Available on the CERN CDS information server **CMS PAS HIG-20-009**

CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2021/03/17

Search for lepton-flavor violating decays of the Higgs boson to $\mu\tau$ and $e\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A search is presented for lepton-flavor violating decays of the Higgs boson to $\mu\tau$ and $e\tau$. The analysis is based on data corresponding to an integrated luminosity of 137 fb^{-1}, collected at the CMS detector of the LHC in proton-proton collisions at a center-of-mass energy of 13 TeV. No significant excess has been found, and the results are interpreted in terms of upper limits on lepton-flavor violating branching fractions of the Higgs boson. The observed and expected upper limits on the branching fractions are, respectively, $\mathcal{B}(\mathrm{H} \to \mu\tau) < 0.15\%$ and $< 0.15\%$ and $\mathcal{B}(\mathrm{H} \to e\tau) < 0.22\%$ and $< 0.16\%$ set at 95% confidence level.

1 Introduction

One of the main goals of the LHC program is to search for processes beyond the standard model (BSM). The lepton-flavor violating (LFV) decays of the Higgs boson [1–3] can provide possible signatures of such processes. The properties and decays of the Higgs boson are thus far consistent with expectations of the standard model (SM) [4–9]. However, there is considerable motivation to search for BSM decays of the Higgs boson. A previous analysis of the combined results from the CMS experiment constrained the branching fraction for these decays $\mathcal{B}(\mathrm{H} \to \mathrm{BSM})$ to < 0.36 at the 95% confidence level (CL), while still leaving the possibility for a large contribution for BSM decays [10].

The LFV decays of the Higgs boson [11], such as $\mathrm{H} \to \mathrm{e}\mu$, $\mathrm{H} \to \mathrm{e}\tau$, or $\mathrm{H} \to \mu\tau$, are forbidden in the SM, but take place through the LFV Yukawa couplings $Y_{\mathrm{e}\mu}$, $Y_{\mathrm{e}\tau}$, or $Y_{\mu\tau}$, respectively. The LFV decays arise in models with more than one Higgs boson doublet [12], certain supersymmetric models [13–15], composite Higgs models [16, 17], models with flavor symmetries [18], the Randall–Sundrum model of extra spatial dimensions [19–23], and other models [24–29].

Here we report a search for LFV decays of the Higgs boson in the $\mu\tau$ and $\mathrm{e}\tau$ channels performed using data collected by the CMS experiment in proton-proton (pp) collisions at a center-of-mass energy of 13 TeV during the 2016-2018 data taking period, corresponding to an integrated luminosity of $137\,\mathrm{fb}^{-1}$. The CMS experiment set upper limits of 0.25% and 0.61% [30] and the ATLAS experiment set upper limits of 0.28% and 0.47% [31] on $\mathcal{B}(\mathrm{H} \to \mu\tau)$ and $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau)$ at 95% CL, respectively, based on the 2016 data set, corresponding to an integrated luminosity of $36\,\mathrm{fb}^{-1}$.

The presence of an LFV Higgs boson coupling leads to processes such as $\mu \to \mathrm{e}$, $\tau \to \mu$, and $\tau \to \mathrm{e}$ to proceed via a virtual Higgs boson [32, 33]. The experimental limits on these decays yield indirect constraints on $\mathcal{B}(\mathrm{H} \to \mathrm{e}\mu)$, $\mathcal{B}(\mathrm{H} \to \mu\tau)$, and $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau)$ [11, 34]. The null result for $\mu \to \mathrm{e}\gamma$ [35] strongly constrains $\mathcal{B}(\mathrm{H} \to \mathrm{e}\mu)$ to $< 10^{-8}$. Searches for rare τ lepton decays [36], such as $\tau \to \mathrm{e}\gamma$ and $\tau \to \mu\gamma$, and the measurement of the electron and muon magnetic moments, have set constraints on $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau)$ and $\mathcal{B}(\mathrm{H} \to \mu\tau)$ of $\approx$10%, which are much less stringent than those from the direct searches.

Our search is performed in the $\mu\tau_{\mathrm{h}}$, $\mu\tau_{\mathrm{e}}$, $\mathrm{e}\tau_{\mathrm{h}}$, and $\mathrm{e}\tau_{\mu}$ channels, where τ_{h}, τ_{e}, and τ_{μ} correspond to the $\tau \to$ hadrons, electron, and muon decay channels of τ leptons, respectively, with each accompanied by corresponding neutrinos. The $\mathrm{e}\tau_{\mathrm{e}}$ and $\mu\tau_{\mu}$ decays are not considered because of the large background contribution from Z boson decays. Our chosen final states are similar to the final states to measure $\mathrm{H} \to \tau\tau$ with some differences in that the muon and electron in the LFV $\mathrm{H} \to \mu\tau$ and $\mathrm{H} \to \mathrm{e}\tau$ decays are produced promptly and tend to have larger momenta than in the SM $\mathrm{H} \to \tau_{\mu}\tau_{\mathrm{h}}$ and $\mathrm{H} \to \tau_{\mathrm{e}}\tau_{\mathrm{h}}$ decays.

Our search significantly improves the sensitivity relative to such previous studies [30, 31, 37]. The analysis makes use of boosted decision tree (BDT) discriminants to distinguish signal from background. Constraints on the branching fractions are extracted under the assumption that only one of the LFV decays contributes additionally to the SM Higgs boson total width, as well as limits on the $Y_{\mathrm{e}\tau}$ and $Y_{\mu\tau}$ LFV Yukawa couplings. An additional analysis based on the collinear mass distribution is also performed to cross-check the results.

This note is organized as follows: a description of the CMS detector is given in Section 2, collision data and simulated events are discussed in Section 3, event reconstruction is described in Section 4, and event selection is described separately for the four decay channels in Section 5. Background estimation and systematic uncertainties are described in Sections 6 and 7, respectively. Results are presented in Section 8 and the note is summarized in Section 9.

2 The CMS detector

The CMS detector consists of a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), a brass and scintillator hadron calorimeter (HCAL), and a muon system composed of gaseous detectors. Each subdetector consists of a barrel and two endcap sections. The central feature of the CMS detector is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. The tracking systems and the calorimeters are contained within the solenoid volume; the muon chambers are embedded in the steel flux-return yoke outside the solenoid. Forward calorimeters extend the pseudorapidity (η) coverage provided by the barrel and endcap detectors.

Events of interest are selected using a two-tiered trigger system. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of $\approx$100 kHz within a fixed latency of $\approx 4\,\mu$s [38]. The second level, the high-level trigger, consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing that reduces the event rate to $\approx$1 kHz before data storage [39]. A more detailed description of the CMS detector, together with a definition of the coordinate system and kinematic variables, can be found in Ref. [40].

3 Collision data and simulated events

The search presented makes use of pp collisions collected at the CMS experiment at a center-of-mass energy of 13 TeV in 2016–2018. The total integrated luminosity amounted to $35.9\,\mathrm{fb}^{-1}$ in 2016, $41.5\,\mathrm{fb}^{-1}$ in 2017, and $59.7\,\mathrm{fb}^{-1}$ in 2018. Single-muon triggers with their isolation criteria are used to collect the data in the $\mu\tau_\mathrm{h}$ channel. Electron-muon triggers are used to collect data in the $\mu\tau_\mathrm{e}$ and $\mathrm{e}\tau_\mu$ channels. Triggers requiring a single isolated electron, or a combination of an electron and τ_h, are used in the $\mathrm{e}\tau_\mathrm{h}$ channel.

Simulated events are used to model signal and background contributions to all the analysis regions using several event generators. In all cases parton showering, hadronization, and underlying event properties are modeled using PYTHIA [41] version 8.212. The PYTHIA parameters affecting the description of the underlying event are set to the CUETP8M1 tune in 2016 [42], except for the $\mathrm{t\bar{t}}$ events that use the CP5 tune. The CP5 tune is used for all the events in 2017 and 2018 [43]. The NNPDF3.0 parton distribution functions (PDFs) are used for all 2016 events and the NNPDF3.1 PDFs are used for the 2017 and 2018 events [44].

This simulation of interactions in the CMS detector is based on GEANT4 [45], using the same reconstruction algorithms as used for data. The Higgs bosons are generated in pp collisions predominantly through gluon fusion (ggH) [46], but also via vector boson fusion (VBF) [47], and in association with a vector boson (W or Z) [48]. Such events are generated at next-to-leading order (NLO) in perturbative quantum chromodynamics (QCD) with the POWHEG v2.0 generator [49–54], using the implementation of Refs. [55, 56]. For the signal, we consider just the Higgs bosons via the ggH and VBF mechanisms as the contribution from associated vector boson production is found to be negligible.

Embedded events are employed for the data-driven estimation of the $\mathrm{Z} \to \tau\tau$ background. These events are obtained from data with well identified $\mathrm{Z} \to \mu\mu$ decays from which muons are removed, and simulated τ leptons are embedded with the same kinematic variables as the replaced muons. The MADGRAPH5_aMC@NLO [57] (version 2.2.2 in 2016, version 2.4.2 in 2017 and 2018) generator is used to simulate the $\mathrm{Z} \to \mathrm{ee}$+jets and $\mathrm{Z} \to \mu\mu$+jets processes, the W+jets background process, and the electroweak (EW) W/Z events. They are simulated at leading order with the MLM jet matching and merging schemes [58].

Diboson production is simulated at NLO using the MADGRAPH5_aMC@NLO generator with the FxFx jet-matching and merging scheme [59]. Top quark-antiquark pair and single top quark production are generated at NLO using POWHEG.

The effect of pileup, where events have multiple pp interactions per bunch crossing, is taken into account in simulated events by generating concurrent minimum bias events. All simulated events are weighted to match the pileup distribution observed in the data.

4 Event reconstruction

The particle flow (PF) algorithm [60] reconstructs and identifies each particle in an event through an optimized combination of information from the various subdetectors of the CMS detector. In this process, identifying the PF candidate type (photons, electrons, muons, charged, and neutral hadrons) plays an important role in determining particle direction and energy. The candidate vertex with the largest value of summed physics object p_{T}^2, where p_{T} is the transverse momentum, is taken to be the primary pp interaction vertex (PV). The physics objects are returned by a jet finding algorithm [61, 62] applied to all charged tracks associated with the vertex, plus the corresponding associated missing transverse momentum.

An electron is identified as a track from the PV combined with one or more ECAL energy clusters. These clusters correspond to the electron and possible bremsstrahlung photons emitted when passing through the tracker. Electrons are accepted in the range $|\eta| < 2.5$, except for the region $1.44 < |\eta| < 1.57$ where the detector's service infrastructure is located. They are identified using a multivariate discriminator that combines observables sensitive to the amount of bremsstrahlung energy deposited along the electron trajectory, the geometric and momentum matching between the electron trajectory and associated clusters, and the distribution in shower energy in the calorimeters, with an efficiency of 80% [63]. Electrons from photon conversions are removed. The electron momentum is estimated by combining the energy measurement in the ECAL with the momentum measurement in the tracker. The momentum resolution for electrons with $p_{\mathrm{T}} \approx 45\,\mathrm{GeV}$ from $\mathrm{Z} \to \mathrm{ee}$ decays ranges from 1.7 to 4.5% depending on the $|\eta|$. It is generally better in the barrel region than in the endcaps [64].

Muons are measured in the $|\eta| < 2.4$ range using the drift tube, cathode strip chamber, and resistive plate chamber technologies. The efficiency to reconstruct and identify muons is greater than 96%. Matching muons to tracks measured in the silicon tracker results in a relative p_{T} resolution for muons with p_{T} up to 100 GeV of 1% in the barrel and 3% in the endcaps [65].

The reconstruction of τ_{h} is performed using the hadrons-plus-strips algorithm, which combines the signature for charged hadrons composed of tracks left in the tracker and energy depositions in the HCAL with the signature for electrons or photons from neutral pion decays that are reconstructed as electromagnetic "strips" in η–ϕ space [66], where ϕ is the azimuth in radians. The combination of these signatures provides the four-vector for the parent τ_{h}. Based on the overall neutral versus charged contents of the τ_{h} reconstruction, a decay mode is assigned as $\mathrm{h}^{\pm}$, $\mathrm{h}^{\pm}\pi^0$, $\mathrm{h}^{\pm}\mathrm{h}^{\mp}\mathrm{h}^{\pm}$, or $\mathrm{h}^{\pm}\mathrm{h}^{\mp}\mathrm{h}^{\pm}\pi^0$, where $\mathrm{h}^{\pm}$ denotes a charged hadron.

