FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Assignment of master's thesis

Title:	Search for tbH+(tautau) with Performance Optimisation for Signal and Background Separation Using Machine Learning with ATLAS Data
Student:	Bc. Martin Rameš
Supervisor:	doc. Dr. André Sopczak
Study program:	Informatics
Branch / specialization:	Knowledge Engineering
Department:	Department of Applied Mathematics
Validity:	until the end of summer semester 2023/2024

Instructions

The particle accelerator at CERN produces a large number of so-called events, describing the collision products and its properties. The task is to recognize the events of interest automatically using the techniques of machine learning, and possibly deep learning, and thus increase the ratio of correctly identified events. The project is a part of the effort to search for electrically charged Higgs bosons. Simulated signal and background events and recorded data are available for the tbH+(tautau) search.

Tasks:
(1) Familiarise yourself with the existing application of the tbH+(ττ) analysis, and tools of collaborative work like gitlab, Mattermost and Indico.
(2) Survey and optimise machine learning algorithms to separate signal and background events.
(3) Implement and apply the algorithms on the provided simulated data.
(4) Study the performance of the algorithms and compare them to the original analysis performance.
(5) Determine the feature importance ranking and study the effect of feature reduction on the performance. Take correlations of the features into account.
(6) Study the uncertainty on the signal and background separation and optimize the signal and background separation including statistical and systematic uncertainties.

Electronically approved by Ing. Magda Friedjungová, Ph.D. on 15 October 2022 in Prague.

Master's thesis

SEARCH FOR TBH$^+$($\tau\tau$) WITH PERFORMANCE OPTIMISATION FOR SIGNAL AND BACKGROUND SEPARATION USING MACHINE LEARNING WITH ATLAS DATA

Bc. Martin Rameš

Faculty of Information Technology
Department of Applied Mathematics
Supervisor: doc. Dr. André Sopczak
May 4, 2023

Citation of this thesis: Rameš Martin. *Search for $tbH^+(\tau\tau)$ with Performance Optimisation for Signal and Background Separation Using Machine Learning with ATLAS Data.* Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2023.

Contents

List of Figures

List of Tables

List of code listings

I would like to thank my supervisor, doc. Dr. André Sopczak, for his support and expert advice. I would also like to thank the CERN organization, for providing the computational resources used for the thesis. Finally, I would like to greatly thank my family for giving me trust and encouragement for the whole duration of my studies.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis. I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as a school work under the provisions of Article 60 (1) of the Act.

In Prague on May 4, 2023 ....

Abstract

The search for charged Higgs bosons, predicted by the Two Doublet Higgs Model and the Minimal Supersymmetric extension of the Standard Model, is challenging because of a large number of background processes and the unknown mass of the charged Higgs bosons. This thesis proposes to use machine learning to separate signal $tbH^+ \to tbWh \to tbW\tau\tau$ from $t\bar{t}h$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$, VV and other background processes. A multi-model approach is proposed, where each model is sensitive in a certain mass range to achieve large significance in its dedicated mass section. Four different model types are optimized and the best model is selected for each mass of the charged Higgs boson analysis. Permutation feature ranking is used for each best model to determine the most important features. Based on the highest-ranking features, feature reduction is demonstrated to reduce the sensitivity only slightly. Results are expressed as expected 95% CL limits.

Keywords ATLAS, CERN, classification, cross-section, machine learning, neural networks, Keras, particle physics, ROOT, tbH$^+$

Abstrakt

Hledání nabitých Higgsových bosonů, které předpovídá model označovaný jako Two Doublet Higgs Model a Minimální supersymetrické rozšíření Standardního modelu, je náročné kvůli velkému množství procesů v pozadí a neznámé hmotnosti nabitých Higgsových bosonů. Tato práce navrhuje použít strojové učení k oddělení signálu $tbH^+ \to tbWh \to tbW\tau\tau$ od $t\bar{t}h$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$, VV a dalších procesů na pozadí. Je navržen vícemodelový přístup, kde je každý model citlivý v určitém rozsahu hmotností, aby dosáhl velké významnosti ve svém vyhrazeném hmotnostním úseku. Jsou optimalizovány čtyři různé typy modelů a pro každou hmotnost analýzy nabitého Higgsova bosonu je vybrán nejlepší model. Pro každý nejlepší model je použito permutační řazení příznaků k určení nejdůležitějších vstupů modelu. Na základě nejlépe hodnocených příznaků je prokázáno, že redukce počtu příznaků snižuje citlivost jen nepatrně. Výsledky jsou vyjádřeny jako očekávané limity na 95% CL.

Klíčová slova ATLAS, CERN, klasifikace, cross-section, strojové učení, neuronové sítě, Keras, částicová fyzika, ROOT, tbH$^+$

Introduction

The Standard Model of particle physics is one of the most important theories of today. It unifies three fundamental forces – electromagnetic, strong, and weak interaction, leaving only gravity not included. One of the last big events in the physics field is the discovery of the Higgs boson, which was predicted as part of the Standard Model. But even with the Standard Model, our understanding of particle physics is still incomplete. There are still measurements, which we cannot fully explain. It is time to go beyond Standard Model.

There are already numerous models that try to explain some of the observed phenomena. Among them, some even predict a new kind of Higgs-boson-like particle with charge. The presence or absence of the charged Higgs boson provides a way to validate or disprove some of the key hypotheses of today.

The observation of the charged Higgs boson is quite challenging. It is expected to be even harder to observe, than the original Higgs boson. While the Large Hadron Collider, which produced the original Higgs boson, should also be able to produce the charged one, it will be hidden among the numerous background events. LHC produces hundreds of millions of proton-proton collision events per second, but it is expected to take tens of minutes if not hours to produce the charged Higgs boson. In such a scenario, automatic filtering of the events and machine learning is key to removing most of the uninteresting events, so that only a few interesting events remain – the ones caused by charged Higgs boson (with as few background events as possible mixed in).

This thesis aims to provide a machine-learning-based filter, which would filter out background events, and estimate, how many events would have to get through the filter out of the fixed number of events so that it would prove that the charged Higgs boson was present among the filtered events.

Chapter 1
Goals

The goal of this thesis is to optimize the separation of signal (charged Higgs bosons) and background events using machine learning algorithms. This can be aided by previous work on the subject, which can serve as a baseline for the performance of the algorithms.

To deal with a high number of features produced by the ATLAS detector (and its simulated variant), the effects of each of the features on the machine learning models will be studied. Feature ranking will be used to get the list of the most important features for the models. Feature reduction will be used to test if the models work better with fewer features. The correlations of the features may also play a role and will be taken into account.

Lastly, the statistical uncertainty of signal and background separation will be analyzed with the goal to optimize the separation based on this uncertainty, as well as systematic uncertainties.

Chapter 2
Background

2.1 ATLAS Detector

The Large Hadron Collider (LHC) at CERN collides bunches of up to 10^{11} protons (p) at a design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. The high interaction rates (collisions 40 million times per second), radiation doses, particle multiplicities, and energies require a specially designed particle detector − ATLAS (A Toroidal LHC ApparatuS) [1].

Before delving into the detector itself, let us review the basic terminology for the measurements. The azimuthal angle ϕ is measured around the beam axis. The polar angle θ is the angle from the beam axis. Pseudorapidity, defined as $\eta = -\ln\tan(\theta/2)$, is often used instead of the polar angle. The transverse momentum p_T is measured in the plane perpendicular to the beam (transverse plane). The pseudorapidity-azimuthal angle distance is defined as equation 2.1.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{2.1}$$

The ATLAS detector 2.1 consists of several layers of detectors. Below is the list of detectors, sorted from the innermost to the outermost detector:

- Pixel detector
- Barrel semiconductor tracker (SCT)
- Barrel transition radiation tracker (TRT)
- End-cap transition radiation tracker (TRT)
- End-cap semiconductor tracker (SCT)
- Liquid Argon (LAr) Calorimeter
- Tile Hadronic Calorimeter
- Muon spectrometer

Pixel detectors and SCT cover the region $|\eta| < 2.5$. TRT enables track following for $|\eta| \leq 2$. The calorimeters cover the range $|\eta| < 4.9$. This list is important to understand the ranges of possible low-level feature values. The available features will be further examined in the “Analysis and Design” chapter.

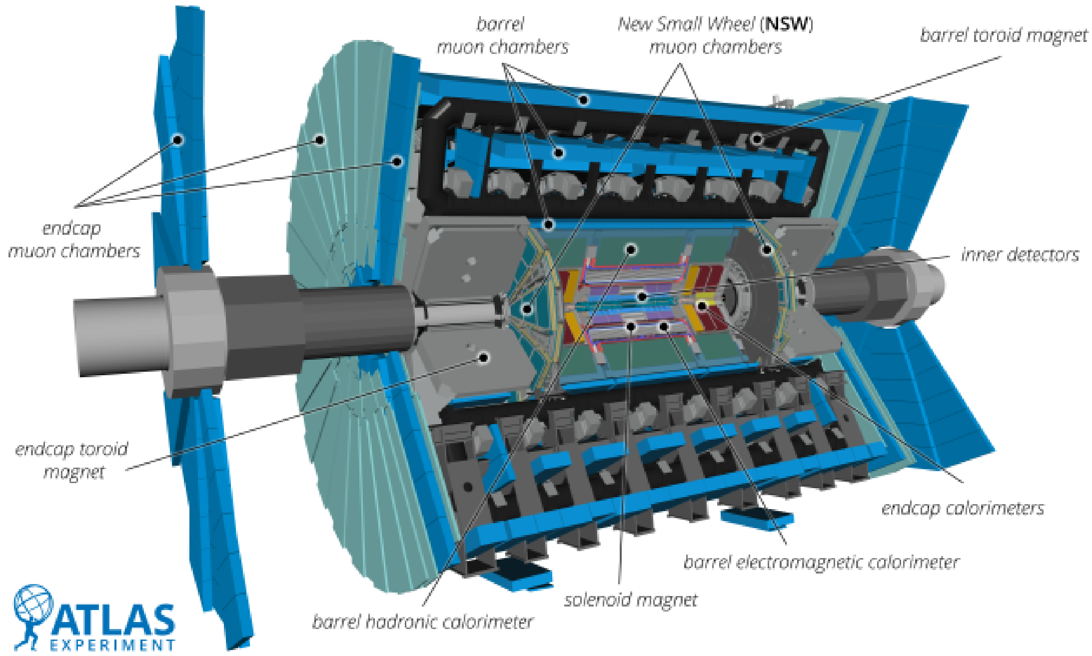


Figure 2.1 ATLAS detector layout [2]

2.2 Particle Physics

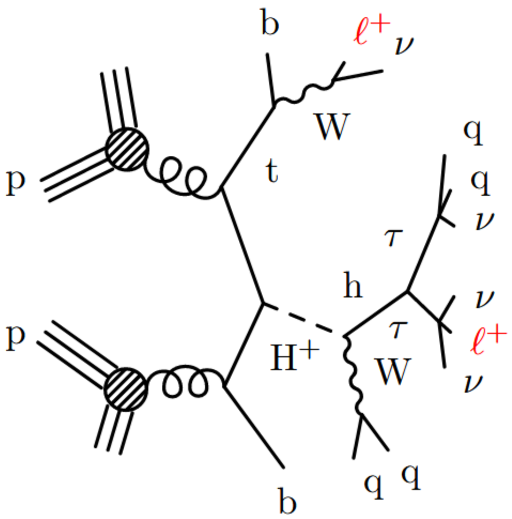
Below is a short overview of what is the charged Higgs boson, what are the most prominent background processes and what metrics are used to evaluate the quality of the separation. This section also includes explanations of the terminology used in the thesis.

2.2.1 Charged Higgs Boson

The Charged Higgs boson appears in several Standard Model (SM) extensions. One of those extensions is the 2 Doublet Higgs Model (2DHM), which predicts 5 physical Higgs bosons – two neutral CP-even scalars (Standard Model Higgs h and heavy Higgs H), two charged Higgs bosons $\mathrm{H}^{\pm}$ and one neutral CP-odd pseudoscalar A. The model has free parameters – the remaining masses of the Higgs bosons and the value of $\tan\beta = \nu_2/\nu_1$, where ν_1 and ν_2 are the vacuum expectation values. Additionally, type II 2DHM is included in the Minimal Supersymmetric SM (MSSM). Type II means, that the fermions couple to ϕ_1 for the down quark and leptons and ϕ_2 for the up quark, where (ϕ_1, ϕ_2) is the second Higgs doublet introduced by the model [3].

Based on the mass of the charged Higgs boson, two production regions can be distinguished – charged Higgs boson with lower mass than the top quark ($m_{H^{\pm}} < m_t$) and charged Higgs boson with higher mass ($m_{H^{\pm}} > m_t$) [3]. For reference, the top quark mass has been measured to be $m_t = 172.13^{+0.76}_{-0.77}$ GeV [4]. The lower mass charged Higgs boson is mainly produced by the decay of a top quark to $\mathrm{H}^{\pm}$b in $t\bar{t}$ production. The higher mass Higgs boson ($m_{H^{\pm}} > m_t$) is produced by the fusion of top-bottom quarks. It has with two possible semi-final states – $\mathrm{H}^{\pm}$tb and $\mathrm{H}^{\pm}$t (depending on flavor scheme). The most important decay channels of high mass charged Higgs boson by contribution are $\mathrm{H}^{\pm} \to \tau\nu$ and $\mathrm{H}^{\pm} \to$ tb [3].

This thesis will be focused on the $tbH^{+} \to tbWh \to tbW\tau\tau$ decay channel.



■ **Figure 2.2** tbH^+ Feynman diagram, leading to the 2lSS1tau final state [5]

2.2.2 Signal and Background

In particle physics, the decay channel of interest is called a signal. Other channels are collectively called background. The channel of interest, in this case, is the $tbH^+ \to tbWh \to tbW\tau\tau$ decay process. The task is therefore a binary classification into signal and background classes. Since the mass of the charged Higgs boson is not known, it is sampled at multiple values. The original analysis worked with 4 mass points – 300 GeV, 800 GeV, 1500 GeV, and 2000 GeV charged Higgs boson [5]. In this thesis, the performance on additional files for three different masses is also evaluated – 250 GeV, 800 GeV, and 3000 GeV.

The background consists of decay processes, which are not easily separated from the signal. These processes include the $t\bar{t}h$ process (figure 2.3), which bears great resemblance to the tbH^+ process. The separation is made harder by the fact, that only some of the final products of the decay are observed by the detector, as the ATLAS detector is not able to detect neutrinos. Other background processes, used in this analysis, are $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$, VV and Others. The same processes were used in the original analysis by Jiří Pospíšil [5].

2.2.3 Weights

The data consist of events. An event is a snapshot of a collision in the Large Hadron Collider (LHC) [6]. The distribution of events in the simulated data does not match the real-world distribution. Weights have to be applied to the simulated data – each event weight is the number of times the event would happen in the real data. The event weight is computed by equation 2.2, where w is the weight of the event, `RunYear` is the year the data were simulated for and the real data were measured in (the year corresponding to an apparatus configuration). The value of x_0 is the LHC luminosity at the time. The `cross_section` is known for background processes, but only estimated for the tbH^+ process. When the cross-section is said to be scaled to some value, for example, 0.1pb, x_6 acts as a free parameter and is set to this value (in this case $x_6 = 0.1$). The source values of x_i – that is `RunYear`, `custTrigSF_LooseID_FCLooseIso_DLT`, `weight_pileup`, `bTagSF_weight_DL1r_85`, `weight_mc`, `lep_SF_CombinedTight_0`, `lep_SF_CombinedTight_1`, `cross_section`, `jvtSF_customOR`, along with `lepSF_PLIV_Prompt_0`, `lepSF_PLIV_Prompt_1` and `totalEventsWeighted` – are available in the simulated data for each event. The weight equation sometimes produces negative-weighted

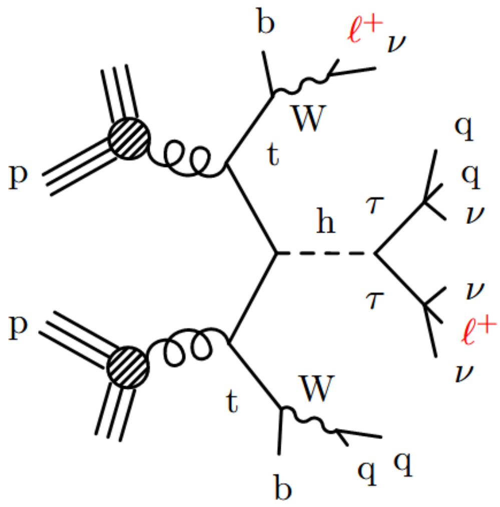


Figure 2.3 $t\bar{t}h$ Feynman diagram from analysis by Jiří Pospíšil [5]

events, as an artifact of the simulation. By the consensus of the 2lSS1tau research group, such events are removed from the training set and left in the validation and testing sets. This is done to minimize the impact of the negative-weighted events on training. The negative weighted events are left in the validation and testing sets because they are needed to get the expected number of events in the real data. This weight equation is used for version 8 of the simulated data. A similar equation, but without x_9 (`lepSF_PLIV_Prompt_0`) and x_{10} (`lepSF_PLIV_Prompt_1`) and with less strict `bTagSF_weight_DL1r_70` used instead of `bTagSF_weight_DL1r_85` for x_4 was used for an older version of the simulated data in the original thesis [5].

$$\begin{gathered}
w = \frac{\prod_{i=0}^{10} x_i}{x_{11}}, \\
x_0 = \begin{cases} 36646.74 & \text{iffRunYear} \in \{2015, 2016\} \\ 44630.6 & \text{iffRunYear} = 2017 \\ 58791.6 & \text{iffRunYear} = 2018 \end{cases} \\
x_1 = \text{custTrigSF_LooseID_FCLooseIso_DLT} \\
x_2 = \text{weight_pileup} \\
x_3 = \text{jvtSF_customOR} \\
x_4 = \text{bTagSF_weight_DL1r_85} \\
x_5 = \text{weight_mc} \\
x_6 = \text{cross_section} \\
x_7 = \text{lep_SF_CombinedTight_0} \\
x_8 = \text{lep_SF_CombinedTight_1} \\
x_9 = \text{lepSF_PLIV_Prompt_0} \\
x_{10} = \text{lepSF_PLIV_Prompt_1} \\
x_{11} = \text{totalEventsWeighted}
\end{gathered} \tag{2.2}$$

2.2.4 Evaluation Metrics

The quality of the separation is evaluated based on accuracy, statistical significance, and sensitivity. In the calculation of all these metrics, the expected number of events is used (weighted w from equation 2.2). The accuracy is the fraction of correctly classified weighted events out of all the weighted events. The significance is part of hypothesis testing, which is used at CERN to establish discoveries or exclusions. The null hypothesis H_0, in this case, is the Standard Model (the charged Higgs boson does not exist) and the alternate hypothesis H_A is the existence of the charged Higgs boson. The statistical test is defined by [7] as q_0 in equation 2.3 and its relation to significance Z in equation 2.5. L is profile-likelihood, $\hat{\mu}$ is the parameter of interest, $\hat{\theta}$ and $\hat{\hat{\theta}}$ are the nuisance parameters, p_0 is the p-value of the test. The unit of significance is σ. The significance of 5σ is needed to establish a discovery and a 2σ corresponds to 95% confidence level (CL) for exclusions. Under the assumption that the Higgs boson detections are independent and that they are from a Poisson distribution, the significance can be approximated by equation 2.9, where $S = w_s \cdot \mathrm{TP}$ is the weighted number of true positive events and $B = w_s \cdot \mathrm{FP}$ is the weighted number of false positive events. The w_s parameter is the event scaled weight, computed by simple equation 2.4, where f is the scaling factor. The scaling factor scales the events of a subset of the dataset to the weights of the full dataset. The scaling factor is computed as $f = |\mathrm{data}|/|\mathrm{data\ subset}|$, so if the testing set is 20% of the data, the scaling factor $f = 5$. If $S \ll B$, the equation can be simplified to equation 2.6. Equations 2.7 and 2.8 are alternatives if B is close to zero. All these equations are used for evaluation of the previous analysis by Jiří Pospíšil [5], the naming is the same as in his thesis for easier reference.

$$q_0 = \begin{cases} -2\frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})} & \text{if} \quad \hat{\mu} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

$$w_s = f \cdot w \tag{2.4}$$

$$Z = \sqrt{q_0} = \Phi^{-1}(1 - p_0) \tag{2.5}$$

$$Z_0 = \frac{S}{\sqrt{B}} \tag{2.6}$$

$$Z_1 = \frac{S}{\sqrt{S + B}} \tag{2.7}$$

$$Z_2 = \frac{S}{\sqrt{B} + 1.5} \tag{2.8}$$

$$Z_3 = \sqrt{((S + B) \cdot \log(1 + \frac{S}{B}) - S)} \tag{2.9}$$

2.3 Used Technologies

2.3.1 Service for Web based ANalysis (SWAN)

SWAN is a cloud data analysis platform. SWAN follows the trend towards *web-based interactive analysis*, where the user interacts with a web-based service, instead of an installed software. The platform uses a browser-based Jupyter Notebook interface with access to the CERN storage, CERNBox, to greatly simplify the access of users to CERN data and computational resources using the web-based “software as a service” provisioning model. After the users enter the CERN account credentials on the authentication page, they are given access to the SWAN cloud environment. There, they can access their CERNBox storage, where they can create Jupyter Notebooks, perform complex data analysis and share the results, as well as the models, with other researchers [8].

2.3.2 ROOT Library

The ROOT system is a framework for large-scale data analysis. ROOT provides a basic set of tools for data acquisition, detector simulation, event generation, and data analysis. ROOT is written in C++ and provides an effective way to store objects in a tree hierarchy, load them, analyze them, and visualize the results [9, 10].

2.3.3 Optuna

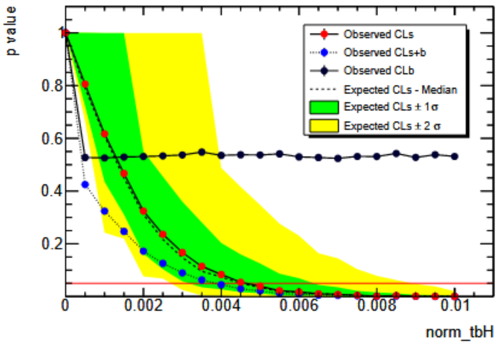
Optuna is a next-generation optimisation framework. Its architecture allows *define-by-run* programming, which allows the users to dynamically construct the search space. It is easy to setup and deploy for tasks ranging from light-weight experiments to heavy-weight distributed computations. Optuna has an efficient sampling and pruning algorithm, allowing efficient automatic hyperparameter tuning. The optimisation framework features an independent sampling algorithm TPE, as well as a relational sampling method CMA-ES. Pruning stops unpromising trials for better cost-effectiveness. Optuna uses a pruning algorithm based on Successive Halving [11].

2.4 Previous Analyses

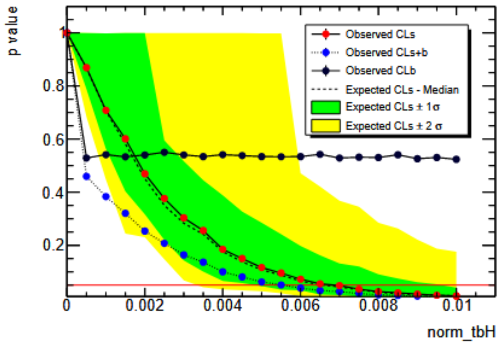
The thesis by Jiří Pospíšil used two model types − MLP and TabNet. The MLP architecture consisted of a series of blocks, each with its own linear layer, activation function, and dropout layer (placed in this order). The sub-type MLPs additionally had feed-forward shortcuts inspired by their usage in CNN architectures.

TabNet, an attentive transformer architecture for tabular data, is the second model type used in the previous thesis on this charged Higgs boson search channel. The best model of the thesis was an MLP with a weighted loss function with event weights with $Z_0 \doteq 16.6$ and $Z_1 \doteq 4.1$ [5].

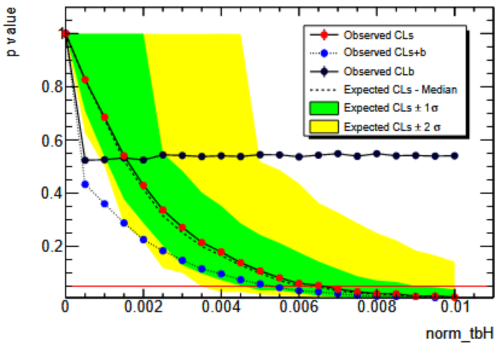
The later evaluation by Niklas Düser used only the MLP neural network model type with the 10 most important features, according to permutation feature importance. His CERN summer student report additionally evaluated the model output using Toy Model Monte Carlo to get the expected upper limit at 95% confidence level on the cross-section. The results for each of the four signal masses are shown in figure 2.4 [12].



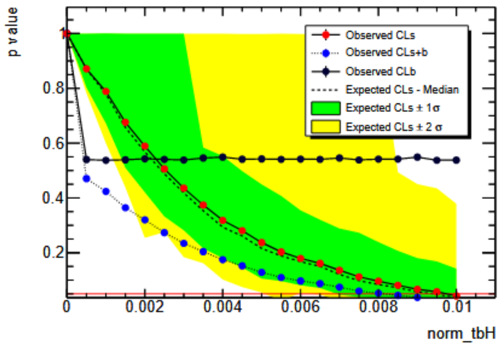
(a) P-value from Toy Model for a 300 GeV signal.



(b) P-value from Toy Model for a 800 GeV signal.



(c) P-value from Toy Model for a 1500 GeV signal.



(d) P-value from Toy Model for a 2000 GeV signal.

Figure 2.4 The 95% CL limit on the cross-section is set where the CLs curve crosses the 0.05 horizontal line (norm_tbH marks the cross-section in pb), plots taken from the report by Niklas Düser [12]

Chapter 3

Analysis and Design

3.1 Input Data

3.1.1 Data Format

CERN uses the ROOT software framework for data analysis and I/O operations (using standardized .root data files). The framework has a C++ interpreter and interface to Python [13]. The dataset consists of several .root files, each containing just one process. Each file is identified by a process DSID in its name. One DSID can belong to only one process, but one process can consist of multiple DSIDs. Full mapping between files and processes is listed in table 3.1.

tbH_300	510374_AF;
tbH_800	510375_AF;
tbH_1500	510376_AF;
tbH_2000	510377_AF;
tbH_250_new	512185;
tbH_800_new	512187;
tbH_3000_new	512186;
ttH	346343, 346344, 346345;
tt	410470;
ttW	700168, 700205;
ttZ	700309;
VV	364250, 364253, 364254, 364255, 364283, 364284, 364285, 364286, 364287, 363355, 363356, 363357, 363358, 363359, 363360, 363489;
Others	410560, 410408, 410646, 410647, 304014, 410080, 345705, 345706, 345715, 345718, 345723, 364242, 364243, 364244, 364245, 364246, 364247, 364248, 364249, 342284, 342285, 410081, 346799_AF, 346678_AF;

Table 3.1 Mapping between DSID and processes

There are multiple datasets. This thesis focuses on version 8 of simulated data for multilepton-ttW-ttH, further referred to as “v8 data”. The v8 data was produced during the creation of this thesis and made available to the team. The majority of the experiments in this thesis are done on v8 data – when a version of data is not specified in the context, the experiment uses v8 data.

