

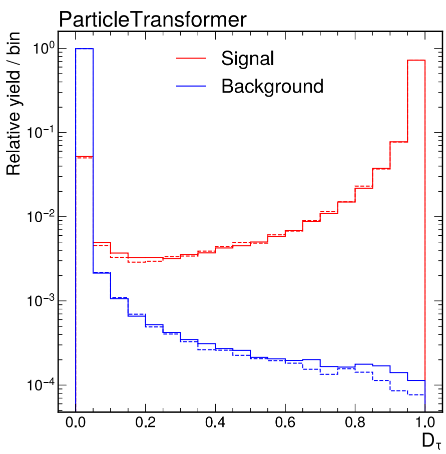


Fig. 3. Distribution in the discriminant $\mathcal{D}_\tau$ for the ParticleTransformer algorithm. The solid curves refer to the test dataset and the dashed curves to the training dataset.

constituents of the jet are considered when computing the per-particle and pairwise features.

The algorithm is trained in batches of 128 jets for a maximum of 100 epochs. The $2.1 \cdot 10^6$ trainable parameters are updated after each epoch, using the AdamW [72] optimizer. The other training parameters are the same as for the LorentzNet algorithm. The final DNN parameters are taken to be those that minimize the loss on the validation dataset, which is computed after each epoch.

Distributions in the discriminant $\mathcal{D}_\tau$ computed by the ParticleTransformer algorithm are shown in Fig. 3. A moderate amount of overtraining, similar in magnitude and shape effect to that of the LorentzNet algorithm, can be seen in the figure.

4. Results

The "receiver operating characteristic" (ROC) curves of the LorentzNet and ParticleTransformer algorithms is compared to the one of the

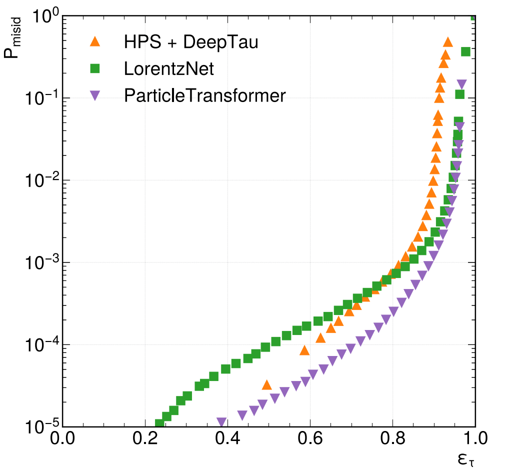


Fig. 4. Misidentification rate for quark and gluon jets as function of the τ_h identification efficiency for the LorentzNet and ParticleTransformer algorithms, compared to the combination of the HPS + DeepTau algorithm.

DeepTau algorithm in Fig. 4. The ROC curves show the τ_h identification efficiencies ε_τ for the ZH, $H \to \tau\tau$ signal sample on the axis of abscissas and the misidentification rates P_{misid} for the $Z/\gamma^* \to q\bar{q}'$ background sample on the ordinate. The curves are constructed by varying the threshold imposed on $\mathcal{D}_\tau$ in 1000 steps within the range 0 to 1 and computing ε_τ and P_{misid} according to Eq. (3) for each such threshold. Points on the left side of the curve correspond to a tighter selection on the output $\mathcal{D}_\tau$ of the τ_h identification algorithm, while points on the right side correspond to a looser selection. All three algorithms achieve misidentification rates on the level of a permille or below for τ_h identification efficiencies in the range 50-80%, the range we expect to be most relevant for physics analyses. Numerical values for P_{misid} for ε_τ values of 50, 60, 70, and 80% are given in Table 1.

The ε_τ obtained for the $Z/\gamma^* \to \tau\tau$ signal sample agrees with the one obtained for the ZH, $H \to \tau\tau$ sample within 5-10%, depending on the threshold on $\mathcal{D}_\tau$, where the quoted values refer to relative differences. The differences between the signal samples are rather small, indicating that the reweighting described in Section 2 works as intended and the algorithms do not exploit differences in event kinematics.

The ParticleTransformer algorithm is seen to outperform the LorentzNet algorithm as well as the DeepTau algorithm, achieving P_{misid} of 2.1×10^{-5} and 2.5×10^{-4} for ε_τ of 50 and 80%, respectively. This performance demonstrates the potential of applying the advanced DL algorithms originally developed for jet-flavour tagging to the task of τ_h identification.

We remark that the performance of the HPS + DeepTau algorithm in our study is significantly higher than the performance reported in Ref. [51]. The higher performance is reflected by the P_{misid} values for the DeepTau algorithm shown in Fig. 4, which are about an order of magnitude lower compared to the misidentification rates shown in Fig. 4 of Ref. [51], for similar τ_h identification efficiencies. We believe the main reason for this difference to be the "cleaner" experimental environment of e^+e^- compared to pp collisions, which considerably simplifies the task of discriminating τ_h from quark and gluon jets. The ROC curve shown in Fig. 4 of Ref. [51] was made for typical experimental conditions during LHC Run 2, which, besides the general higher hadronic activity arising from the production of extra jets and the underlying event, also included about 50 minimum bias overlay interactions, referred to as "pileup" [77]. To illustrate this point, we show in Fig. 5 the distributions in the number of particles N_{iso} in the isolation cone and in the observable $I_\tau = I_{\text{ch5}} + I_{\gamma 5}$, where I_{ch5} and

Table 1
Average misidentification rates, computed according to Eq. (3), for average τ_h identification efficiencies $\langle \varepsilon_\tau \rangle$ of 50, 60, 70, and 80%. The numbers given in the table correspond to the inverse, $1/\langle P_{misid} \rangle$, of the average misidentification rates, and to the statistical uncertainties on these values. Bold numbers highlight the best-performing algorithm.