The τ_{h} reconstructed using the hadrons-plus-strips algorithm must be well identified to reject jets, muons, and electrons misidentified as τ_{h}. A deep neural network (DNN) discriminator is used for τ_{h} identification [67]. The input variables to the DNN include τ_{h} lifetime, isolation, and information of PF candidates reconstructed within the τ lepton signal or isolation cones. A p_{T} dependent threshold on the output of the DNN is used to distinguish τ_{h} from jets. The chosen working point (WP) has a τ_h identification efficiency of 70% with a misidentification probability of 1%. The DNN can reject electrons and muons misidentified as τ_h using dedicated criteria based on the consistency between the tracker, calorimeter, and muon detector measurements. In the $\mu\tau_\mathrm{h}$ and $\mathrm{e}\tau_\mathrm{h}$ channel, we use a WP that has an efficiency of 97.5% and 87.5% with a misidentification probability of 1–2% and 0.2–0.3% to discriminate τ_h against electrons, and we use a WP that has an efficiency of 99.6% and 99.8% with a misidentification probability of 0.04% and 0.06% to discriminate τ_h against muons, respectively.

Charged hadrons are defined as PF tracks from the PV not reconstructed as electrons, muons, or τ_h leptons. Neutral hadrons are identified as HCAL energy clusters not assigned to any charged hadron or as excesses in ECAL or HCAL energies relative to the small charged-hadron energy deposit. All the PF hadron candidates are clustered into jets using the infrared- and collinear-safe anti-k_T algorithm [61] with a distance parameter of 0.4. Jet momentum is determined as the vectorial sum of all particle momenta in the jet. It is found from simulation to be, on average, within 5 to 10% of the true momentum over the entire p_T spectrum and detector acceptance [68]. The reconstructed jets must have a $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 4.7$. Data collected in the high $|\eta|$ region of the ECAL endcaps were affected by noise during the 2017 data taking. This is mitigated by discarding events containing jets with $p_\mathrm{T} < 50\,\mathrm{GeV}$ and $2.65 < |\eta| < 3.14$ in the 2017 data. Jets that contain b quarks are tagged using a DNN-based DEEPCSV algorithm, using a WP with efficiency of 70% for a misidentification probability for light-flavor jets of 1% [69].

The interactions from pileup add more tracks and calorimetric energy depositions, thereby increasing the apparent jet momenta. To mitigate this effect, tracks identified as originating from pileup vertices are discarded, and an offset correction is applied to correct the remaining contributions [70]. Jet energy corrections are obtained from simulation studies so that the average measured energy of jets matches that of particle level jets. In-situ measurements of the momentum balance in photon+jet, Z+jet, and multijet events are used to determine any residual differences between the jet energy scale in data and simulation, and appropriate corrections are applied [71]. Additional selection criteria are applied to each jet to remove jets potentially dominated by instrumental effects or reconstruction failures. When combining information from the entire detector, the jet energy resolution typically amounts to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV. The variable $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is used to measure the separation between reconstructed objects in the detector. Any jet within $\Delta R = 0.5$ of identified leptons is removed.

The missing transverse momentum vector $\vec{p}_\mathrm{T}^\text{miss}$ is computed as the negative of the vector sum of the p_T of all the PF candidates in an event, and its magnitude is denoted as p_T^miss [72]. The $\vec{p}_\mathrm{T}^\text{miss}$ is modified to account for corrections to the reconstructed jets' energy scale in the event. Anomalous high-p_T^miss events can originate from various reconstruction failures, detector malfunctions, or backgrounds not from beam-beam sources. Such events are rejected using event filters designed to identify more than 85–90% of the spurious high-p_T^miss events with a mistag rate of less than 0.1% [72]. In addition to the event-filtering algorithms, we require the jets to have a neutral hadron energy fraction smaller than 0.9, which rejects more than 99% of jets due to detector noise, independent of jet p_T, with a negligible mistag rate. Corrections applied to the $\vec{p}_\mathrm{T}^\text{miss}$ reduce the mismodeling of $\vec{p}_\mathrm{T}^\text{miss}$ in simulated Z, W, and Higgs boson events. The corrections are applied to simulated events based on the vectorial difference in the measured $\vec{p}_\mathrm{T}^\text{miss}$ and total p_T of neutrinos originating from the decay of the Z, W, or Higgs bosons. Their average effect is the reduction of the magnitude of the $\vec{p}_\mathrm{T}^\text{miss}$ obtained from the simulation by a few GeV.

The muon or electron isolations are measured relative to its p_{T}^{ℓ}, where ℓ is either μ or e, values by summing over the scalar p_{T} of PF particles in a cone of $\Delta R = 0.4$ or 0.3 around the lepton:

$$I_{\mathrm{rel}}^{\ell} = \left(\sum p_{\mathrm{T}}^{\text{PV charged}} + \mathrm{MAX}\left[0, \sum p_{\mathrm{T}}^{\text{neutral}} + \sum p_{\mathrm{T}}^{\gamma} - p_{\mathrm{T}}^{\mathrm{PU}}(\ell)\right]\right) \Big/ \, p_{\mathrm{T}}^{\ell},$$

where $p_{\mathrm{T}}^{\text{PV charged}}$, $p_{\mathrm{T}}^{\text{neutral}}$, and p_{T}^{γ} indicate the p_{T} of a charged hadron, a neutral hadron, and a photon within the cone, respectively. The neutral particle contribution to isolation from pileup, $p_{\mathrm{T}}^{\mathrm{PU}}(\ell)$, is estimated from the area of jet and its median energy density in the event [73] for the electron. For the muon, half of the sum of p_{T} of charged hadrons within the isolation cone, not originating from the PV, is used instead. The charged-particle contribution to isolation from the pileup is rejected by requiring the tracks to originate from the PV.

5 Event selection

The signal topology consists of an isolated muon or electron and an oppositely charged isolated τ lepton (τ_{μ}, τ_{e}, or τ_{h}). Jets misidentified as electrons or muons are suppressed by imposing isolation requirements. The events in the $\mu\tau$ and $\mathrm{e}\tau$ channels are further divided into leptonic and hadronic channels based on the τ lepton decay mode. A set of loose selection criteria, known as the 'preselection', is first defined in each channel's respective signature. Events with more than two jets are not considered in the analysis. Each channel's events are then divided into categories based on the number of jets in the event (0-, 1-, or 2-jet) to enhance different Higgs boson production mechanisms. The dominant production mechanism contributing to the signal yield in the 0-jet category is ggH, while in the 1-jet category, it is ggH with initial-state radiation. The 2-jet category is further split into two based on the invariant mass of the two jets (m_{jj}). The optimization resulted in a threshold of 550 GeV and 500 GeV on m_{jj} for the $\mu\tau$ and $\mathrm{e}\tau$ channels, respectively, to give the best sensitivity. The dominant production mechanism is ggH for events with $m_{\mathrm{jj}} < 550\,\mathrm{GeV}$ and $< 500\,\mathrm{GeV}$, while it is VBF for events with $m_{\mathrm{jj}} > 550\,\mathrm{GeV}$ and $> 500\,\mathrm{GeV}$ for the $\mu\tau$ and $\mathrm{e}\tau$ channels, respectively.

The collinear mass m_{col} variable provides an estimate of m_{H} using the observed decay products of the Higgs boson. It is reconstructed using the "collinear approximation" based on the observation that, since $m_{\mathrm{H}} \gg m_{\tau}$, the τ lepton decay products are produced in a relativistic manner in the direction of the τ lepton [74]. The momentum of neutrino(s) from the τ lepton decay can be approximated to have the same direction as the visible decay products of the τ lepton ($\vec{\tau}^{\mathrm{vis}}$). The component of the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the direction of the $\vec{\tau}^{\mathrm{vis}}$ is used to estimate the transverse component of the neutrino momentum ($p_{\mathrm{T}}^{\vec{\nu},\mathrm{est}}$). The collinear mass can then be extracted from the visible mass of the τ-μ or τ-e system (m_{vis}) as $m_{\mathrm{col}} = m_{\mathrm{vis}} / \sqrt{x_{\tau}^{\mathrm{vis}}}$, where x_{τ}^{vis} is the fraction of the τ lepton p_{T} carried by the $\vec{\tau}^{\mathrm{vis}}$ ($x_{\tau}^{\mathrm{vis}} = p_{\mathrm{T}}^{\vec{\tau}^{\mathrm{vis}}} / (p_{\mathrm{T}}^{\vec{\tau}^{\mathrm{vis}}} + p_{\mathrm{T}}^{\vec{\nu},\mathrm{est}})$).

The transverse mass $m_{\mathrm{T}}(\ell)$ is a variable constructed from the lepton p_{T} and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ vectors: $m_{\mathrm{T}}(\ell) = \sqrt{2|\vec{p}_{\mathrm{T}}^{\ell}||\vec{p}_{\mathrm{T}}^{\mathrm{miss}}|(1 - \cos\Delta\phi_{\ell,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}})}$, where $\Delta\phi_{\ell,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}}$ is the angle in the transverse plane between the lepton and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, used to discriminate the Higgs boson signal from the W+jets background. The $m_{\mathrm{T}}(\ell)$ distribution for the signal defined using visible decay products of the τ lepton peaks at lower values, while it peaks at higher values for the W+jets background.

To improve discrimination between signal and background events, a BDT is trained, using the TMVA toolkit of the ROOT analysis package [75]. A BDT is trained in each channel using a mixture of simulated signal events comprising the ggH and VBF processes, weighted according to their expected yield from SM production cross sections. In hadronic channels, the dominant sources of background come from the $\mathrm{Z} \to \tau\tau$ process and events with misidentified leptons. The background used for training a BDT in the hadronic channels is obtained from data containing misidentified lepton events of the same electric charge for both the leptons and $\mathrm{Z} \to \ell\ell$ simulated events with their applied signal selections. In leptonic channels, the dominant sources of background come from the $\mathrm{Z} \to \tau\tau$ process, the $\mathrm{t\bar{t}}$ process, and events with misidentified leptons. The background used for training a BDT in the leptonic channels is obtained from $\mathrm{t\bar{t}}$ and $\mathrm{Z} \to \ell\ell$ simulated events mixed and weighted according to their expected yield from SM production cross sections. Additional background for training comes from events with misidentified leptons in a control region (CR) in data, where the isolation requirements are inverted with the same electric charge for both the leptons. A detailed description of the different background processes and their estimation is given in Section 6.

The input variables to the BDT are mentioned separately for each channel below. The input variables are chosen based on their separation power as observed during training the BDT. The trained BDT is validated in a dedicated background enriched CR for each channel. In all the channels, events containing additional electrons, muons, or τ_h candidates are vetoed. Also, events with at least one b-tagged jet are rejected to suppress the $\mathrm{t\bar{t}}$ background. The distributions in all the channels are shown after determining the best fit values of the uninteresting parameters from the maximum likelihood fit to the signal-plus-background hypothesis, as discussed later in Section 7.

5.1 $H \to \mu\tau_h$

In this channel, the preselection requires a muon and τ_h of opposite electric charge with a separation of $\Delta R > 0.5$. The trigger is based on the presence of an isolated muon with a p_T threshold of 24 GeV. This trigger is not fully efficient in 2017 and is used in conjunction with the trigger based on the presence of an isolated muon with a p_T threshold of 27 GeV. The muon is required to have $p_\mathrm{T} > 26\,\mathrm{GeV}$, $|\eta| < 2.1$, and $I^{\mu}_{\mathrm{rel}} < 0.15$. The τ_h is required to have $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.3$. The selections for the $\mu\tau_\mathrm{h}$ channel are summarized in Table 1.

The input variables to the BDT are p^{μ}_T, $p^{\tau_\mathrm{h}}_\mathrm{T}$, m_col, $\vec{p}^{\,\mathrm{miss}}_\mathrm{T}$, $m_\mathrm{T}(\tau, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$, $\Delta\eta(\mu, \tau_\mathrm{h})$, $\Delta\phi(\mu, \tau_\mathrm{h})$, and $\Delta\phi(\tau_\mathrm{h}, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$. The neutrino is assumed to be collinear with τ_h, which motivates using the $\Delta\phi(\tau_\mathrm{h}, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$ variable as one of the input variables to the BDT. The two leptons are usually produced in opposite directions of the azimuthal plane, which motivates using the $\Delta\phi(\mu, \tau_\mathrm{h})$ variable as one of the input variables to the BDT. The BDT discriminant distributions of simulated signal, data, and backgrounds in each category of the $\mu\tau_\mathrm{h}$ channel are shown in Fig. 1.