The previous thesis [5] and summer student report [12] used data without an assigned version, created as a combination of existing simulated background data files, and added simulated signal files. This data will be further referred to as “v5 data”. This version of the data will be analyzed to better understand the previous results and to evaluate any changes in the model performance between v5 and v8 based on data quality. There is also version 6 of simulated data (“v6 data”), which was used for a similar analysis of $t\bar{t}H$ (where $t\bar{t}H$ is the signal process, the rest of the background processes stays the same) [14].

3.1.2 Features

Both the v5 data and the v8 data contain a high number of features. Not all of the features can be used by the model. There are also features detailing the settings of the generator and the truth value of the process. Such features must not be used for prediction, so only a subset of features is loaded from the dataset. These features are then preprocessed, creating a few derived features in the process.

In general, the preselected data contains events with information about two leptons, one hadronically decaying tau, and a various number of jets. The leptons have the same sign, meaning either are both particles or antiparticles. Events, where one lepton is a particle and one an antiparticle are not used (per the directions for the 2lSS1tau group). Leptons and jets have assigned indexes based on their transverse momentum. The particle/jet with the highest transverse momentum is said to be “leading” [15] (index 0), with the second highest “subleading” (index 1), and so on. Transverse momentum index is written as part of the feature name to identify the particle/jet (along with “lep” – meaning lepton, “taus” – meaning the tau, “jet” – meaning jet). Below are feature explanations, grouped by their assigned normalization. Imputation of the missing values is done before normalization.

3.1.2.1 Features with Imputed Values

Features `taus_passJVT_0`, `taus_width_0` and `taus_DL1r_0` represent if the tau passed jet vertex tagging, the width of the tau and the output of the DL1r algorithm (a jet is considered b-tagged if its DL1r score is above a certain threshold) [16] for the tau, respectively. All of these features have missing values (-99, -2, and -999, respectively) – roughly 2% of the data are missing. Imputation of these values is selected to not lose the data.

The imputation is done separately for the categorical feature, `taus_passJVT_0`, and for both continuous features (`taus_width_0` and `taus_DL1r_0`). The categorical feature is imputed using a decision tree classifier. The continuous features are imputed using a decision tree regressor. The models are fitted on known values of the imputed feature and then predict the unknown values. The prediction is based on the rest of the features as the input values.

3.1.2.2 Features without Normalization

This section lists basic features and derived features, for which the normalization is not necessary, mostly because the features are already in the range $\langle -1, 1\rangle$, range $\langle 0, 1\rangle$, or they are indicators with values $\in \{0, 1\}$.

The feature `lep_ID_0` indicates whether the leading lepton is electron (abs(`lep_ID_0`) $= 11$) or muon (abs(`lep_ID_0`) $= 13$), as well as if it is a particle (`lep_ID_0` > 0) or an antiparticle (`lep_ID_0` < 0). A similar feature exists for the subleading lepton – `lep_ID_1`. Since both leptons must be of the same sign, the features provide redundant information. To remove the redundancy, the features are split into three binary features – `lep_0_is_muon` – an indicator if the leading lepton is a muon, `lep_1_is_muon` – an indicator if the subleading lepton is a muon and `lep_tau_opposite_charge` – an indicator if both leptons have opposite charge relative to the tau. The two original features `lep_ID_0` and `lep_ID_1` are removed. The feature `total_charge` is also removed, because it contains the sum of charges of both light leptons (tau charge is not included), which strictly depends on if both leptons have an opposite charge, relative to the tau – it, therefore, provides duplicate information and it is left out.

Feature name	Unique values	Min. value	Max. value	Explanation
lep_0_is_muon	2	0	1	Indicator of leading lepton being a muon
lep_1_is_muon	2	0	1	Indicator of subleading lepton being a muon
lep_tau_opposite_charge	2	0	1	Indicator of both leptons having opposite charges relative to the tau
taus_decayMode_[value]	2	0	1	Indicator of tau decay mode belonging to [value] class
nTaus_OR_Pt25	2	0	1	Number of taus with at least 25 GeV transverse momentum
taus_passJVT_0	2	0	1	Indicator of tau passing jet vertex tagging
taus_RNNJetScoreSigTrans_0	-	0.25	1.00	Output of RNN tau identification
taus_JetRNNSigTight_0	2	0	1	Output of RNN tau identification for tight (highest background rejection working point)
taus_charge_0	2	-1	1	Charge of the tau
taus_fromPV_0	2	0	1	Indicator of the tau originating in primary vertex
taus_passEleOLR_0	2	0	1	Indicator of tau not overlapping with a good electron reconstruction

Table 3.2 List of features with no needed normalization (after missing value imputation)

The categorical feature `taus_decayMode_0` distinguishes the decay modes of the tau. Since the decay modes do not have a simple meaningful order in the context of this analysis, one-hot encoding on the feature is performed (replacing the feature with indicators of each of the decay modes happening – the new features are named `taus_decayMode_[value]`, where `[value]` is one of the possible values of the original feature).

Table 3.2 shows the full list of features without further normalization. The binary feature `nTaus_OR_Pt25` represents the number of taus with at least 25 GeV transverse momentum [5]. It can attain only two values because the number of taus is set to 1 (and only some pass the transverse momentum threshold). Another four features – `taus_passJVT_0`, `taus_RNNJetScoreSigTrans_0`, `taus_JetRNNSigTight_0` and `taus_passEleOLR_0` – belong to the tau reconstruction (tau is not directly measured, measured are only the products of its decay). The use of these features incorporates the certainty of the tau reconstruction into the tbH^+ classifier. The feature `taus_fromPV_0` indicates if the tau originated at the point at which a proton-proton (pp) interaction occurred (primary vertex) [17]. Along with `taus_charge_0`, there is no need for normalization, as these features are already in the acceptable ranges and further normalization might impact their interpretability.

3.1.2.3 Features with Logarithm Z-score Normalization

Features, which are listed below, are assumed to have a log-normal or similarly skewed distribution.A logarithm is applied to these features, before using the z-score normalization, to get closer to the normal distribution of the features and to get more even distribution of feature values.

Features representing measurements of transverse momentum (p_T) of particles/jets often have distribution close to log-normal (as can be seen in figure 3.1). These features are `jet_pt[id]`, which belong to jets with $(\mathrm{id}+1)$ highest p_T in an event. They are created from items at index `id` of a vector feature `jets_pt` after preselection. The vector feature is removed afterward, as it almost never contains more than six items after the preselection cut. The preselection requires the first three highest to be non-zero (the jet to be present). A similar approach is used for jet energy features (`jet_e[id]`). Log-z-score is also used for transverse momentum and energy of all three particles (features `lep_E_0`, `lep_E_1`, `lep_Pt_0`, `lep_Pt_1`, `taus_pt_0`), as well as p_T sums (`HT_inclFwdJets`, `HT_fwdJets`, `HT_jets`, `HT`, `HT_lep`).

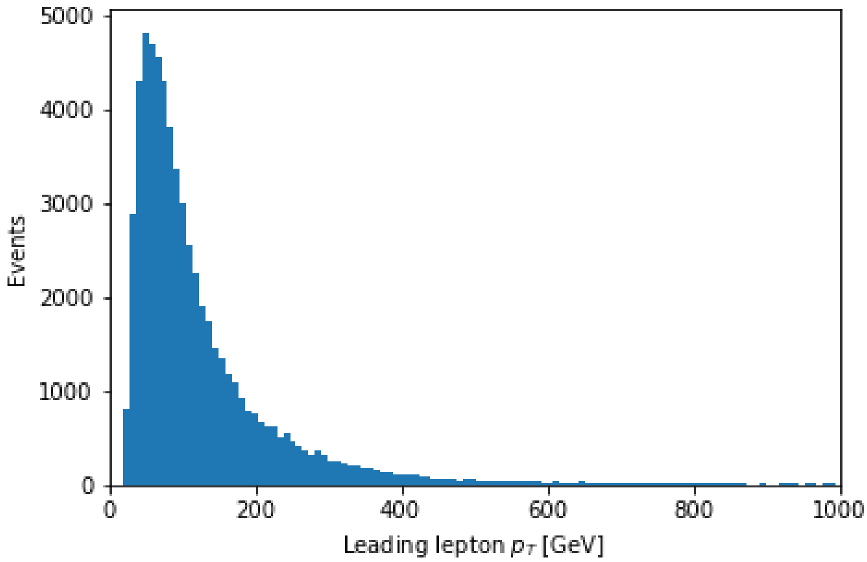


Figure 3.1 Histogram of the leading lepton transverse momentum

Feature name	Description
eta_frwdjet	Pseudorapidity of the forward jet
lep_Eta_0	Pseudorapidity of the leading lepton
lep_Eta_1	Pseudorapidity of the subleading lepton
taus_eta_0	Pseudorapidity of the tau
lep_EtaBE2_0	–
lep_EtaBE2_1	–
lep_Z0SinTheta_0	Longitudinal impact parameter of the leading lepton
lep_Z0SinTheta_1	Longitudinal impact parameter of the subleading lepton
jet_eta[id]	Pseudorapidity of `id`-th jet

Table 3.3 List of features with normalization of standard deviation

3.1.2.4 Features with Z-score Normalization

The features with z-score normalization follow. These features are `taus_DL1r_0`, `taus_width_0` (both features with imputed values were already explained previously), `minDeltaR_LJ_0`, feature `minDeltaR_LJ_1` and `minDeltaR_LJ_2` have the values of minimum ΔR, as defined by equation 2.1. Also undergoing this normalization are features `max_eta`, `dEta_maxMjj_frwdjet`, `DRll01`, `nFwdJets_OR`, `nJets_OR` (number of jets), `sumPsbtag` (sum of jet b-tags), leading and subleading lepton track multiplicity (`lep_nTrackParticles_0` and `lep_nTrackParticles_1`) and `taus_numTrack_0`.

3.1.2.5 Features with Standard Deviation Normalization

These features (table 3.3) are normalized only by division by the standard deviation (mean is assumed to be zero or the feature value distribution is mostly symmetric around $x = 0$). This type of normalization applies to pseudorapidity feature (η), namely `lep_Eta_0`, `lep_Eta_1` and `taus_eta_0`, as well as `jet_eta[id]` created from `jets_eta` vector feature. The normalization is also applied to features `eta_frwdjet`, `lep_EtaBE2_0` and `lep_EtaBE2_1`.

Features `lep_Z0SinTheta_0` and `lep_Z0SinTheta_1` refer to the longitudinal impact param-

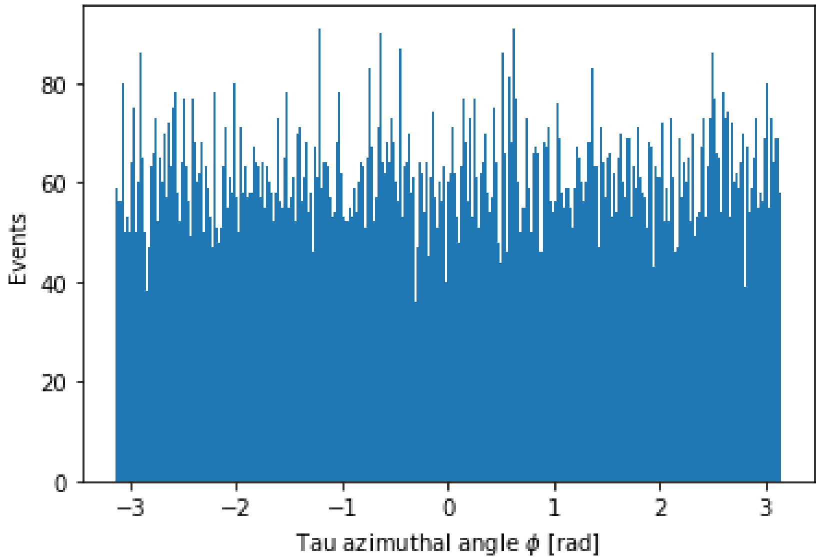


Figure 3.2 Azimuthal angle of the tau

Feature name	Explanation
lep_Phi_0	Azimuthal angle of the leading lepton
lep_Phi_1	Azimuthal angle of the subleading lepton
met_phi	Azimuthal angle of the missing transverse energy
taus_phi_0	Azimuthal angle of the tau
jet_phi[id]	Azimuthal angle of the `id`-th jet

Table 3.4 Feature explanation for the features with angle normalization

eter of the leading and subleading leptons, respectively. This parameter is defined as the distance of the track to the primary vertex in the longitudinal plane at the point of closest approach in r-ϕ. Due to the long lifetime of b-hadrons, tracks generated from b-hadron decay products tend to have large impact parameters enabling their contribution to be separated from the contribution of tracks from the primary vertex [18].

3.1.2.6 Features with Angle Normalization

Azimuthal angle features (table 3.4) have an almost uniform distribution in the range $(-\pi, \pi)$ (figure 3.2). For the purpose of normalization, they are simply divided by π. Features `jet_phi[id]` are once again extracted from vector feature `jets_phi`, before being normalized.

3.1.2.7 Removed Features

The features are deleted for two reasons – because the feature has only one value after preselection (group 1) or because the feature has only redundant information (group 2 – including the source features for one-hot encoding and similar transformations). Both lists of removed features are given in two groups:

Group 1 best_Z_Mll, nTaus_OR, minOSMll, minOSSFMll, Mlll012, Mllll0123, total_leptons, taus_JetRNNSigLoose_0, taus_passEleBDT_0, best_Z_other_Mll, taus_JetRNNSigMedium_0, best_Z_other_MtLepMet

New	Campaign	Mass (GeV)	Filter Eff. (%)	Created	Ntuple	Preselected	Ratio (‰)	Eff. (‰)
Yes	MC16a	250	32.14	300000	24027	1110	3.70	1.19
Yes	MC16a	800	40.36	300000	34138	2761	9.20	3.71
Yes	MC16a	3000	49.57	300000	16378	552	1.84	0.91
Yes	MC16d	250	32.14	398000	29629	1340	3.37	1.08
Yes	MC16d	800	40.36	399000	43746	3372	8.45	3.41
Yes	MC16d	3000	49.57	399000	21449	705	1.77	0.88
Yes	MC16e	250	32.14	499000	37380	1718	3.44	1.11
Yes	MC16e	800	40.36	498000	54246	4316	8.67	3.50
Yes	MC16e	3000	49.57	499000	26631	851	1.71	0.85
No	MC16e	300	59.57	1200000	74423	2595	2.16	1.29
No	MC16e	800	67.73	800000	72027	4035	5.04	3.42
No	MC16e	1500	73.49	600000	48461	2523	4.20	3.09
No	MC16e	2000	75.71	400000	25334	972	2.43	1.84

■ **Table 3.5** Table of preselection efficiencies. The efficiency is the product of filter efficiency and ratio.

Group 2 dilep_type, lep_ID_0, lep_ID_1, taus_decayMode_0, total_charge

3.1.3 Preselection

The preselection pre-filters the data, leaving only the events meeting the preselection condition. The preselection selects events with 2 leptons of the same sign and 1 hadronically decaying tau (as per the directions of the 2lSS1tau group − using such preselection cuts, each such group works on a separate dataset). Additionally, the preselection removes events with fewer than 4 jets and removes irrelevant events by additional conditions. The full preselection formula is noted in appendix A.1.

It should be noted that this is not the only preselection the data go through. The simulated data are first filtered during the event generation and then again before being provided to the working group in the “Ntuple” format. The counts of the data are listed in table 3.5. The column `New` tells if the data are part of newly simulated signal masses (these data were not available for the previous thesis and summer student report). Sub-campaign matches a period of recording of the real data, which is being simulated. Sub-campaign `MC16a` matches years 2015 and 2016, sub-campaign `MC16d` matches year 2017 and `MC16e` matches year 2018 [19]. The mass refers to the mass of the charged Higgs boson. Filter efficiency refers to the fraction of selected events out of all generated events (the output of the simulation is therefore already filtered). The filter criterion in the first signal simulation was the presence of one light lepton, and in the new signal simulation, two light leptons were required. Created events are events after this first filtering. The events are then filtered again, and the resulting Ntuples are provided for data analysis. The `Ntuple` column shows the full size of the signal part of the v8 dataset. The preselection, described in the first paragraph, is then applied to the data, leaving the `Preselected` number of events. The last column shows the ratio between `Created` and `Preselected` events.

3.1.4 Checking Linear Separability

The purpose of the linear separability check is to ensure the consistency of the signal and background event generation. Any linear separability found in the dataset must therefore be looked into to determine its cause. The check is done on only two sets of data − training set and validation set, having 70% and 30% of the data, respectively. The data for the sets is selected pseudo-randomly. The analysis uses a simple binary Linear SVM classifier model with $C = 1$ (not expecting a perfect separability). Both v5 and v8 data are checked. Firstly, the test is performed with just 9 most important features according to feature ranking in the previous thesis

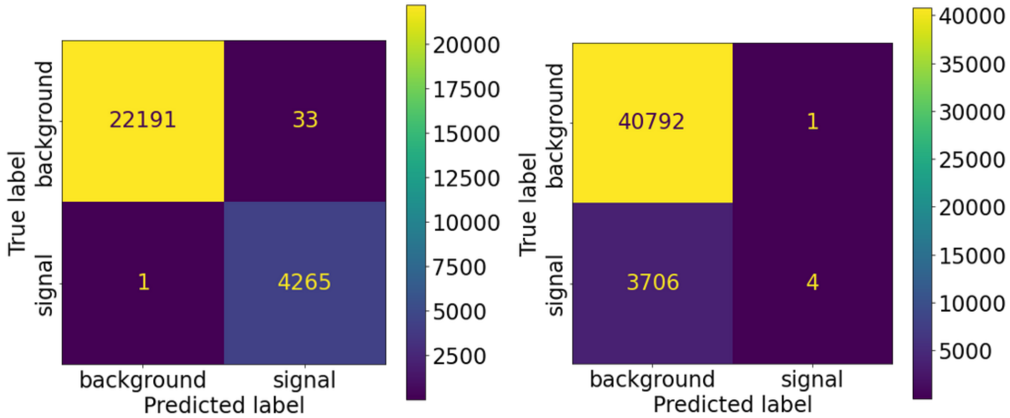


Figure 3.3 Confusion matrix for v5 data (left) and v8 data (right)

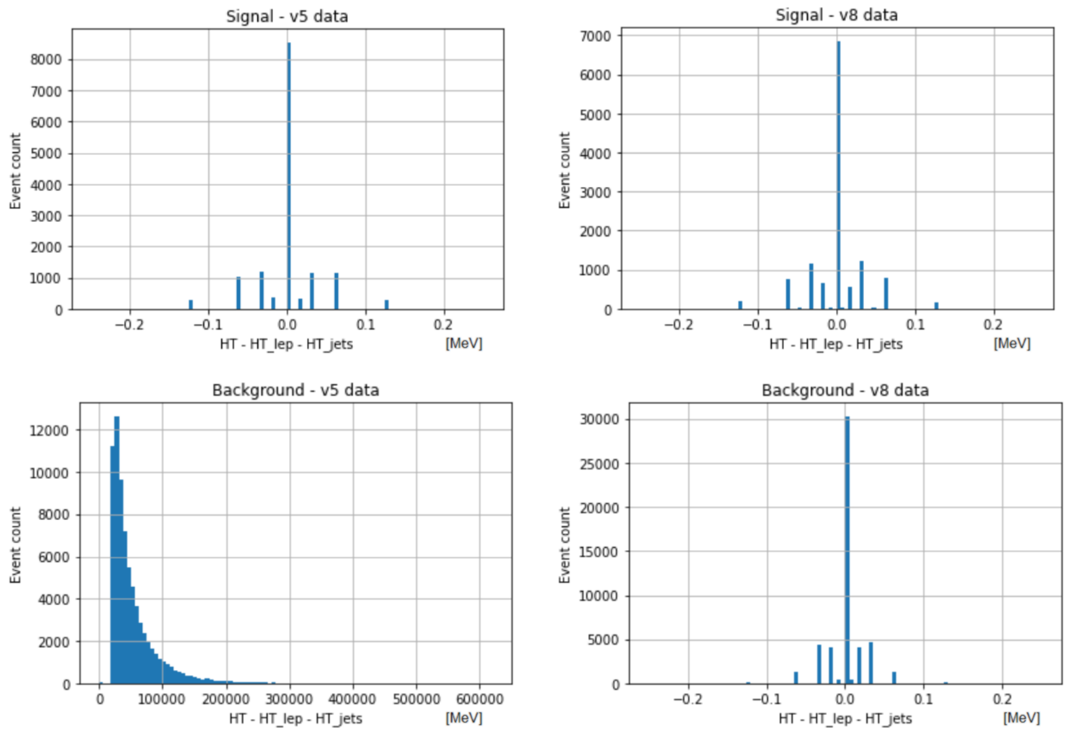
by Jiří Pospíšil [5] and CERN summer student report by Niklas Düser [12] – the features being `nTaus_OR_Pt25`, `jet_pt0`, `lep_Pt_1`, `HT_fwdJets`, `HT_lep`, `lep_Pt_0`, `HT_inclFwdJets`, `HT` and `HT_jets`. The results in figure 3.3 show that v5 data is almost perfectly linearly separable, while v8 data is not. The quality of the linear separation for the v5 data indicates a likely difference in signal and background file generation.

$$\begin{aligned} ax + by + cz + d &= 0 \\ a &\doteq 0.00023257 \\ b &\doteq -0.00022694 \\ c &\doteq 0.00022889 \\ d &\doteq 2.65083051 \cdot 10^{-7} \end{aligned} \tag{3.1}$$

Next, each of the features is tested, if it is necessary for linear separation. The testing is done by fitting the SVM model on the training set with a reduced number of features and looking at its performance on the validation set with the matching set of features. Using this method, features `nTaus_OR_Pt25`, `jet_pt0`, `HT_inclFwdJets`, `lep_Pt_0`, `HT_fwdJets` and `lep_Pt_1` have been removed. The model separates the remaining features (`HT_lep`, `HT`, `HT_jets`) using a plane described by equation 3.1.4, where d is the model intercept and a, b, c are the coefficients of the model (for features `HT_lep` $= x$, `HT` $= y$, `HT_jets` $= z$).

Based on this information, the relation of the features has been simplified by approximating the equation 3.1.4 by equation 3.2, where d' is the new intercept, which can be expressed as d' = `HT` − `HT_lep` − `HT_jets`. To visualize the linear separability, two histograms are plotted for v5 data and two histograms for v8 data for comparison (figure 3.4). The histograms show values `HT` − `HT_lep` − `HT_jets` – one for signal events and one for background events. The v5 histograms indicate, that for signal, feature `HT` is the sum of features `HT_lep` and `HT_jets`, while for background, feature `HT` consists of one more additional part. This discrepancy in feature values is likely caused by separate signal and background data generation, with the more complicated feature dependency getting through the previous data quality checks. The data generation problem is not present in the new v8 data production, so no action had to be taken in this regard. Technically, as the signal Ntuples are produced after the background Ntuples, likely in the ntuple production code the `HT` definition changed for the signal production.

While no action is needed for this analysis with v8 data, the issue in v5 data has a clear impact on the previous thesis [5] and the summer student report [12].



■ **Figure 3.4** Histogram of `HT - HT_lep - HT_jets` for v5 data (left) and v8 data (right)

$$y - x - z - d' = 0 \tag{3.2}$$

3.2 Architecture of Used Machine Learning Models

3.2.1 Multilayer Perceptron

The multilayer perceptron (MLP) neural network is an often-used machine learning algorithm. Given its high flexibility in terms of model complexity and usage in the previous thesis on the subject, it is also selected to be one of the models used in this thesis.

The Stochastic Gradient Descent (SGD) has been selected as the optimizer of the weights of the network, based on paper [20], which evaluated the Stochastic Gradient Method and the Stochastic Gradient Descent. It showed that in a broad range of settings, if the number of iterations is linear, the generalization error is bounded by a vanishing function of the sample size. Going by this principle, an early stopping criterion is set – ending the model fitting if the validation loss of the model does not improve for 20 epochs.

The output layer of the network has one neuron with a sigmoid activation function, which predicts if a given sample belongs to the signal class. The binary cross-entropy has been chosen as the loss function for the network. The full schema of a five-hidden-layer network is shown in figure 3.5. The number of hidden layers will be one of the optimized hyperparameters. The activation function of the hidden layers in the figure is Rectified Linear Unit (ReLU), though, in the model selection experiment, a sigmoid activation function is sometimes evaluated in place of ReLU. Other notable hyperparameters of an MLP model are – batch size, dropout rate, and learning rate.

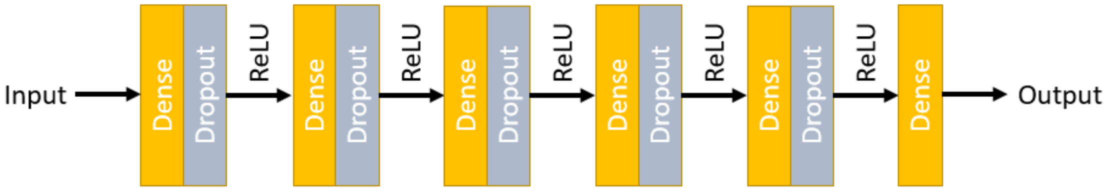


Figure 3.5 Network architecture with 5 hidden layers

3.2.2 Support Vector Machine

The Support Vector Machine binary classifier uses a hyperplane to separate the events into two classes. The model can efficiently separate high-dimensional spaces by only focusing on the data points closest to the separation boundary – the so-called support vectors. The model can use different kernels to transform the feature space before performing the separation. Notable is the linear kernel and RBF kernel. The first leaves the features without change, allowing the model to perform a linear separation, and the second transforms the feature space to infinite dimensions, which is achievable thanks to the kernel trick [21].

Linear separation is necessary to check that no further leak of truth information into the data has taken place. The previous model performance during data quality checks (figure 3.3) shows no leak for the v8 data, but a final check is necessary on the preprocessed data. The infinite-dimensional feature space is a unique property of the SVM classifier among the chosen models.

The model hyperparameters are `C` – the penalization constant for incorrect classification. Small values of `C` lead to a greater margin between the hyperplane and data, possibly attaining greater generalization power. Higher values of `C` lead to a smaller-margin hyperplane, which better separates the training data [22],

Some implementations of the SVM also allow the model to output class probability instead of binary output. In particular, the scikit-learn uses probabilities calibrated by Platt scaling. The model uses logistic regression on the scores of the SVM, fit by additional cross-validation on training data [23] [24].