Algorithm	$\langle \varepsilon_\tau \rangle = 50\%$	$\langle \varepsilon_\tau \rangle = 60\%$	$\langle \varepsilon_\tau \rangle = 70\%$	$\langle \varepsilon_\tau \rangle = 80\%$
HPS + DeepTau	31034 ± 3849	11796 ± 902	3564 ± 150	1308 ± 33
LorentzNet	10137 ± 719	5634 ± 298	3006 ± 116	1450 ± 39
ParticleTransformer	$\mathbf{48028 \pm 7411}$	$\mathbf{24904 \pm 2767}$	$\mathbf{10617 \pm 770}$	$\mathbf{4042 \pm 181}$

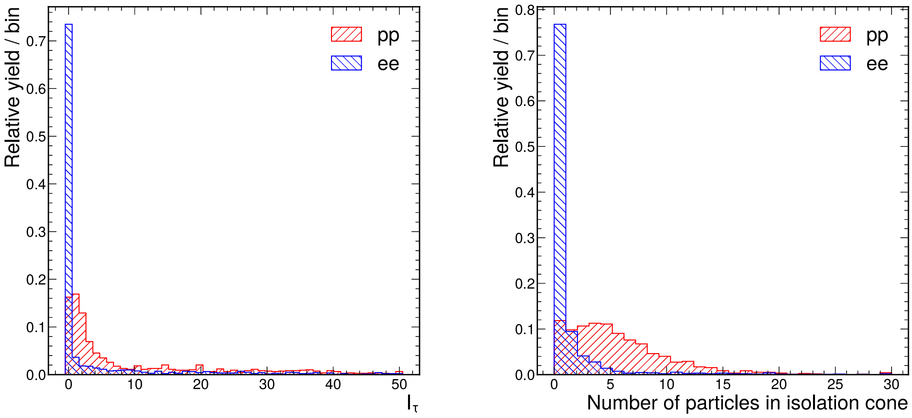


Fig. 5. Distributions in the isolation I_τ of the τ_h and in the multiplicity N_{iso} of particles in the isolation cone of the τ_h, for hadronic τ decays in ZH, $H \to \tau\tau$ signal events produced in pp (red) and in e^+e^- (blue) collisions.

$I_{\gamma 5}$ are computed according to Eq. (A.1) in the appendix, for τ_h in ZH, $H \to \tau\tau$ events produced in pp collisions at $\sqrt{s} = 13$ TeV with 50 pileup interactions and in ZH, $H \to \tau\tau$ events produced in e^+e^- collisions at $\sqrt{s} = 380$ GeV. Particles that are matched to τ decay products on generator-level are excluded from the computation of the observables N_{iso} and I_τ. The observable I_τ represents one of the main handles to separate τ_h from quark and gluon jets and significantly benefits from the "cleaner" experimental environment.

5. Summary and outlook

The main result of this paper is that the advancements in modern deep-learning techniques, which drove the recent progress in jet-flavour tagging, can similarly be applied to the task of identifying hadronic τ decays. Of the two jet-flavour tagging algorithms that we studied, LorentzNet and ParticleTransformer, the ParticleTransformer algorithm provides the superior performance. It achieves a misidentification rate of 2.1×10^{-5} (2.5×10^{-4}) for a τ_h identification efficiency of 50 (80%). A low misidentification rate is of particular importance for measurements of τ lepton branching fractions via the "tag-and-probe" method, as discussed in Section 1 of Ref. [5]. Remarkably, the ParticleTransformer algorithm achieves this level of performance while performing τ_h reconstruction and identification in a single step without any manual tuning by domain experts. We believe this result, that algorithms originally developed for jet-flavour tagging and repurposed for the task of τ_h identification provide a performance that is as good as or better than state-of-the-art τ_h identification algorithms to be applicable to pp collisions at the LHC also. We remark that the performance of the DL-based algorithms is achieved in an end-to-end approach without manual tuning by domain experts, while in particular the HPS algorithm benefitted from substantial tuning by domain experts.

We believe the numerical values for ε_τ and P_{misid} given in Table 1 may be useful in the context of sensitivity studies of physics analyses with τ leptons at future high-energy e^+e^- experiments, similar to how the parametrizations of particle reconstruction and identification performances provided in the Delphes [78] fast detector simulation software by the ATLAS and CMS experiments for pp collisions at the LHC have been used. We remark that ε_τ and P_{misid} depend on p_T and θ. This dependency is not reflected in the numbers given in Table 1, which represent averages over the p_T and θ spectrum of τ_h and jets in the signal and background samples. Values of ε_τ and P_{misid} as function of p_T and θ can be obtained from the authors upon request.

Owing to the "cleaner" experimental environment, the misidentification rates are substantially lower in e^+e^- collisions compared to pp collisions at the LHC, for similar τ_h identification efficiencies.

Future work includes to study in detail the effect of the $\gamma\gamma \to$ hadrons overlay background on the τ_h identification performance and to extend the ParticleTransformer algorithm to reconstruct individual hadronic τ decay modes. The capability to distinguish individual hadronic τ decay modes is important for measurements of the τ spin polarization [44–46,4,79]. Furthermore, in analogy with jet-flavour tagging, it may be important to study the robustness of the taggers to theoretical and experimental systematic effects and to design ML models that are robust against such effects. Another important aspect for future work is hardware portability, latency and throughput, to ensure that ML-based reconstruction models are resource-efficient and can be deployed at the trigger level, if needed.

CRediT authorship contribution statement

Torben Lange: Investigation, Software, Writing – original draft. **Saswati Nandan:** Investigation, Software, Visualization, Writing – original draft. **Joosep Pata:** Conceptualization, Project administration, Soft-