5.2 $H \to \mu\tau_e$

In this channel, the preselection requires a muon and electron of opposite electric charge with a separation of $\Delta R > 0.3$. The triggers require both a muon and an electron, where the muon has p_T above 23 GeV, and the electron has p_T above 12 GeV. The muon is required to have $p_\mathrm{T} > 24\,\mathrm{GeV}$, $|\eta| < 2.4$, and $I^{\mu}_{\mathrm{rel}} < 0.15$. The electron is required to have $p_\mathrm{T} > 13\,\mathrm{GeV}$, $|\eta| < 2.5$, and $I^{\mathrm{e}}_{\mathrm{rel}} < 0.1$. The selections for the $\mu\tau_\mathrm{e}$ channel are summarized in Table 1.

The input variables to the BDT are p^{μ}_T, $p^{\mathrm{e}}_\mathrm{T}$, m_col, $m_\mathrm{T}(\mu, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$, $m_\mathrm{T}(\mathrm{e}, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$, $\Delta\phi(\mathrm{e}, \mu)$, $\Delta\phi(\mu, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$, and $\Delta\phi(\mathrm{e}, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$. The neutrinos are assumed to be collinear with the electron, which motivates using the $\Delta\phi(\mathrm{e}, \vec{p}^{\,\mathrm{miss}}_\mathrm{T})$ variable as one of the input variables to the BDT. The two leptons are usually produced in opposite directions of the azimuthal plane, which motivates using the $\Delta\phi(\mathrm{e}, \mu)$ variable as one of the input variables to the BDT. The BDT discriminant distributions of simulated signal, data, and backgrounds in each category of the $\mu\tau_\mathrm{e}$ channel are shown in

Table 1: Event selection criteria for the $\mathrm{H} \to \mu\tau$ channels.

Variable	$\mu\tau_\mathrm{h}$	$\mu\tau_\mathrm{e}$
p_T^e	—	$> 13\,\mathrm{GeV}$
p_T^μ	$> 26\,\mathrm{GeV}$	$> 24\,\mathrm{GeV}$
$p_\mathrm{T}^{\tau_\mathrm{h}}$	$> 30\,\mathrm{GeV}$	—
$\lvert\eta\rvert^\mathrm{e}$	—	< 2.5
$\lvert\eta\rvert^\mu$	< 2.1	< 2.4
$\lvert\eta\rvert^{\tau_\mathrm{h}}$	< 2.3	—
$I_\mathrm{rel}^\mathrm{e}$	—	< 0.1
I_rel^μ	< 0.15	< 0.15
Trigger requirement	μ [24] (all years) μ [27] (2017)	e [12] and μ [23] (all years)

Fig. 2.

5.3 $H \to e\tau_h$

In this channel, the preselection requires an electron and τ_h of opposite electric charge with a separation of $\Delta R > 0.5$. The triggers are based on the presence of an isolated electron with a p_T threshold of 25 GeV (2016), 27 GeV (2017), or 32 GeV (2018). In 2017 and 2018, the analysis acceptance is increased by selecting events where the electron has p_T above 24 GeV and the τ_h has p_T above 30 GeV. The electron is required to have $p_\mathrm{T} > 27\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.1$, and $I_\mathrm{rel}^\mathrm{e} < 0.15$. The τ_h is required to have $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $\lvert\eta\rvert < 2.3$. The selections for the $\mathrm{e}\tau_\mathrm{h}$ channel are summarized in Table 2.

The input variables to the BDT are p_T^e, $p_\mathrm{T}^{\tau_\mathrm{h}}$, m_col, m_vis, $m_\mathrm{T}(\tau, \vec{p}_\mathrm{T}^\mathrm{miss})$, $\Delta\eta(\mathrm{e}, \tau_\mathrm{h})$, $\Delta\phi(\mathrm{e}, \tau_\mathrm{h})$, and $\Delta\phi(\tau_\mathrm{h}, \vec{p}_\mathrm{T}^\mathrm{miss})$. As can be seen, the input variables are similar to $\mu\tau_\mathrm{h}$ channel except for the addition of the variable m_vis and removing $\vec{p}_\mathrm{T}^\mathrm{miss}$. The variable m_vis has better separation power as the $\mathrm{e}\tau_\mathrm{h}$ channel has more $\mathrm{Z} \to \mathrm{ee}$+jets background than the $\mathrm{Z} \to \mu\mu$+jets background in the $\mu\tau_\mathrm{h}$ channel. The BDT discriminant distributions of simulated signal, data, and backgrounds in each category of the $\mathrm{e}\tau_\mathrm{h}$ channel are shown in Fig. 3.

5.4 $H \to e\tau_\mu$

In this channel, the preselection requires an electron and muon of opposite electric charge with a separation of $\Delta R > 0.4$. The triggers require both an electron and a muon, where the electron has p_T above 23 GeV, and the muon has p_T above 8 GeV. The electron is required to have $p_\mathrm{T} > 24\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.5$, and $I_\mathrm{rel}^\mathrm{e} < 0.1$. The muon is required to have $p_\mathrm{T} > 10\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.4$, and $I_\mathrm{rel}^\mu < 0.15$. The selections for the $\mathrm{e}\tau_\mu$ channel are summarized in Table 2.

The input variables to the BDT are p_T^μ, p_T^e, m_col, m_vis, $m_\mathrm{T}(\mu, \vec{p}_\mathrm{T}^\mathrm{miss})$, $\Delta\phi(\mathrm{e}, \mu)$, $\Delta\phi(\mu, \vec{p}_\mathrm{T}^\mathrm{miss})$, and $\Delta\phi(\mathrm{e}, \vec{p}_\mathrm{T}^\mathrm{miss})$. As can be seen, the input variables are similar to $\mu\tau_\mathrm{e}$ channel except for the addition of the variable m_vis and removing $m_\mathrm{T}(\mathrm{e}, \vec{p}_\mathrm{T}^\mathrm{miss})$. The BDT discriminant distributions of simulated signal, data, and backgrounds in each category of the $\mathrm{e}\tau_\mu$ channel are shown in Fig. 4.

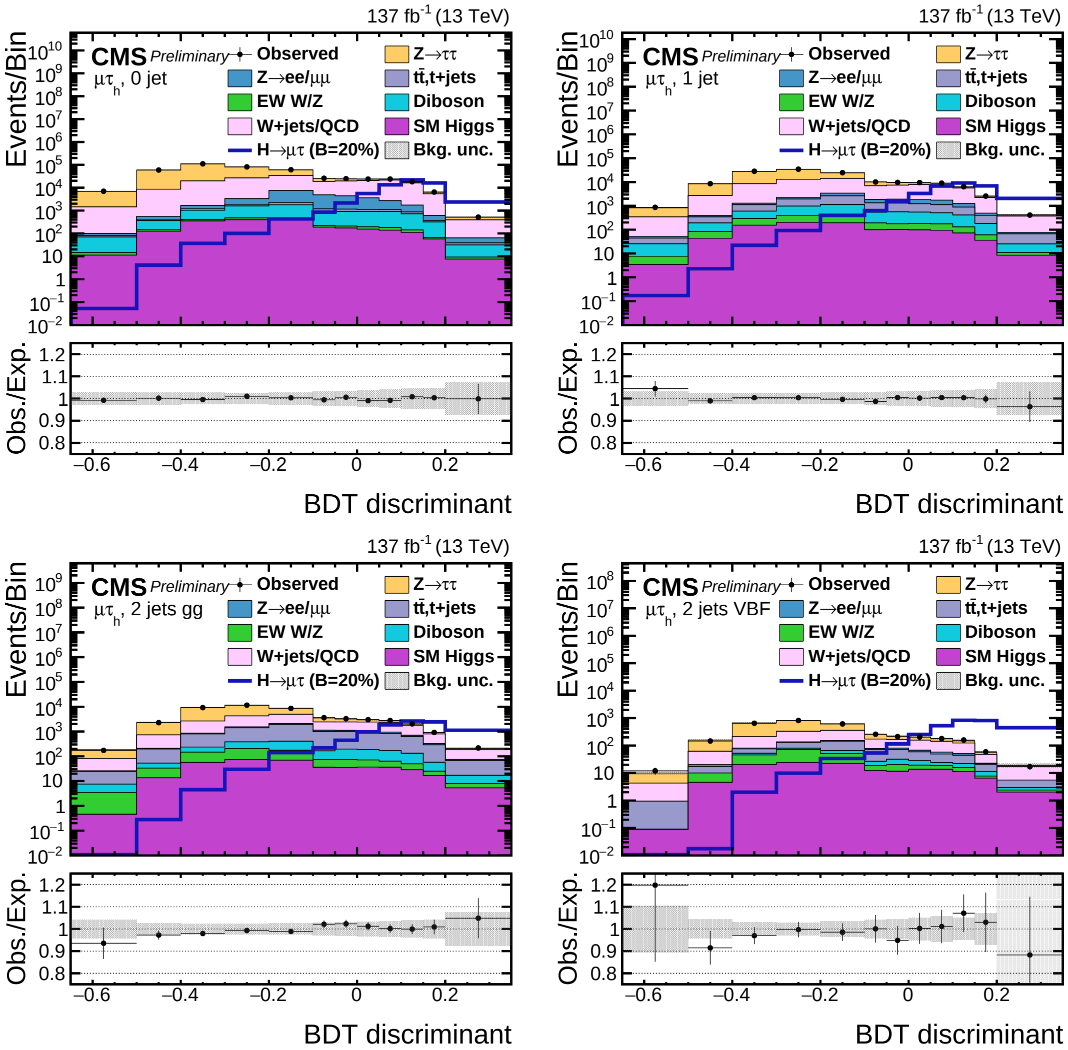

Figure 1: BDT discriminant distributions for the data and background processes in the $H \to \mu\tau_h$ channel. A $\mathcal{B}(H \to \mu\tau) = 20\%$ is assumed for the signal. The channel categories are 0 jets (upper row left), 1 jet (upper row right), 2 jets ggH (lower row left), and 2 jets VBF (lower row right). The lower panel in each plot shows the ratio of data and estimated background. The uncertainty band corresponds to the post-fit statistical and systematic uncertainties added in quadrature.

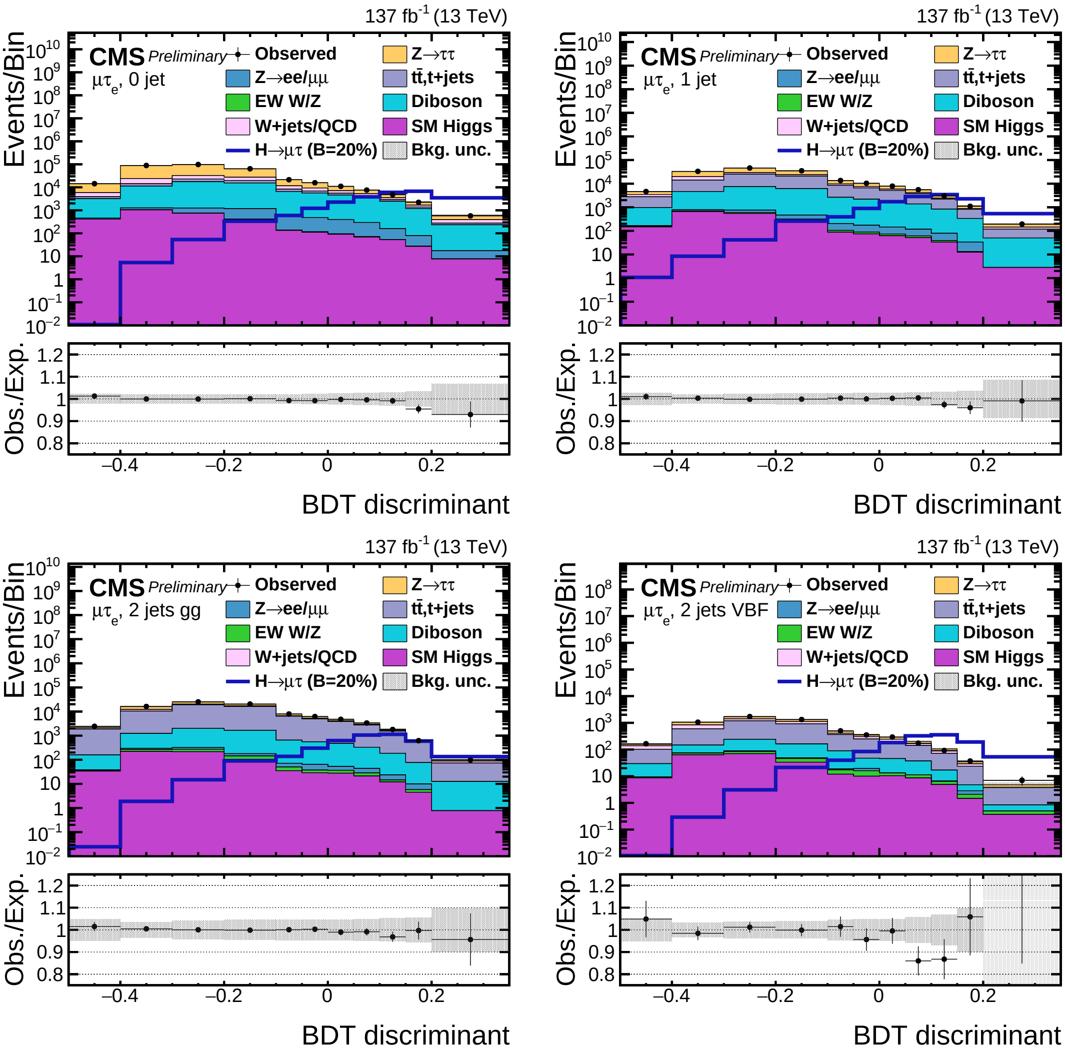

Figure 2: BDT discriminant distributions for the data and background processes in the H → $\mu\tau_e$ channel. A $\mathcal{B}(\mathrm{H} \to \mu\tau) = 20\%$ is assumed for the signal. The channel categories are 0 jets (upper row left), 1 jet (upper row right), 2 jets ggH (lower row left), and 2 jets VBF (lower row right). The lower panel in each plot shows the ratio of data and estimated background. The uncertainty band corresponds to the post-fit statistical and systematic uncertainties added in quadrature.