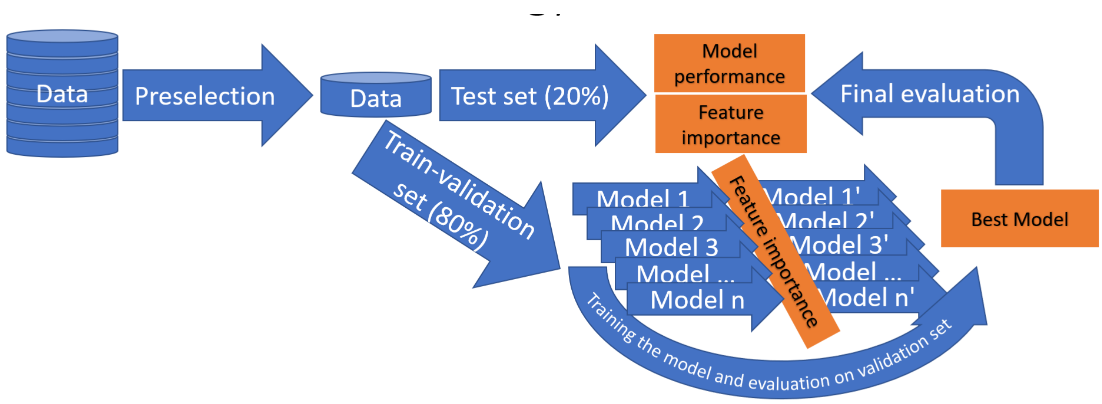
3.2.3 Decision Tree

The decision tree model has a large advantage in its explainability. The tree consists of binary decisions, located in a tree hierarchy and traversed top-to-bottom (from root to leaves). The decision tree can also be used for feature importance ranking (later referred to as tree feature ranking).

The decision tree is created using the greedy CART (Classification and Regression Trees) algorithm, which constructs binary trees according to inequality condition-based node splitting. The algorithm either selects the best split out of `max_features`, or the best split out of several random splits. The algorithm uses either entropy or the Gini index to evaluate split quality. Other hyperparameters include maximum tree depth and minimum number of samples in a leaf. Each leaf is assigned a probability based on the fraction of expected signal events in it [23].

3.2.4 Random Forest

The random forest classifier is an ensemble model, which uses multiple decision trees to improve its performance. The random forest predicts the probability that a given event belongs to the signal class, based on the mean of the outputs of its trees. As such, the model has all



■ **Figure 3.6** Model Evaluation Pipeline

the hyperparameters of the decision tree model with additional hyperparameters related to the ensemble, such as the number of trees in the forest [23].

The random forest, by default, uses the bootstrap technique to construct the training datasets for its trees. Bootstrap is a sampling technique with replacement, which increases the ability of the model to learn different aspects of the dataset and further improves the performance of the model [23].

3.3 Feature Importance Ranking

There are two main types of feature importance ranking – the first is based on the internal structure of the model and is available only for some models – in our case mainly just for decision trees and random forests. The second is based on feature information removal through permutation. Permutation feature ranking is defined as the decrease in model performance if a feature is randomly shuffled. Given that different random shuffles can lead to different decreases in model performance, the mean and standard deviation of the feature importance can be also computed [23].

3.4 Model Type Selection

For this analysis, four different types of models have been selected – decision tree, random forest, support vector machine (SVM), and multilayer perceptron (MLP). An evaluation pipeline has been built to measure the performance of each of the models in a controlled environment.

The pipeline, shown in figure 3.6, leads the data files through preselection into a single dataset. During preprocessing, weights are computed for the events, according to equation 2.2. The signal cross-section is set to 0.1 pb The dataset is then split into the train, validation, and test sets in the ratio 64:16:20. The train set is used to train the models, the validation set is used to compare the models and the test set is used for the final evaluation of the best model performance. Since there are multiple possible masses of the charged Higgs boson and the real charged Higgs boson would have only one mass, a separate model is trained for each charged Higgs boson mass. An 800 GeV mass is a special case, since – for comparison purposes – additional files were generated for it. To test the differences between the original and the new 800 GeV data, two models are trained – one for the original 800 GeV files and one for the new 800 GeV files.

The model comparison trains 4 model types, each having multiple model instances for different hyperparameters, for each of the charged Higgs boson masses (and two models for 800 GeV mass). The models are then compared based on the significance of their output on the validation set.

Based on the significance, the best model is selected for each mass. This defines the seven best models.

3.5 Feature Reduction

After selecting the seven models, which are best suited for signal and background separation at the given masses, the permutation feature importance is evaluated for each of the seven models. Based on the result, the 5, 10, and 20 most important features are selected for each model and features with large correlation are removed (replaced with the next most important feature for the model). Finally, new models are trained on the reduced feature datasets and their significance is compared on the validation set. Based on the results, the best-performing model in terms of significance for each charged Higgs boson mass (two for 800 GeV) is selected to be used in the final evaluation.

3.6 Best Model Evaluation

The best-performing model for each mass is evaluated using the TRExFitter (Top Related Experiment Fitter), a framework for binned template profile likelihood fits [25]. For the evaluation in TRExFitter, the signal cross-section is scaled to 1 pb and multiplied by filter efficiency (table 3.5). The correction by filter efficiency is done to take into account the different preselection efficiencies caused by stricter event generator filtering of the new charged Higgs boson files. TRExFitter is used to display histograms for the 10 most important features of the best models. A histogram of the model outputs of each of the best models will be analyzed. Additionally, plots of the upper expected limit at the 95% confidence level (CL) on the cross-section will be created.

The expected upper limit is the upper limit at 95% CL of the signal cross-section one could obtain if the background-only hypothesis is true. It is therefore important, that it is lower than the actual signal cross-section (in this case scaled to 1 pb for TRExFitter evaluation). There are two main methods of computation – the Asymptotics method and the Toy Monte Carlo method [25]. The methods differ in the computation strategy – the Asymptotics method directly computes integrals, while the Toy Monte Carlo uses an ensemble of pseudo-experiments to evaluate the necessary integrals numerically [26].

3.7 Data preprocessing

The data aggregation process, which applies the preselection function (available in appendix A.1) and copies the data from multiple process ROOT files into a single PyTorch `.pt` file is managed using modified utility files from Jiří Pospíšil, created for his thesis [5]. Once the events are aggregated into a single dataset file, they are loaded by the new code, which was created for this thesis – namely the `preprocessing.py` file and the Jupyter Notebook `.ipynb` files.

For the aggregation of the data, the code of the thesis uses the following modified code from Jiří Pospíšil [5]:

utils/root_utils.py functions for loading the `.root` files

utils/file_utils.py utility functions for creating and searching directories and files

utils/dataset_utils.py the definition of the Dataset class, used to save data to a single `.pt` file

utils/utils.py short general utility functions

create_dataset.py loads the data from multiple ROOT files into a single dataset file

Chapter 4

Experiments

4.1 Experiment Setup

All the tests in this chapter are performed in a SWAN environment [8]. The tests were performed using the following software specifications:

- Python 3.9.12
- Numpy, version 1.22.3
- Pandas, version 1.2.2
- TensorFlow and Keras, version 2.8.0
- Scikit-learn, version 0.24.2
- Optuna, version 3.1.1

All the models use sample weights for training, validation, and testing. The training, validation, and testing set have associated scaled weights, computed by equation 2.2 and scaled by equation 2.4 with $f = 0.64^{-1}$ for the training set, $f = 0.16^{-1}$ for the validation set and $f = 0.2^{-1}$ for the testing set. The following sections describe the experiment-dependent setup and results of the experiments.

4.1.1 Model Selection Preparation

As part of the experiment preparation, the models have been compared on the validation set based on the simplified significance formula $Z_0 = S/\sqrt{B}$ (equation 2.6). Results for $B = 0$ have been filtered out as irrelevant since the equation is only suited for $S \ll B$. However, this still led to the selection of models with almost no background events and few signal events. A good example is the best decision tree selected by the Z_0 criterion (whole tree in figure 4.1). The tree was chosen from models created by a grid search in the form later used in section 4.2.2. The model managed to separate approximately 13% of the expected signal events and just a small fraction of a background event (achieved by the background events having decimal weights), getting a high Z_0 score on the 3000 GeV mass. Based on this result, The more precise Z_3 (equation 2.9) significance approximation was selected to compare the models. The result is also a good indicator that the tau transverse momentum (`taus_pt_0`) is an important feature for separating the signal from the background at high H^+ masses.

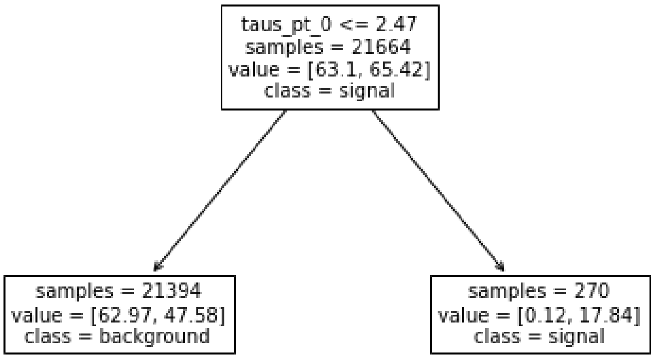


Figure 4.1 Model selected for highest Z_0 score among the decision trees

4.2 Model Selection

For the selection of the best model, the four types of models described previously are used – Support Vector Machine, Decision Tree, Random Forest, and Multilayer Perceptron. Each of the models is optimized on a grid of possible hyperparameter values. Grid search is used to find the best combination, if computationally viable. Otherwise, the Optuna optimizer is employed to search the hyperparameter space.

4.2.1 Suport Vector Machine

Two hyperparameter grids are used for the SVM model – one for the linear kernel and one for the RBF kernel. The SVM with the linear kernel is there to show, how much of the signal class can be linearly separated. Too good separation would be an indicator of a data generation mistake, but a small amount of linear separability is expected (as was already demonstrated by the decision tree in the model selection preparation section 4.1.1). The radial basis function kernel is there to provide transformation into infinite-dimensional space.

Grid for the linear kernel:

- $\mathrm{C} \in \{0.01, 0.1, 1\}$

Grid for the RBF kernel:

- $\mathrm{C} \in \{0.01, 0.1, 1\}$
- $\mathrm{gamma} \in \{\mathrm{scale}, 0.01, 0.001, 0.0001\}$

In this context, gamma value `scale` is defined as $1/(\mathrm{n\_features} \cdot \mathrm{X.var()})$ [23].

The model uses Support Vector Classifier (`SVC`) from scikit-learn with enabled probability prediction. When evaluating the significance, approximation from equation 2.9 is used on thresholded model output. The evaluation function tries 9 different thresholds, sampled evenly at 0.1 intervals, starting at 0.1 and ending at 0.9. The result with the highest significance is returned, along with the threshold. Some thresholds lead to the separation of a small part of the signal and no background (i.e. 0 expected background events). This causes division by 0 in the significance approximation. Those are automatically not considered viable thresholds for the model and are given minimum priority. If all the thresholds lead to no expected background events, the Z_3 score of the model is set to `np.NaN` and the model is not considered during the selection of the best model.

order	mass	Z_3 $[\sigma]$	C	kernel	gamma
1	tbH_300	2.38 (5.21)	0.10	RBF	0.0001
2	tbH_300	2.34	1.00	RBF	0.001
3	tbH_300	2.28	0.10	RBF	0.001
1	tbH_800	8.89 (8.82)	0.10	linear	
2	tbH_800	8.52	1.00	linear	
3	tbH_800	8.44	1.00	RBF	scale
1	tbH_1500	16.63 (14.92)	1.00	linear	
2	tbH_1500	16.52	0.01	linear	
3	tbH_1500	16.13	0.10	linear	
1	tbH_2000	14.91 (17.51)	0.10	linear	
2	tbH_2000	13.23	0.01	linear	
3	tbH_2000	11.13	1.00	RBF	0.001
1	tbH_250_new	5.24 (6.50)	0.10	RBF	0.0001
2	tbH_250_new	5.19	0.01	RBF	0.001
3	tbH_250_new	5.15	0.01	RBF	0.0001
1	tbH_800_new	15.09 (10.57)	0.01	RBF	0.01
2	tbH_800_new	14.96	1.00	RBF	scale
3	tbH_800_new	14.95	0.10	RBF	scale
1	tbH_3000_new	7.33 (15.31)	1.00	linear	
2	tbH_3000_new	7.27	0.10	linear	
3	tbH_3000_new	7.17	0.01	linear	

Table 4.1 Best SVM models and their validation set results with signal cross-section 0.1 pb

The validation set is used for all model evaluations. For each signal mass m (and separately for original and new 800 GeV events), a function creates a shallow copy of the training set with the given signal mass and all background events. This shallow copy of the training set is used for the training of all models in the grid search. A similar approach is used to get a shallow copy of the validation set for model evaluation. In total, the grid search is run 7 times – once for each of the masses (counting original and new 800 GeV separately) – the SVM models are trained for each mass separately and the best model is selected for classification of each mass.

The top 3 models and their results for each of the masses are written in table 4.1. The table shows, that the RBF kernel is best suited for lower signal masses, while the linear kernel performs very well on masses 1500 GeV and above. The models obtain much better results than in the linearity check (see right confusion matrix in figure 3.3). This demonstrates the need for task separation for each of the signal masses – at least in the case of such simple models as the SVM.

Notable is the increase in the significance of the new 800 GeV signal mass. The reason for the discrepancy becomes easily visible when looking at a confusion matrix – see figure 4.2. The different results are caused by the higher weight of the new 800 GeV mass events. While simply multiplying the weights of the original and new 800 GeV might seem like the best thing to do, this would disrupt the normalization to the cross-section of 0.1 pb. The cross-section relates to the frequency of occurrence of a particular process before any filter is applied and therefore does not guarantee the weights in the dataset after preselection will be equal. To deal with the

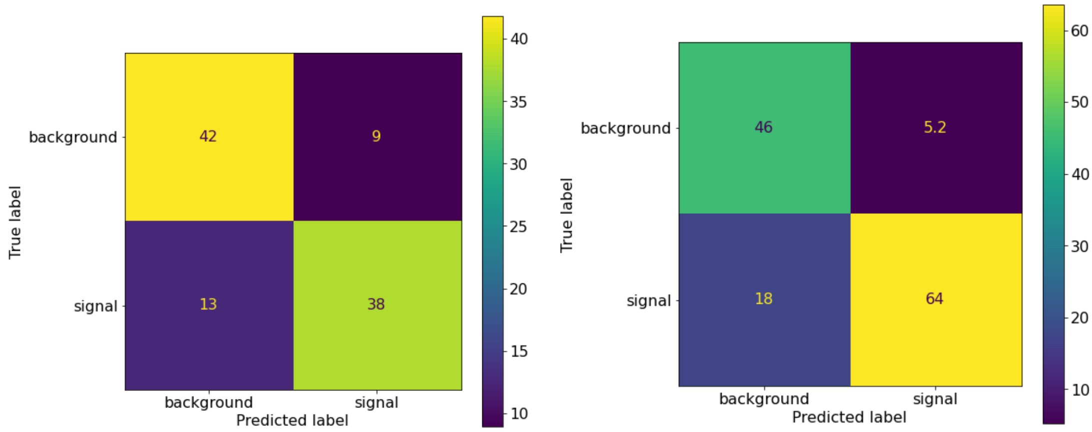


Figure 4.2 Weighted confusion matrix of the best SVM models for the original 800 GeV (left) and new 800 GeV (right) simulated data

	0.1 pb signal weights	normalized signal weights
tbH_300	19.05	50.00
tbH_800	50.50	50.00
tbH_1500	58.83	50.00
tbH_2000	38.24	50.00
tbH_250_new	38.73	50.00
tbH_800_new	81.11	50.00
tbH_3000_new	15.54	50.00
ttH	12.94	12.94
tt	4.44	4.44
ttW	10.55	10.55
ttZ	12.59	12.59
VV	4.49	4.49
Others	5.78	5.78

Table 4.2 Table of expected validation set signal and background events before and after normalization

issue, an additional value in parentheses will be added to the results of the best model for each mass – the significance measured on the validation dataset with the weighted sum of expected signal events set to 50. This approach keeps the original significance scores while allowing the comparison of results for different signal masses. The number of expected background events stays the same. Table 4.2 lists the number of expected events on the validation set for each process.

The best model for new 800 GeV normalized data still achieves better significance than the old model, even after scaling the weights of the validation set for evaluation. But the results are much more similar – the difference between the significance 8.82 and 10.57 could now be reasonably caused by the different weights during the training process and more than double the data points for the new 800 GeV mass.

The significance after normalization behaves mostly as expected – the lower signal weights are in general harder to distinguish. The exceptions are new signal masses 250 GeV and 3000 GeV. The model for mass 250 GeV achieves better results than a similar model for mass 300 GeV. Though given the relatively small difference between the masses and different filter efficiency for

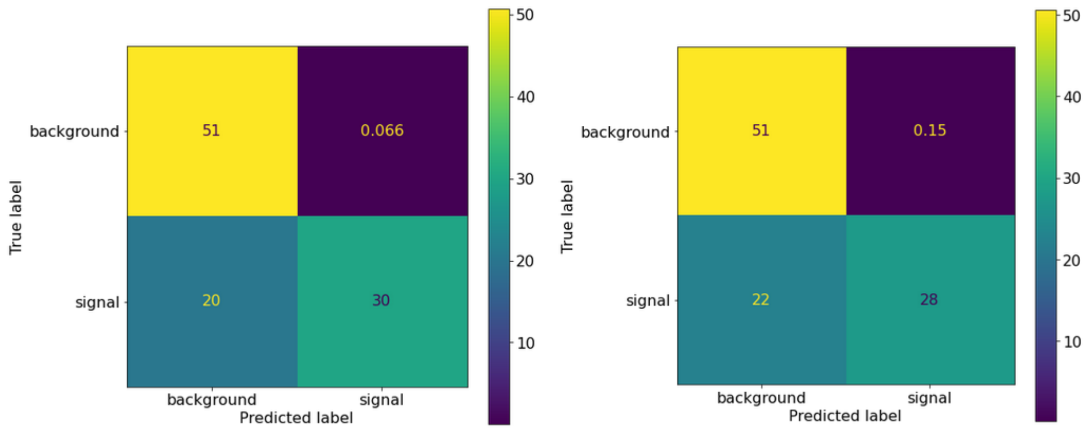


Figure 4.3 Weighted confusion matrix of the best SVM models for 2000 GeV (left) and 3000 GeV (right) signal

the new data, which caused a similar difference between the original 800 GeV and new 800 GeV data, it is reasonable to assume a similar effect also happened in this case.

Another noticeable decrease in significance occurs between the 2000 GeV mass and the new 3000 GeV mass. The outputs of the models are displayed in figure 4.3. Similarly to the models in preliminary testing, these models use linearly separable signal regions to remove almost all of the background. Unlike the studied models in the preliminary testing, these models manage to separate over half of the signal. In terms of unweighted events, 19 entries remain as false positives for the 2000 GeV signal, and 5 entries remain as false positives for the 3000 GeV signal, out of 5767 total unweighted background events.

At such low data counts, discrepancies in computed significancies are expected, which explains the likely cause for the lower significance of the 3000 GeV signal. To further mitigate the effect of the minimal background, the final model evaluation will be done on the best model output without thresholding. TRExFitter performs its own binning, which allows more precise results.

4.2.2 Decision Tree

The decision tree model uses the `DecisionTreeClassifier` from scikit-learn. A fast training time of the models allows us to perform the full grid search through the hyperparameter space. The grid search is performed to get the best decision tree model for each signal mass. The models use the `predict_proba` function, which outputs the probability that a given event belongs to the signal class. Each model is once again trained on the training set with only one charged Higgs boson mass and evaluated on the validation set with only one signal mass (the same one used in the training). The best threshold is selected on the validation set, from the thresholds $t \in \{0.1, 0.2, \ldots, 0.9\}$. The lower number of thresholds is selected to reduce the chance of overfitting the threshold on the validation set.

The following grid is used for each mass separately to get the seven best decision trees:

- criterion $\in \{\text{gini}, \text{entropy}\}$
- splitter $\in \{\text{best}, \text{random}\}$
- max_features $\in \{\text{None}, 32, 16, 4, 1\}$
- max_depth $\in \{\text{None}, 1, 2, 3, 4, 8, 16\}$
- min_samples_leaf $\in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$

order	mass	Z_3 [σ]	criterion	splitter	max_features	max_depth	min_samples_leaf
1	tbH_300	4.00	entropy	best	16	16	2
2	tbH_300	3.96	entropy	best	32	4	4
3	tbH_300	3.95	gini	best	32	4	32
1	tbH_800	10.15	entropy	best	32	8	4
2	tbH_800	10.06	entropy	best	16	8	2
3	tbH_800	9.71	gini	best	32	8	256
1	tbH_1500	15.86	gini	best		8	128
2	tbH_1500	15.86	gini	best		16	128
3	tbH_1500	15.86	gini	best			128
1	tbH_2000	12.41	entropy	random		8	32
2	tbH_2000	12.16	entropy	best	32		4
3	tbH_2000	12.03	entropy	best	32	16	128
1	tbH_250_new	6.37	gini	best	16	3	16
2	tbH_250_new	6.37	entropy	random	32	8	1
3	tbH_250_new	6.31	gini	random	16	8	16
1	tbH_800_new	14.59	entropy	best	32	8	128
2	tbH_800_new	14.30	gini	random		8	8
3	tbH_800_new	14.06	gini	random		16	64
1	tbH_3000_new	9.14	entropy	best	16	16	256
2	tbH_3000_new	7.74	entropy	best	32		1
3	tbH_3000_new	7.66	entropy	best			4

Table 4.3 Best decision tree results on the validation set with signal cross-section 0.1 pb

The results of the grid search, sorted based on the Z_3 significance, are shown in table 4.3. The results indicate the preference for the `best` splitter, which selects the best split out of all available features. Criterion `entropy` achieves five out of seven best results, demonstrating a small advantage over the Gini index. Trees with larger maximum depth are predictably preferred, with a maximum depth of 8 being the most common for the three best models of each mass. Still, the trees with a maximum depth of 4 achieved decent scores on the 300 GeV mass, and the best tree for the new 250 GeV mass only required a depth of 3 to achieve the best Z_3 score on the validation set. The minimum samples per leaf do not show a clear advantage for any of the values or masses, all the values of the hyperparameter are used, spread over the best three models of each of the masses, with the exception of the 1500 Gev mass. All three best 1500 GeV decision tree models have the same number of minimum samples per leaf. The likely cause is that the `best` tree splitter and high minimum samples per leaf led to the creation of the same (or very similar) model. This is further supported by all three best 1500 GeV decision trees having the same score (15.859203).

Given the small maximum depth of the best 250 GeV decision tree, it can be easily visualized. The tree is shown in figure 4.4. The tree consists of inner nodes, which have the decision condition at the top, and leaves. All the leaves are located in the rightmost part of the plot. Each node also contains the following information (top to bottom) – number of unweighted training samples in the node, number of expected background (`value` left) and signal events (`value` right) and name of the weighted majority class (the class with a larger number of expected events in the node). When the model needs to make a prediction, the rules are applied from left to right, until a leaf is reached. Once a leaf is reached, the model predicts probability based on the fraction of the expected signal in the leaf (based on the training set samples).

It is clear the model is very simple, the cuts made use only the energy and transverse momentum. No features with pseudorapidity or azimuthal angle are used. The main selection of the tree, in terms of the expected number of both signal and background events, is the topmost leaf. This leaf is reached by satisfying all the conditions on the path from the root (leftmost

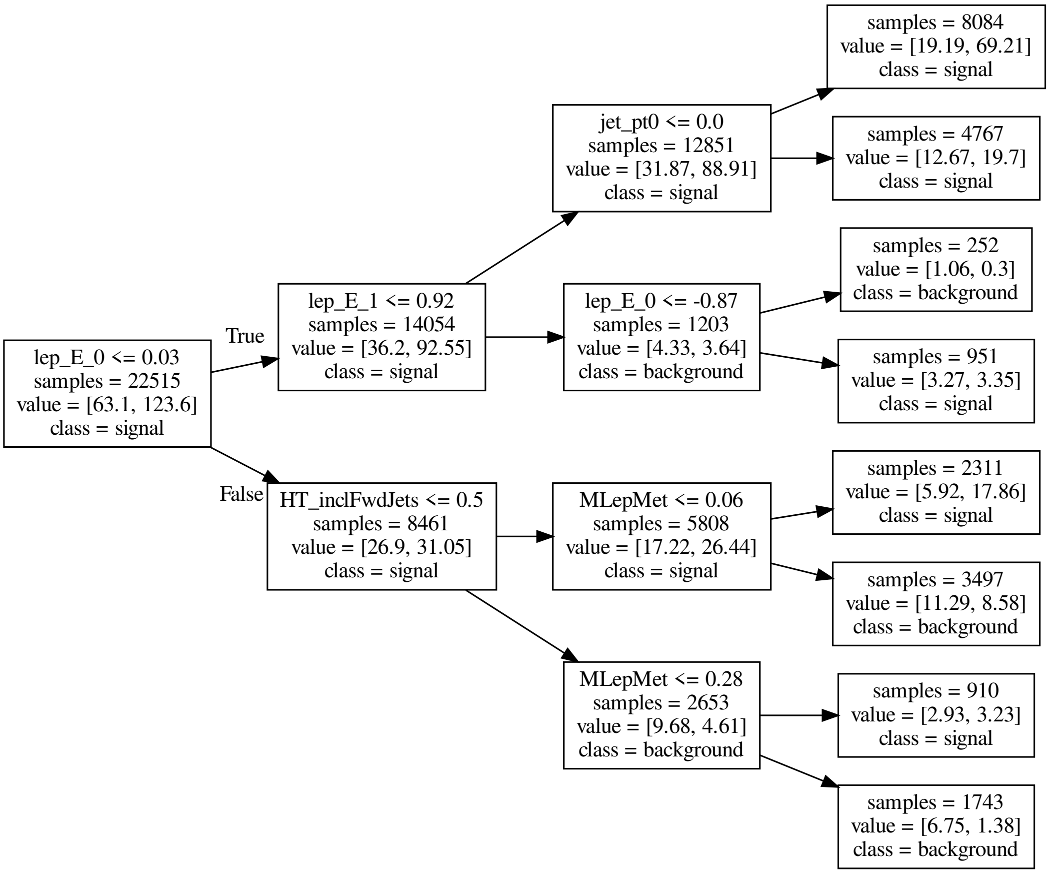


Figure 4.4 Best decision tree for the 250 GeV mass, feature values are normalized

node) of the tree to the leaf. By itself, the leaf contains over half of the expected training signal events. The conditions test each of the two leptons and the leading jet and remove the background based on the detected excess energy. The expected number of events is from the training set, so the results are too optimistic and the validation performance is worse (Z_3 significance on the validation set is 6.37). Still, this model shows that for low charged Higgs masses at 250 GeV and below, even simple models can achieve decent performance.