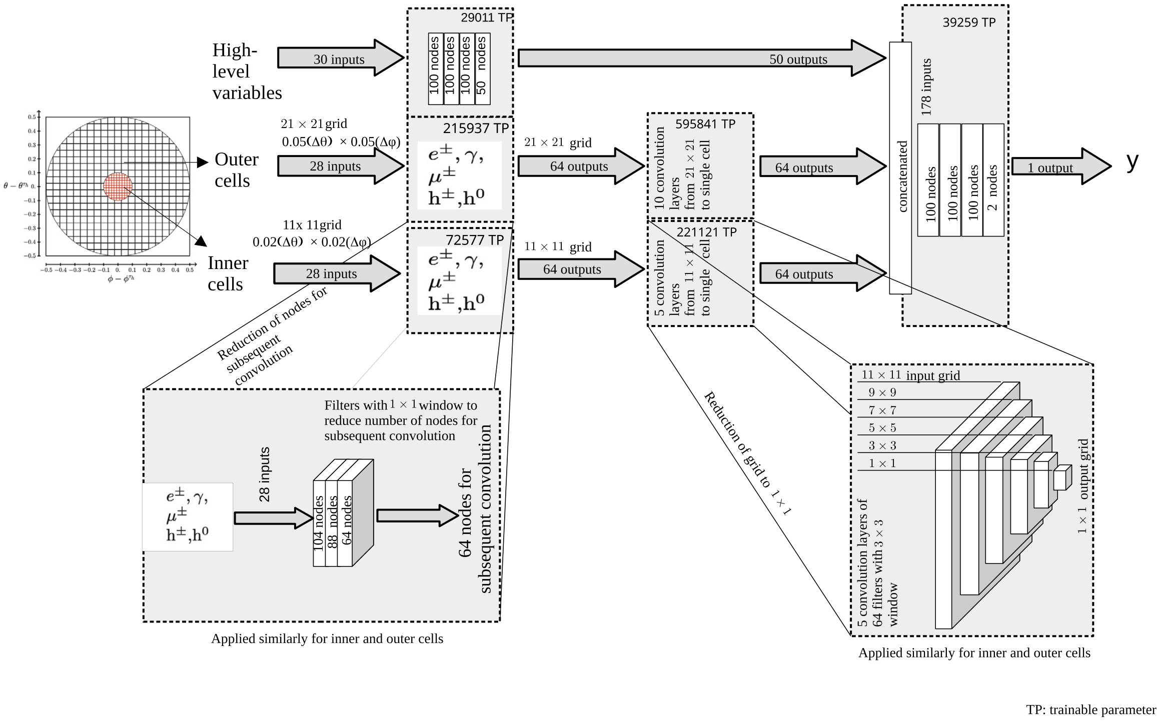


Fig. 6. The DNN architecture of the DeepTau algorithm. The numbers of trainable parameters (TP) for different components of the network are given in the figure.

ware, Supervision, Writing – original draft, Writing – review & editing. **Laurits Tani:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Christian Veelken:** Conceptualization, Funding acquisition, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Christian Veelken reports financial support was provided by Estonian Research Council grant PRG445. Joosep Pata reports financial support was provided by Estonian Research Council grant PSG864. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code used to produce the samples and perform the analysis is available at https://doi.org/10.5281/zenodo.8113344.

Acknowledgements

The authors would like to thank Patrizia Azzi for suggesting to study the topic of τ_h identification at future e^+e^- colliders and the authors of Ref. [51] for help with the preparation of Fig. 6. This work was supported by the Estonian Research Council grants PRG445 and PSG864. Computing resources were provided by the NICPB HEP cluster.

Appendix A

A.1. HPS algorithm

The HPS algorithm is published in Refs. [49,50]. The algorithm aims to reconstruct individual hadronic τ decay modes. The decay modes targeted by the HPS algorithm are $\tau^- \to h^- \nu_\tau$, $\tau^- \to h^- \pi^0 \nu_\tau$, $\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$, $\tau^- \to h^- h^+ h^- \nu_\tau$, $\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$, and the charge conjugates of these decays for τ^+.

The π^0 mesons are reconstructed by clustering photons of $p_T >$ 1.0 GeV into "strips". Electrons and positrons of $p_T > 0.5$ GeV are included in the clustering, assuming that they originate from photon conversions within the tracking detector. The clustering proceeds via an iterative procedure. The procedure is seeded by the γ or e of highest p_T that is not yet included in a strip, where we use the symbol e to refer to e^- and e^+ irrespective of their electric charge. The θ and ϕ of the seed defines the initial location of the strip. The γ or e of next highest p_T, which is within an $\theta \times \phi$ window of size 0.05×0.20 around the strip location, is added to the strip. The strip momentum is recomputed as the momentum sum of all particles in the strip and the position of the window is updated accordingly. The energy of the strip is chosen such that its mass matches the π^0 meson mass. The clustering continues until there are no more unclustered γ or e within the $\eta \times \phi$ window. The algorithm then proceeds by choosing a new seed and building the next cluster. The reconstruction of π^0 mesons ends when all jet constituents of type γ or e are clustered into strips.

The strips are combined with jet constituents of types $h^{\pm}$ in the next step of the HPS algorithm. Following the implementation of the HPS algorithm in CMS, particles of type electron are also considered as "charged hadrons" in this step. Potential double-counting of reconstructed electrons is resolved at a later stage of the algorithm. The motivation for considering reconstructed e as $h^{\pm}$ is that these e may

Table 2
Properties of particles used as input to the two subnetworks that process the inner and outer grids in the DeepTau algorithm. The variables $\Delta\theta$ and $\Delta\phi$ are defined as $\Delta\theta = \theta_i - \theta_\tau$ and $\Delta\phi = \phi_i - \phi_\tau$, where the subscript τ refers to the direction of the τ_h and i to that of a particle in the inner or outer grid. The variables d_{xy}, σ_{dxy}, d_z, and σ_{dz} are computed as described in Ref. [65]. They are set to zero if the particle is of type γ or h^0.

Variable	Description
p_T^i/p_T^τ	p_T of particle i in relation to τ_h
$\Delta\theta$, $\Delta\phi$	distance between particle i and τ_h in polar and azimuthal direction
M_i	mass of particle i
charge	electric charge of particle i
d_{xy}, σ_{dxy}	transverse impact parameter and its uncertainty
d_z, σ_{dz}	longitudinal impact parameter and its uncertainty
$\mathcal{O}_e$, $\mathcal{O}_\mu$, $\mathcal{O}_\gamma$, $\mathcal{O}_{ch}$, $\mathcal{O}_{nh}$	type of particle e, μ, γ, $h^\pm$, h^0 in one-hot-encoded format

result from the overlap of a charged hadron of low energy with a high energetic π^0 meson. The experimental signature of such an overlap is a track that spatially matches a cluster in the electromagnetic calorimeter (ECAL) whose energy is significantly higher than the track momentum and little energy in the hadronic calorimeter (HCAL). This case often gets reconstructed as one particle of type electron by the PF algorithm in CMS.

The combination of $h^\pm$ with strips proceeds via a combinatorial approach. A set of τ_h candidates corresponding to combinations of either one $h^\pm$ with up to two strips or three $h^\pm$ with up to one strip, representing the decay modes mentioned above, are constructed in parallel. In case there exist multiple possibilities for choosing the $h^\pm$ among the jet constituents or for choosing the strips among the set of strips reconstructed in the previous step, the HPS algorithm constructs all possible combinations among the 6 highest p_T $h^\pm$ and the 6 highest p_T strips. The restriction to the 6 highest p_T objects is imposed to reduce the computational complexity of the algorithm.