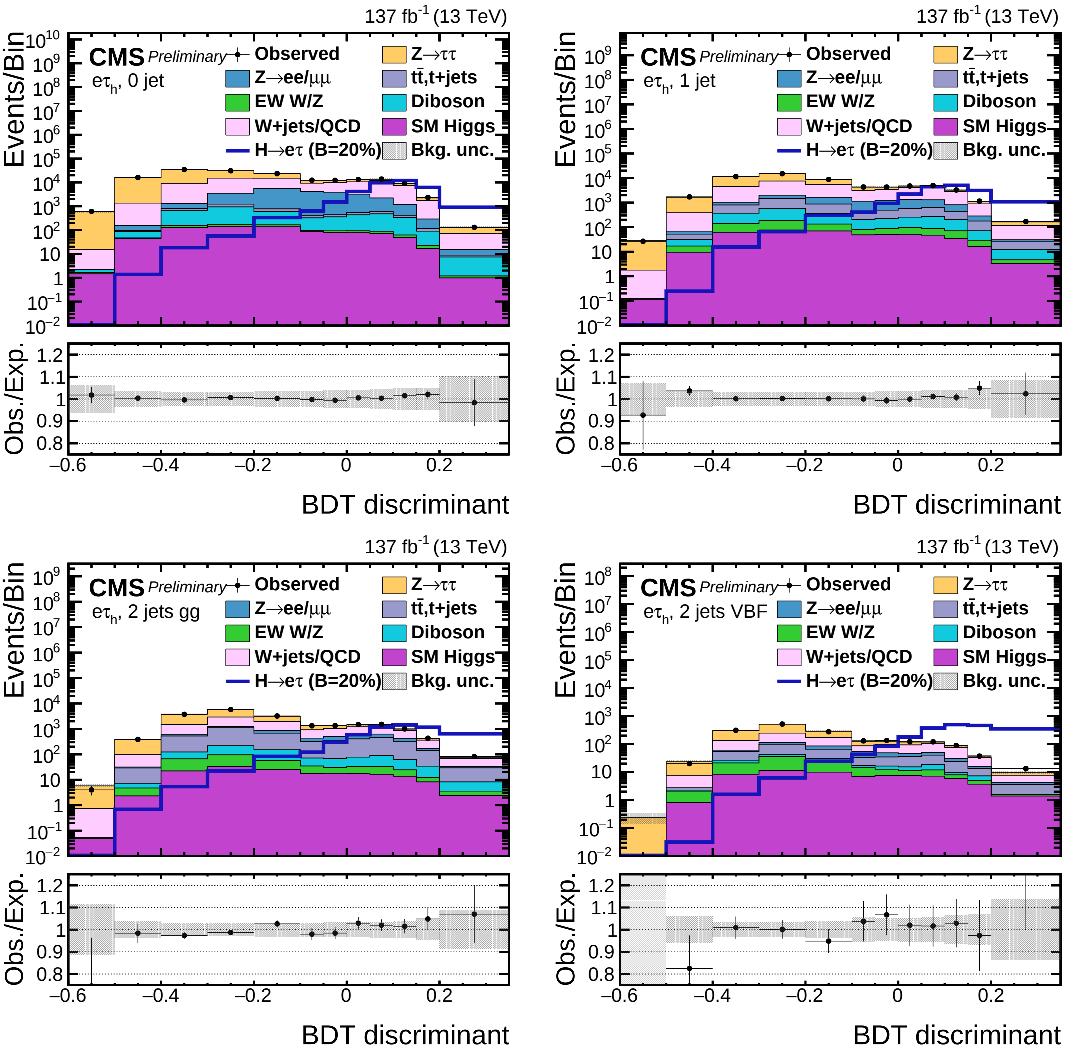

Figure 3: BDT discriminant distributions for the data and background processes in the $H \to e\tau_h$ channel. A $\mathcal{B}(H \to e\tau) = 20\%$ is assumed for the signal. The channel categories are 0 jets (upper row left), 1 jet (upper row right), 2 jets ggH (lower row left), and 2 jets VBF (lower row right). The lower panel in each plot shows the ratio of data and estimated background. The uncertainty band corresponds to the post-fit statistical and systematic uncertainties added in quadrature.

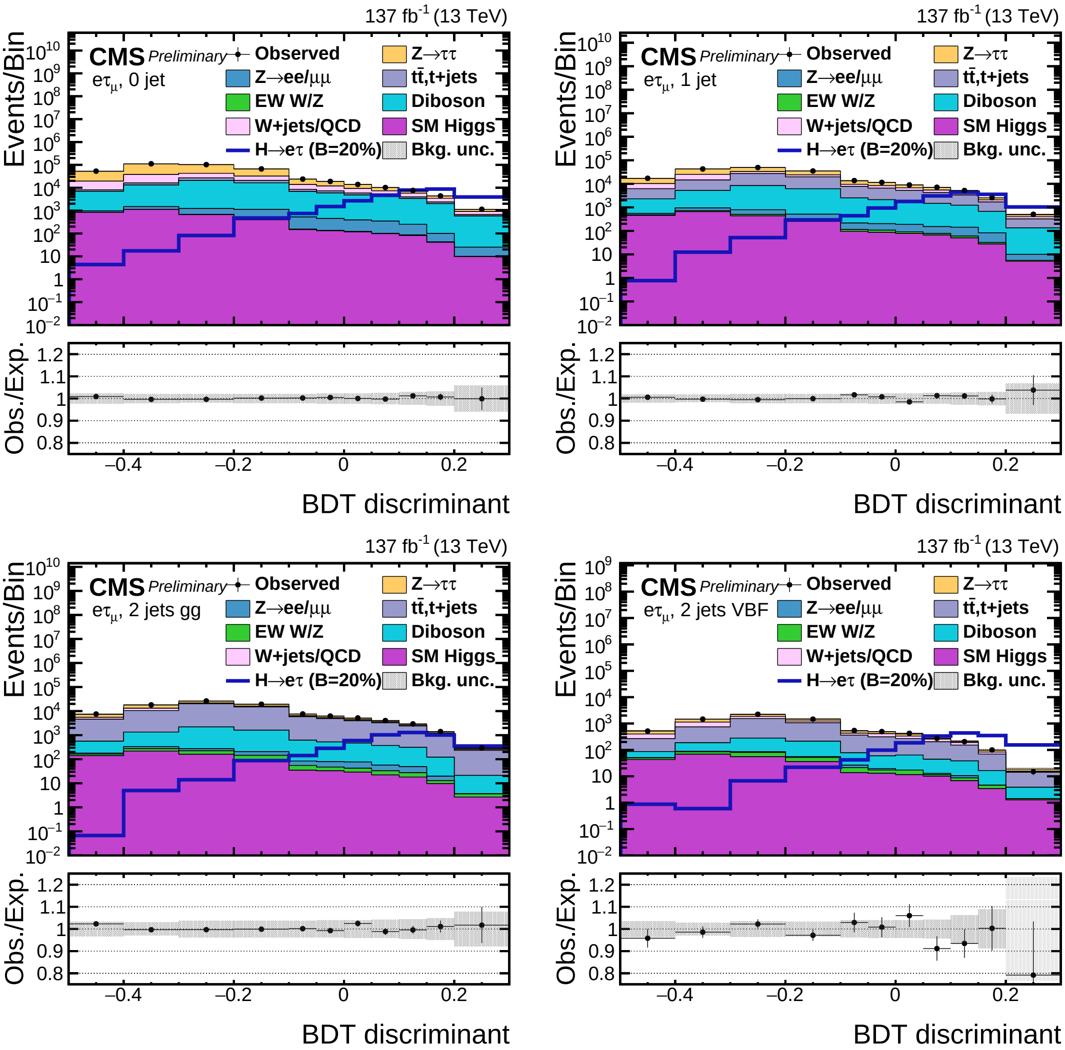

Figure 4: BDT discriminant distributions for the data and background processes in the H → eτ_μ channel. A $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau) = 20\%$ is assumed for the signal. The channel categories are 0 jets (upper row left), 1 jet (upper row right), 2 jets ggH (lower row left), and 2 jets VBF (lower row right). The lower panel in each plot shows the ratio of data and estimated background. The uncertainty band corresponds to the post-fit statistical and systematic uncertainties added in quadrature.

Table 2: Event selection criteria for the $\mathrm{H} \to \mathrm{e}\tau$ channels.

Variable	$\mathrm{e}\tau_\mathrm{h}$	$\mathrm{e}\tau_\mu$
p_T^e	$> 27\,\mathrm{GeV}$	$> 24\,\mathrm{GeV}$
p_T^μ	—	$> 10\,\mathrm{GeV}$
$p_\mathrm{T}^{\tau_\mathrm{h}}$	$> 30\,\mathrm{GeV}$	—
$\lvert\eta\rvert^\mathrm{e}$	< 2.1	< 2.5
$\lvert\eta\rvert^\mu$	—	< 2.4
$\lvert\eta\rvert^{\tau_\mathrm{h}}$	< 2.3	—
$I_\mathrm{rel}^\mathrm{e}$	< 0.15	< 0.1
I_rel^μ	—	< 0.15
Trigger requirement	e [25] (2016) e [27] (2017) e [32] (2018) e [24] and τ_h [30] (2017, 2018)	e [23] and μ [8] (all years)

6 Background estimation

One of the major background contributions comes from the $\mathrm{Z} \to \tau\tau$ process, in which the muon or electron arises from a τ lepton decay. The other major background contributions arise from the W+jets process and from multijets events produced through the strong interaction (referred to as QCD multijet events hereafter), where one or more of the jets are misidentified as leptons. These backgrounds are estimated from data either fully or with the aid of simulation. The $\mathrm{t\bar{t}}$ and single top quark background contributes substantially in leptonic channels and is estimated using simulated events along with the other backgrounds. The background estimates are validated in different orthogonal CRs constructed to have enhanced contributions from specific backgrounds.

6.1 $Z \to \tau\tau$ background

The $\mathrm{Z} \to \tau\tau$ background is estimated from data using an embedding technique [76]. This technique allows for an estimation of the genuine $\tau\tau$ SM backgrounds from data with reduced simulation input. This minimizes the uncertainties that arise from using simulation, which has poor event description. Events with a pair of oppositely charged muons are selected in data so that $\mathrm{Z} \to \mu\mu$ events largely dominate. These data events are selected independently of the event selection criteria described in Section 5. The muons are removed from the selected events and replaced with simulated τ leptons with the same kinematic properties as those of the replaced muon. In that way, a set of hybrid events is obtained that relies on simulation only for the decay of the τ leptons. The description of the underlying event or the production of associated jets is taken entirely from data. This technique results in a more accurate description of the $\vec{p}_\mathrm{T}^\mathrm{miss}$ and jet-related variables than simulation and an overall reduction in the systematic uncertainties. Embedded events cover all backgrounds with two genuine τ leptons, and this includes a small fraction of $\mathrm{t\bar{t}}$, diboson, and EW W/Z events. The simulated events from the $\mathrm{t\bar{t}}$, diboson, and EW W/Z where both τ candidates match to τ leptons at the generator level are removed to avoid any double counting.