4.2.3 Random Forest

The random forest model uses the `RandomForestClassifier` from scikit-learn. As an ensemble model for the decision tree, a random forest has the hyperparameters of its tree components on top of its own hyperparameters. All trees, which are part of the random forest, have the same hyperparameters. The hyperparameters for the trees are chosen to reflect the parameter grid from the decision tree model but with some notable changes. The hyperparameter grid for the trees of the ensemble is reduced in size, removing values for the minimum samples required to be in a leaf node (`min_samples_leaf`) and the maximum tree depth (`max_depth`). Minimum samples per leaf are reduced due to little effect seen for the decision tree model. Maximum depth is reduced, removing the unlimited depth option, as well as the maximum depth of 16. Both

order	mass	Z_3 $[\sigma]$	n_estimators	criterion	max_features	max_depth	min_samples_leaf
1	tbH_300	4.25	100	gini	32	8	16
2	tbH_300	4.14	10	gini	16	8	16
3	tbH_300	4.07	1000	entropy	16	8	4
1	tbH_800	9.71	10	entropy		8	16
2	tbH_800	9.40	100	entropy	16	8	4
3	tbH_800	9.31	10	entropy	32	8	4
1	tbH_1500	17.90	1000	entropy		8	16
2	tbH_1500	17.74	1000	entropy	32	8	16
3	tbH_1500	17.49	100	entropy		8	16
1	tbH_2000	14.17	1000	entropy	32	8	1
2	tbH_2000	13.40	100	entropy	32	8	4
3	tbH_2000	13.00	10	entropy	32	8	16
1	tbH_250_new	6.77	100	gini		4	1
2	tbH_250_new	6.71	10	gini		4	16
3	tbH_250_new	6.45	10	gini		8	16
1	tbH_800_new	15.35	100	entropy	16	8	16
2	tbH_800_new	15.04	100	entropy	32	8	1
3	tbH_800_new	14.47	1000	entropy	16	8	16
1	tbH_3000_new	9.17	100	entropy	16	8	4
2	tbH_3000_new	8.58	10	entropy	16	8	1
3	tbH_3000_new	8.45	10	gini	32	8	16

Table 4.4 Best random forest models and their validation set results with signal cross-section 0.1 pb

depths are considered to produce too complex trees for the ensemble. Finally, the `splitter` is removed from the grid because it is not accessible for the `RandomForestClassifier` class (splitter value `best` is used). From the ensemble hyperparameters, only the number of trees is chosen to be optimized with three possible values based on the base 10 logarithmic scale. The random forest is set to use the bootstrap technique on the data to improve its performance.

The full hyperparameter grid is written below:

- n_estimators $\in \{1000, 100, 10\}$
- criterion $\in \{\text{gini}, \text{entropy}\}$
- max_features $\in \{\text{None}, 32, 16, 4, 1\}$
- max_depth $\in \{8, 4, 3, 2, 1\}$
- min_samples_leaf $\in \{1, 4, 16\}$

The hyperparameters have been optimized using grid search, the same as for the previous two model types. The grid is searched for each of the signal masses – a training dataset with one signal mass and all background processes is created and the random forest model is fitted on it. Then it is evaluated on a similarly created validation set. The output of the ensemble has been set to the prediction of signal probability through the `predict_proba` function. For each model, the best threshold $t \in \{0.1, 0.2, \ldots, 0.9\}$ for the probability prediction is optimized to achieve the highest Z_3 significance approximation. The model predicts an event as signal if the value of the `predict_proba` function is higher than the threshold. Otherwise, the event is predicted as background. Since the significance considers only the events predicted as signal, the models predicting only background or only a few signal events with no background are not considered in the model selection. After the grid search is complete, the models are sorted by Z_3 significance, separately for each of the signal masses – based on the mass the model was trained on.

The three best random forest models for each signal mass are listed in table 4.4. Models with the `entropy` criterion function achieved the best score on five out of seven signal masses, which further shows that entropy is slightly better suited as a branching criterion for building decision trees on the charged Higgs boson dataset. The maximum depth of the trees has the highest value for all but one signal mass − 250 GeV. This can reflect the relatively good separation a simple decision tree achieved on this mass with a depth of 3. The random forest achieves the significance $Z_3 = 6.77\sigma$, which is an improvement over the $Z_3 = 6.37\sigma$, at the cost of a significantly more complex model, which uses a hundred trees to achieve the better score.

4.2.4 Multilayer Perceptron

The multilayer perceptron model uses the TensorFlow Keras `Sequential` class to construct the neural network. The `Sequential` class accepts a varying number of layers, constructed using the `Dense` class (perceptron layer), `Dropout` layer after each dense layer, and `Input` layer at the start (the output layer is another `Dense` layer, with only one sigmoid output) [27]. The exact layout of the model is given by hyperparameters, but the model always starts with an `Input` layer, followed by `Dense` hidden layers (with `Dropout` layer after each hidden layer, if `dropout_rate` is not 0). Last is the output `Dropout` layer with one neuron, which predicts the probability that a given event belongs to the signal class. An example layout with five hidden layers is available in figure 3.5, in the Analysis and Design chapter. The network uses the Stochastic Gradient Descent parameter optimizer and binary cross-entropy loss function for training the parameters of the model.

Besides the number of hidden layers, combined with the number of perceptrons per hidden layer into the `hidden_layer_sizes` hyperparameter, which has a direct impact on the network architecture and the number of trained parameters, other hyperparameters are also optimized. The hidden layer activation function provides a non-linear transformation of the output of the hidden layers. The batch size determines how many samples will be used at once to update the parameters of the model. The learning rate affects the step size in model parameter updates. The dropout rate is the fraction of neuron outputs per hidden layer set to zero during model training. The dropout helps prevent overfitting [27].

The Optuna optimizer is used to get the best hyperparameters for multilayer perceptron because the classical grid-search would be too slow in this case. The optimizer is run for 25 trials for each of the signal masses (25 models with various hyperparameters, selected by Optuna from the grid below are trained on the training set with only one signal mass and all background processes and evaluated on validation data, limited to one signal mass as is done for the previous model types).

The hyperparameter grid for multilayer perceptron (used by Optuna):

- hidden_layer_sizes $\in \{(64_1, 64_2), (128_1, \ldots, 128_3), (512_1, \ldots, 512_5), (32_1, ..., 32_5), (32_1, ..., 32_8)\}$
- batch_size $\in \{65536, 2048, 64\}$
- hidden_layers_activation_function $\in \{\text{relu}, \text{sigmoid}\}$
- dropout_rate $\in \{0, 0.1, 0.5, 0.8\}$
- learning_rate $\in (0.00001, 0.01)$, sampled from the log-uniform distribution

Once all the models are trained, their performance is measured using the Z_3 significance on the validation set, as was done with the previous model types. The best three models for each signal mass are shown in table 4.5. The most often used hidden layer sizes are two layers with 64 neurons each, $(64_1, 64_2)$. The best hidden layer activation function is ReLU.

Given the high flexibility of the MLP, the models achieve noticeably lower Z_3 scores than expected. There are some factors, which could lower the model performance. The first is the small number of trials, which was chosen in accordance with the time available for this section of experiments, but which might lead to not enough hyperparameter combinations being tested.

order	mass	Z_3 $[\sigma]$	layer_sizes	batch size	activation function of the hidden layers	dropout rate	learning rate
1	tbH_300	2.90	$(64_1, 64_2)$	2048	relu	0.5	0.000101
2	tbH_300	2.58	$(64_1, 64_2)$	2048	relu	0.5	0.000115
3	tbH_300	2.56	$(64_1, 64_2)$	64	relu	0.5	0.000176
1	tbH_800	6.42	$(32_1, \ldots, 32_8)$	2048	relu	0.1	0.003709
2	tbH_800	6.29	$(64_1, 64_2)$	2048	sigmoid	0.8	0.000049
3	tbH_800	6.23	$(32_1, \ldots, 32_8)$	2048	relu	0.0	0.000084
1	tbH_1500	16.27	$(512_1, \ldots, 512_5)$	64	relu	0.5	0.003010
2	tbH_1500	16.26	$(64_1, 64_2)$	64	relu	0.1	0.005337
3	tbH_1500	15.55	$(64_1, 64_2)$	64	relu	0.1	0.002770
1	tbH_2000	12.61	$(64_1, 64_2)$	64	relu	0.5	0.009379
2	tbH_2000	12.32	$(64_1, 64_2)$	64	relu	0.5	0.000697
3	tbH_2000	12.02	$(32_1, \ldots, 32_5)$	64	relu	0.5	0.005041
1	tbH_250_new	5.03	$(64_1, 64_2)$	65536	relu	0.5	0.000264
2	tbH_250_new	5.03	$(64_1, 64_2)$	64	relu	0.5	0.008654
3	tbH_250_new	4.99	$(64_1, 64_2)$	2048	sigmoid	0.5	0.000041
1	tbH_800_new	9.73	$(512_1, \ldots, 512_5)$	2048	relu	0.0	0.000179
2	tbH_800_new	9.55	$(64_1, 64_2)$	64	relu	0.0	0.001427
3	tbH_800_new	9.52	$(64_1, 64_2)$	64	relu	0.0	0.000163
1	tbH_3000_new	6.92	$(64_1, 64_2)$	64	sigmoid	0.1	0.001871
2	tbH_3000_new	6.85	$(64_1, 64_2)$	64	sigmoid	0.1	0.004646
3	tbH_3000_new	6.25	$(64_1, 64_2)$	64	sigmoid	0.1	0.003340

Table 4.5 Best MLP model results on the validation set with signal cross-section 0.1 pb

Smaller charged Higgs boson dataset, caused by training a separate model for each signal mass, is also a candidate factor.

4.2.5 Best Models

Based on the validation results, the overall best model is selected from the four model types for each of the signal masses. The summary of the best model performances on the validation dataset is available in table 4.6. The model selection is done based on the Z_3 score column in tables 4.1, 4.3, 4.4, 4.5. The results are measured for a signal cross-section set to 0.1 pb. The corresponding numbers of expected events after preselection and at the working point for each mass are listed in 4.7, together with the number of background events.

Charged Higgs boson mass	Model type	Z_3 $[\sigma]$	Threshold
tbH_300	Random Forest	4.25	0.6
tbH_800	Decision Tree	10.15	0.9
tbH_1500	Random Forest	17.90	0.9
tbH_2000	SVM	14.91	0.6
tbH_250_new	Random Forest	6.77	0.7
tbH_800_new	Random Forest	15.35	0.9
tbH_3000_new	Random Forest	9.17	0.9

Table 4.6 Best models, trained on all features, and their validation set results and thresholds with signal cross-section 0.1 pb

H$^+$ mass [GeV]	300	800	1500	2000	250 new	800 new	3000 new
Preselection	11.35/50.79	34.20/50.79	43.23/50.79	28.95/50.79	12.45/50.79	32.74/50.79	7.70/50.79
Working point	8.72/7.86	27.39/6.69	28.63/0.24	17.30/0.066	10.88/15.63	18.73/1.56	2.68/0.00083

■ **Table 4.7** The expected values for signal/background on the validation set, after preselection and for each of the best models at the working point threshold cut. The signal is normalized to 0.1 pb.

Out of the four evaluated model types, three are among the best models – SVM, decision tree, and random forest. The random forest is the most successful model type, being selected as the best model for five out of seven signal masses. It does not provide the highest significance on the validation set for only two signal masses, original 800 and 2000 GeV. For the original 800 GeV, a decision tree performs better than the other three model types. The best decision tree is created with the `entropy` splitting criterion, `best` splitter, 32 features used in the search for the best split at each iteration, maximum depth of the tree 8 and 4 minimum samples per leaf. The 2000 GeV mass is best separated by the SVM, with $C = 0.10$ and linear kernel.

The table 4.7 shows how much each of the best models reduced the expected number of signal/background events, compared to the preselection. The working point is the best threshold, selected on the validation set by maximizing Z_3 significance and displayed in table 4.6. The fraction of the correctly classified charged Higgs boson events is lower for higher signal masses. The 3000 GeV model correctly classifies only about a third of the signal events. The most likely reason is the significance function (equation 2.9), which uses only the true positive and false positive values. This is a potentially desirable behavior, as a model capable of separating signal from the background with minimum false positives would allow easier testing for the charged Higgs boson presence. The downside is the higher uncertainty connected with the very low number of background events and the low number of signal events.

4.3 Feature Importance Ranking

After preprocessing, the dataset contains 104 features. This large number can cause issues for some of the models, also known as the curse of dimensionality. Even for some of the models, which can effectively deal with high-dimensional spaces (for instance the SVM classifier), it can be beneficial to reduce the feature space for the sake of better explainability and smaller storage requirements for the input data, depending on the drop of model performance.

Before reducing the number of features, a feature importance ranking is constructed first. The ranking is performed using permutation feature importance on the validation set – measured for each of the seven best models separately. Since the permutation feature importance performs a random permutation of a feature, before measuring the decrease of performance on the modified dataset, multiple measurements were done for each model to gain not only the mean feature importance but also the standard deviation.

The permutation feature importance is measured using `permutation_importance` function from scikit-learn. For each model, 20 permutations and measurements are done for each feature. Significance approximation Z_3 (equation 2.9) is used as the score function for the feature importance measurement – the `permutation_importance` function, therefore, measures the decrease in Z_3 significance after a permutation of a feature in the validation set. The features are then sorted based on their mean decrease in significance.

The three most important features are `MLepMet`, `MtLepMet`, and `taus_pt_0`, each being selected as the most important for two signal masses. The tau transverse momentum (`taus_pt_0`) additionally is the second most important feature for the three remaining masses, making it the most important feature, overall. The invariant transverse mass of all leptons and missing transverse energy (`MtLepMet`) is the second most important feature (additionally ranked third and fifth) and the invariant mass of all leptons and missing transverse energy (`MLepMet`) is the third most important feature (additionally ranked fourth and eighth). The normalized distributions of the three most important features are available in figure 4.5. The distributions are based on the testing dataset and all signal masses. The importance of the features is clearly shown in the distribution of the signal, which extends beyond background processes.

The plots in figures 4.6 and 4.7 show 10 features with the highest importance for each of the seven best models. The box plot shows a line at the median, the box extends from the Q_1 quartile to the Q_3 quartile. The whiskers extend to the farthest data point within the $1.5(Q_3-Q_1)$ distance from the edges of the box. The outliers are marked by small circles, drawn separately [28]. The feature importances are sorted based on the mean decrease in Z_3 significance. The full list of feature importance scores for each model is available in the appendix A.2.

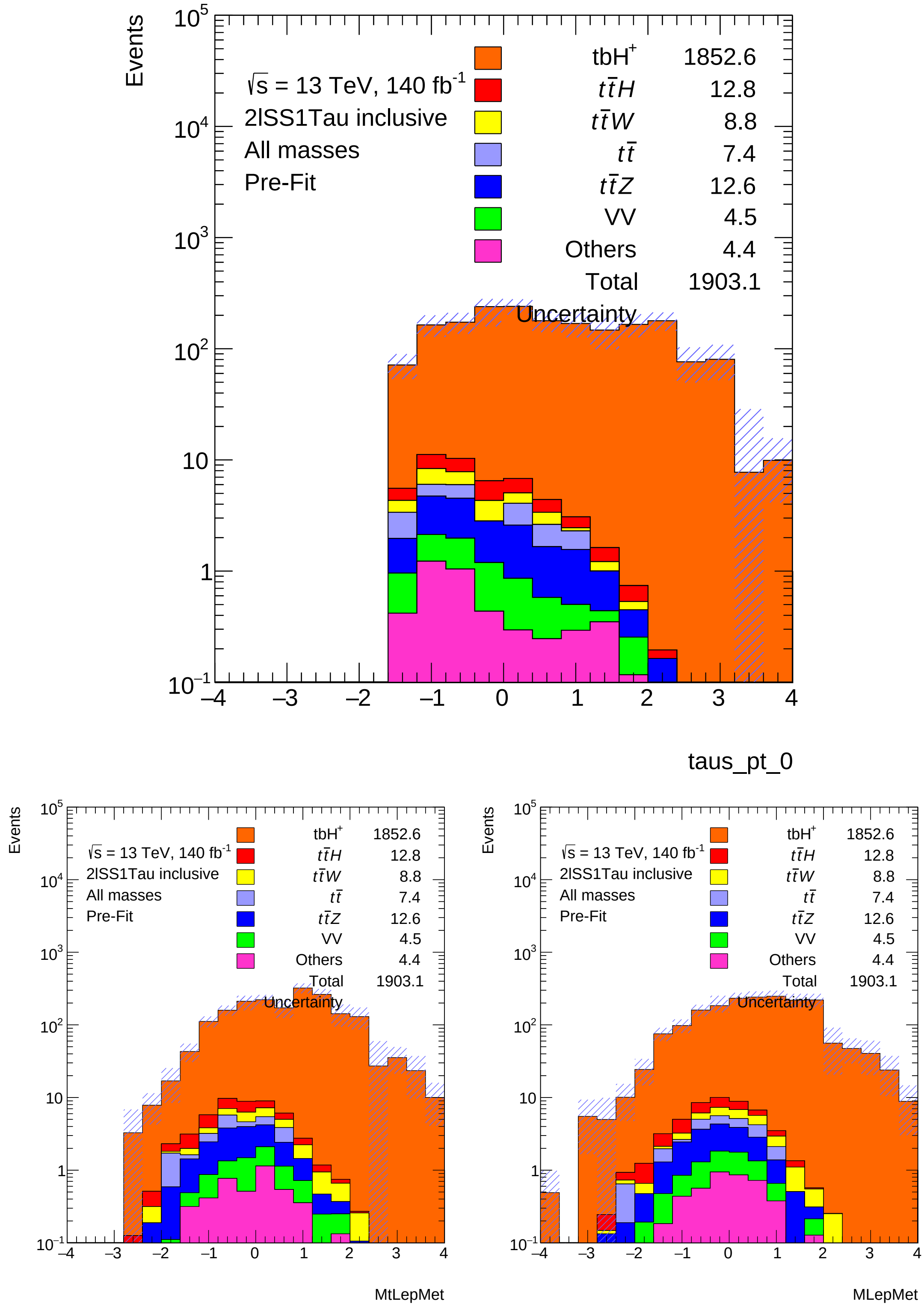
The division of feature importance into separate plots based on the signal masses and their associated models allows us to examine the effect of mass on significance score and, by proxy, the quality of separation of signal and background. It is noted that the composition of the feature importance varies with mass. The most important feature of the 300 GeV mass model is the invariant mass of all leptons and missing transverse energy. The tau transverse momentum, a similar feature – in the sense of being dependent on the Higgs mass – has taken second place in the feature ranking.

The feature ranking of the 250 GeV model is quite similar to the 300 GeV model. The first feature is also `MLepMet`. The second is the sum of the transverse momentum of all jets, instead of the tau transverse momentum, though both features are ranked among the top 10 most important for both the 250 GeV model and 300 GeV model. A likely cause of the exchange in ranking placement is the shared information between the features.

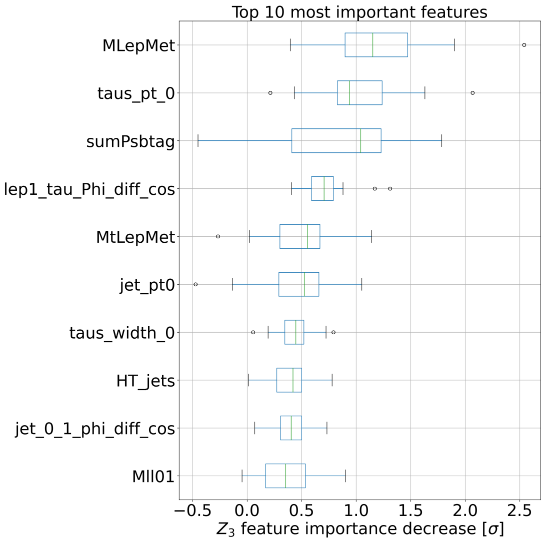
Another interesting result is the different feature importance ranking for the original and new 800 GeV signal masses. The most important feature of the original 800 GeV mass model is the momentum of the tau (`taus_pt_0`), which, when its information is removed by random permutation, causes a decrease in significance by approximately 3σ. A similar loss of significance happens when the `taus_pt_0` feature undergoes permutation in the new 800 GeV dataset. Still, the new 800 GeV model performance depends on the `MtLepMet` feature, which is more important than the tau transverse momentum. One of the possible reasons for this discrepancy is the different type of model selected for the original 800 GeV mass – decision tree, compared to the random forest for the new 800 GeV mass. Another reason might be the greedy algorithm selecting the features, which might prioritize different features in a different set, which might lead to different overall feature importance. This difference could be further increased by the correlation of the features, though the Pearson correlation coefficient between `taus_pt_0` and `MtLepMet` is only 0.23.

To properly take the correlation of the features into account, the Pearson correlation coefficient is applied to the 25 most important features of each of the seven best models. For the original 800 GeV model, the heatmap of the Pearson coefficients (figure 4.8), shows that `MtLepMet` is quite strongly correlated to multiple other features – namely `lep_Pt_0` (0.84), `HT` (0.70), `HT_lep` (0.86) and others, which are not present in the top 10 most important features of the new 800 GeV model. The increased correlation coefficient indicates the presence of shared information. Under the assumption that the decision of the model is based on this shared information, the model should perform similarly with the correlated features removed. This fact might be beneficial for feature reduction because out of two strongly correlated features, only one has to be included in the dataset and the model should perform similarly well. In the feature reduction phase, this effect will be explored further.

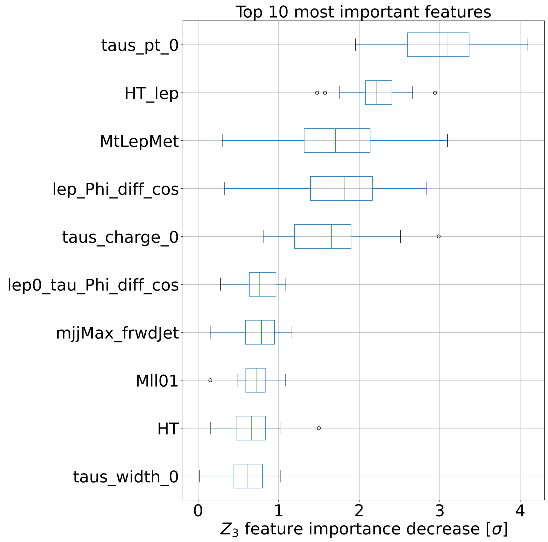
In the final part of the feature importance analysis, higher signal masses come into focus. Unsurprisingly, for the highest mass, 3000 GeV charged Higgs boson, a transverse momentum feature – namely the transverse momentum of the subleading jet – is the most important feature by far. However, the number of angle-based features among the top 10 ranked features is more unexpected. Given the difference in importance of the transverse momentum and most of the angle-based features, the model uses the angle-based features for residual separation, after the information about high mass has already been used.



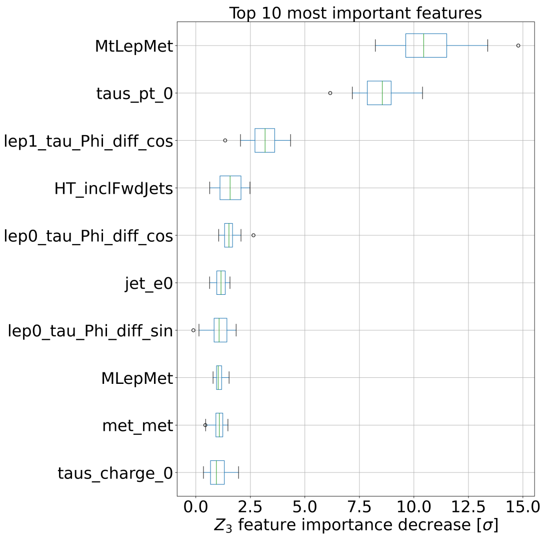
■ **Figure 4.5** Distributions of the three most important features after normalization. Top: tau transverse momentum, left: invariant transverse mass of all leptons and missing transverse energy, right: invariant mass of all leptons and missing transverse energy. The signal cross-section is 1 pb.



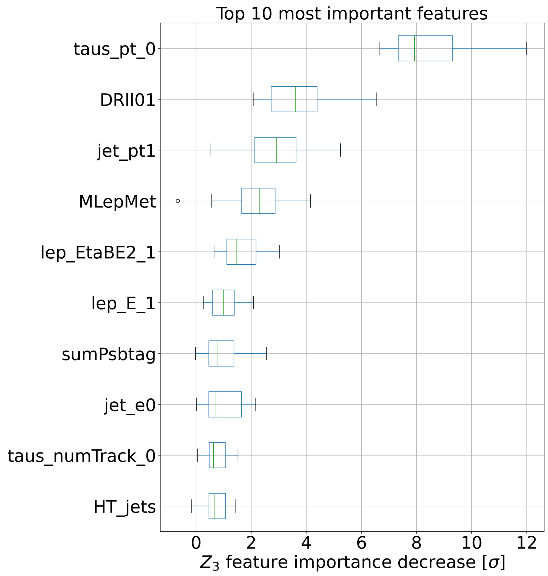
300 GeV



800 GeV

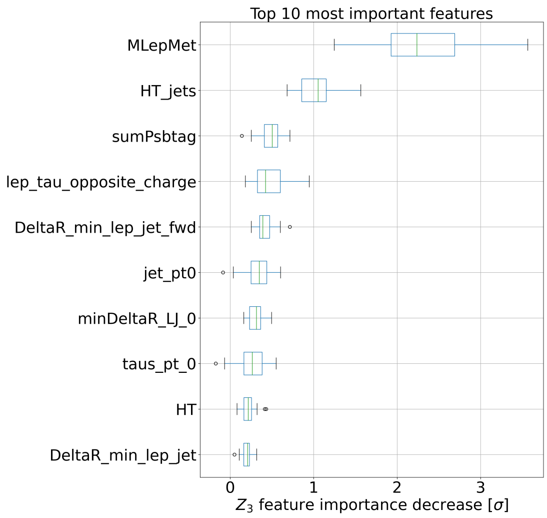


1500 GeV

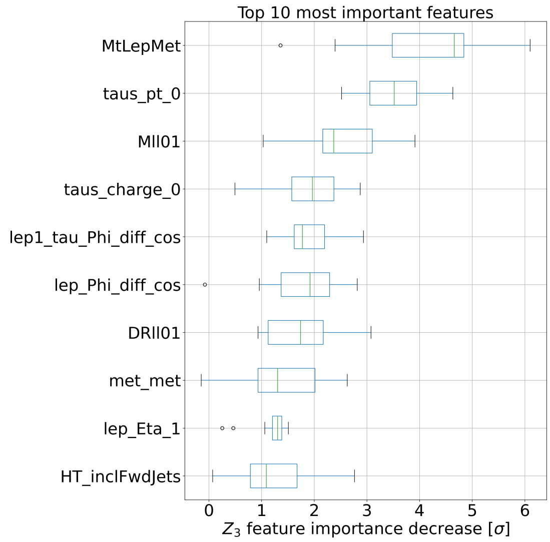


2000 GeV

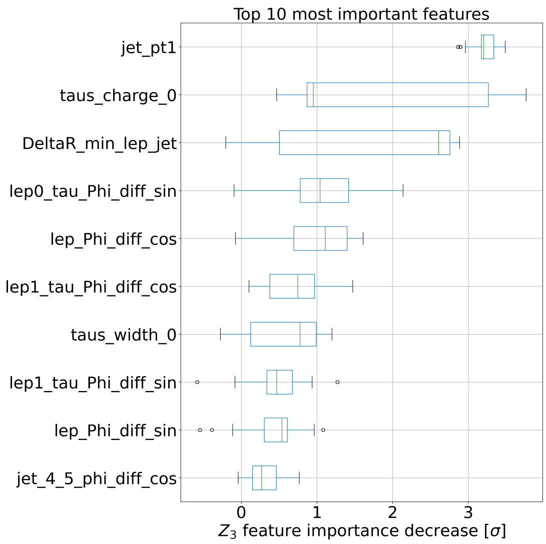
Figure 4.6 Feature importance of the models trained on the original signal masses