The constructed τ_h candidates are subject to preselection criteria, which demand the sum of $h^\pm$ charges to be equal to ± 1, all $h^\pm$ and strips to be within a signal cone of radius $\Delta R = 3.0/(p_T\,\mathrm{GeV}^{-1})$ (limited to a minimum of 0.05 and a maximum of 0.10), and the mass of the τ_h candidate to be within a certain mass window [49]. The signal cone is centred on the momentum vector of the τ_h candidate. The distance between the τ_h candidate and the $h^\pm$ or strip is computed according to Eq. (1).

The four-vector of the τ_h candidate is computed by summing the four-vectors of its constituent $h^\pm$ and strips. The energy of electrons and positrons that are considered as $h^\pm$ is adjusted such that their mass matches the $\pi^\pm$ meson mass when summing the four-vectors. In order to avoid double-counting of e in case they are considered as $h^\pm$ and have been clustered into strips, the e considered as $h^\pm$ are removed from the strips and the strip momentum is recomputed. In case there remains no particle in the strip after removing these e, the τ_h candidate is discarded.

In case multiple τ_h candidates pass the preselection criteria, the τ_h candidate of highest p_T is retained and all other τ_h candidates corresponding to the same jet are discarded.

A.2. DeepTau

The DeepTau algorithm is published in Ref. [51]. The purpose of the algorithm is to identify the τ_h that are reconstructed by the HPS algorithm as described in the previous section.

The DNN architecture of the DeepTau algorithm is illustrated in Fig. 6. It is composed of three subnetworks. Two subnetworks process the information about individual particles near the τ_h, while the third subnetwork processes a set of high-level features of the τ_h.

The first and second subnetworks have similar structure. The information about location, type, and other properties of the particles near the τ_h are discretised in two grids in the θ–ϕ plane, an inner grid of 11×11 cells of size 0.02×0.02 and an outer grid of 21×21 cells of size 0.05×0.05. The finer segmentation of the inner grid reflects the fact that the particles produced in hadronic τ decays are typically highly collimated (cf. Fig. 1) and furthermore helps to resolve the dense core of high-energetic quark and gluon jets [51]. The inner and outer grids are centred on the direction of the τ_h. The grids are populated by iterating over all particles reconstructed in the event and computing their distance in θ and ϕ with respect to the τ_h direction. Particles falling into the same cell are sorted in the order of decreasing p_T regardless of their type. The location, type, and other properties of the two particles of highest p_T are concatenated to a vector of size 28. The particle properties used to build this vector are given in Table 2. We use 14 properties per particle. Cells in the outer grid that overlap with the inner grid are skipped when populating the grids, in order to avoid redundancy of information between the two grids.[2] The features in each cell of the inner and outer grids are preprocessed by four fully-connected layers of size 104, 88, 64. The preprocessed information is then passed through a stack of 5 convolutional layers for the inner grid and 10 for the outer grid. Each convolutional layer uses 64 filters and a kernel of size 3×3. Because the convolutional layers use no padding, the size of the grid decreases by two units in θ and two units in ϕ per convolutional layer. The output of the two subnetworks for the inner and outer grids are two single cells that each hold a vector of size 64, resulting from the application of the 64 filters.

[2] The CMS implementation of the DeepTau algorithm uses a few more particle properties as inputs, which are not available in our simulation. The missing inputs concern mainly detector-level observables, which are used to improve the separation of τ_h from electrons and muons in CMS. We expect these observables to have little effect on the performance in separating τ_h from quark and gluon jets. Our implementation of the DeepTau algorithm differs from the implementation in CMS in three further aspects: First, we compute distances between particles in the θ–ϕ plane, while CMS computes the distances in the η–ϕ plane, where the symbol $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$ denotes the pseudo-rapidity. Second, CMS builds three separate inner grids and three separate outer grids for particles of types e/γ, μ, and $h^\pm/h^0$ and for each type considers only the particle of highest p_T when building the vector of particle properties. We find that we get better performance if we use a single inner and a single outer grid for particles of any type and instead consider up to two particles (of highest p_T) per cell. Third, CMS does not remove the overlap between the inner and outer grids. We find that, besides avoiding redundancy of information, removing this overlap also improves the performance (by a small amount). We have adjusted the size of subsequent layers in the network according to these differences.

Table 3
High-level features used as input to the DeepTau algorithm.

Variable	Description
p_T^τ, θ_τ, ϕ_τ, M_τ	p_T, θ, ϕ, and mass of τ_h
charge	τ_h charge (equal to sum of $h^\pm$ charges)
I_{ch5}, $I_{\gamma 5}$, I_{nh5}	isolation of the τ_h with respect to charged particles, γ, and h^0, computed for an isolation cone of size $\Delta R = 0.5$
I_{ch3}, $I_{\gamma 3}$, I_{nh3}	isolation of the τ_h with respect to charged particles, γ, and h^0, computed for an isolation cone of size $\Delta R = 0.3$
N_{ch}, N_γ	multiplicity of τ_h constituents of type $h^\pm$ and γ
$\max(p_T^{ch})$	maximum p_T among τ_h constituent of type $h^\pm$
$\sum p_T^\gamma / p_T^\tau$	p_T-sum of τ_h constituents of type γ, divided by p_T of τ_h
$p_T^{\gamma,out}$	p_T-sum of γ, which are in strips, but outside signal cone
$\langle r^\gamma \rangle$, $\langle r_\theta^\gamma \rangle$, $\langle r_\phi^\gamma \rangle$	p_T-weighted distances in ΔR, θ, and ϕ between τ_h and γ
$\langle r^{\gamma,out} \rangle$	p_T-weighted distance in ΔR between τ_h and γ, which are in strips, but outside signal cone
N_e^i, N_μ^i, N_γ^i, N_{ch}^i, N_{nh}^i	number of particles of type e, μ, γ, $h^\pm$, h^0 in inner grid
N_e^o, N_μ^o, N_γ^o, N_{ch}^o, N_{nh}^o	number of particles of type e, μ, γ, $h^\pm$, h^0 in outer grid

The high-level features processed by third subnetwork are given in Table 3. We use 30 high-level features in total. The variable I_{ch3} (I_{ch5}) refers to the isolation of the τ_h with respect to charged particles (e, μ, $h^\pm$), which are within an "isolation cone" of size $\Delta R = 0.3$ (0.5) and were not used to build the τ_h object by the HPS algorithm:

$$I_{ch} = \sum_{charged} p_T \,. \tag{A.1}$$

The variables $I_{\gamma 3}$ and $I_{\gamma 5}$ (I_{nh3} and I_{nh5}) are computed similarly by summing the p_T of all particles of type γ (h^0) within the isolation cone. The variable $\langle r^\gamma \rangle$ is computed using Eq. (2), with the sum extending over all τ_h constituents of type γ and distance ΔR computed between the τ_h constituent and the direction of the τ_h. The variables $\langle r_\theta^\gamma \rangle$ and $\langle r_\phi^\gamma \rangle$ are computed analogously, but taking only differences in either θ or ϕ into account. The high-level features are processed by four fully-connected layers, with a size of 100 for the first three layers and a size of 50 for the fourth layer.