6.2 Misidentified lepton background

The misidentified lepton background corresponds to events where jets are misidentified as leptons. They mostly arise from two sources: W+jets and QCD multijet events. In W+jets background events, one of the leptons is from the W boson decay while the other is a jet misidentified as a lepton. In QCD multijet events, both the leptons are misidentified jets. In the $\mu\tau_{\mathrm{h}}$ and $\mathrm{e}\tau_{\mathrm{h}}$ channels, the contributions from misidentified lepton backgrounds have been estimated using a ″misidentification rate″ approach. In the $\mu\tau_{\mathrm{e}}$ and $\mathrm{e}\tau_{\mu}$ channels, an ″extrapolation factor″ approach is adopted, which is consistent with the ″misidentification rate″ approach, and is used because of limited statistical precision in the leptonic channels.

6.2.1 Misidentification rate approach

The misidentified lepton background in the signal region (SR) is estimated using misidentification rates measured in a Z+jets CR and applied to a background-enriched region from collision data. The misidentification rates are evaluated using events with a Z boson and at least one jet that can be misidentified as a lepton. The probabilities with which jets are misidentified as an electron, muon, or τ_{h} are labeled as f_{e}, f_{μ}, and $f_{\tau_{\mathrm{h}}}$, respectively. The Z boson is formed using two muons with $p_{\mathrm{T}} > 26\,\mathrm{GeV}$, $|\eta| < 2.4$, and $I^{\mu}_{\mathrm{rel}} < 0.15$ for measuring the jet $\to \tau_{\mathrm{h}}, \mu, \mathrm{e}$ misidentification rate. The muons are required to be oppositely charged and have invariant mass between 70 and 110 GeV. The contribution from diboson events, where there is a genuine lepton, is subtracted using simulation.

The jet is required to pass the same lepton identification criteria as used in the SR. A “signal-like” and “background-like” regions are defined. The isolation for the electron and muon is required to have $I^{\ell}_{\mathrm{rel}} < 0.15$ and the τ_{h} discriminated against jets at a WP that has an identification efficiency of about 70% for the “signal-like” region. For the “background-like” region, lepton isolation is required to be $0.15 < I^{\mu}_{\mathrm{rel}} < 0.25$ for the muon, $0.15 < I^{\mathrm{e}}_{\mathrm{rel}} < 0.50$ for the electron, and the τ_{h} is discriminated against jets at a WP that has an identification efficiency of about 80% and not pass the WP that has an identification efficiency of about 70%. After the “signal-like” and “background-like” regions are defined, the misidentification rates are computed as functions of the lepton p_{T}. The misidentification rates f_{e}, f_{μ}, and $f_{\tau_{\mathrm{h}}}$ are estimated following:

$$f_i = \frac{N_i(\text{signal-like})}{N_i(\text{background-like}) + N_i(\text{signal-like})}$$

where $N_i(\text{signal-like})$ is the number of events in the “signal-like” region, while $N_i(\text{background-like})$ is the number of events in the “background-like” region. The τ_{h} misidentification rate shows a p_{T} dependence that depends on the τ_{h} decay mode and $|\eta|$ and is therefore evaluated as a function of p^{τ}_{T} for the different decay modes and two η regions ($|\eta| < 1.5$ or $|\eta| > 1.5$).

In the $\mathrm{e}\tau_{\mathrm{h}}$ channel, the τ_{h} misidentification rate is evaluated using events with a Z boson formed using two electrons with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$, $|\eta| < 2.5$, and $I^{\mathrm{e}}_{\mathrm{rel}} < 0.15$. The electrons must be oppositely charged and have an invariant mass between 70 and 110 GeV. The reason for using $\mathrm{Z} \to \mathrm{ee}$ events for evaluating the τ_{h} misidentification rate in $\mathrm{e}\tau_{\mathrm{h}}$ channel is that the DNN WPs used for discriminating τ_{h} against electrons and muons are different in this channel compared to the $\mu\tau_{\mathrm{h}}$ channel as described in Section 4. The misidentification rates evaluated using this CR are compatible with the misidentification rates measured in $\mathrm{Z} \to \mu\mu$ events.

The computed misidentification rates f_i depend on the lepton p_{T} for electrons and muons or p_{T}, η, and decay mode for the τ_{h} candidates. They are used to estimate the background yields and obtain the distributions of the misidentified lepton background. This is accomplished through the following procedure. Each event in the background-enriched region, defined using the collision data with the same selection as the SR, but loosening the isolation requirements on one of the leptons, is weighted by a factor $f_i/(1-f_i)$. Events with the possibility of being double counting because of two misidentified leptons are subtracted. For example, events with a misidentified muon or electron and a misidentified τ_h are subtracted in the $\mu\tau_\mathrm{h}$ or $\mathrm{e}\tau_\mathrm{h}$ channel using a weight, $f_\tau f_\ell/[(1-f_\tau)(1-f_\ell)]$, where $\ell = \mu$ or e.

The background estimate is validated in a CR by requiring the two leptons to have the same electric charge, enhancing the misidentified lepton background. Figure 5 (left) shows the comparison of data with background estimates in this CR for the $\mu\tau_\mathrm{h}$ channel. The background estimate is also validated in a W boson enriched CR, as shown in Fig. 5 (middle). This CR is obtained by applying the preselection, $m_\mathrm{T}(\ell, \vec{p}_\mathrm{T}^\mathrm{miss}) > 60\,\mathrm{GeV}$ (ℓ = e or μ), and $m_\mathrm{T}(\tau_\mathrm{h}, \vec{p}_\mathrm{T}^\mathrm{miss}) > 80\,\mathrm{GeV}$. As detailed in Section 7, uncertainty in the misidentified lepton background is taken from the W boson enriched CR, which is defined orthogonally to the SR.

6.2.2 Extrapolation factor approach

In the $\mathrm{e}\tau_\mu$ and $\mu\tau_\mathrm{e}$ channels, the QCD multijet background is estimated from the data using events with an electron and a muon with the same electric charge. Contributions from other processes are estimated from simulation and subtracted from the data. Extrapolation factors from the CR requiring the two leptons to have the same electric charge to the SR are measured in data as a function of the jet multiplicity and the ΔR separation between the electron and the muon.

The extrapolation factors are estimated using events with a muon failing the isolation requirement and an isolated electron. The contribution from $\mathrm{b\bar{b}}$ events to the QCD multijet background gives rise to the ΔR dependence and is parameterized with a linear function. The extrapolation factors are higher for events with low ΔR separation between the electron and muon, decreasing as the ΔR separation increases. The extrapolation factors also depend on the electron and muon p_T. This p_T dependence comes from the leptons arising from the semileptonic c quark decay. These leptons tend to be softer in p_T and less isolated, resulting in a reduction in the number of such events passing the p_T and isolation requirements [77].

As the extrapolation factors are measured in a CR where the muon fails the isolation requirement, an additional correction is applied to cover a potential mismodeling. This correction is calculated by measuring the extrapolation factors in two different CRs. The first CR has events where the muon is isolated, and the electron fails the isolation requirement. The second CR has events where both the electron and the muon fail the isolation requirement. The ratio of the extrapolation factors measured in these two CRs is taken as the correction to account for the potential mismodeling induced by requiring the muon to fail the isolation requirement.

6.3 Other backgrounds

Other background contributions come from processes in which a lepton pair is produced from the weak decays of quarks and vector bosons. These include $\mathrm{t\bar{t}}$, WW, WZ, and ZZ events. There are nonnegligible contributions from processes such as $\mathrm{W}\gamma^{(*)}$+jets, single top quark production, and $\mathrm{Z} \to \ell\ell$ ($\ell = \mathrm{e}, \mu$). Figure 5 (right) shows the comparison of data with background estimates in the $\mathrm{t\bar{t}}$ CR for the $\mu\tau_\mathrm{e}$ channel. This CR is defined by requiring the presence of at least one b-tagged jet in the event in addition to the preselection. The SM Higgs boson production contribution mainly comes from $\mathrm{H} \to \tau\tau$ and $\mathrm{H} \to \mathrm{WW}$ decays.

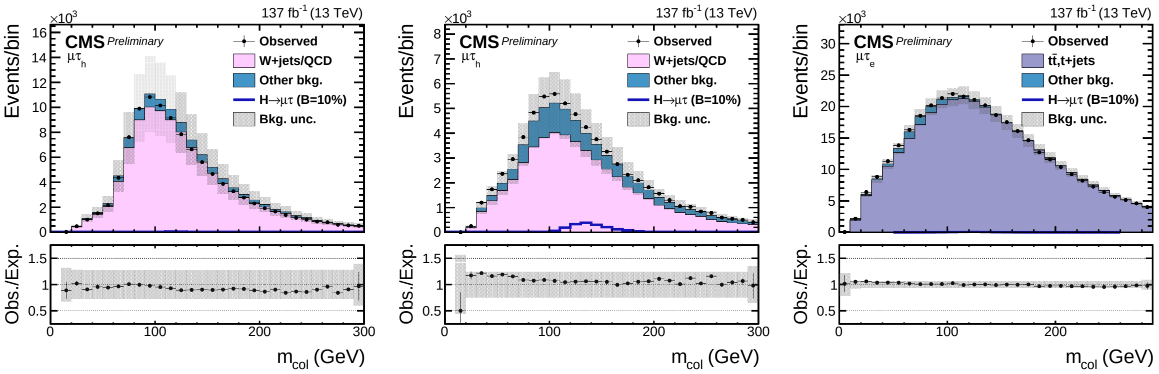

Figure 5: The m_{col} distribution in CR with same electric charge for both leptons (left), W+jets CR (middle), and $\mathrm{t\bar{t}}$ CR (right). In each distribution, the CR's dominant background is shown, and all the other backgrounds are grouped into "Other bkg.". A $\mathcal{B}(\mathrm{H} \to \mu\tau) = 10\%$ is assumed for the signal. The lower panel in each plot shows the ratio of data and estimated background. The uncertainty band shows the statistical and systematic uncertainties added in quadrature.

7 Systematic uncertainties

Several sources of experimental and theoretical systematic uncertainties are taken into account in the statistical analysis. In the statistical analysis, the different systematic uncertainties are incorporated in the likelihood as uninteresting parameters. The maximum likelihood and profile likelihood with asymptotic approximation are then computed using the defined likelihood to obtain the best fit branching fraction and upper limits on the branching fraction for the LFV Higgs boson decays. These uncertainties affect both the normalization and the distribution of the different processes. As the analysis is categorized into different final states, partial and complete correlations between the uncertainties in different categories are taken into account and are summarized in Table 3.

The uncertainties to reconstruct a τ_{h} and estimate its identification efficiency for different p_{T} ranges are measured using a tag-and-probe method [78] and found to be in the range of 2–3%. The uncertainties for different ranges of p_{T} are treated as uncorrelated. These uncertainties are also considered for the embedded $\tau\tau$ background, where they are treated as 50% correlated with the simulation uncertainties. For the embedded events, triggering on muons before being replaced by τ leptons leads to an uncertainty in the trigger efficiency of about 4%, which is treated as uncorrelated between the three years due to different triggering criteria. The embedded events have higher track reconstruction efficiency because of reconstruction in an empty detector environment and have event migration effects for τ_{h} decay modes with a π^0 because of the footprint of replaced muons. Scale factors cover these effects with corresponding systematic uncertainties.

Uncertainties arising from an electron or a muon misidentified as τ_{h} correspond to 40% or between 10–70%, respectively, for different bins of p_{T}, η, and τ_{h} decay modes. The uncertainty in the τ_{h} energy scale is treated as uncorrelated for different decay modes and 50% correlated between embedded and simulated backgrounds and ranges from 0.7 to 1.2%. The uncertainty in the electron energy scale and the muon momentum scale for misidentified leptons is independent of the τ_{h} energy scale and amounts to 7% and 1%, respectively. The effect of lepton energy resolution is found to be negligible.

The jet energy scale is affected by several sources, and its uncertainty is evaluated as a function of p_{T} and η. The jet energy scale's effect is propagated to the BDT discriminant and varies from

3% to 20%. The uncertainties in jet energy resolution are also taken into account and mostly impact the m_{jj}-defined categories. The jets with $p_{\mathrm{T}} < 10\,\mathrm{GeV}$ fall under unclustered energy. The unclustered energy scale is considered independently for charged particles, neutral hadrons, photons, and very forward particles, that affect both the distributions and the total yields and are treated as uncorrelated. The efficiency to classify a jet as b-tagged is different in data and simulation, and scale factors that depend on jet p_{T} are used to correct the simulation. The uncertainties in the measured values of these scale factors are taken as sources of systematic uncertainties.