New 250 GeV

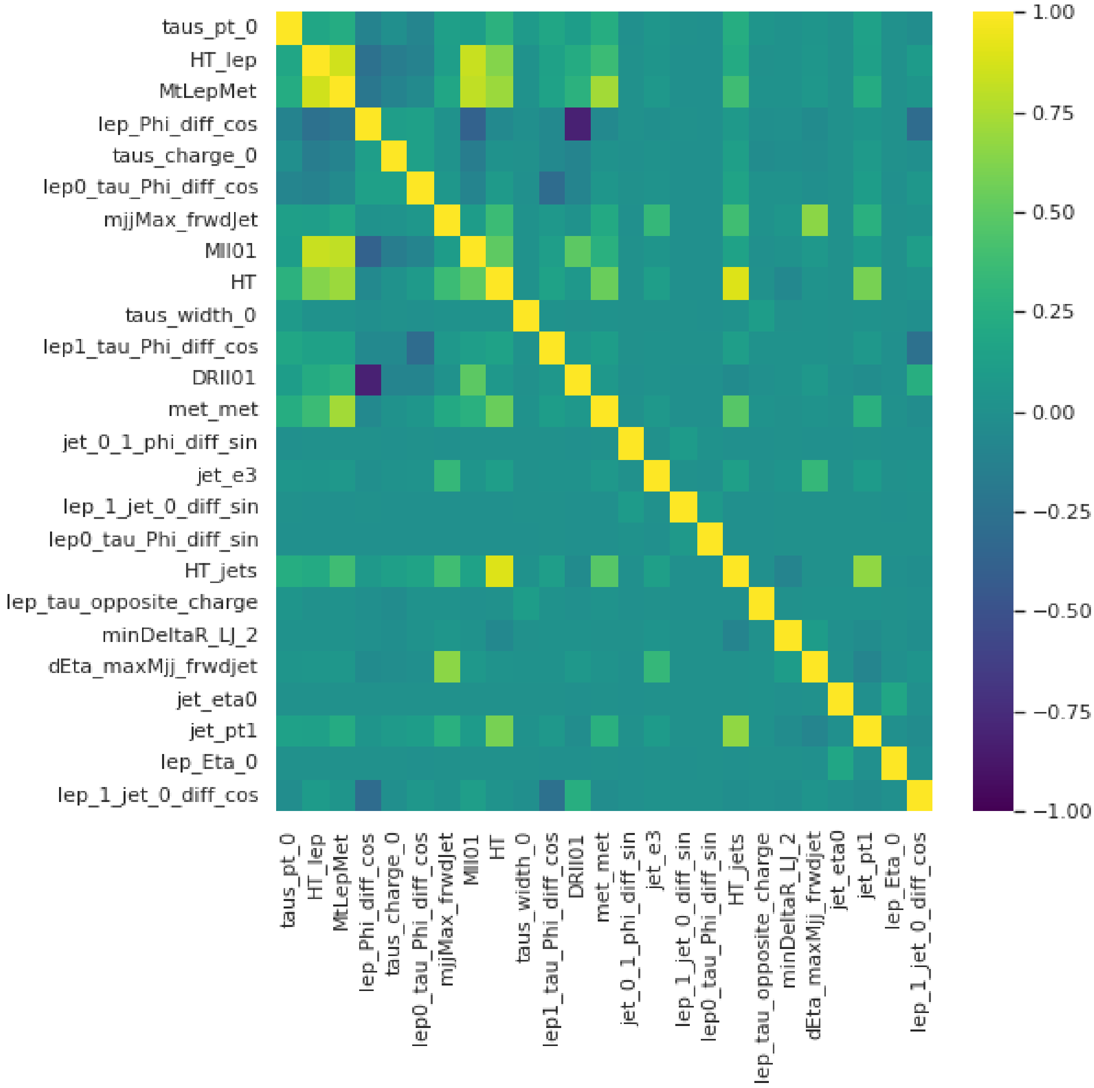


New 800 GeV



New 3000 GeV

■ **Figure 4.7** Feature importance of the models trained on the new signal masses



■ **Figure 4.8** Pearson correlation coefficients for 25 most important features of the original 800 GeV model, correlations measured on full dataset with all signal masses and background

4.4 Feature Reduction Performance Analysis

Based on the feature importance ranking, the number of features is reduced. To stay comparable with the thesis of Jiří Pospíšil, models are trained with the number of features reduced to 20, 10, and 5. The features in the subset are selected based on the feature ranking – i.e. from the list of features sorted by the mean decrease of Z_3 significance, the first 20, 10, and 5 features are selected, respectively. Since the number of features after reduction is quite small (especially in the case of 5 features), the strongly correlated features are removed from the sorted feature ranking list before reduction to 20, 10, and 5 features to clear the way for additional non-correlated features. A Pearson correlation coefficient above 0.6 is considered a strong correlation. The full list of features used in the feature reduction experiment is available in the appendix A.4.

The creation of the list of features for the feature reduction experiment is done in two steps. First, for each of the best models, a feature list is sorted according to the permutation feature ranking – from the most important to the least important feature. The list is stored as `[mass_name]_most_important_features` for all best models (for instance, the feature list for the SVM model on the 2000 GeV mass is assigned name `tbH_2000_most_important_features`). Second, a filtering function is applied to the feature list, removing the strongly correlated features. The filtering function is implemented in the `Feature_importances_final.ipynb` file and its code is available in figure 1.

The function `filter_out_correlated_features` takes the first `n_initial_features` out of the given `[mass_name]_most_important_features` and measures the Pearson correlation coefficient for each possible pair of features on the whole dataset (labeled `df2`). The pairs with a correlation above 0.6 and the first feature with more importance than the second feature are labeled as the `list_of_correlated_features`. The list of pairs is then traversed and the less important feature (the second feature of each pair) is removed, if neither feature of the pair is already removed when the traversal encounters the pair.

```
def filter_out_correlated_features(
        most_important_features,  # array of features sorted by importance
        correlation_threshold=0.6,  # threshold for Pearson correlation coefficient
        n_initial_features=25):  # number of features to enter the filtering process
    # take the first n_initial_features from most_important_features
    selected_features = [feature for feature in most_important_features[:n_initial_features]]
    # compute Pearson correlation coefficient for each of the selected_features
    correlation_matrix = df2[selected_features].corr()
    # mask out correlations on and below the diagonal
    # also mask out correlations with absolute value below the correlation_threshold
    triu_corr = pd.DataFrame(
        np.triu(correlation_matrix.abs(), 1) > correlation_threshold,
        columns=selected_features,
        index=selected_features
    )
    # apply the mask and get pairs of correlated features with correlation above the threshold
    list_of_correlated_features = correlation_matrix[triu_corr].stack().index.tolist()
    removed_features = set()
    # add the less important feature from each pair to the removed features
    # unless the other feature of the pair is already removed
    for some_feature, feature_to_remove in list_of_correlated_features:
        if some_feature not in removed_features and feature_to_remove not in removed_features:
            removed_features.add(feature_to_remove)

    print('Removed features: ')
    print(removed_features)
    # return the selected_features, which are not in removed_features
    return [feature for feature in selected_features if feature not in removed_features]
```

Code listing 1 Function to filter out features correlated above a certain threshold

Z_3 [σ]	All features	20 features	10 features	5 features
tbH_300	4.25	4.09	3.86	3.69
tbH_800	10.15	8.77	8.15	7.27
tbH_1500	17.90	16.88	17.47	17.77
tbH_2000	14.91	12.77	13.90	14.46
tbH_250_new	6.77	6.24	6.55	6.24
tbH_800_new	15.35	13.74	14.39	13.66
tbH_3000_new	9.17	5.22	5.20	4.80

Table 4.8 Comparison of the significance of the best models trained on all the features with models trained on a subset of features. The signal is normalized to the cross-section 0.1 pb

After generating the lists of features for the feature reduction experiment, a model generator is constructed. The model generator produces models, which have the same model type and hyperparameters as the best models measured on the full dataset. Since the feature importance was measured on these best models, feature reduction performance should be measured on the same model base to make the test outputs comparable.

In the experiment, the model generator produces three models for each mass, which are fitted to the training data, which only has 20, 10, and 5 features for the first, second, and third models, respectively. The feature subsets, assigned to each mass, are available in the appendix A.4. The best Z_3 significance is then determined for each model on thresholds $t \in \{0.1, 0.2, \ldots, 0.9\}$. The resulting Z_3 significances are listed in table 4.8 for the reduced-dataset models. The column `All features` is taken from table 4.6 and it is not generated again for this experiment. Table 4.8 shows a large decrease in significance for the 3000 GeV mass. In other cases, a smaller decrease in performance is noted. For the 1500 and 2000 GeV signal masses, the classifier with 5 features achieves nearly the same significance as the classifier with all features. Nonetheless, the models, which use all the available features achieve higher performance.

4.5 Analysis of Best Model Performance

The significance of each of the models was measured on the testing set with the threshold value between 0 and 1. The working point is defined by a threshold cut with the highest significance. The results are shown in figure 4.10.

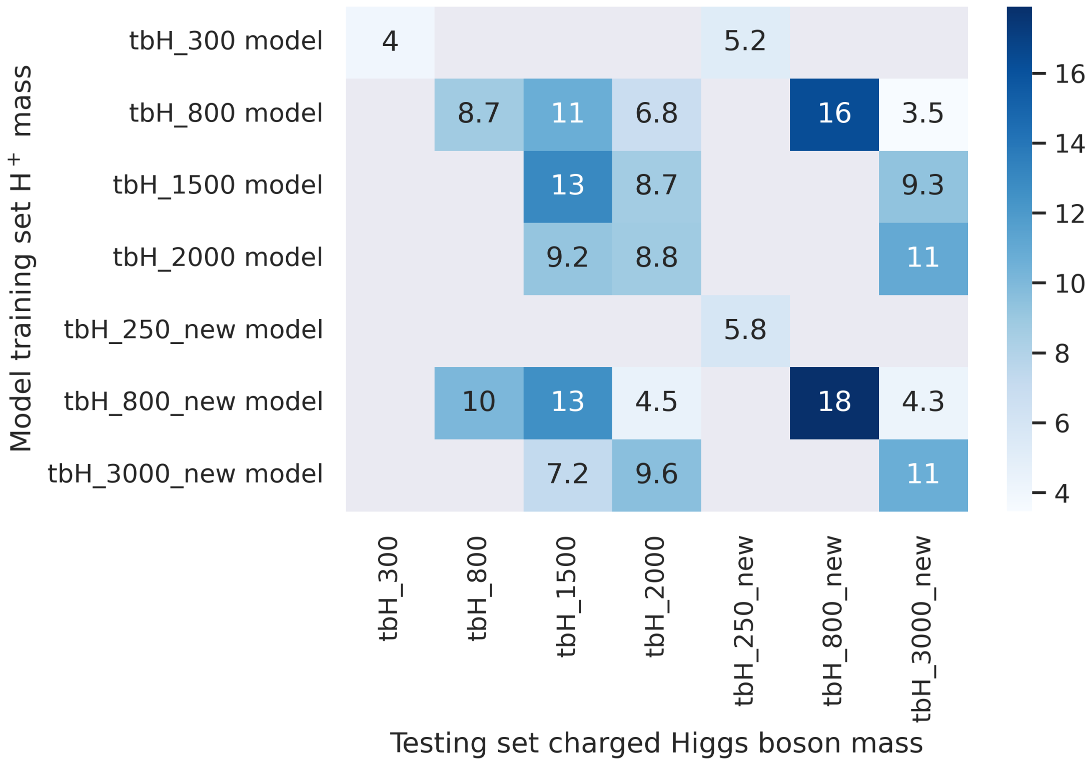
The threshold selection was performed on values $\{0.1, 0.2, \ldots, 0.9\}$ on the validation set. Figure 4.10 shows the significance distribution for the testing set. The plots show that the thresholds are a small distance from the peak of the significance curve. Note that there are only nine sampling points of the threshold on the validation set. Table 4.9 shows that the model performance is better for higher charged Higgs boson masses.

The model performance on non-trained masses has been studied, and the results are shown in figure 4.9. The 300 GeV model generalizes well on the 250 GeV mass and the high-mass models generalize quite well for other high masses.

4.6 Expected Limits

4.6.1 Asymptotic

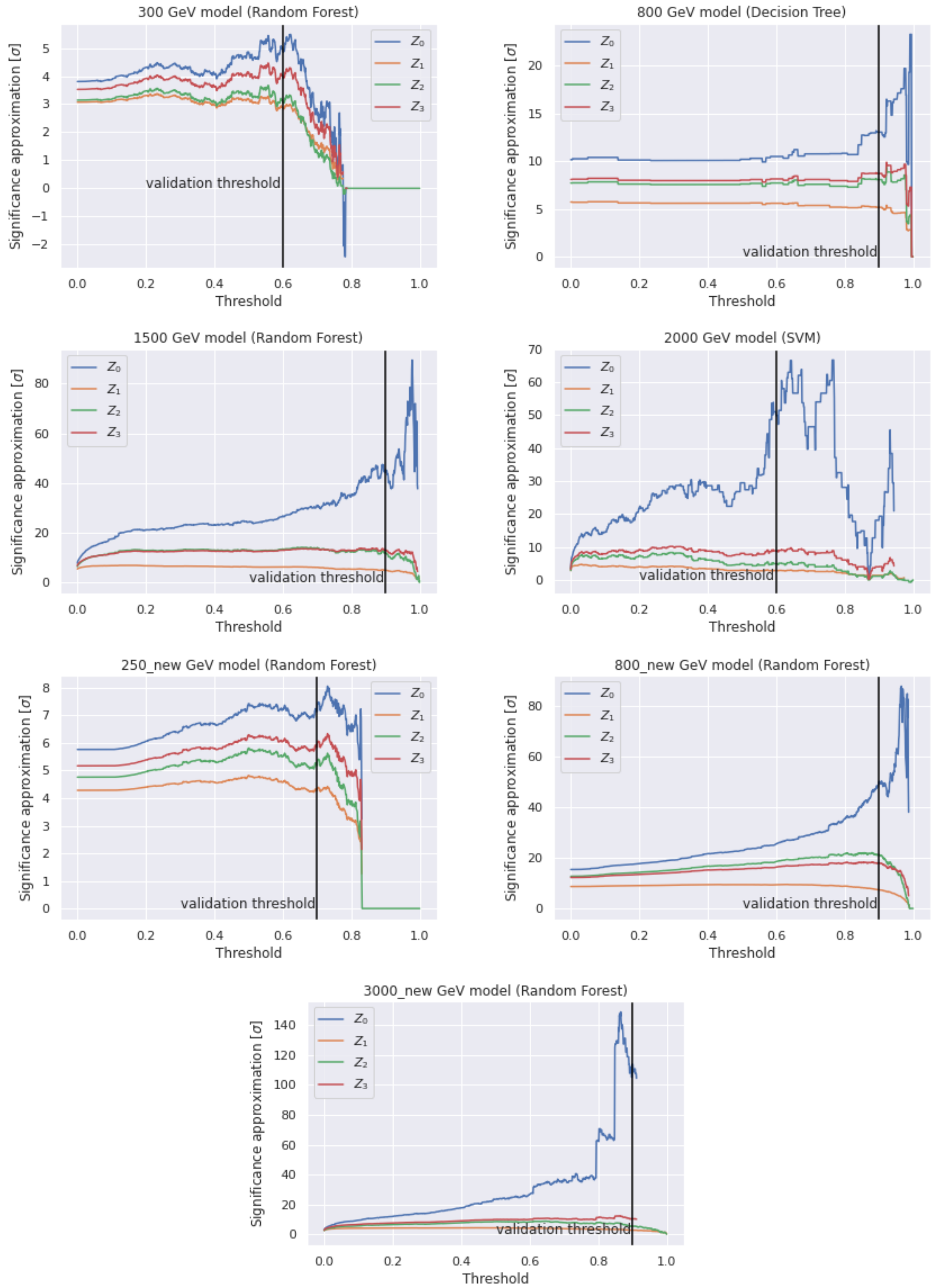
The output of the best model for each of the signal masses is evaluated using TRExFitter to obtain the asymptotic expected upper limit at 95% CL on cross-section (figure 4.11). The limit plot shows the minimum cross-section of the signal, at which the classifier can still detect it with



■ **Figure 4.9** Significance of the best models for each mass on the testing sets with assigned signal masses. In the gray areas the efficiency is less than the preselection efficiency.

H^+ mass [GeV]	300	800	1500	2000	250 new	800 new	3000 new
Preselection	16.12/50.51	42.14/50.51	39.56/50.51	18.77/50.51	13.16/50.51	44.00/50.51	11.51/50.51
Preselection Z_0	2.27	5.93	5.57	2.64	1.85	6.19	1.62
Preselection Z_1	1.97	4.38	4.17	2.26	1.65	4.53	1.46
Preselection Z_2	1.87	4.9	4.6	2.18	1.53	5.11	1.34
Preselection Z_3	2.16	5.3	5.01	2.5	1.78	5.52	1.56
Working point	7.33/6.26	21.55/5.95	17.60/0.28	6.19/0.027	9.10/15.72	22.14/1.21	4.55/0.01
Working point Z_0	2.93	8.84	33.32	37.71	2.3	20.14	56.34
Working point Z_1	1.99	4.11	4.16	2.48	1.83	4.58	2.13
Working point Z_2	1.83	5.47	8.68	3.72	1.67	8.52	2.88
Working point Z_3	2.53	6.41	10.66	7.44	2.12	9.7	7.11

■ **Table 4.9** Expected values for signal/background of the testing set, after preselection and for each of the best models at the working point threshold cut. The signal cross-section is set to 0.1 pb. The significances are also given.



■ **Figure 4.10** Dependence of significance approximation on the selected threshold on the testing set for each of the best models, also shown is the threshold selected per model on the validation set

H$^+$ mass [GeV] / Cross-section [pb]	300	800	1500	2000	250 new	800 new	3000 new
Expected upper limit	0.1019	0.0242	0.0152	0.0271	0.1019	0.0207	0.0305
Expected upper limit + σ	0.1571	0.0371	0.0241	0.0487	0.1516	0.0305	0.0484
Expected upper limit + 2σ	0.2427	0.0573	0.0402	0.1086	0.2238	0.0447	0.0837
Expected upper limit - σ	0.0735	0.0175	0.0109	0.0195	0.0734	0.0149	0.0220
Expected upper limit - 2σ	0.0547	0.0130	0.0081	0.0146	0.0547	0.0111	0.0164

■ **Table 4.10** Asymptotic expected upper limit at 95% CL on cross-section, with 68% and 95% confidence intervals

H$^+$ mass [GeV] / Cross-section [pb]	300	800	1500	2000	250 new	800 new	3000 new
Expected upper limit	0.1018	0.0249	0.0156	0.0293	0.1016	0.0214	0.0328
Expected upper limit + σ	10.0000	0.0366	0.0244	0.0507	10.0000	0.0306	0.0463
Expected upper limit + 2σ	10.0000	0.0574	0.0405	0.1336	10.0000	0.0434	0.0724
Expected upper limit - σ	0.0686	0.0172	0.0127	0.0205	0.0668	0.0147	0.0271
Expected upper limit - 2σ	0.0444	0.0133	0.0083	0.0150	0.0580	0.0126	0.0196

■ **Table 4.11** Toy Model expected upper limit at 95% CL on cross-section, with 68% and 95% confidence intervals

95% CL. The 68% and 95% confidence intervals for the limit on cross-section are also included in the plot. The filter efficiencies (table 3.5) are taken into account. The combined plot in figure 4.11 does not contain the original 800 GeV mass, the mass can be compared in the comparison plots showing original and new asymptotics results separately (figure 4.12).

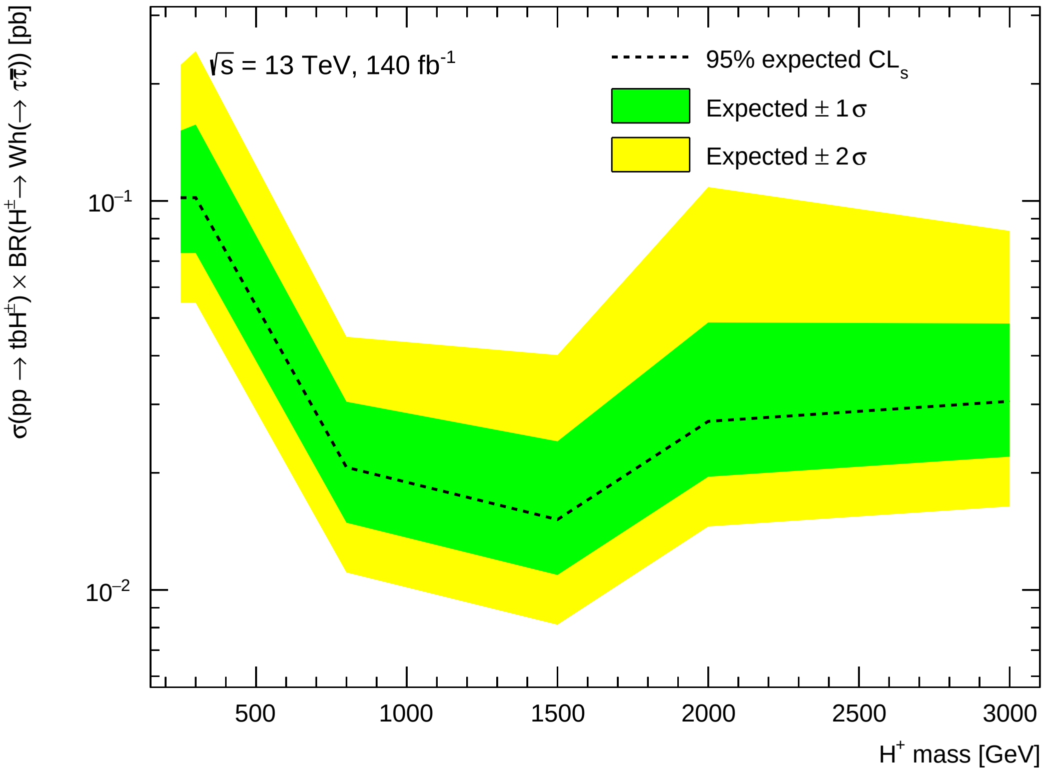
4.6.2 Toy Model

In addition, the expected cross-section limits are computed using the Toy model. The results are consistent with the asymptotics and are shown in figure 4.14 and table 4.11. The results of the Toy model and the Asymptotic model are consistent with each other.

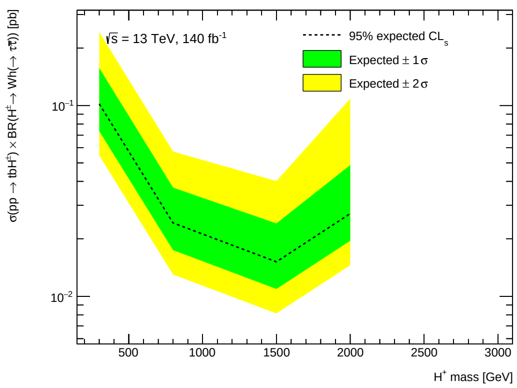
4.6.3 Comparison with Previous Results

The expected limits of this analysis are significantly weaker than the previous limits (figure 2.4) [5, 12]. The likely reason is a difference in signal and background file generation. A further improvement in this analysis is the inclusion of the filter efficiency of the event generation. Because of these significant differences, a detailed comparison for individual masses is not suitable.

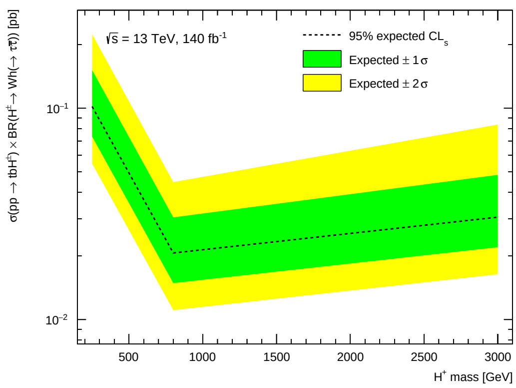
The results of this thesis are also compared with the results of the CMS collaboration. Figure 4.13 shows the expected upper limits at 95% CL on the product of cross section and branching fraction $\sigma_{\mathrm{H}^\pm}(\mathrm{H}^\pm \to \mathrm{HW}^\pm, \mathrm{H} \to \tau\tau)$ as a function of $m_{\mathrm{H}^\pm}$ from the CMS charged Higgs boson decaying into a heavy neutral Higgs boson and a W boson analysis [29]. The expected limit (figure 4.11) in the mass range 300 to 700 GeV obtained in this thesis is similar; however, the CMS analysis includes systematic uncertainties which reduce the sensitivity and were not addressed in this analysis [29].