The output of the fully-connected layers that processed the high-level features is concatenated with the outputs of the two subnetworks for the inner and outer grids. The resulting vector of size 178 is passed through three fully-connected layers of size 100. The discriminant $\mathcal{D}_\tau$ of the network is computed by a layer of size 1. The softmax activation function [80] is used for this last layer, while all other fully-connected layers and the convolutional layers use the PReLU activation function [81].

The network is implemented in PyTorch [71]. It has $1.6 \cdot 10^6$ trainable parameters, which are trained in batches of 500 jets for a maximum of 200 epochs, using the AdamW [72] optimizer. A fixed learning rate of 10^{-4} is used throughout the training. For the loss function, we use the focal loss [74], with a value $\gamma = 2$ for the focusing parameter γ. The robustness of the training is increased by applying layer normalisation [82] to the inputs of each fully-connected and convolutional layer. The loss on the validation dataset is monitored throughout the training and the model with the minimum validation loss is retained for further study.

We show the distribution in the discriminant $\mathcal{D}_\tau$ computed by the DeepTau algorithm in Fig. 7, separately for the training and test datasets.

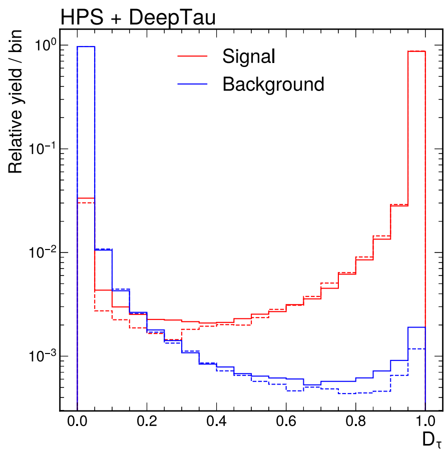


Fig. 7. Distribution in the discriminant $\mathcal{D}_\tau$ for the DeepTau algorithm. The solid curves refer to the test dataset and the dashed curves to the training dataset.

References

[1] A.J. Larkoski, I. Moult, B. Nachman, Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning, Phys. Rep. 841 (2020) 1, https://doi.org/10.1016/j.physrep.2019.11.001, arXiv:1709.04464.

[2] R. Kogler, et al., Jet substructure at the Large Hadron Collider: experimental review, Rev. Mod. Phys. 91 (4) (2019) 045003, https://doi.org/10.1103/RevModPhys.91.045003, arXiv:1803.06991.

[3] Particle Data Group, Review of particle physics, PTEP 2022 (2022) 083C01, https://doi.org/10.1093/ptep/ptac097.

[4] M. Dam, The τ challenges at FCC-ee, Eur. Phys. J. Plus 136 (9) (2021) 963, https://doi.org/10.1140/epjp/s13360-021-01894-y, arXiv:2107.12832.

[5] A. Pich, Challenges for tau physics at the TeraZ, Eur. Phys. J. Plus 136 (11) (2021) 1117, https://doi.org/10.1140/epjp/s13360-021-02077-5, arXiv:2012.07099.

[6] ATLAS Collaboration, Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, J. High Energy Phys. 01 (2018) 055, https://doi.org/10.1007/JHEP01(2018)055, arXiv:1709.07242.

[7] ATLAS Collaboration, Search for high-mass resonances decaying to $\tau\nu$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Phys. Rev. Lett. 120 (2018) 161802, https://doi.org/10.1103/PhysRevLett.120.161802, arXiv:1801.06992.

[8] A. Tumasyan, et al., Search for new physics in the τ lepton plus missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 09 (2023) 051, https://doi.org/10.1007/JHEP09(2023)051, arXiv:2212.12604.

[9] ATLAS Collaboration, Searches for third-generation scalar lepto-quarks in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, J. High Energy Phys. 06 (2019) 144, https://doi.org/10.1007/JHEP06(2019)144, arXiv:1902.08103.

[10] ATLAS Collaboration, Search for the direct production of charginos and neutralinos in final states with tau leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, Eur. Phys. J. C 78 (2018) 154, https://doi.org/10.1140/epjc/s10052-018-5583-9, arXiv:1708.07875.

[11] ATLAS Collaboration, Search for top squarks decaying to tau sleptons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Phys. Rev. D 98 (2018) 032008, https://doi.org/10.1103/PhysRevD.98.032008, arXiv:1803.10178.

[12] ATLAS Collaboration, Search for squarks and gluinos in final states with hadronically decaying τ leptons, jets, and missing transverse momentum using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Phys. Rev. D 99 (2019) 012009, https://doi.org/10.1103/PhysRevD.99.012009, arXiv:1808.06358.

[13] CMS Collaboration, Search for direct pair production of supersymmetric partners to the τ lepton in proton-proton collisions at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C 80 (2020) 189, https://doi.org/10.1140/epjc/s10052-020-7739-7, arXiv:1907.13179.

[14] CMS Collaboration, Search for supersymmetry with a compressed mass spectrum in events with a soft τ lepton, a highly energetic jet, and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. Lett. 124 (2020) 041803, https://doi.org/10.1103/PhysRevLett.124.041803, arXiv:1910.01185.

[15] ATLAS Collaboration, Search for direct stau production in events with two hadronic τ leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, Phys. Rev. D 101 (2020) 032009, https://doi.org/10.1103/PhysRevD.101.032009, arXiv:1911.06660.