The uncertainties in the reconstruction of electrons and muons, along with their isolation criteria, are measured using the tag-and-probe method in data in $\mathrm{Z} \to \mathrm{ee}$ and $\mathrm{Z} \to \mu\mu$ events and sum up to about 2% [64, 79, 80]. The uncertainty in the measurement of the muon momentum scale is in the range 0.4–2.7% for different $|\eta|$ ranges, while for the electron momentum scale, it is less than 1%. The selection of events using electron- and muon-based triggers results in an additional 2% uncertainty in the yield of simulated processes. In the $\mathrm{e}\tau_{\mathrm{h}}$ channel, an additional 5% uncertainty is associated with using the trigger requiring the presence of both an electron and τ_{h} in 2017 and 2018. The uncertainties related to the lepton identification, momentum scale, and triggering are treated as correlated between the three years.

The misidentification rates in the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states are parameterized using a linear function dependent on τ_{h} p_{T}, where two uncertainties are ascribed per fit function. The normalization uncertainties in the estimates of the misidentified lepton backgrounds (jet $\to \tau_{\mathrm{h}}, \mu, \mathrm{e}$) from data are taken from the CR, as described in Section 6. Additional uncertainty is estimated for the misidentified lepton background in the W boson enriched CR. It is parameterized as a function of $\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\text{miss}})$ for the $\mu\tau_{\mathrm{h}}$ channel and as a function of $\Delta\phi(\mathrm{e}, \vec{p}_{\mathrm{T}}^{\text{miss}})$ for the $\mathrm{e}\tau_{\mathrm{h}}$ channel. Discriminants with different signal-to-background ratios are used to differentiate τ_{h} against electrons and muons, which entails an additional 3% uncertainty for the $\mathrm{e}\tau_{\mathrm{h}}$ channel.

The misidentified lepton background in the $\mathrm{e}\tau_{\mu}$ and $\mu\tau_{\mathrm{e}}$ final states is affected by different uncertainties. The statistical uncertainties arising from both fits of the extrapolation factors as a function of the lepton p_{T} and the spatial separation between electron and muon are taken into account. The uncertainty in extrapolation factors resulting from inverting the muon isolation is taken into account. These uncertainties have a combined effect of about 20% on the normalization. The dominant source of uncertainty in the simulated background processes, $\mathrm{Z} \to \mathrm{ee}$, $\mathrm{Z} \to \mu\mu$, $\mathrm{Z} \to \tau\tau$, WW, ZZ, $\mathrm{W}\gamma$, $\mathrm{t\bar{t}}$, and single top quark production is the measurement of the cross section for these processes and is treated as correlated between the three years.

The theoretical uncertainties affecting the Higgs boson production cross section's measurement are the QCD scales (renormalization and factorization scales), choice of the PDFs, and the strong coupling constant (α_{S}) evaluated at the mass of the Z boson. These uncertainties affect the signal's normalization and are treated as correlated between the three years [81]. The changes made in QCD scales provide 3.9, 0.5, 0.9, and 0.8% uncertainties in the ggH, VBF, ZH, WH cross sections, respectively, while changes in the PDFs and α_{S} result in 3.2, 2.1, 1.3, and 1.9% uncertainties, respectively. The acceptance is taken into account when changes are made in QCD scales and the PDFs and α_{S}.

The measured $\mathrm{H} \to \tau\tau$ signal strength ($\mu = 0.85^{+0.12}_{-0.11}$) is used to normalize the event yield for $\mathrm{H} \to \tau\tau$ taking account of the uncertainties [77]. The uncertainty in the $\mathcal{B}(\mathrm{H} \to \tau\tau)$ includes a 1.7% uncertainty due to missing higher-order corrections, a 0.99% uncertainty in the quark masses, and a 0.62% uncertainty on α_{S}. The normalization of the event yield for $\mathrm{H} \to \mathrm{WW}$ is taken from simulation. The uncertainty in the $\mathcal{B}(\mathrm{H} \to \mathrm{WW})$ includes a 0.99% uncertainty due to missing higher-order corrections, a 0.99% uncertainty in the quark masses, and a 0.66%

uncertainty in α_S.

The bin-by-bin uncertainties account for the statistical uncertainties in each bin of the distributions of every process. The Barlow–Beeston Lite [82] approach is used, assigning a single parameter to scale the sum of the process yields in each bin, constrained by the total uncertainty, instead of requiring separate parameters, one per process. This is useful to reduce the number of parameters required in the statistical analysis. They are treated as uncorrelated between bins, categories, and channels.

The integrated luminosities of the 2016, 2017, and 2018 data taking periods are individually known to have uncertainties in the 2.3–2.5% range [83–85], while the total integrated luminosity has an uncertainty of 1.8%, the improvement in precision reflecting the uncorrelated time evolution of some systematic effects. The uncertainty in the integrated luminosity affects all processes, with the normalization taken directly from the simulation. Uncertainty related to pileup is evaluated through changes made in the weights applied to the simulation and is treated as correlated between the three years. The dependence on weight is obtained through a 5% change in the total inelastic cross section used to estimate the number of pileup events in data. Other minimum-bias event modeling and initial- and final-state radiation uncertainties are estimated to be much smaller than those on the rate and are therefore neglected.

During the 2016 and 2017 data taking periods, a gradual shift in the timing of the inputs from the ECAL first-level trigger in the region of $|\eta| > 2.0$ caused a specific trigger inefficiency. For events containing an electron or a jet with respective $p_T > 50\,\text{GeV}$ or $> 100\,\text{GeV}$, in the region $2.5 < |\eta| < 3.0$ the efficiency loss is 10–20%, depending on p_T, η, and time. Correction factors are computed from data and applied to the acceptance evaluated through simulation. Uncertainty due to this correction factor is $\approx$1% and is treated as correlated between the two years.

8 Results

A maximum likelihood estimate of the signal strength is performed by fitting distributions of signal and background to data using the BDT discriminant distribution. Uncertainties are treated as uninteresting parameters for which log-normal a priori distributions are assumed, and distribution variations are taken into account via continuous morphing [86], as indicated in Table 3. The fits are performed simultaneously in all channels and categories. Upper limits on the branching fraction of Higgs boson decay are computed using the modified frequentist approach for CL_s, taking the profile likelihood as a test statistic [87–89] in the asymptotic approximation.

No significant excess has been found for the LFV Higgs boson decays in both channels, and upper limits have been placed. The observed and expected upper limits on the Higgs boson branching fractions are 0.15% and 0.15% for $H \to \mu\tau$ and 0.22% and 0.16% for $H \to e\tau$, respectively, at the 95% CL. The upper limits and the best fit branching fractions, for $\mathcal{B}(H \to \mu\tau)$ and $\mathcal{B}(H \to e\tau)$, are reported in Tables 4 and 5. The limits are also summarized graphically in Fig. 6 and shown in Table 6. The limits are significantly improved from previous results [30]. The systematic uncertainties have a dominant contribution to the results.

The upper limits on $\mathcal{B}(H \to \mu\tau)$ and $\mathcal{B}(H \to e\tau)$ are subsequently used to put constraints on LFV Yukawa couplings [11]. The LFV decays $e\tau$ and $\mu\tau$ arise at tree level from the assumed flavor violating Yukawa interactions, $Y_{\ell^\alpha\ell^\beta}$, where ℓ^α, ℓ^β are the leptons of different flavors ($\ell^\alpha \neq$

Table 3: Systematic uncertainties in the expected event yields. All uncertainties are treated as correlated among categories, except those with two values separated by the $\oplus$ sign. In this case, the first value is the correlated uncertainty and the second value is the uncorrelated uncertainty for each category.

Systematic uncertainty	$\mu\tau_{\mathrm{h}}$	$\mu\tau_{\mathrm{e}}$	$\mathrm{e}\tau_{\mathrm{h}}$	$\mathrm{e}\tau_{\mu}$
Muon ident. and iso.	2%	2%	—	2%
Electron ident. and iso.	—	2%	2%	2%
Trigger	2%	2%	2%	2%
τ_{h} ident.	p_{T} dep. (2 – 3%)	—	p_{T} dep. (2 – 15%)	—
$\mu \to \tau_{\mathrm{h}}$ misid.	10–70%	—	—	—
$\mathrm{e} \to \tau_{\mathrm{h}}$ misid.	—	—	40%	—
b tagging efficiency	$< 6.5\%$	$< 6.5\%$	$< 6.5\%$	$< 6.5\%$
Embedded bkg.	4%	4%	4%	4%
$\mathrm{Z} \to \mu\mu$, ee bkg.	4% $\oplus$ 5%	4% $\oplus$ 5%	4% $\oplus$ 5%	4% $\oplus$ 5%
EW bkg.	4% $\oplus$ 5%	4% $\oplus$ 5%	4% $\oplus$ 5%	4% $\oplus$ 5%
W+jets bkg.	—	10%	—	10%
Diboson bkg.	5% $\oplus$ 5%	5% $\oplus$ 5%	5% $\oplus$ 5%	5% $\oplus$ 5%
$\mathrm{t\bar{t}}$ bkg.	6% $\oplus$ 5%	6% $\oplus$ 5%	6% $\oplus$ 5%	6% $\oplus$ 5%
Single top quark bkg.	5% $\oplus$ 5%	5% $\oplus$ 5%	5% $\oplus$ 5%	5% $\oplus$ 5%
Jet $\to \tau_{\mathrm{h}}$ bkg.	30% $\oplus$ 10%	—	30% $\oplus$ 10%	—
Jet energy scale	3 – 20%	3 – 20%	3 – 20%	3 – 20%
τ_{h} energy scale	0.7 – 1.2%	—	0.7 – 1.2%	—
$\mathrm{e} \to \tau_{\mathrm{h}}$ energy scale	1 – 7%	—	1 – 7%	—
$\mu \to \tau_{\mathrm{h}}$ energy scale	1%	—	1%	—
Electron energy scale	—	1 – 2.5%	1 – 2.5%	1 – 2.5%
Muon energy scale	0.4 – 2.7%	0.4 – 2.7%	—	0.4 – 2.7%
Trigger timing inefficiency	0.2 – 1.3%	0.2 – 1.3%	0.2 – 1.3%	0.2 – 1.3%
Integrated luminosity	1.8%	1.8%	1.8%	1.8%
QCD scales (ggH)	3.9%			
QCD scales (VBF)	0.5%			
PDF + α_{S} (ggH)	3.2%			
PDF + α_{S} (VBF)	2.1%			
QCD acceptance (ggH)	-10.3% – +5.9%			
QCD acceptance (VBF)	-2.7% – +2.3%			
PDF + α_{S} acceptance (ggH)	-0.8% – +2.8%			
PDF + α_{S} acceptance (VBF)	-1.7% – +2.3%			

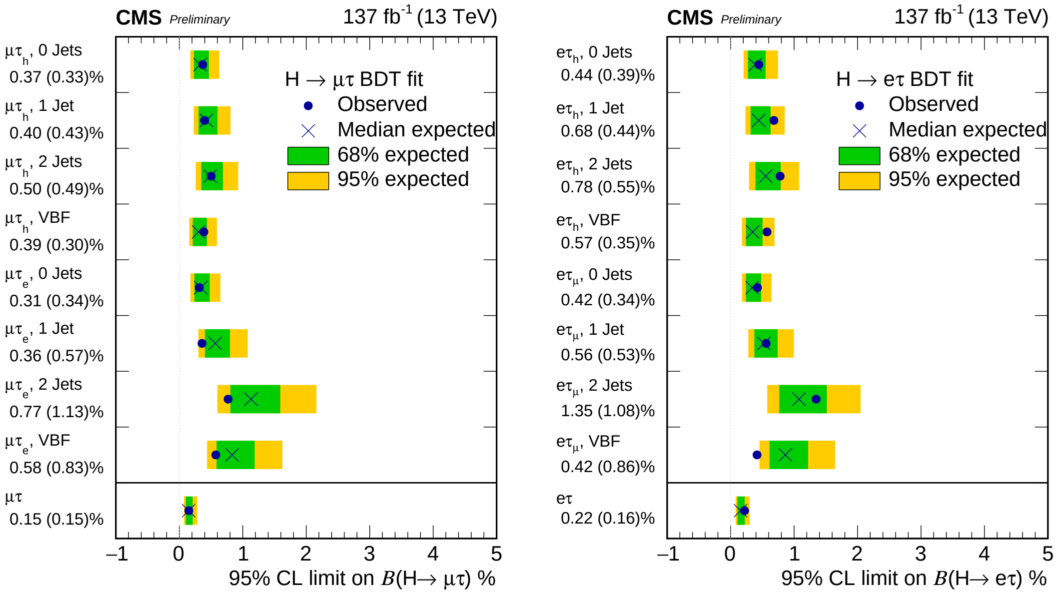

Figure 6: Observed (expected) 95% CL upper limits on the $\mathcal{B}(\mathrm{H} \to \mu\tau)$ (left) and $\mathcal{B}(\mathrm{H} \to \mathrm{e}\tau)$ (right) for each individual category and combined.