■ **Figure 4.11** Asymptotic expected upper limit at 95% CL on cross-section, with 68% and 95% confidence intervals as function of charged Higgs boson mass



Original simulated signal data production



New simulated signal data production

■ **Figure 4.12** Asymptotic expected upper limit at 95% CL on cross-section, with 68% and 95% confidence intervals as function of charged Higgs boson mass

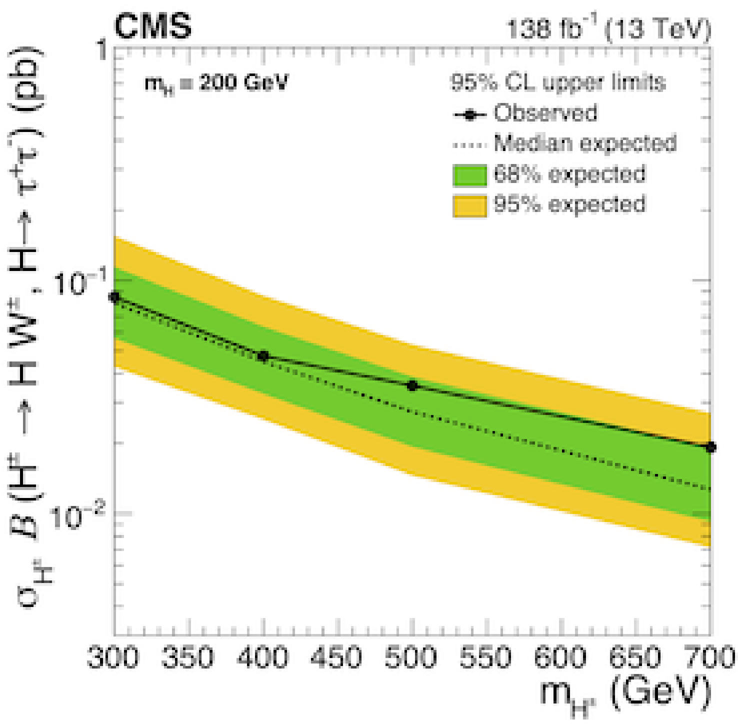
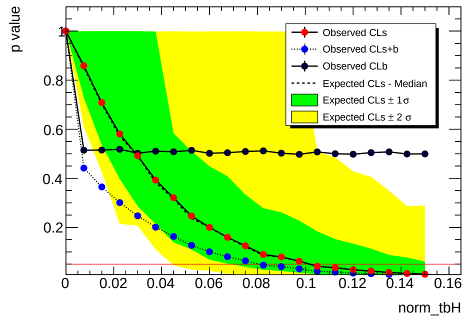
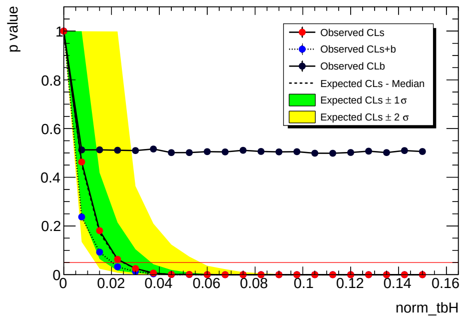


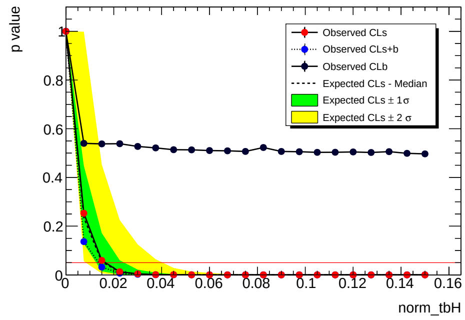
Figure 4.13 Expected and observed upper limits at 95% CL on the product of cross-section and branching fraction $\sigma_{H^\pm}(H^\pm \rightarrow HW^\pm, H \rightarrow \tau\tau)$ as a function of $m_{H^\pm}$ and assuming $m_H = 200$ GeV for the combination of all final states considered. The observed upper limits are represented by a solid black line and circle markers. The median expected limit (dashed line), 68% (inner green band), and 95% (outer yellow band) confidence intervals are also shown. Taken from [29]



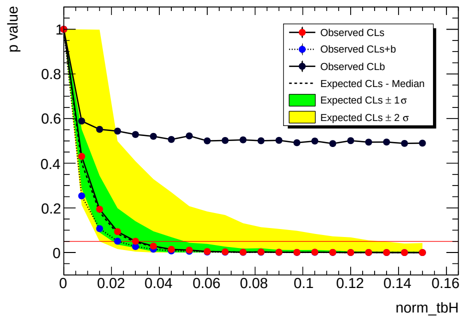
300 GeV



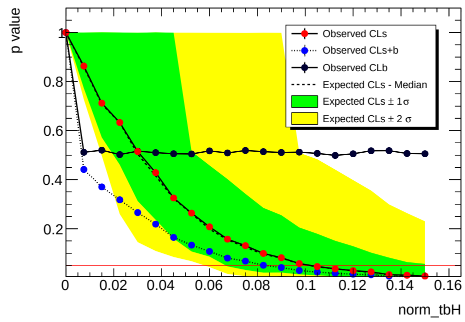
800 GeV



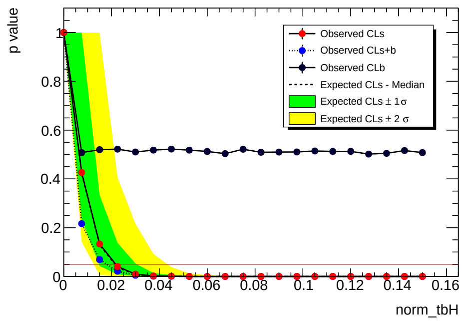
1500 GeV



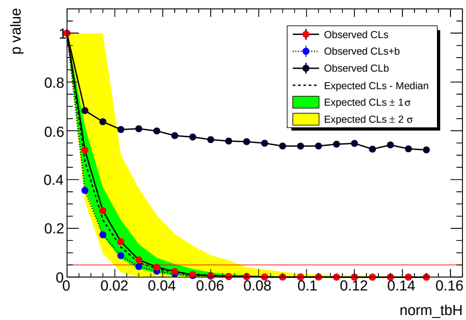
2000 GeV



New 250 GeV



New 800 GeV



New 3000 GeV

Figure 4.14 P-value for different cross-sections, produced by Toy Monte Carlo for all signal masses. The label norm_tbH corresponds to the cross-section in [pb].

Chapter 5
Conclusion

This thesis focuses on the analysis and separation of the charged Higgs process using machine learning. To achieve this, a thorough feature analysis is first conducted, focusing on the feature origins and matching normalizations in order to more optimally use the information contained in the features.

Following this analysis, a multi-model approach is proposed, where each model is sensitive in a certain mass range to achieve large significance in its dedicated mass section. Four different model types are chosen for the model selection. Support Vector Machines are chosen to check the linear separability of the problem and data quality. Decision Trees are chosen for their explainability. Random Forests are selected as an ensemble model with better performance over explainability. The multilayer perceptron is chosen as the universal machine learning model.

Next, each model type is optimized on the problem domain via the selection of the best-suited set of hyperparameters. Each model type is assigned a grid of viable hyperparameters. The SVM, decision tree, and random forest models use grid search for hyperparameter optimisation, while the multilayer perceptron uses Optuna optimizer to search the hyperparameter state space. After the hyperparameters have been selected for each model type and each mass, the best model for each mass is selected out of the four model types. The criterion for model comparison during hyperparameter selection, as well as the selection of the best model for each mass, is the significance function approximation.

Feature ranking was computed for each of the best models, using permutation feature importance. The most important feature is the tau transverse energy, the second is the invariant transverse mass of all leptons and missing transverse energy and the third is the invariant mass of all leptons and missing transverse energy. Based on the feature ranking, feature reduction was performed. The reduction from 104 features to 20, 10, and 5 most important features, after the removal of strongly correlated features, led to a small decrease in significance for all masses.

The best models were tested on the testing set. The upper expected limit at 95% CL on cross-section was computed for each model, using the Asymptotic and the Toy Model methods. The expected limits at 95% Cl are in the range of 0.1 pb to 0.02 pb, depending on the charged Higgs boson mass. Each of the best models was then tested on other signal masses, to give an estimate of the generalization power on the neighborhood of its assigned mass point. Models with assigned charged Higgs mass of 1500 GeV and more have good generalization power on high masses, while the 300 GeV and 800 GeV masses are successfully predicted only by their assigned models. Future work should focus on training the lower-mass models on all masses while keeping one mass as the focus of the model to keep the benefit of high significance while broadening the neighborhood the model can focus on. While statistical uncertainties are included in the analysis, systematic uncertainty inclusion is a future project.

A goal of this thesis concerning the optimisation related to systematic uncertainties has been moved to future work with the agreement of the thesis supervisor. The implementation of the analysis using the version 8 dataset took significantly longer than expected.

Appendix A
Appendix

A.1 Preselection Details

```
(custTrigMatch_LooseID_FCLooseIso_DLT )
& ( (dilep_type > 0 ) & ( (lep_ID_0*lep_ID_1)>0) )
& ( ((lep_Pt_0>=10e3) & (lep_Pt_1>=10e3))
    & ((fabs(lep_Eta_0)<=2.5) & (fabs(lep_Eta_1)<=2.5))
    & ((( abs(lep_ID_0) == 13 ) & ( lep_isMedium_0 )
            & ( lep_isolationLoose_VarRad_0 ) & ( passPLIVTight_0 ))
        | ((( abs(lep_ID_0) == 11 ) & ( lep_isTightLH_0 )
                & ( lep_isolationLoose_VarRad_0 ) & ( passPLIVTight_0 )
                & ( lep_ambiguityType_0 == 0 )
                & ( lep_chargeIDBDTResult_recalc_rel207_tight_0>0.7 ) )
            & ( ((~((~((lep_Mtrktrk_atConvV_CO_0<0.1)
                        & (lep_Mtrktrk_atConvV_CO_0>=0) & (lep_RadiusCO_0>20)))
                    & ((lep_Mtrktrk_atPV_CO_0<0.1) & (lep_Mtrktrk_atPV_CO_0>=0)))))
                & (~((lep_Mtrktrk_atConvV_CO_0<0.1)
                    & (lep_Mtrktrk_atConvV_CO_0>=0)
                    & (lep_RadiusCO_0>20))))))
    & ( (( abs(lep_ID_1) == 13 ) & ( lep_isMedium_1 )
            & ( lep_isolationLoose_VarRad_1 ) & ( passPLIVTight_1 ) )
        | ( (( abs(lep_ID_1) == 11 ) & ( lep_isTightLH_1 )
            & ( lep_isolationLoose_VarRad_1 ) & ( passPLIVTight_1 )
            & ( lep_ambiguityType_1 == 0 )
            & ( lep_chargeIDBDTResult_recalc_rel207_tight_1>0.7 ))
            & (((~((~((lep_Mtrktrk_atConvV_CO_1<0.1)
                        & (lep_Mtrktrk_atConvV_CO_1>=0)
                        & (lep_RadiusCO_1>20)))
                    & ((lep_Mtrktrk_atPV_CO_1<0.1)
                        & (lep_Mtrktrk_atPV_CO_1>=0)))))
                & (~((lep_Mtrktrk_atConvV_CO_1<0.1)
                    & (lep_Mtrktrk_atConvV_CO_1>=0)
                    & (lep_RadiusCO_1>20)))))) )
& ( nTaus_OR==1 ) & ( nJets_OR_DL1r_85>=1 ) & ( nJets_OR>=4 )
& ( ((dilep_type==2) ) | ( abs(Mll01-91.2e3)>10e3))
```

A.2 Feature Importances

feature name	mean	std
MLepMet	1.204666	0.519208
taus_pt_0	1.008721	0.430085
sumPsbtag	0.893841	0.585750
lep1_tau_Phi_diff_cos	0.731140	0.213217
MtLepMet	0.483445	0.360567
jet_pt0	0.470068	0.371581
taus_width_0	0.445110	0.185848
HT_jets	0.409812	0.196070
jet_0_1_phi_diff_cos	0.400350	0.160209
Mll01	0.382739	0.263802
DeltaR_min_lep_jet_fwd	0.356089	0.199653
HT	0.340320	0.191305
taus_RNNJetScoreSigTrans_0	0.315878	0.131832
dEta_maxMjj_frwdjet	0.295173	0.255849
lep_Z0SinTheta_1	0.267763	0.100456
minDeltaR_LJ_1	0.245780	0.120746
minDeltaR_LJ_0	0.244420	0.167706
taus_eta_0	0.242310	0.064653
lep_Eta_1	0.233946	0.101889
DeltaR_min_lep_jet	0.230378	0.173996
lep_E_1	0.228396	0.115587
jet_e3	0.213672	0.110805
jet_pt2	0.206834	0.094651
jet_phi0	0.199095	0.080845
lep0_tau_Phi_diff_cos	0.185602	0.070840
max_eta	0.178412	0.148207
HT_inclFwdJets	0.170817	0.188063
jet_e4	0.163412	0.113039
lep1_tau_Phi_diff_sin	0.160184	0.079864
jet_phi3	0.157790	0.091352
nJets_OR	0.154386	0.089577
jet_2_3_phi_diff_sin	0.153142	0.125117
jet_eta0	0.146965	0.087675
met_met	0.143584	0.113738
taus_JetRNNSigTight_0	0.133116	0.062966
lep_tau_opposite_charge	0.129573	0.162534
lep_Z0SinTheta_0	0.128401	0.071509
jet_0_1_phi_diff_sin	0.126915	0.057736
met_phi	0.122310	0.055427
lep_E_0	0.120263	0.124146
jet_4_5_phi_diff_cos	0.112262	0.085581
lep_0_jet_0_diff_cos	0.110264	0.079467
jet_4_5_phi_diff_sin	0.108127	0.043194
jet_eta3	0.106398	0.058805
lep_Pt_1	0.104907	0.080214
lep_Pt_0	0.101797	0.064730
minDeltaR_LJ_2	0.101389	0.089827
jet_3_4_phi_diff_sin	0.097972	0.077278
jet_pt4	0.095206	0.078774
jet_e1	0.086095	0.054609
lep_Phi_1	0.083563	0.059243
lep_Phi_0	0.079984	0.044790
jet_phi2	0.075939	0.062576
jet_2_3_phi_diff_cos	0.075861	0.063870
Ptll01	0.075385	0.078555
taus_phi_0	0.074004	0.102417
jet_phi1	0.071491	0.072050
lep_Phi_diff_cos	0.071047	0.085204
lep_1_jet_0_diff_cos	0.070558	0.053512
DRll01	0.068938	0.092523
jet_e0	0.066847	0.057849
lep_Phi_diff_sin	0.066205	0.065122
lep_EtaBE2_1	0.063605	0.091847
jet_eta2	0.058937	0.057114
eta_frwdjet	0.058780	0.064995
jet_1_2_phi_diff_sin	0.056033	0.072168
lep_EtaBE2_0	0.047071	0.042403
lep_1_jet_0_diff_sin	0.044687	0.054415
jet_1_2_phi_diff_cos	0.034687	0.104052
nTaus_OR_Pt25	0.033282	0.080073
jet_e5	0.031225	0.034053
taus_charge_0	0.028225	0.038833
jet_eta5	0.027397	0.035993
jet_pt5	0.026283	0.033325
jet_e2	0.022814	0.058528
jet_pt3	0.022081	0.064260
lep_Eta_0	0.019188	0.088002
taus_decayMode_2.0	0.018617	0.033386
HT_lep	0.017839	0.106307
jet_3_4_phi_diff_cos	0.014893	0.050558
lep_nTrackParticles_1	0.013208	0.053871
nFwdJets_OR	0.009860	0.024082
jet_phi4	0.006178	0.027946
taus_fromPV_0	0.005804	0.017575
HT_fwdJets	0.005379	0.022044
jet_eta4	0.004871	0.071016
taus_decayMode_1.0	0.000947	0.000972
lep_0_jet_0_diff_sin	0.000532	0.055584
lep_1_is_muon	0.000471	0.001657
taus_decayMode_3.0	0.000000	0.000000
taus_decayMode_4.0	0.000000	0.000000
taus_decayMode_0.0	0.000000	0.000000
lep_0_is_muon	0.000000	0.000000
taus_decayMode_6.0	0.000000	0.000000
taus_passJVT_0	0.000000	0.000000
taus_passEleOLR_0	0.000000	0.000000
taus_numTrack_0	-0.000560	0.002712
lep_nTrackParticles_0	-0.008264	0.036919
jet_phi5	-0.016534	0.058973
taus_DL1r_0	-0.017680	0.079668
mjjMax_frwdJet	-0.019912	0.037627
jet_pt1	-0.040167	0.121113
jet_eta1	-0.043942	0.046512
lep0_tau_Phi_diff_sin	-0.044328	0.033088

Table A.1 300 GeV model feature importances

feature name	mean	std	feature name	mean	std
taus_pt_0	3.027258	0.573703	lep_0_is_muon	0.000000	0.000000
HT_lep	2.198179	0.362655	jet_4_5_phi_diff_cos	0.000000	0.000000
MtLepMet	1.733756	0.711090	lep_0_jet_0_diff_cos	0.000000	0.000000
lep_Phi_diff_cos	1.728527	0.638963	jet_4_5_phi_diff_sin	0.000000	0.000000
taus_charge_0	1.623596	0.581598	jet_3_4_phi_diff_sin	0.000000	0.000000
lep0_tau_Phi_diff_cos	0.763291	0.233275	jet_2_3_phi_diff_sin	0.000000	0.000000
mjjMax_frwdJet	0.756985	0.257327	jet_1_2_phi_diff_sin	0.000000	0.000000
Mll01	0.698736	0.201294	lep_nTrackParticles_1	0.000000	0.000000
HT	0.685869	0.300255	taus_passEleOLR_0	0.000000	0.000000
taus_width_0	0.595840	0.290110	DeltaR_min_lep_jet	0.000000	0.000000
lep1_tau_Phi_diff_cos	0.282686	0.344255	lep_nTrackParticles_0	0.000000	0.000000
DRll01	0.130723	0.091047	jet_pt4	0.000000	0.000000
met_met	0.104296	0.053771	jet_phi4	0.000000	0.000000
jet_0_1_phi_diff_sin	0.081140	0.039321	jet_phi3	0.000000	0.000000
jet_e3	0.071902	0.056585	jet_phi1	0.000000	0.000000
lep_1_jet_0_diff_sin	0.064105	0.071973	jet_e5	0.000000	0.000000
lep0_tau_Phi_diff_sin	0.058734	0.047207	jet_e4	0.000000	0.000000
HT_jets	0.054727	0.051226	jet_e0	0.000000	0.000000
lep_tau_opposite_charge	0.053503	0.142050	jet_pt5	0.000000	0.000000
minDeltaR_LJ_2	0.038387	0.037697	jet_pt3	0.000000	0.000000
dEta_maxMjj_frwdjet	0.027189	0.012067	sumPsbtag	0.000000	0.000000
jet_eta0	0.024812	0.079171	jet_eta5	0.000000	0.000000
jet_pt1	0.024245	0.151940	jet_eta4	0.000000	0.000000
lep_Eta_0	0.020416	0.197564	jet_eta3	0.000000	0.000000
lep_1_jet_0_diff_cos	0.014734	0.072746	jet_eta1	0.000000	0.000000
jet_e2	0.012873	0.051629	HT_inclFwdJets	0.000000	0.000000
jet_phi5	0.011395	0.027232	HT_fwdJets	0.000000	0.000000
jet_phi2	0.008115	0.010461	eta_frwdjet	0.000000	0.000000
jet_eta2	0.006781	0.035808	lep_E_0	0.000000	0.000000
taus_RNNJetScoreSigTrans_0	0.005846	0.008278	taus_decayMode_6.0	0.000000	0.000000
lep_E_1	0.003533	0.089856	lep_Z0SinTheta_1	0.000000	0.000000
jet_3_4_phi_diff_cos	0.002891	0.003251	DeltaR_min_lep_jet_fwd	0.000000	0.000000
met_phi	0.000762	0.073631	lep_EtaBE2_1	0.000000	0.000000
jet_pt0	0.000739	0.043079	nTaus_OR_Pt25	0.000000	0.000000
jet_2_3_phi_diff_cos	0.000546	0.009158	lep_Phi_0	0.000000	0.000000
jet_phi0	0.000495	0.000213	lep_Phi_1	0.000000	0.000000
jet_pt2	0.000210	0.001149	nJets_OR	0.000000	0.000000
taus_JetRNNSigTight_0	0.000000	0.000000	lep_Pt_1	0.000000	0.000000
taus_numTrack_0	0.000000	0.000000	lep_Z0SinTheta_0	0.000000	0.000000
taus_decayMode_3.0	0.000000	0.000000	max_eta	0.000000	0.000000
taus_decayMode_1.0	0.000000	0.000000	nFwdJets_OR	0.000000	0.000000
taus_decayMode_2.0	0.000000	0.000000	MLepMet	0.000000	0.000000
taus_eta_0	0.000000	0.000000	jet_e1	-0.000246	0.002704
taus_fromPV_0	0.000000	0.000000	jet_0_1_phi_diff_cos	-0.004231	0.053859
taus_decayMode_4.0	0.000000	0.000000	lep_0_jet_0_diff_sin	-0.006969	0.016299
taus_phi_0	0.000000	0.000000	lep_EtaBE2_0	-0.014520	0.066667
taus_decayMode_0.0	0.000000	0.000000	Ptll01	-0.037246	0.035838
taus_passJVT_0	0.000000	0.000000	minDeltaR_LJ_1	-0.054835	0.090502
lep_1_is_muon	0.000000	0.000000	lep1_tau_Phi_diff_sin	-0.055395	0.034278
taus_DL1r_0	0.000000	0.000000	lep_Eta_1	-0.061548	0.062033
jet_1_2_phi_diff_cos	0.000000	0.000000	lep_Pt_0	-0.228115	0.118430
lep_Phi_diff_sin	0.000000	0.000000	minDeltaR_LJ_0	-0.292790	0.181685

Table A.2 800 GeV model feature importances

feature name	mean	std	feature name	mean	std
MtLepMet	10.715283	1.699835	jet_phi2	0.072993	0.123990
taus_pt_0	8.449437	0.973712	lep_E_1	0.068996	0.237466
lep1_tau_Phi_diff_cos	3.110103	0.785966	jet_4_5_phi_diff_sin	0.065326	0.082103
HT_inclFwdJets	1.575046	0.509917	jet_eta0	0.063077	0.144645
lep0_tau_Phi_diff_cos	1.574229	0.360287	jet_phi1	0.061180	0.076876
jet_e0	1.139809	0.269509	taus_decayMode_2.0	0.061174	0.127476
lep0_tau_Phi_diff_sin	1.052009	0.473405	jet_e4	0.051880	0.105878
MLepMet	1.047756	0.187420	jet_eta4	0.049436	0.077477
met_met	1.044287	0.273011	jet_eta1	0.049142	0.077025
taus_charge_0	1.034181	0.462671	taus_DL1r_0	0.048809	0.076554
lep_Pt_0	0.950642	0.229079	met_phi	0.041748	0.074200
lep_Phi_diff_cos	0.741358	0.302607	jet_pt5	0.024718	0.060369
max_eta	0.697128	0.419684	jet_0_1_phi_diff_cos	0.024399	0.104762
HT	0.689389	0.577100	jet_phi3	0.023835	0.058212
HT_lep	0.683032	0.237516	lep_Phi_0	0.018877	0.087607
DeltaR_min_lep_jet	0.505881	0.200264	lep_EtaBE2_0	0.008211	0.240807
Mll01	0.493561	0.244676	taus_phi_0	0.005659	0.060446
lep1_tau_Phi_diff_sin	0.474708	0.808762	jet_eta2	0.002402	0.093643
taus_width_0	0.461149	0.313287	lep_0_is_muon	0.000000	0.000000
DeltaR_min_lep_jet_fwd	0.456321	0.222852	taus_decayMode_4.0	0.000000	0.000000
HT_jets	0.437248	0.195598	taus_decayMode_0.0	0.000000	0.000000
mjjMax_frwdJet	0.370948	0.207650	taus_decayMode_3.0	0.000000	0.000000
jet_e1	0.369101	0.289666	eta_frwdjet	0.000000	0.000000
DRll01	0.356922	0.335037	taus_decayMode_1.0	0.000000	0.000000
lep_Phi_1	0.351716	0.237533	jet_4_5_phi_diff_cos	0.000000	0.000000
lep_Eta_0	0.347595	0.426330	lep_1_is_muon	0.000000	0.000000
taus_RNNJetScoreSigTrans_0	0.332905	0.262329	taus_passEleOLR_0	0.000000	0.000000
jet_e2	0.330293	0.153850	HT_fwdJets	0.000000	0.000000
jet_pt0	0.296598	0.087400	lep_nTrackParticles_0	0.000000	0.000000
minDeltaR_LJ_0	0.286989	0.291209	jet_e5	0.000000	0.000000
taus_eta_0	0.265660	0.172754	jet_phi5	0.000000	0.000000
lep_E_0	0.249435	0.194823	jet_eta5	0.000000	0.000000
jet_pt1	0.238495	0.133509	nFwdJets_OR	0.000000	0.000000
sumPsbtag	0.238267	0.198148	taus_passJVT_0	0.000000	0.000000
minDeltaR_LJ_1	0.238215	0.248822	nTaus_OR_Pt25	0.000000	0.000000
jet_1_2_phi_diff_cos	0.237665	0.116224	nJets_OR	0.000000	0.000000
lep_Eta_1	0.218804	0.155981	taus_fromPV_0	0.000000	0.000000
jet_phi0	0.190458	0.137340	taus_JetRNNSigTight_0	0.000000	0.000000
dEta_maxMjj_frwdjet	0.166433	0.221035	taus_numTrack_0	0.000000	0.000000
lep_Phi_diff_sin	0.159833	0.254794	taus_decayMode_6.0	0.000000	0.000000
jet_1_2_phi_diff_sin	0.149189	0.074703	jet_2_3_phi_diff_cos	-0.002277	0.147958
jet_e3	0.146590	0.176247	lep_Z0SinTheta_1	-0.022189	0.118975
jet_2_3_phi_diff_sin	0.130153	0.122105	jet_pt2	-0.045257	0.157184
lep_1_jet_0_diff_sin	0.117234	0.174617	lep_nTrackParticles_1	-0.081830	0.083956
jet_pt3	0.105715	0.172181	jet_3_4_phi_diff_cos	-0.089895	0.139071
minDeltaR_LJ_2	0.103570	0.144179	jet_pt4	-0.091104	0.083551
lep_0_jet_0_diff_sin	0.090963	0.100504	jet_phi4	-0.096007	0.081233
jet_3_4_phi_diff_sin	0.086512	0.108708	lep_tau_opposite_charge	-0.108422	0.121204
lep_0_jet_0_diff_cos	0.084298	0.126115	lep_1_jet_0_diff_cos	-0.137783	0.094847
jet_eta3	0.082394	0.084534	lep_EtaBE2_1	-0.142432	0.144104
lep_Pt_1	0.076758	0.218631	Ptll01	-0.143678	0.120547
lep_Z0SinTheta_0	0.073573	0.084650	jet_0_1_phi_diff_sin	-0.253841	0.067365