[16] ATLAS Collaboration, Search for bottom squark pair production in pp collision events at $\sqrt{s} = 13$ TeV with hadronically decaying τ leptons, b jets and missing transverse momentum using the ATLAS detector, Phys. Rev. D 104 (2021) 032014, https://doi.org/10.1103/PhysRevD.104.032014, arXiv:2103.08189.

[17] CMS Collaboration, Search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 04 (2022) 147, https://doi.org/10.1007/JHEP04(2022)147, arXiv:2106.14246.

[18] CMS Collaboration, Search for top squark pair production in a final state with at least one hadronically decaying tau lepton in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 07 (2023) 110, https://doi.org/10.1007/JHEP07(2023)110, arXiv:2304.07174.

[19] ATLAS Collaboration, Search for charged Higgs bosons decaying via $H^\pm \to \tau^\pm \nu_\tau$ in the τ+jets and τ+lepton final states with 36 fb^{-1} of pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS experiment, J. High Energy Phys. 09 (2018) 139, https://doi.org/10.1007/JHEP09(2018)139, arXiv:1807.07915.

[20] CMS Collaboration, Search for charged Higgs bosons in the $H^\pm \to \tau^\pm \nu_\tau$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 07 (2019) 142, https://doi.org/10.1007/JHEP07(2019)142, arXiv:1903.04560.

[21] ATLAS Collaboration, Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. Lett. 125 (2020) 051801, https://doi.org/10.1103/PhysRevLett.125.051801, arXiv:2002.12223.

[22] CMS Collaboration, Searches for additional Higgs bosons and for vector lepto-quarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 07 (2023) 073, https://doi.org/10.1007/JHEP07(2023)073, arXiv:2208.02717.

[23] CMS Collaboration, Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau$bb final state at 13 TeV, J. High Energy Phys. 11 (2021) 057, https://doi.org/10.1007/JHEP11(2021)057, arXiv:2106.10361.

[24] M. Dam, Tau-lepton physics at the FCC-ee circular e^+e^- collider, SciPost Phys. Proc. 1 (2019) 041, https://doi.org/10.21468/SciPostPhysProc.1.041, arXiv:1811.09408.

[25] ATLAS Collaboration, A search for lepton-flavor-violating decays of the Z boson into a τ lepton and a light lepton with the ATLAS detector, Phys. Rev. D 98 (2018) 092010, https://doi.org/10.1103/PhysRevD.98.092010, arXiv:1804.09568.

[26] R. Harnik, J. Kopp, J. Zupan, Flavor violating Higgs decays, J. High Energy Phys. 03 (2013) 026, https://doi.org/10.1007/JHEP03(2013)026, arXiv:1209.1397.

[27] CMS Collaboration, Search for lepton flavour violating decays of the Higgs boson to $\mu\tau$ and $e\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 06 (2018) 001, https://doi.org/10.1007/JHEP06(2018)001, arXiv:1712.07173.

[28] ATLAS Collaboration, Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector, J. High Energy Phys. 04 (2015) 117, https://doi.org/10.1007/JHEP04(2015)117, arXiv:1501.04943.

[29] CMS Collaboration, Evidence for the direct decay of the 125 GeV Higgs boson to fermions, Nat. Phys. 10 (2014) 557, https://doi.org/10.1038/nphys3005, arXiv:1401.6527.

[30] ATLAS Collaboration, Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector, J. High Energy Phys. 07 (2023) 040, https://doi.org/10.1007/JHEP07(2023)040, arXiv:2209.10910.

[31] CMS Collaboration, Search for non-resonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Lett. B 842 (2023) 137531, https://doi.org/10.1016/j.physletb.2022.137531, arXiv:2206.09401.

[32] A. Elagin, P. Murat, A. Pranko, A. Safonov, A new mass reconstruction technique for resonances decaying to di-tau, Nucl. Instrum. Methods Phys. Res., Sect. A 654 (2011) 481, https://doi.org/10.1016/j.nima.2011.07.009, arXiv:1012.4686.

[33] L. Bianchini, B. Calpas, J. Conway, A. Fowlie, L. Marzola, C. Veelken, L. Perrini, Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of τ leptons using matrix element techniques, Nucl. Instrum. Methods Phys. Res., Sect. A 862 (2017) 54, https://doi.org/10.1016/j.nima.2017.05.001, arXiv:1603.05910.

[34] L. Linssen, A. Miyamoto, M. Stanitzki, H. Weerts, Physics and detectors at CLIC: CLIC conceptual design report, arXiv:1202.5940, https://doi.org/10.5170/CERN-2012-003.

[35] CLIC Collaboration, CLICdet: the post-CDR CLIC detector model, CLICdp note, https://cds.cern.ch/record/2254048, 2017.

[36] J.S. Marshall, M.A. Thomson, The Pandora software development kit for particle flow calorimetry, J. Phys. Conf. Ser. 396 (2012) 022034, https://doi.org/10.1088/1742-6596/396/2/022034.

[37] J.S. Marshall, A. Münnich, M.A. Thomson, Performance of particle flow calorimetry at CLIC, Nucl. Instrum. Methods Phys. Res., Sect. A 700 (2013) 153, https://doi.org/10.1016/j.nima.2012.10.038, arXiv:1209.4039.

[38] J.S. Marshall, M.A. Thomson, The Pandora software development kit for pattern recognition, Eur. Phys. J. C 75 (2015) 439, https://doi.org/10.1140/epjc/s10052-015-3659-3, arXiv:1506.05348.

[39] CMS Collaboration, CMS strategies for tau reconstruction and identification using particle-flow techniques, in: CMS Physics Analysis Summary, 2008, http://cds.cern.ch/record/1198228.

[40] ATLAS Collaboration, Tau reconstruction and identification performance in ATLAS, ATLAS note, http://cds.cern.ch/record/1298857, 2010.

[41] A. Muennich, TauFinder: a reconstruction algorithm for τ leptons at linear colliders, LCD note, http://cds.cern.ch/record/1443551, 2010.

[42] M. Dong, et al., CEPC conceptual design report: volume 2 – physics & detector, arXiv:1811.10545.

[43] S. Giagu, L. Torresi, M. Di Filippo, Tau lepton identification with graph neural networks at future electron–positron colliders, Front. Phys. 10 (2022) 909205, https://doi.org/10.3389/fphy.2022.909205.

[44] H. Abramowicz, et al., The international linear collider technical design report – volume 4: detectors, arXiv:1306.6329.