ℓ^{β}). The decay widths $\Gamma(\mathrm{H} \to \ell^{\alpha}\ell^{\beta})$ in terms of the Yukawa couplings are given by:

$$\Gamma(\mathrm{H} \to \ell^{\alpha}\ell^{\beta}) = \frac{m_{\mathrm{H}}}{8\pi}(|\mathrm{Y}_{\ell^{\alpha}\ell^{\beta}}|^2 + |\mathrm{Y}_{\ell^{\beta}\ell^{\alpha}}|^2),$$

and the branching fractions are given by:

$$\mathcal{B}(\mathrm{H} \to \ell^{\alpha}\ell^{\beta}) = \frac{\Gamma(\mathrm{H} \to \ell^{\alpha}\ell^{\beta})}{\Gamma(\mathrm{H} \to \ell^{\alpha}\ell^{\beta}) + \Gamma_{\mathrm{SM}}}.$$

The SM Higgs boson decay width is assumed to be $\Gamma_{\mathrm{SM}} = 4.1\,\mathrm{MeV}$ [90] for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. The 95% CL upper limit on the Yukawa couplings obtained from the expression for the branching fraction above is shown in Table 6. The limits on the Yukawa couplings are shown in Fig. 7.

9 Summary

This section summarizes a search for lepton-flavor violations in the $\mu\tau$ and $\mathrm{e}\tau$ channels of the Higgs boson in data collected by the CMS experiment. The data correspond to an integrated luminosity of $137\,\mathrm{fb}^{-1}$ of pp collisions at a center-of-mass energy of 13 TeV. The results are extracted through a maximum likelihood fit to a boosted decision tree discriminant output, trained to distinguish the expected signal from backgrounds. The observed and expected upper limits on the branching fraction of the Higgs boson to $\mu\tau$ are 0.15% and 0.15% and to $\mathrm{e}\tau$ are 0.22% and 0.16%, respectively, at 95% confidence level. Upper limits on the off-diagonal $\mu\tau$ and $\mathrm{e}\tau$ couplings are derived from these constraints, $\sqrt{|\mathrm{Y}_{\mu\tau}|^2 + |\mathrm{Y}_{\tau\mu}|^2} < 1.11\times10^{-3}$ and $\sqrt{|\mathrm{Y}_{\mathrm{e}\tau}|^2 + |\mathrm{Y}_{\tau\mathrm{e}}|^2} < 1.35\times10^{-3}$. These results constitute an improvement over the previous limits from CMS and ATLAS.

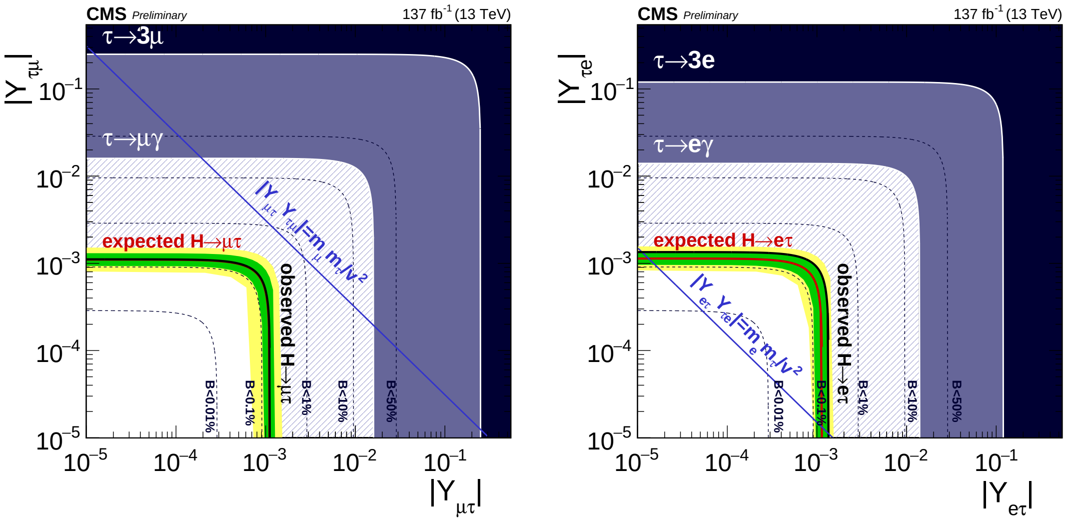

Figure 7: Expected (red line) and observed (black solid line) 95% CL upper limits on the LFV Yukawa couplings, $|\mathrm{Y}_{\mu\tau}|$ vs. $|\mathrm{Y}_{\tau\mu}|$ (left) and $|\mathrm{Y}_{\mathrm{e}\tau}|$ vs. $|\mathrm{Y}_{\tau\mathrm{e}}|$ (right). In the left plot, the expected limit is covered by the observed limit as they have similar values. The flavor diagonal Yukawa couplings are approximated by their SM values. The green (yellow) band indicates the range that is expected to contain 68 (95)% of all observed limit variations from the expected limit. The shaded regions are constraints obtained from null searches for $\tau \to 3\mu$ or $\tau \to 3\mathrm{e}$ (dark blue) [91] and $\tau \to \mu\gamma$ or $\tau \to \mathrm{e}\gamma$ (purple). The blue diagonal line is the theoretical naturalness limit $|\mathrm{Y}_{ij}\mathrm{Y}_{ji}| = m_i m_j / v^2$ [11]

.

Table 4: Observed and expected upper limits at 95% CL and best fit branching fractions for each individual jet category, and their combinations, in the $\mathrm{H} \to \mu\tau$ channel.

	Expected limits (%)				
	0-jet	1-jet	2-jets	VBF	Combined
$\mu\tau_\mathrm{e}$	< 0.34	< 0.57	< 1.13	< 0.83	< 0.27
$\mu\tau_\mathrm{h}$	< 0.33	< 0.43	< 0.49	< 0.30	< 0.18
$\mu\tau$					< 0.15

	Observed limits (%)				
	0-jet	1-jet	2-jets	VBF	Combined
$\mu\tau_\mathrm{e}$	< 0.31	< 0.36	< 0.77	< 0.58	< 0.19
$\mu\tau_\mathrm{h}$	< 0.37	< 0.40	< 0.50	< 0.39	< 0.24
$\mu\tau$					< 0.15

	Best fit branching fractions (%)				
	0-jet	1-jet	2-jets	VBF	Combined
$\mu\tau_\mathrm{e}$	-0.03 ± 0.17	-0.40 ± 0.28	-0.66 ± 0.56	-0.41 ± 0.39	-0.14 ± 0.13
$\mu\tau_\mathrm{h}$	$+0.05 \pm 0.17$	-0.05 ± 0.22	$+0.02 \pm 0.25$	$+0.10 \pm 0.16$	$+0.07 \pm 0.09$
$\mu\tau$					$+0.00 \pm 0.07$

Table 5: Observed and expected upper limits at 95% CL and best fit branching fractions for each individual jet category, and their combinations, in the $H \to e\tau$ channel.

	Expected limits (%)				
	0-jet	1-jet	2-jets	VBF	Combined
$e\tau_{\mu}$	< 0.34	< 0.53	< 1.08	< 0.86	< 0.26
$e\tau_{h}$	< 0.39	< 0.44	< 0.55	< 0.35	< 0.20
$e\tau$					< 0.16
	Observed limits (%)				
	0-jet	1-jet	2-jets	VBF	Combined
$e\tau_{\mu}$	< 0.42	< 0.56	< 1.35	< 0.42	< 0.22
$e\tau_{h}$	< 0.44	< 0.68	< 0.78	< 0.57	< 0.37
$e\tau$					< 0.22
	Best fit branching fractions (%)				
	0-jet	1-jet	2-jets	VBF	Combined
$e\tau_{\mu}$	$+0.11 \pm 0.17$	$+0.04 \pm 0.27$	$+0.35 \pm 0.55$	-1.04 ± 0.44	-0.07 ± 0.13
$e\tau_{h}$	$+0.07 \pm 0.20$	$+0.29 \pm 0.23$	$+0.27 \pm 0.29$	$+0.27 \pm 0.17$	$+0.20 \pm 0.10$
$e\tau$					$+0.08 \pm 0.08$

Table 6: Summary of observed and expected upper limits at 95% CL, best fit branching fractions and corresponding constraints on Yukawa couplings for the $H \to \mu\tau$ and $H \to e\tau$ channels.

	Observed (Expected) upper limits (%)	Best fit branching fractions (%)	Yukawa coupling constraints
$H \to \mu\tau$	$< 0.15\,(0.15)$	0.00 ± 0.07	$< 1.11\,(1.10) \times 10^{-3}$
$H \to e\tau$	$< 0.22\,(0.16)$	0.08 ± 0.08	$< 1.35\,(1.14) \times 10^{-3}$

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, `doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[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, `doi:10.1016/j.physletb.2012.08.021`, `arXiv:1207.7235`.

[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, `doi:10.1007/JHEP06(2013)081`, `arXiv:1303.4571`.

[4] ATLAS Collaboration, “Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment”, *Eur. Phys. J. C* **76** (2016) 6, `doi:10.1140/epjc/s10052-015-3769-y`, `arXiv:1507.04548`.

[5] CMS Collaboration, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”, *Eur. Phys. J. C* **75** (2015) 212, `doi:10.1140/epjc/s10052-015-3351-7`, `arXiv:1412.8662`.

[6] CMS Collaboration, “Study of the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs”, *Phys. Rev. Lett.* **110** (2013) 081803, `doi:10.1103/PhysRevLett.110.081803`, `arXiv:1212.6639`.

[7] ATLAS Collaboration, “Evidence for the spin-0 nature of the Higgs boson using ATLAS data”, *Phys. Lett. B* **726** (2013) 120, `doi:10.1016/j.physletb.2013.08.026`, `arXiv:1307.1432`.

[8] CMS Collaboration, “Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV”, *Phys. Rev. D* **92** (2015) 012004, `doi:10.1103/PhysRevD.92.012004`, `arXiv:1411.3441`.

[9] CMS Collaboration, “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **11** (2017) 047, `doi:10.1007/JHEP11(2017)047`, `arXiv:1706.09936`.

[10] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$”, *Eur. Phys. J. C* **79** (2019) 421, `doi:10.1140/epjc/s10052-019-6909-y`, `arXiv:1809.10733`.

[11] R. Harnik, J. Kopp, and J. Zupan, “Flavor violating Higgs decays”, *JHEP* **03** (2013) 026, `doi:10.1007/JHEP03(2013)026`, `arXiv:1209.1397`.

[12] J. Bjorken and S. Weinberg, “A mechanism for nonconservation of muon number”, *Phys. Rev. Lett.* **38** (1977) 622, `doi:10.1103/PhysRevLett.38.622`.

[13] L. Diaz-Cruz and J. Toscano, “Lepton flavor violating decays of Higgs bosons beyond the standard model”, *Phys. Rev. D* **62** (2000) 116005, `doi:10.1103/PhysRevD.62.116005`, `arXiv:hep-ph/9910233`.

[14] T. Han and D. Marfatia, “h —> mu tau at hadron colliders”, *Phys. Rev. Lett.* **86** (2001) 1442, `doi:10.1103/PhysRevLett.86.1442`, `arXiv:hep-ph/0008141`.

[15] A. Arhrib, Y. Cheng, and O. Kong, “Comprehensive analysis on lepton flavor violating Higgs boson to $\mu^{\mp}\tau^{\pm}$ decay in supersymmetry without R parity”, *Phys. Rev. D* **87** (2013) 015025, `doi:10.1103/PhysRevD.87.015025`, `arXiv:1210.8241`.

[16] K. Agashe and R. Contino, “Composite Higgs-mediated FCNC”, *Phys. Rev. D* **80** (2009) 075016, `doi:10.1103/PhysRevD.80.075016`, `arXiv:0906.1542`.