Table A.3 1500 GeV model feature importances

feature name	mean	std	feature name	mean	std
taus_pt_0	8.615530	1.720347	taus_passEleOLR_0	-0.014623	0.065395
DRll01	3.752879	1.239824	nFwdJets_OR	-0.029527	0.293938
jet_pt1	2.888815	1.199397	lep_0_jet_0_diff_cos	-0.036898	0.385761
MLepMet	2.195968	1.167438	lep_Phi_0	-0.037205	0.300278
lep_EtaBE2_1	1.711264	0.745173	minDeltaR_LJ_0	-0.040662	0.503578
lep_E_1	1.042407	0.548502	nJets_OR	-0.042281	0.385127
sumPsbtag	0.995894	0.763880	taus_decayMode_0.0	-0.043868	0.107140
jet_e0	0.946223	0.665401	taus_decayMode_2.0	-0.045721	0.151161
taus_numTrack_0	0.742801	0.429691	taus_RNNJetScoreSigTrans_0	-0.046069	0.112515
HT_jets	0.723015	0.485177	jet_eta1	-0.058491	0.120021
Mll01	0.713279	0.413549	lep_1_jet_0_diff_cos	-0.061527	0.126251
mjjMax_frwdJet	0.569055	0.335925	lep_Z0SinTheta_1	-0.067481	0.157983
lep_nTrackParticles_1	0.547799	0.272317	taus_decayMode_3.0	-0.073114	0.129927
minDeltaR_LJ_2	0.517844	0.541345	jet_2_3_phi_diff_cos	-0.076782	0.136444
jet_phi3	0.390835	0.504706	nTaus_OR_Pt25	-0.091217	0.239497
minDeltaR_LJ_1	0.353564	0.538258	lep0_tau_Phi_diff_cos	-0.092000	0.241645
jet_pt0	0.323973	0.415357	lep_Eta_0	-0.098619	0.612710
jet_e2	0.280397	0.408662	lep_Phi_diff_sin	-0.110258	0.359199
DeltaR_min_lep_jet	0.224285	0.287723	taus_DL1r_0	-0.130702	0.236693
jet_0_1_phi_diff_sin	0.191185	0.356540	jet_0_1_phi_diff_cos	-0.131206	0.208164
lep_0_is_muon	0.157038	0.161118	taus_decayMode_1.0	-0.131605	0.149275
jet_eta4	0.149207	0.434847	lep_E_0	-0.140741	0.447463
lep_0_jet_0_diff_sin	0.126241	0.237290	max_eta	-0.177617	0.427983
taus_eta_0	0.120991	0.220898	lep1_tau_Phi_diff_sin	-0.181202	0.251129
taus_JetRNNSigTight_0	0.115757	0.226690	jet_eta0	-0.197470	0.207073
taus_charge_0	0.109927	0.153696	jet_pt5	-0.199699	0.354427
lep_Z0SinTheta_0	0.103367	0.233886	jet_e1	-0.206448	0.292721
lep_1_is_muon	0.102853	0.198942	met_phi	-0.227810	0.320667
dEta_maxMjj_frwdjet	0.087330	0.371171	eta_frwdjet	-0.232698	0.257725
jet_e3	0.083058	0.219595	lep_EtaBE2_0	-0.240449	0.271309
jet_e4	0.081270	0.241907	jet_3_4_phi_diff_cos	-0.240478	0.252233
DeltaR_min_lep_jet_fwd	0.076477	0.323092	lep0_tau_Phi_diff_sin	-0.245567	0.238571
jet_pt4	0.054296	0.257118	taus_phi_0	-0.248360	0.269206
lep_Pt_0	0.044547	0.348978	jet_2_3_phi_diff_sin	-0.257526	0.376637
HT_fwdJets	0.032716	0.263313	lep_tau_opposite_charge	-0.276414	0.094531
taus_width_0	0.028931	0.307223	HT_lep	-0.292456	0.000000
jet_eta3	0.012535	0.532795	jet_pt2	-0.293396	0.097979
jet_phi4	0.000000	0.000000	jet_e5	-0.299406	0.339448
jet_3_4_phi_diff_sin	0.000000	0.000000	jet_eta2	-0.314715	0.207428
taus_decayMode_4.0	0.000000	0.000000	Ptll01	-0.315171	0.259804
jet_4_5_phi_diff_cos	0.000000	0.000000	lep_Pt_1	-0.352309	0.333186
taus_fromPV_0	0.000000	0.000000	jet_phi2	-0.358609	0.159542
jet_phi1	0.000000	0.000000	jet_phi0	-0.395453	0.512549
jet_4_5_phi_diff_sin	0.000000	0.000000	jet_eta5	-0.399272	0.217492
taus_decayMode_6.0	0.000000	0.000000	jet_1_2_phi_diff_sin	-0.448631	0.249804
lep_1_jet_0_diff_sin	0.000000	0.000000	jet_pt3	-0.450471	0.362664
taus_passJVT_0	0.000000	0.000000	lep_Eta_1	-0.465609	0.181855
lep_Phi_diff_cos	0.000000	0.000000	lep_nTrackParticles_0	-0.526188	0.322363
jet_phi5	0.000000	0.000000	HT	2.346587	1.348885
lep1_tau_Phi_diff_cos	0.000000	0.000000	HT_inclFwdJets	2.484161	1.466220
lep_Phi_1	-0.008853	0.311810	met_met	4.562813	1.673473
jet_1_2_phi_diff_cos	-0.009042	0.245671	MtLepMet	5.340643	2.091216

Table A.4 2000 GeV model feature importances

feature name	mean	std	feature name	mean	std
MLepMet	2.418832	0.674713	HT_lep	0.010380	0.034175
HT_jets	1.035409	0.225443	jet_3_4_phi_diff_cos	0.009770	0.019542
sumPsbtag	0.489840	0.152324	taus_fromPV_0	0.009032	0.023512
lep_tau_opposite_charge	0.471623	0.221563	jet_eta5	0.008930	0.009872
DeltaR_min_lep_jet_fwd	0.417646	0.118190	lep0_tau_Phi_diff_sin	0.008361	0.012988
jet_pt0	0.334748	0.174384	jet_2_3_phi_diff_sin	0.005898	0.014034
minDeltaR_LJ_0	0.312349	0.101245	lep_Phi_diff_sin	0.004230	0.022454
taus_pt_0	0.246662	0.186970	lep_Pt_1	0.002785	0.009855
HT	0.226881	0.091848	lep_Phi_0	0.001098	0.013259
DeltaR_min_lep_jet	0.204507	0.067271	lep_Pt_0	0.000860	0.000371
taus_charge_0	0.181631	0.118237	lep_Z0SinTheta_1	0.000304	0.000799
taus_width_0	0.169516	0.102500	jet_phi0	0.000248	0.000182
HT_inclFwdJets	0.156448	0.149894	jet_pt5	0.000226	0.000282
DRll01	0.156105	0.083876	lep1_tau_Phi_diff_sin	0.000150	0.000669
MtLepMet	0.129373	0.086069	jet_e1	0.000126	0.000117
lep1_tau_Phi_diff_cos	0.120296	0.044008	HT_fwdJets	0.000018	0.000079
dEta_maxMjj_frwdjet	0.112454	0.054778	lep_1_jet_0_diff_cos	0.000000	0.000000
jet_0_1_phi_diff_cos	0.102652	0.071930	jet_pt3	0.000000	0.000000
lep_0_jet_0_diff_cos	0.075275	0.030755	taus_decayMode_3.0	0.000000	0.000000
jet_pt1	0.075047	0.057822	jet_eta4	0.000000	0.000000
lep0_tau_Phi_diff_cos	0.070173	0.039476	taus_decayMode_2.0	0.000000	0.000000
jet_phi3	0.060823	0.056990	taus_decayMode_4.0	0.000000	0.000000
jet_phi1	0.055549	0.046414	lep_0_jet_0_diff_sin	0.000000	0.000000
lep_1_jet_0_diff_sin	0.054632	0.027337	lep_0_is_muon	0.000000	0.000000
jet_eta0	0.053854	0.028168	lep_1_is_muon	0.000000	0.000000
taus_eta_0	0.046255	0.043002	taus_decayMode_0.0	0.000000	0.000000
lep_Eta_0	0.044236	0.014752	taus_decayMode_1.0	0.000000	0.000000
max_eta	0.043144	0.044614	jet_eta1	0.000000	0.000000
jet_3_4_phi_diff_sin	0.041417	0.035220	minDeltaR_LJ_2	0.000000	0.000000
Ptll01	0.039420	0.036971	jet_e2	0.000000	0.000000
jet_eta2	0.035702	0.019254	taus_JetRNNSigTight_0	0.000000	0.000000
lep_EtaBE2_1	0.034359	0.040291	nFwdJets_OR	0.000000	0.000000
met_phi	0.033932	0.018697	nJets_OR	0.000000	0.000000
Mll01	0.032285	0.035935	nTaus_OR_Pt25	0.000000	0.000000
minDeltaR_LJ_1	0.030965	0.023792	lep_nTrackParticles_0	0.000000	0.000000
lep_nTrackParticles_1	0.030595	0.016972	taus_passEleOLR_0	0.000000	0.000000
jet_pt2	0.029710	0.046383	jet_phi5	0.000000	0.000000
jet_pt4	0.027061	0.025978	taus_decayMode_6.0	0.000000	0.000000
lep_E_1	0.027033	0.036913	taus_numTrack_0	0.000000	0.000000
jet_4_5_phi_diff_cos	0.022852	0.036502	jet_phi4	0.000000	0.000000
taus_RNNJetScoreSigTrans_0	0.021823	0.031102	taus_passJVT_0	0.000000	0.000000
jet_1_2_phi_diff_cos	0.017870	0.041328	lep_Eta_1	-0.000191	0.059669
eta_frwdjet	0.014546	0.033648	jet_4_5_phi_diff_sin	-0.000282	0.000197
jet_eta3	0.014542	0.020552	jet_e4	-0.000343	0.000848
jet_1_2_phi_diff_sin	0.014237	0.019222	lep_E_0	-0.000569	0.000878
jet_e3	0.013966	0.027236	jet_2_3_phi_diff_cos	-0.002302	0.018232
jet_phi2	0.013092	0.017023	jet_0_1_phi_diff_sin	-0.003285	0.002433
taus_DL1r_0	0.012206	0.033058	jet_e5	-0.007172	0.014183
met_met	0.011868	0.056822	lep_Phi_diff_cos	-0.009912	0.021144
jet_e0	0.011377	0.050886	lep_EtaBE2_0	-0.020465	0.022503
mjjMax_frwdJet	0.010774	0.018371	lep_Phi_1	-0.033422	0.026443
lep_Z0SinTheta_0	0.010508	0.019220	taus_phi_0	-0.040058	0.009087

Table A.5 New 250 GeV model feature importances

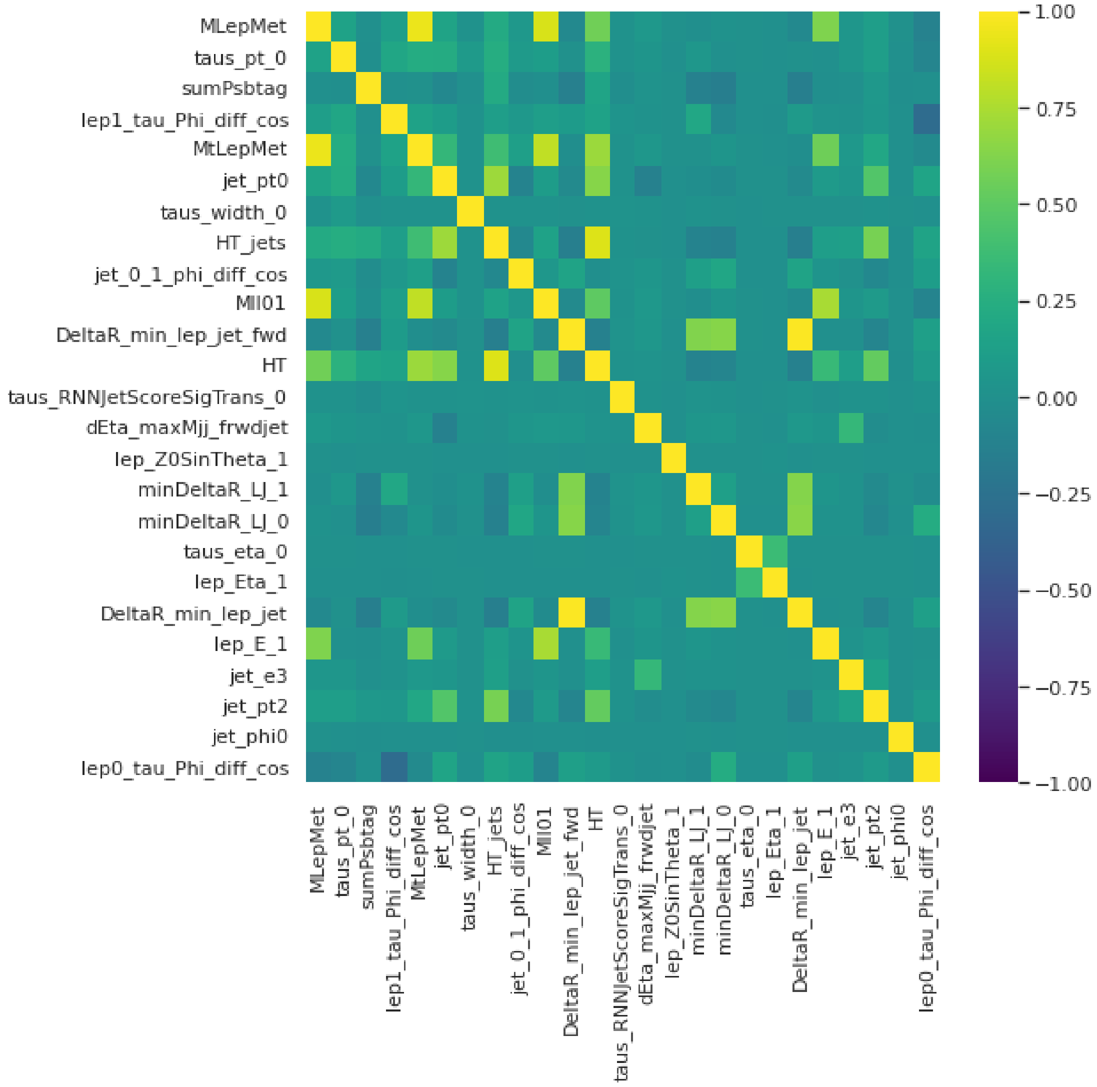
feature name	mean	std	feature name	mean	std
MtLepMet	4.253547	1.223928	met_phi	0.008890	0.051758
taus_pt_0	3.543661	0.618822	taus_passEleOLR_0	0.007248	0.032261
Mll01	2.498689	0.775792	taus_decayMode_1.0	0.007157	0.005388
taus_charge_0	1.940902	0.608596	mjjMax_frwdJet	0.003887	0.208944
lep1_tau_Phi_diff_cos	1.871623	0.495560	taus_decayMode_6.0	0.000000	0.000000
lep_Phi_diff_cos	1.808359	0.707600	taus_decayMode_0.0	0.000000	0.000000
DRll01	1.765589	0.651490	lep_1_is_muon	0.000000	0.000000
met_met	1.374030	0.783516	taus_passJVT_0	0.000000	0.000000
lep_Eta_1	1.216830	0.315184	taus_decayMode_3.0	0.000000	0.000000
HT_inclFwdJets	1.206561	0.684414	taus_decayMode_4.0	0.000000	0.000000
HT	1.127910	0.603420	jet_4_5_phi_diff_sin	-0.000554	0.038870
lep_Pt_0	1.046766	0.377972	lep_nTrackParticles_1	-0.003080	0.011554
MLepMet	0.966165	0.505554	taus_fromPV_0	-0.003947	0.027795
taus_width_0	0.954986	0.650333	jet_phi1	-0.004349	0.082743
max_eta	0.853395	0.547894	lep_0_is_muon	-0.010819	0.008145
lep0_tau_Phi_diff_cos	0.829265	0.342363	taus_JetRNNSigTight_0	-0.012675	0.026893
HT_lep	0.761826	0.409014	jet_0_1_phi_diff_sin	-0.014023	0.048014
HT_jets	0.501460	0.417280	lep_nTrackParticles_0	-0.014056	0.036553
lep0_tau_Phi_diff_sin	0.407068	0.443335	taus_numTrack_0	-0.016893	0.032967
jet_e0	0.401048	0.227662	jet_3_4_phi_diff_cos	-0.018532	0.088004
sumPsbtag	0.370411	0.283221	lep_0_jet_0_diff_sin	-0.019597	0.078953
minDeltaR_LJ_1	0.250766	0.211469	jet_phi3	-0.019844	0.106839
jet_0_1_phi_diff_cos	0.197565	0.239095	nJets_OR	-0.020632	0.036114
DeltaR_min_lep_jet_fwd	0.181012	0.372854	jet_pt2	-0.024386	0.145480
nTaus_OR_Pt25	0.179716	0.244783	minDeltaR_LJ_2	-0.025154	0.117855
lep_tau_opposite_charge	0.175364	0.202908	lep_EtaBE2_0	-0.028563	0.138122
lep_Phi_0	0.171128	0.107597	jet_phi4	-0.028754	0.063964
lep1_tau_Phi_diff_sin	0.168244	0.223407	jet_pt5	-0.029698	0.103083
jet_pt0	0.155189	0.210403	lep_1_jet_0_diff_sin	-0.030926	0.081880
jet_e3	0.148571	0.092144	jet_eta0	-0.033001	0.116095
Ptll01	0.130187	0.234685	dEta_maxMjj_frwdjet	-0.038130	0.154607
jet_e4	0.117050	0.163406	HT_fwdJets	-0.042183	0.103673
jet_pt4	0.090766	0.090060	lep_EtaBE2_1	-0.043468	0.059673
jet_e5	0.085447	0.077553	jet_eta5	-0.046727	0.036081
lep_Pt_1	0.082733	0.434060	lep_Phi_1	-0.047367	0.051334
jet_eta1	0.065920	0.056580	jet_3_4_phi_diff_sin	-0.052107	0.071314
jet_4_5_phi_diff_cos	0.058590	0.090916	lep_Eta_0	-0.078416	0.133209
DeltaR_min_lep_jet	0.053645	0.244912	jet_1_2_phi_diff_cos	-0.078656	0.106325
nFwdJets_OR	0.051400	0.113956	jet_pt3	-0.078878	0.163509
lep_E_0	0.047518	0.315501	lep_1_jet_0_diff_cos	-0.082607	0.126492
eta_frwdjet	0.046237	0.043097	lep_0_jet_0_diff_cos	-0.090893	0.230160
lep_Z0SinTheta_0	0.031069	0.060082	jet_1_2_phi_diff_sin	-0.122530	0.074840
lep_E_1	0.030175	0.120215	taus_phi_0	-0.136058	0.123771
jet_phi5	0.029815	0.040056	jet_eta2	-0.137441	0.105901
taus_decayMode_2.0	0.028223	0.033272	minDeltaR_LJ_0	-0.151761	0.376662
jet_eta4	0.024698	0.054877	lep_Z0SinTheta_1	-0.168849	0.097048
jet_2_3_phi_diff_cos	0.024642	0.080818	jet_2_3_phi_diff_sin	-0.179997	0.075464
jet_phi0	0.022449	0.074093	taus_DL1r_0	-0.188369	0.109462
jet_eta3	0.019251	0.036946	lep_Phi_diff_sin	-0.203268	0.155079
jet_phi2	0.017606	0.112266	jet_e2	-0.211553	0.104913
taus_RNNJetScoreSigTrans_0	0.014583	0.200560	jet_pt1	-0.325665	0.140454
taus_eta_0	0.009694	0.193198	jet_e1	-0.345910	0.072104

Table A.6 New 800 GeV model feature importances

feature name	mean	std	feature name	mean	std
jet_pt1	3.210628	0.187634	jet_phi5	-0.241944	0.157023
taus_charge_0	1.838014	1.287685	lep_nTrackParticles_1	-0.286850	0.067517
DeltaR_min_lep_jet	1.759808	1.264008	dEta_maxMjj_frwdjet	-0.289487	0.191154
lep0_tau_Phi_diff_sin	1.019533	0.553750	jet_3_4_phi_diff_cos	-0.475499	0.189475
lep_Phi_diff_cos	0.999391	0.473570	jet_3_4_phi_diff_sin	-inf	NaN
lep1_tau_Phi_diff_cos	0.718046	0.383171	lep_0_jet_0_diff_cos	-inf	NaN
taus_width_0	0.562124	0.518252	lep_1_jet_0_diff_sin	-inf	NaN
lep1_tau_Phi_diff_sin	0.480318	0.397294	lep_Eta_0	-inf	NaN
lep_Phi_diff_sin	0.428976	0.415628	lep_E_0	-inf	NaN
jet_4_5_phi_diff_cos	0.257770	0.215380	jet_phi3	-inf	NaN
jet_phi2	0.180898	0.152634	jet_pt0	-inf	NaN
jet_eta4	0.179490	0.136367	jet_1_2_phi_diff_cos	-inf	NaN
jet_eta2	0.158635	0.261976	jet_2_3_phi_diff_sin	-inf	NaN
nTaus_OR_Pt25	0.106798	0.161459	jet_2_3_phi_diff_cos	-inf	NaN
met_phi	0.097367	0.171885	Mll01	-inf	NaN
jet_e4	0.093791	0.206445	lep_Eta_1	-inf	NaN
lep_0_jet_0_diff_sin	0.062564	0.153599	jet_phi1	-inf	NaN
minDeltaR_LJ_0	0.060917	0.441272	jet_phi0	-inf	NaN
minDeltaR_LJ_2	0.048797	0.174175	jet_e5	-inf	NaN
jet_4_5_phi_diff_sin	0.039574	0.096653	jet_pt2	-inf	NaN
lep_Phi_0	0.027810	0.094706	jet_e3	-inf	NaN
taus_decayMode_6.0	0.000000	0.000000	jet_e1	-inf	NaN
nFwdJets_OR	0.000000	0.000000	taus_decayMode_2.0	-inf	NaN
jet_eta5	0.000000	0.000000	jet_e0	-inf	NaN
taus_passEleOLR_0	0.000000	0.000000	jet_pt5	-inf	NaN
taus_fromPV_0	0.000000	0.000000	DRll01	-inf	NaN
taus_JetRNNSigTight_0	0.000000	0.000000	HT	-inf	NaN
taus_numTrack_0	0.000000	0.000000	MLepMet	-inf	NaN
taus_passJVT_0	0.000000	0.000000	lep0_tau_Phi_diff_cos	-inf	NaN
HT_fwdJets	0.000000	0.000000	mjjMax_frwdJet	-inf	NaN
nJets_OR	0.000000	0.000000	jet_eta3	-inf	NaN
taus_decayMode_0.0	0.000000	0.000000	Ptll01	-inf	NaN
lep_0_is_muon	0.000000	0.000000	sumPsbtag	-inf	NaN
taus_decayMode_4.0	0.000000	0.000000	lep_nTrackParticles_0	-inf	NaN
taus_decayMode_3.0	0.000000	0.000000	met_met	-inf	NaN
taus_decayMode_1.0	0.000000	0.000000	max_eta	-inf	NaN
lep_tau_opposite_charge	0.000000	0.000000	lep_Z0SinTheta_1	-inf	NaN
jet_0_1_phi_diff_cos	-0.004703	0.146598	jet_eta0	-inf	NaN
taus_phi_0	-0.035230	0.119436	lep_Z0SinTheta_0	-inf	NaN
jet_eta1	-0.057914	0.102915	taus_eta_0	-inf	NaN
eta_frwdjet	-0.081784	0.120240	taus_pt_0	-inf	NaN
lep_1_jet_0_diff_cos	-0.100217	0.244032	lep_Pt_1	-inf	NaN
jet_0_1_phi_diff_sin	-0.115024	0.152500	lep_Pt_0	-inf	NaN
lep_1_is_muon	-0.124278	0.140964	HT_lep	-inf	NaN
jet_pt4	-0.148602	0.112420	taus_DL1r_0	-inf	NaN
lep_E_1	-0.149568	0.237834	taus_RNNJetScoreSigTrans_0	-inf	NaN
jet_1_2_phi_diff_sin	-0.150974	0.154896	lep_EtaBE2_1	-inf	NaN
minDeltaR_LJ_1	-0.152091	0.253347	HT_jets	-inf	NaN
jet_phi4	-0.158298	0.155070	HT_inclFwdJets	-inf	NaN
DeltaR_min_lep_jet_fwd	-0.173770	0.323339	lep_EtaBE2_0	-inf	NaN
lep_Phi_1	-0.184521	0.143270	jet_pt3	-inf	NaN
jet_e2	-0.236792	0.236940	MtLepMet	-inf	NaN

Table A.7 New 3000 GeV model feature importances

A.3 Pearson Correlation Coefficients



■ **Figure A.1** Pearson correlation coefficients for 25 most important features of the 300 GeV mass model, correlations measured on full dataset with all signal masses and background

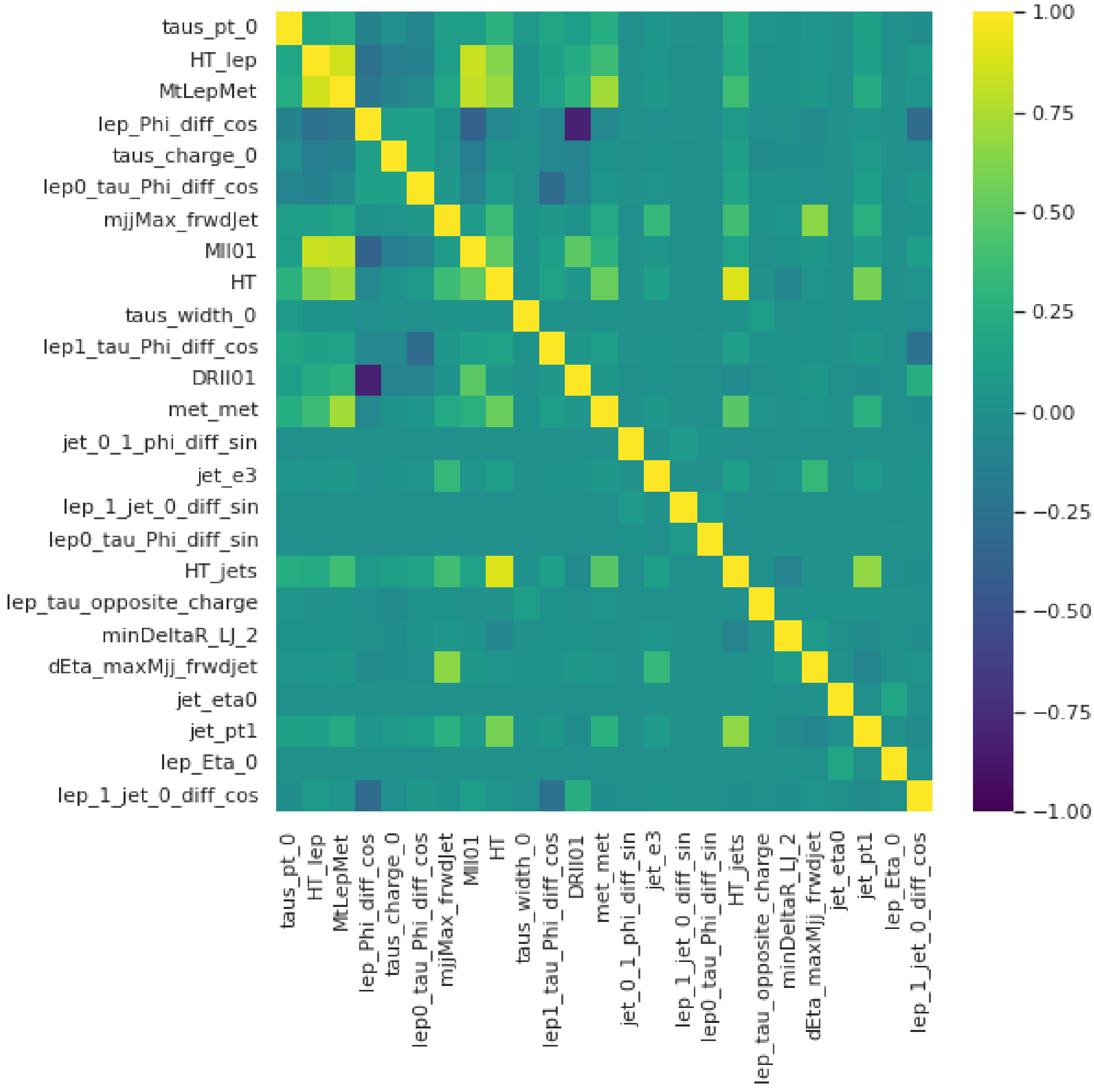
A.4 Feature reduction

Below are the features used to fit the feature reduction models. Three levels of feature reduction are used, matching up to three lines in the following description – 5 features (first line), 10 features (first and second line) and 20 (first, second and third line)