[45] T.H. Tran, V. Balagura, V. Boudry, J.-C. Brient, H. Videau, Reconstruction and classification of tau lepton decays with ILD, Eur. Phys. J. C 76 (8) (2016) 468, https://doi.org/10.1140/epjc/s10052-016-4315-2, arXiv:1510.05224.

[46] B. Xu, Detectors and physics at a future linear collider, Ph.D. thesis, Cambridge U., 12 2017.

[47] S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, W. Du, Z.-M. Ma, T.-Y. Liu, An efficient Lorentz equivariant graph neural network for jet tagging, J. High Energy Phys. 07 (2022) 030, https://doi.org/10.1007/JHEP07(2022)030, arXiv:2201.08187.

[48] H. Qu, C. Li, S. Qian, Particle transformer for jet tagging, in: International Conference on Machine Learning, PMLR, 2022, p. 18281, arXiv:2202.03772.

[49] CMS Collaboration, Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS, J. Instrum. 11 (2016) P01019, https://doi.org/10.1088/1748-0221/11/01/P01019, arXiv:1510.07488.

[50] CMS Collaboration, Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV, J. Instrum. 13 (2018) P10005, https://doi.org/10.1088/1748-0221/13/10/P10005, arXiv:1809.02816.

[51] CMS Collaboration, Identification of hadronic tau lepton decays using a deep neural network, J. Instrum. 17 (2022) P07023, https://doi.org/10.1088/1748-0221/17/07/P07023, arXiv:2201.08458.

[52] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, p. 1492, arXiv:1611.05431.

[53] P.T. Komiske, E.M. Metodiev, J. Thaler, Energy flow networks: deep sets for particle jets, J. High Energy Phys. 01 (2019) 121, https://doi.org/10.1007/JHEP01(2019)121, arXiv:1810.05165.

[54] H. Qu, L. Gouskos, ParticleNet: jet tagging via particle clouds, Phys. Rev. D 101 (2020) 056019, https://doi.org/10.1103/PhysRevD.101.056019, arXiv:1902.08570.

[55] FCC-ee FCC Collaboration, The lepton collider: future circular collider conceptual design report volume 2, Eur. Phys. J. Spec. Top. 228 (2019) 261, https://doi.org/10.1140/epjst/e2019-900045-4.

[56] H. Baer, T. Barklow, K. Fujii, Y. Gao, A. Hoang, S. Kanemura, J. List, H.E. Logan, A. Nomerotski, M. Perelstein, et al., The international linear collider technical design report – volume 2: physics, arXiv:1306.6352.

[57] C. Bierlich, et al., A comprehensive guide to the physics and usage of PYTHIA 8.3, SciPost Physics Codebases, arXiv:2203.11601, https://doi.org/10.21468/SciPostPhysCodeb.8.

[58] Y. Amhis, M. Hartmann, C. Helsens, D. Hill, O. Sumensari, Prospects for $B_c^+ \to \tau^+\nu_\tau$ at FCC-ee, J. High Energy Phys. 12 (2021) 133, https://doi.org/10.1007/JHEP12(2021)133, arXiv:2105.13330.

[59] GEANT4 Collaboration, Geant4: a simulation toolkit, Nucl. Instrum. Methods Phys. Res., Sect. A 506 (2003) 250, https://doi.org/10.1016/S0168-9002(03)01368-8.

[60] F. Gaede, Marlin and LCCD: software tools for the ILC, Nucl. Instrum. Methods Phys. Res., Sect. A 559 (2006) 177, https://doi.org/10.1016/j.nima.2005.11.138.

[61] G. Ganis, C. Helsens, V. Völkl, Key4hep, a framework for future HEP experiments and its use in FCC, Eur. Phys. J. Plus 137 (2022) 149, https://doi.org/10.1140/epjp/s13360-021-02213-1, arXiv:2111.09874.

[62] H. Abramowicz, et al., Higgs physics at the CLIC electron–positron linear collider, Eur. Phys. J. C 77 (2017) 475, https://doi.org/10.1140/epjc/s10052-017-4968-5, arXiv:1608.07538.

[63] N. Bacchetta, et al., CLD – a detector concept for the FCC-ee, arXiv:1911.12230.

[64] J.D. Jackson, Classical Electrodynamics, John Wiley & Sons, 1999.

[65] T. Lange, S. Nandan, J. Pata, L. Tani, C. Veelken, HEP-KBFI/ml-tau-reco: submission, https://doi.org/10.5281/zenodo.8113344, Jul 2023.

[66] M. Boronat, J. Fuster, I. Garcia, P. Roloff, R. Simoniello, M. Vos, Jet reconstruction at high-energy electron–positron colliders, Eur. Phys. J. C 78 (2018) 144, https://doi.org/10.1140/epjc/s10052-018-5594-6, arXiv:1607.05039.

[67] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv:1409.0473.

[68] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, advances in neural information processing systems 30, arXiv:1706.03762.

[69] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825.

[70] S. Villar, D.W. Hogg, K. Storey-Fisher, W. Yao, B. Blum-Smith, Scalars are universal: equivariant machine learning, structured like classical physics, in: Advances in Neural Information Processing Systems, 2023, arXiv:2106.06610.

[71] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: an imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, 2019, arXiv:1912.01703.

[72] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv:1711.05101.

[73] L.N. Smith, N. Topin, Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates, Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, vol. 11006, SPIE, 2019, p. 369, arXiv:1708.07120.

[74] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, p. 2980, arXiv:1708.02002.

[75] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, arXiv:1810.04805.

[76] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding by generative pre-training, Tech. Rep., OpenAI, 2012, https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

[77] CMS Collaboration, Luminosity determination using Z boson production at the CMS experiment, in: CMS Physics Analysis Summary, 2021, http://cds.cern.ch/record/2852905.

[78] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, M. Selvaggi, DELPHES 3, a modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057, https://doi.org/10.1007/JHEP02(2014)057, arXiv:1307.6346.

[79] S. Atag, A.A. Billur, Possibility of determining τ lepton electromagnetic moments in $\gamma\gamma \to \tau^+\tau^-$ process at the CERN-LHC, J. High Energy Phys. 11 (2010) 060, https://doi.org/10.1007/JHEP11(2010)060, arXiv:1005.2841.

[80] J.S. Bridle, Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition, in: F.F. Soulié, J. Hérault (Eds.), Neurocomputing, Springer Berlin Heidelberg, Berlin, Heidelberg, 1990, p. 227.