[17] A. Azatov, M. Toharia, and L. Zhu, “Higgs mediated FCNC’s in warped extra dimensions”, *Phys. Rev. D* **80** (2009) 035016, `doi:10.1103/PhysRevD.80.035016`, `arXiv:0906.1990`.

[18] H. Ishimori et al., “Non-Abelian discrete symmetries in particle physics”, *Prog. Theor. Phys. Suppl.* **183** (2010) 1, `doi:10.1143/PTPS.183.1`, `arXiv:1003.3552`.

[19] G. Perez and L. Randall, “Natural neutrino masses and mixings from warped geometry”, *JHEP* **01** (2009) 077, `doi:10.1088/1126-6708/2009/01/077`, `arXiv:0805.4652`.

[20] S. Casagrande et al., “Flavor physics in the Randall-Sundrum model: I. Theoretical setup and electroweak precision tests”, *JHEP* **10** (2008) 094, `doi:10.1088/1126-6708/2008/10/094`, `arXiv:0807.4937`.

[21] A. J. Buras, B. Duling, and S. Gori, “The impact of Kaluza-Klein fermions on standard model fermion couplings in a RS model with custodial protection”, *JHEP* **09** (2009) 076, `doi:10.1088/1126-6708/2009/09/076`, `arXiv:0905.2318`.

[22] M. Blanke et al., “$\Delta F = 2$ observables and fine-tuning in a warped extra dimension with custodial protection”, *JHEP* **03** (2009) 001, `doi:10.1088/1126-6708/2009/03/001`, `arXiv:0809.1073`.

[23] M. E. Albrecht et al., “Electroweak and flavour structure of a warped extra dimension with custodial protection”, *JHEP* **09** (2009) 064, `doi:10.1088/1126-6708/2009/09/064`, `arXiv:0903.2415`.

[24] G. F. Giudice and O. Lebedev, “Higgs-dependent Yukawa couplings”, *Phys. Lett. B* **665** (2008) 79, `doi:10.1016/j.physletb.2008.05.062`, `arXiv:0804.1753`.

[25] J. A. Aguilar Saavedra, “A minimal set of top-Higgs anomalous couplings”, *Nucl. Phys. B* **821** (2009) 215, `doi:10.1016/j.nuclphysb.2009.06.022`, `arXiv:0904.2387`.

[26] A. Goudelis, O. Lebedev, and J. h. Park, “Higgs-induced lepton flavor violation”, *Phys. Lett. B* **707** (2012) 369, `doi:10.1016/j.physletb.2011.12.059`, `arXiv:1111.1715`.

[27] D. McKeen, M. Pospelov, and A. Ritz, “Modified Higgs branching ratios versus CP and lepton flavor violation”, *Phys. Rev. D* **86** (2012) 113004, `doi:10.1103/PhysRevD.86.113004`, `arXiv:1208.4597`.

[28] A. Pilaftsis, “Lepton flavor nonconservation in H^0 decays”, *Phys. Lett. B* **285** (1992) 68, `doi:10.1016/0370-2693(92)91301-O`.

[29] J. Korner, A. Pilaftsis, and K. Schilcher, “Leptonic CP asymmetries in flavor changing H^0 decays”, *Phys. Rev. D* **47** (1993) 1080, `doi:10.1103/PhysRevD.47.1080`, `arXiv:hep-ph/9301289`.

[30] CMS Collaboration, “Search for lepton flavour violating decays of the Higgs boson to $\mu\tau$ and eτ in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **06** (2018) 001, `doi:10.1007/JHEP06(2018)001`, `arXiv:1712.07173`.

[31] ATLAS Collaboration, “Searches for lepton-flavour-violating decays of the Higgs boson in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”, *Phys. Lett. B* **800** (2020) 135069, `doi:10.1016/j.physletb.2019.135069`, `arXiv:1907.06131`.

[32] O. U. Shanker, “Flavor violation, scalar particles and leptoquarks”, *Nucl. Phys. B* **206** (1982) 253, `doi:10.1016/0550-3213(82)90534-X`.

[33] B. McWilliams and L.-F. Li, “Virtual effects of Higgs particles”, *Nucl. Phys. B* **179** (1981) 62, `doi:10.1016/0550-3213(81)90249-2`.

[34] G. Blankenburg, J. Ellis, and G. Isidori, “Flavour-changing decays of a 125 GeV Higgs-like particle”, *Phys. Lett. B* **712** (2012) 386, `doi:10.1016/j.physletb.2012.05.007`, `arXiv:1202.5704`.

[35] MEG Collaboration, “New constraint on the existence of the $\mu^+ \to e^+\gamma$ decay”, *Phys. Rev. Lett.* **110** (2013) 201801, `doi:10.1103/PhysRevLett.110.201801`, `arXiv:1303.0754`.

[36] A. Celis, V. Cirigliano, and E. Passemar, “Lepton flavor violation in the Higgs sector and the role of hadronic τ-lepton decays”, *Phys. Rev. D* **89** (2014) 013008, `doi:10.1103/PhysRevD.89.013008`, `arXiv:1309.3564`.

[37] CMS Collaboration, “Search for lepton flavour violating decays of the Higgs boson to eτ and eμ in proton–proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett. B* **763** (2016) 472, `doi:10.1016/j.physletb.2016.09.062`, `arXiv:1607.03561`.

[38] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020) P10017, `doi:10.1088/1748-0221/15/10/P10017`, `arXiv:2006.10165`.

[39] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017) P01020, `doi:10.1088/1748-0221/12/01/P01020`, `arXiv:1609.02366`.

[40] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[41] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159, `doi:10.1016/j.cpc.2015.01.024`, `arXiv:1410.3012`.

[42] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, `doi:10.1140/epjc/s10052-016-3988-x`, `arXiv:1512.00815`.

[43] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, *Eur. Phys. J. C* **80** (2020) 4, `doi:10.1140/epjc/s10052-019-7499-4`, `arXiv:1903.12179`.

[44] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur. Phys. J. C* **77** (2017) 663, `doi:10.1140/epjc/s10052-017-5199-5`, `arXiv:1706.00428`.

[45] GEANT4 Collaboration, "GEANT4 — a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, `doi:10.1016/S0168-9002(03)01368-8`.

[46] H. Georgi, S. Glashow, M. Machacek, and D. Nanopoulos, "Higgs bosons from two gluon annihilation in proton-proton collisions", *Phys. Rev. Lett.* **40** (1978) 692, `doi:10.1103/PhysRevLett.40.692`.

[47] R. Cahn, S. Ellis, R. Kleiss, and J. Stirling, "Transverse momentum signatures for heavy Higgs bosons", *Phys. Rev. D* **35** (1987) 1626, `doi:10.1103/PhysRevD.35.1626`.

[48] S. Glashow, D. Nanopoulos, and Yildiz, "Associated production of Higgs bosons and Z particles", *Phys. Rev. D* **18** (1978) 1724, `doi:10.1103/PhysRevD.18.1724`.

[49] P. Nason, "A new method for combining NLO QCD with shower monte carlo algorithms", *JHEP* **11** (2004) 040, `doi:10.1088/1126-6708/2004/11/040`, `arXiv:hep-ph/0409146`.

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

[51] 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, `doi:10.1007/JHEP06(2010)043`, `arXiv:1002.2581`.

[52] S. Alioli et al., "Jet pair production in POWHEG", *JHEP* **04** (2011) 081, `doi:10.1007/JHEP04(2011)081`, `arXiv:1012.3380`.

[53] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG", *JHEP* **04** (2009) 002, `doi:10.1088/1126-6708/2009/04/002`, `arXiv:0812.0578`.

[54] 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, `doi:10.1007/JHEP02(2012)088`, `arXiv:1111.2854`.

[55] 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, `doi:10.1007/JHEP08(2017)088`, `arXiv:1703.09252`.

[56] 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, `doi:10.1007/JHEP09(2018)057`, `arXiv:1806.05162`.

[57] 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, `doi:10.1007/JHEP07(2014)079`, `arXiv:1405.0301`.

[58] 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, `doi:10.1140/epjc/s10052-007-0490-5`, `arXiv:0706.2569`.

[59] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, `doi:10.1007/JHEP12(2012)061`, `arXiv:1209.6215`.

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

[61] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_T jet clustering algorithm", *JHEP* **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[62] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, `doi:10.1140/epjc/s10052-012-1896-2`, `arXiv:1111.6097`.

[63] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", `arXiv:2012.06888`.

[64] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}$ = 8 TeV", *JINST* **10** (2015) P06005, `doi:10.1088/1748-0221/10/06/P06005`, `arXiv:1502.02701`.

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

[66] 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* **13** (2018) P10005, `doi:10.1088/1748-0221/13/10/P10005`, `arXiv:1809.02816`.

[67] CMS Collaboration, "Identification of tau leptons using Deep Learning techniques at CMS", Technical Report CMS-CR-2019-272, CERN, Geneva, Nov, 2019.

[68] CMS Collaboration, "Jet algorithms performance in 13 TeV data", Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[69] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, `doi:10.1088/1748-0221/13/05/P05011`, `arXiv:1712.07158`.

[70] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119, `doi:10.1016/j.physletb.2007.09.077`, `arXiv:0707.1378`.

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

[72] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019) P07004, `doi:10.1088/1748-0221/14/07/P07004`, `arXiv:1903.06078`.

[73] M. Cacciari, G. P. Salam, and G. Soyez, "The catchment area of jets", *JHEP* **04** (2008) 005, `doi:10.1088/1126-6708/2008/04/005`, `arXiv:0802.1188`.

[74] K. Ellis, I. Hinchliffe, Soldate, and J. van der Bij, "Higgs decay to tau+ tau-: A possible signature of intermediate mass Higgs bosons at the SSC", *Nucl. Phys. B* **297** (1988) 221, `doi:10.1016/0550-3213(88)90019-3`.

[75] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt, “TMVA, the toolkit for multivariate data analysis with ROOT”, in *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*, p. 40. 2007. `arXiv:physics/0703039`. [PoS(ACAT)040]. `doi:10.22323/1.050.0040`.

[76] CMS Collaboration, “An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data”, *JINST* **14** (2019) P06032, `doi:10.1088/1748-0221/14/06/P06032`, `arXiv:1903.01216`.

[77] CMS Collaboration, “Measurement of Higgs boson production in the decay channel with a pair of τ leptons”, Technical Report CMS-PAS-HIG-19-010, CERN, Geneva, 2020.

[78] CMS Collaboration, “Measurements of inclusive w and z cross sections in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **01** (2011) 080, `doi:10.1007/JHEP01(2011)080`, `arXiv:1012.2466`.

[79] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, `doi:10.1088/1748-0221/7/10/P10002`, `arXiv:1206.4071`.

[80] CMS Collaboration, “Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS”, *JINST* **11** (2016) P01019, `doi:10.1088/1748-0221/11/01/P01019`, `arXiv:1510.07488`.

[81] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector”, `doi:10.23731/CYRM-2017-002`, `arXiv:1610.07922`.

[82] R. J. Barlow and C. Beeston, “Fitting using finite monte carlo samples”, *Comput. Phys. Commun.* **77** (1993) 219, `doi:10.1016/0010-4655(93)90005-W`.

[83] CMS Collaboration, “CMS luminosity measurements for the 2016 data-taking period”, Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[84] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”, Technical Report CMS-PAS-LUM-17-004, CERN, 2018.

[85] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV”, Technical Report CMS-PAS-LUM-18-002, CERN, 2019.

[86] J. S. Conway, “Nuisance parameters in likelihoods for multisource spectra”, in *Proceedings of PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, H. Propser and L. Lyons, eds., number CERN-2011-006, p. 115. CERN, 2011.

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

[88] A. L. Read, “Presentation of search results: The $\mathrm{CL_s}$ technique”, *J. Phys. G* **28** (2002) 2693, `doi:10.1088/0954-3899/28/10/313`.

[89] 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, `doi:10.1140/epjc/s10052-011-1554-0`, `arXiv:1007.1727`. [Erratum: `doi:10.1140/epjc/s10052-013-2501-z`].

[90] A. Denner et al., "Standard model Higgs-boson branching ratios with uncertainties", *Eur. Phys. J. C* **71** (2011) 1753, `doi:10.1140/epjc/s10052-011-1753-8`, `arXiv:1107.5909`.

[91] K. Hayasaka et al., "Search for lepton flavor violating τ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs", *Phys. Lett. B* **687** (2010) 139, `doi:10.1016/j.physletb.2010.03.037`, `arXiv:1001.3221`.