[81] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, arXiv:1502.01852.

[82] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, arXiv:1607.06450.

Curriculum Vitae

1. Personal data

Name	Laurits Tani
Date and place of birth	1 May 1994 Tallinn, Estonia
Nationality	Estonian

2. Contact information

Address National Institute of Chemical Physics and Biophysics,
High Energy and Computational Physics,
Rävala pst 10, 10143 Tallinn, Estonia
E-mail laurits.tani@cern.ch

3. Education

2019–2024	Tallinn University of Technology, Applied Physics, PhD studies
2017–2019	ETH Zürich Physics, MSc
2014–2017	Tallinn University of Technology, Engineering Physics, BSc

4. Language competence

Estonian	native
English	fluent
German	fluent
Russian	elementary proficiency

5. Professional employment

2020–. . .	National Institute of Chemical Physics and Biophysics, Junior Researcher

6. Workshops and Schools

2022	The International Doctorate Network in Particle Physics, Astrophysics and Cosmology (IDPASC)
2021	Baltic School of High-Energy Physics and Accelerator Technologies 2021
2021	Seventh Machine Learning in High Energy Physics Summer School 2021 (MLHEP)
2021	PyHEP 2021
2020	CMS Data Analysis Schoold (CMSDAS)
2019	Combine Workshop and Tutorial
2019	Trans-European School of High Energy Physics
2015	CERN Summer School

7. Teaching

2021–2022	Tallinn University of Technology Exercise classes in *Introduction to particle physics (YFX1130)*
2016–2017	Tallinn University of Technology Math and physics courses
2016	Mektory Physics crash course for high school students

8. Computer skills

- Operating systems: GNU/Linux
- Document preparation: LaTeX
- Programming languages: Python, C++, Bash
- Scientific packages: Wolfram Mathematica

9. Defended theses

- 2019, *"Monitoring the optical quality of the FACT Cherenkov telescope"*, MSc, supervisor Prof. Dr. Adrian Biland, ETH Zürich
- 2017, *"The effect of light on the capacitive properties of the solar cell"*, supervisor Dr. Raavo Josepson, Tallinn University of Technology, Institute of Physics

10. Field of research

- High Energy Physics
- Machine learning

11. Scientific work

Papers

1. L. Tani *et al.*, "Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics", *Eur. Phys. J. C*, vol. 81, no. 2, p. 170, 2021. DOI: 10.1140/epjc/s10052-021-08950-y. arXiv: 2011.04434 [hep-ex]
2. CMS collaboration, "Search for Higgs boson pairs decaying to WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s}$ = 13 TeV", *JHEP*, vol. 07, p. 095, 2023. DOI: 10.1007/JHEP07(2023)095. arXiv: 2206.10268 [hep-ex]
3. L. Tani and C. Veelken, "Comparison of Bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning applications in high energy physics", *Comput. Phys. Commun.*, vol. 294, p. 108 955, 2024. DOI: 10.1016/j.cpc.2023.108955. arXiv: 2201.06809 [physics.data-an]
4. T. Lange *et al.*, "Tau lepton identification and reconstruction: A new frontier for jet-tagging ML algorithms", *Comput. Phys. Commun.*, vol. 298, p. 109 095, 2024. DOI: 10.1016/j.cpc.2024.109095. arXiv: 2307.07747 [hep-ex]

Conference presentations

1. **L. Tani**. 'Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics', The 2nd CERN Baltic Conference (CBC 2022): 10–12 October 2022, Vilnius

2. **L. Tani**. 'Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics', The 21st International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2022): 23–28 October 2022, Villa Romanazzi Carducci in Bari, Italy.

3. **L. Tani**. 'FACT-Measuring the Evolution of the Optical Point Spread Function using Muon-Rings', DPG-Frühjahrstagung 2019: 25–29 October 2019, Aachen, Germany

Elulookirjeldus

1. Isikuandmed

Nimi	Laurits Tani
Sünniaeg ja -koht	01.05.1994, Tallinn, Eesti
Kodakondsus	Eesti

2. Kontaktandmed

Aadress	Keemilise ja Bioloogilise Füüsika Instituut, Kõrge Energia ja Arvutusfüüsika Laboratoorium, Rävala pst 10, 10143 Tallinn, Estonia
E-mail	laurits.tani@cern.ch

3. Haridus

2019–2024	Tallinna Tehnikaülikool, Rakendusfüüsika, PhD
2017–2019	ETH Zürich, Füüsika, MSc
2014–2017	Tallinna Tehnikaülikool, Tehniline füüsika, BSc

4. Keelteoskus

eesti keel	emakeel
inglise keel	kõrgtase
saksa keel	kõrgtase
vene keel	algtase

5. Teenistuskäik

2020– . . .	Keemilise ja Bioloogilise Füüsika Instituut, nooremteadur

6. Koolitused ja töötoad

2022	The International Doctorate Network in Particle Physics, Astrophysics and Cosmology (IDPASC)
2021	Baltic School of High-Energy Physics and Accelerator Technologies 2021
2021	Seventh Machine Learning in High Energy Physics Summer School 2021 (MLHEP)
2021	PyHEP 2021
2020	CMS Data Analysis Schoold (CMSDAS)
2019	Combine Workshop and Tutorial
2019	Trans-European School of High Energy Physics
2015	CERNi suvekool

7. Õpetamine

2021–2022	Tallinna Tehnikaülikool Harjutustunnid aines *Introduction to particle physics (YFX1130)*
2016–2017	Tallinna Tehnikaülikool Füüsika ja matemaatika ettevalmistuskursused
2016	Mektory Füüsika kiirkursus gümnasistidele

8. Arvutioskus

- Operatsioonisüsteemid: GNU/Linux
- Kontoritarkvara: LaTeX
- Programmeerimiskeeled: Python, C++, Bash
- Teadustarkvara paketid: Wolfram Mathematica

9. Kaitstud lõputööd

- 2019, *"FACT Cherenkov teleskoobi optilise kvaliteedi monitoorimine"*, MSc, juhendaja Prof. Dr. Adrian Biland, ETH Zürich
- 2017, *"Valguse mõju päikesepatarei mahtuvuslikele omadustele"*, juhendaja Dr. Raavo Josepson, Tallinn University of Technology, Institute of Physics

10. Teadustöö põhisuunad

- Kõrge energia füüsika
- Masinõpe

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-141-3 (PDF)