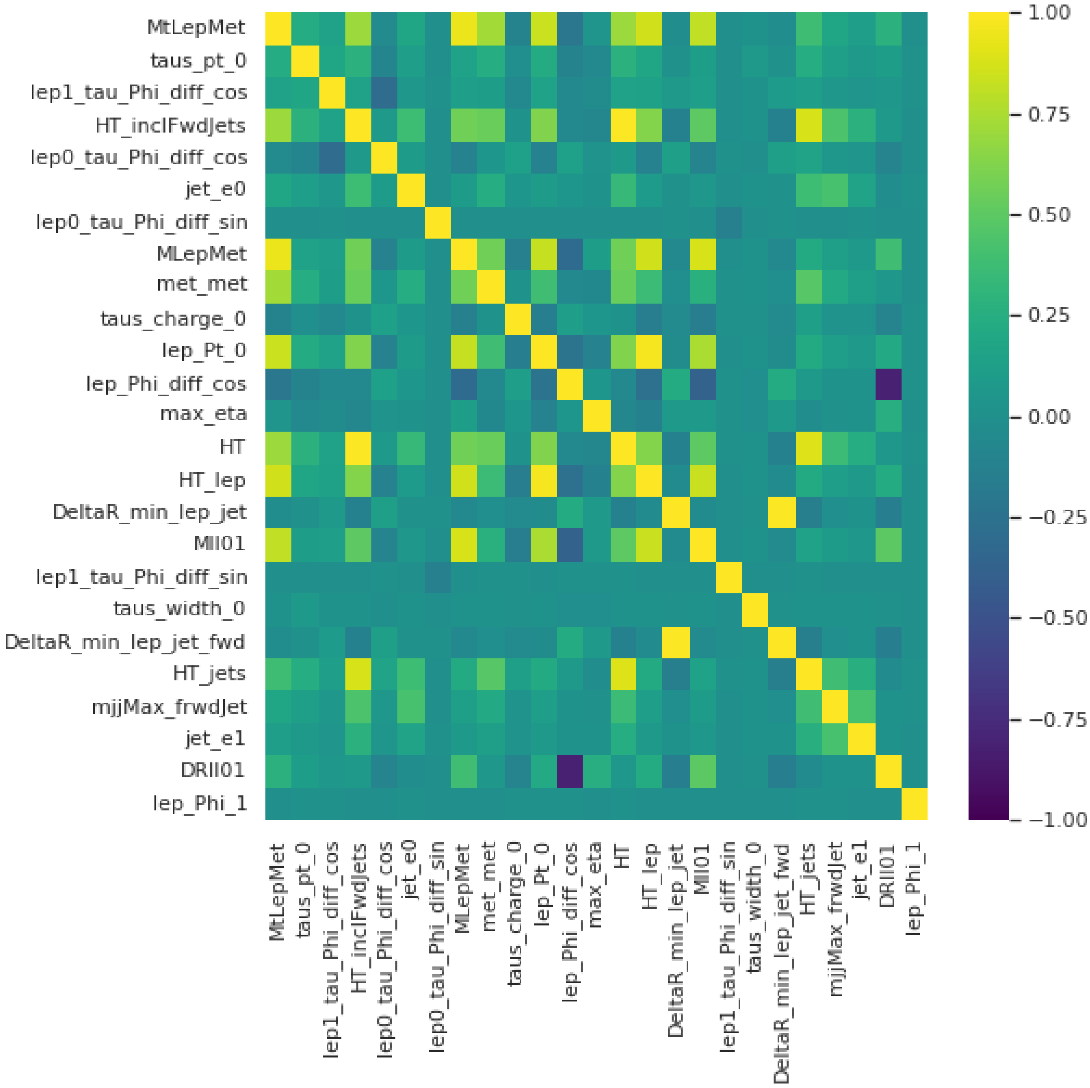
300 GeV model MLepMet, taus_pt_0, sumPsbtag, lep1_tau_Phi_diff_cos, jet_pt0

taus_width_0, jet_0_1_phi_diff_cos, DeltaR_min_lep_jet_fwd, taus_RNNJetScoreSigTrans_0, dEta_maxMjj_frwdjet

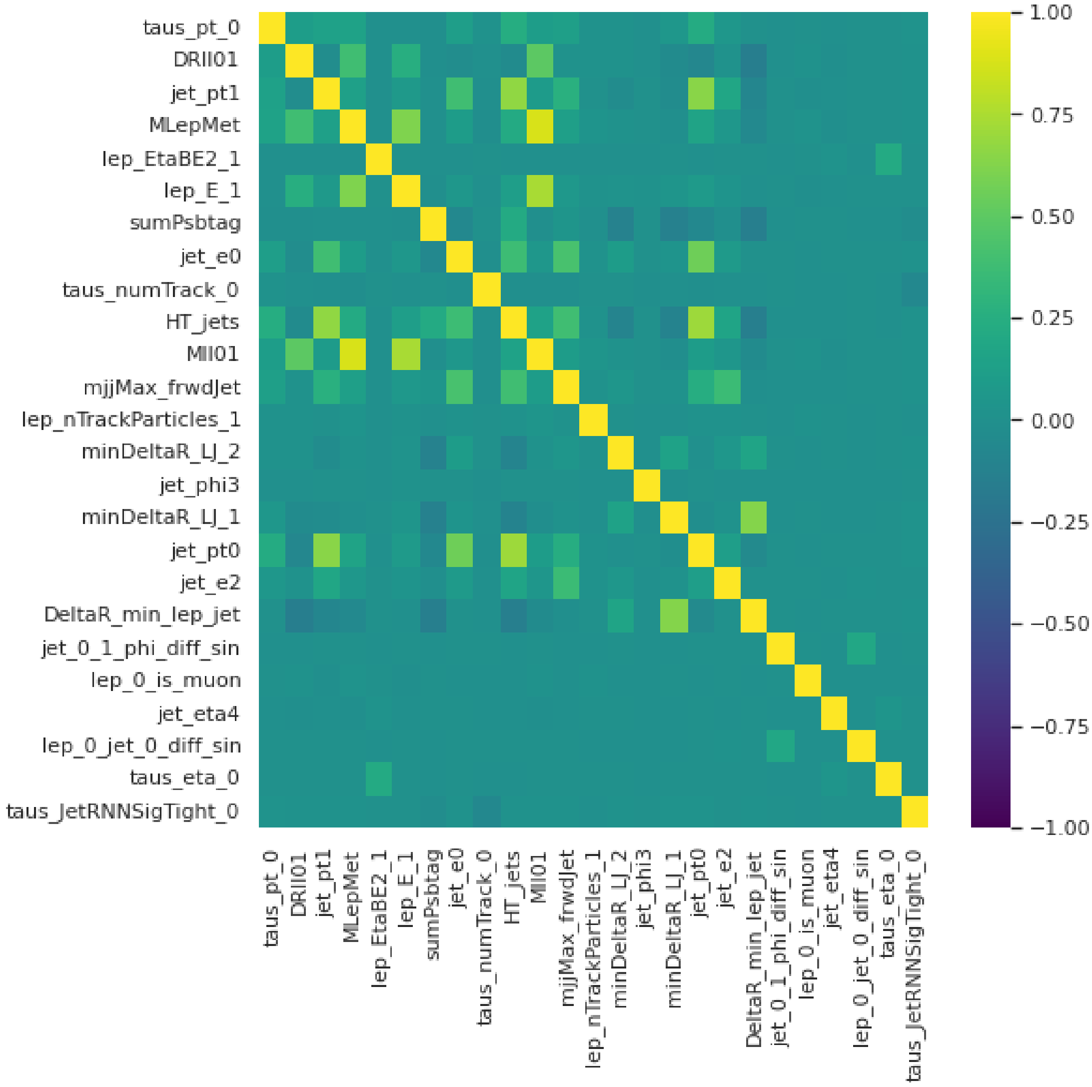
lep_Z0SinTheta_1, taus_eta_0, lep_Eta_1, jet_e3, jet_pt2, jet_phi0, lep0_tau_Phi_diff_cos, max_eta, jet_e4, lep1_tau_Phi_diff_sin



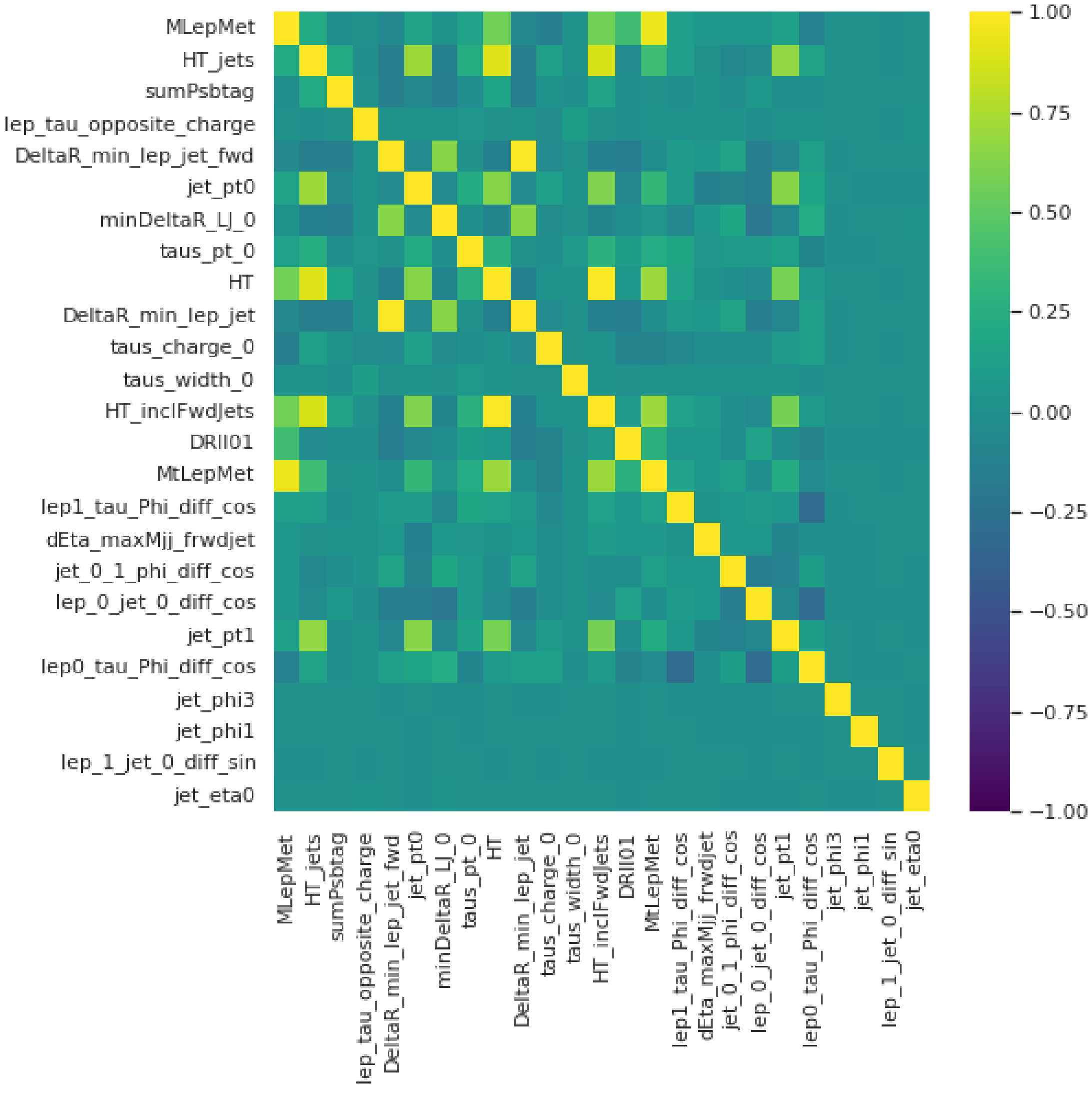
■ **Figure A.2** Pearson correlation coefficients for 25 most important features of the original 800 GeV mass model, correlations measured on full dataset with all signal masses and background



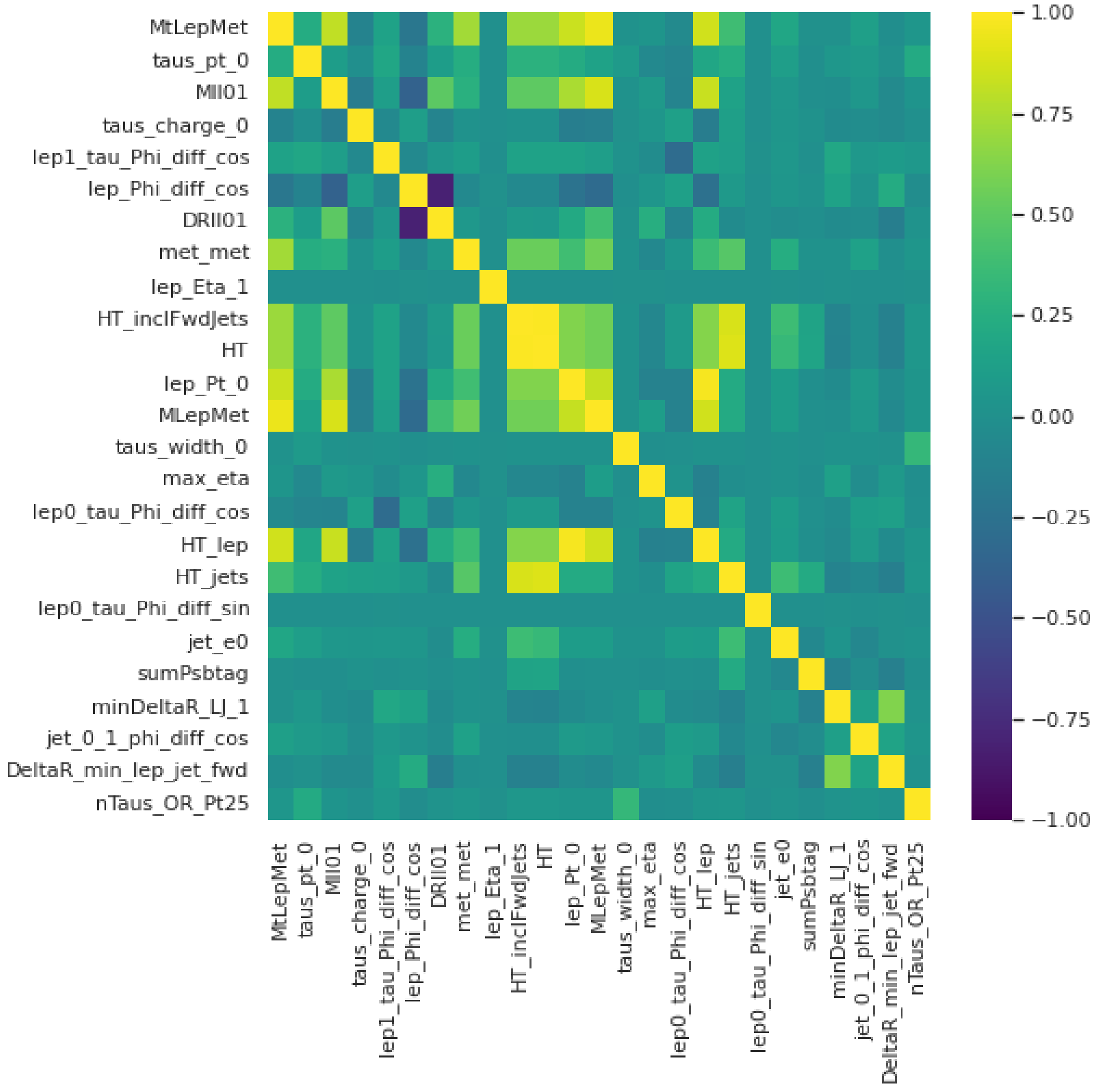
■ **Figure A.3** Pearson correlation coefficients for 25 most important features of the 1500 GeV mass model, correlations measured on full dataset with all signal masses and background



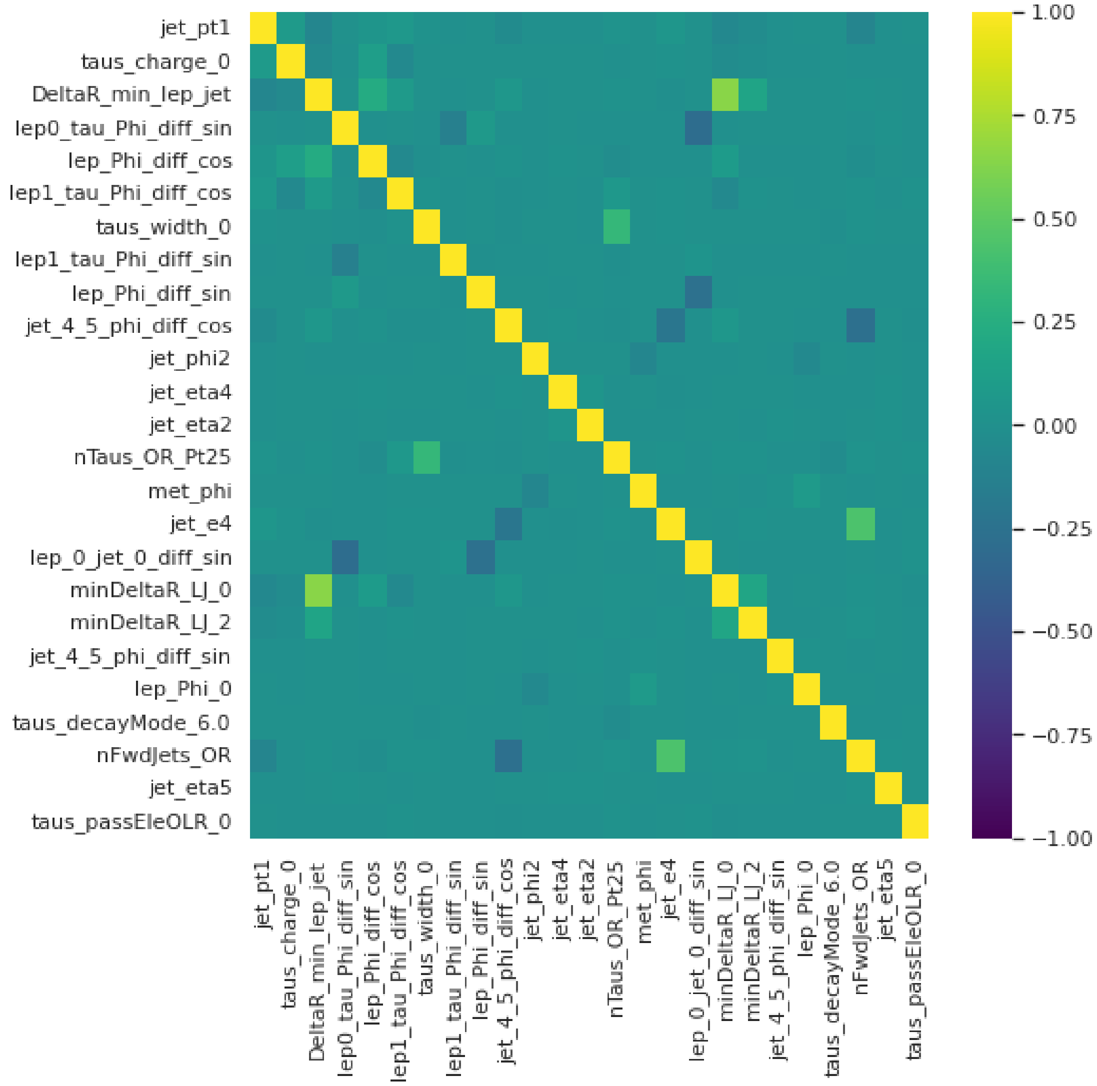
■ **Figure A.4** Pearson correlation coefficients for 25 most important features of the 2000 GeV mass model, correlations measured on full dataset with all signal masses and background



■ **Figure A.5** Pearson correlation coefficients for 25 most important features of the new 250 GeV mass model, correlations measured on full dataset with all signal masses and background



■ **Figure A.6** Pearson correlation coefficients for 25 most important features of the new 800 GeV mass model, correlations measured on full dataset with all signal masses and background



■ **Figure A.7** Pearson correlation coefficients for 25 most important features of the new 3000 GeV mass model, correlations measured on full dataset with all signal masses and background

800 GeV model taus_pt_0, HT_lep, lep_Phi_diff_cos, taus_charge_0, lep0_tau_Phi_diff_cos

mjjMax_frwdJet, taus_width_0, lep1_tau_Phi_diff_cos, met_met, jet_0_1_phi_diff_sin, jet_e3, lep_1_jet_0_diff_sin, lep0_tau_Phi_diff_sin, HT_jets, lep_tau_opposite_charge, minDeltaR_LJ_2, jet_eta0, lep_Eta_0, lep_1_jet_0_diff_cos, jet_e2

1500 GeV model MtLepMet, taus_pt_0, lep1_tau_Phi_diff_cos, lep0_tau_Phi_diff_cos, jet_e0

lep0_tau_Phi_diff_sin, taus_charge_0, lep_Phi_diff_cos, max_eta, DeltaR_min_lep_jet

lep1_tau_Phi_diff_sin, taus_width_0, HT_jets, mjjMax_frwdJet, jet_e1, lep_Phi_1, lep_Eta_0, taus_RNNJetScoreSigTrans_0, jet_e2, taus_eta_0

2000 GeV model taus_pt_0, DRll01, jet_pt1, MLepMet, lep_EtaBE2_1

sumPsbtag, jet_e0, taus_numTrack_0, mjjMax_frwdJet, lep_nTrackParticles_1

minDeltaR_LJ_2, jet_phi3, minDeltaR_LJ_1, jet_e2, jet_0_1_phi_diff_sin, lep_0_is_muon, jet_eta4, lep_0_jet_0_diff_sin, taus_eta_0, taus_JetRNNSigTight_0

New 250 GeV model MLepMet, HT_jets, sumPsbtag, lep_tau_opposite_charge, DeltaR_min_lep_jet_fwd

taus_pt_0, taus_charge_0, taus_width_0, DRll01, lep1_tau_Phi_diff_cos

dEta_maxMjj_frwdjet, jet_0_1_phi_diff_cos, lep_0_jet_0_diff_cos, lep0_tau_Phi_diff_cos, jet_phi3, jet_phi1, lep_1_jet_0_diff_sin, jet_eta0, taus_eta_0, lep_Eta_0

New 800 GeV model MtLepMet, taus_pt_0, taus_charge_0, lep1_tau_Phi_diff_cos, lep_Phi_diff_cos

lep_Eta_1, taus_width_0, max_eta, lep0_tau_Phi_diff_cos, HT_jets

lep0_tau_Phi_diff_sin, jet_e0, sumPsbtag, minDeltaR_LJ_1, jet_0_1_phi_diff_cos, nTaus_OR_Pt25, lep_tau_opposite_charge, lep_Phi_0, lep1_tau_Phi_diff_sin, jet_e3

New 3000 GeV model jet_pt1, taus_charge_0, DeltaR_min_lep_jet, lep0_tau_Phi_diff_sin, lep_Phi_diff_cos

lep1_tau_Phi_diff_cos, taus_width_0, lep1_tau_Phi_diff_sin, lep_Phi_diff_sin, jet_4_5_phi_diff_cos

jet_phi2, jet_eta4, jet_eta2, nTaus_OR_Pt25, met_phi, jet_e4, lep_0_jet_0_diff_sin, minDeltaR_LJ_2, jet_4_5_phi_diff_sin, lep_Phi_0

Bibliography

1. COLLABORATION, The ATLAS; AAD, G; ABAT, E; ABDALLAH, J; ABDELALIM, A A; AL., A Abdesselam et. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*. 2008, vol. 3, no. 08, S08003–S08003. ISSN 1748-0221. Available from DOI: `10.1088/1748-0221/3/08/S08003`.
2. BIANCHI, Riccardo Maria; COLLABORATION, ATLAS. *ATLAS experiment schematic illustration*. 2022. Available also from: `https://cds.cern.ch/record/2837191`. General Photo.
3. EYSERMANS, Jan; PEDRAZA, Isabel. Charged Higgs Analysis in CMS. *Journal of Physics: Conference Series*. 2016, vol. 761, p. 012030. Available from DOI: `10.1088/1742-6596/761/1/012030`.
4. TUMASYAN, A.; ADAM, W.; ANDREJKOVIC, J. W.; BERGAUER, T.; AL, S. Chatterjee et. Measurement of the top quark mass using events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 13\text{TeV}$. *Journal of High Energy Physics*. 2021, vol. 2021, no. 12. Available from DOI: `10.1007/jhep12(2021)161`.
5. POSPISIL, Jiri. *Application of Machine Learning for the Charged Higgs Boson Search Using ATLAS Data*. 2022. Available also from: `https://cds.cern.ch/record/2812372`. Presented 14 Jun 2022.
6. *Glossary*. Geneva Switzerland: CERN, 2023. Available also from: `https://atlas.cern/glossary`.
7. GROSS, E. Practical Statistics for High Energy Physics. *CERN Yellow Rep. School Proc.* 2018, vol. 3, pp. 199–221. Available from DOI: `10.23730/CYRSP-2018-003.199`.
8. PIPARO, Danilo; TEJEDOR, Enric; MATO, Pere; MASCETTI, Luca; MOSCICKI, Jakub; LAMANNA, Massimo. SWAN: A service for interactive analysis in the cloud. *Future Generation Computer Systems*. 2018, vol. 78, pp. 1071–1078. ISSN 0167-739X. Available from DOI: `https://doi.org/10.1016/j.future.2016.11.035`.
9. BRUN, Rene; RADEMAKERS, Fons. ROOT — An object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 1997, vol. 389, no. 1-2, pp. 81–86. ISSN 01689002. Available from DOI: `10.1016/S0168-9002(97)00048-X`.
10. NAUMANN, Axel; CANAL, Philippe; TEJEDOR, Enric; GUIRAUD, Enrico; MONETA, Lorenzo; BELLENOT, Bertrand; COUET, Olivier; TADEL, Alja Mrak; TADEL, Matevz; LINEV, Sergey; GOMEZ, Javier Lopez; REMBSER, Jonas; PADULANO, Vincenzo Eduardo; BLOMER, Jakob; HAHNFELD, Jonas; GRUBER, Bernhard Manfred; VASSILEV, Vassil. *ROOT for the HL-LHC: data format*. 2022. Available from arXiv: `2204.04557 [hep-ex]`.

11. AKIBA, Takuya; SANO, Shotaro; YANASE, Toshihiko; OHTA, Takeru; KOYAMA, Masanori. *Optuna: A Next-generation Hyperparameter Optimization Framework*. 2019. Available from arXiv: 1907.10902 [cs.LG].

12. DUSER, Niklas. Optimization of tbH$^+$ Signal and Background Separation Using Machine Learning in the 2lSS1tau Channel, Comparison of Limit Setting Techniques and Signal Injection Studies - Summer Student Report. 2022. Available also from: https://cds.cern.ch/record/2843046.

13. COUET, Olivier; GUIRAUD, Enrico; PADULANO, Vincenzo Eduardo; HAGEBOECK, Stephan; KABADZHOV, Ivan; LOPEZ-GOMEZ, Javier. *ROOT Primer*. Genève: CERN Organisation Européenne pour la Recherche Nucléaire, 2023. Available also from: https://root.cern/primer/.

14. KONIG, Severin. Optimization of ttH$^-$ Signal and Background Separation Using Machine Learning in the 2lSS1tau Channel. 2022. Available also from: https://cds.cern.ch/record/2836425.

15. KHACHATRYAN, V. et al. Measurement of transverse momentum relative to dijet systems in PbPb and pp collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76$ TeV. *Journal of High Energy Physics*. 2016, vol. 2016, no. 1, p. 6. ISSN 1029-8479. Available from DOI: 10.1007/JHEP01(2016)006.

16. PEREIRA SANCHEZ, Laura. Calibration of flavour tagging algorithms in ATLAS on $t\bar{t}$ and Z+jets final states. *PoS*. 2022, vol. EPS-HEP2021, p. 434. Available from DOI: 10.22323/1.398.0434.

17. MELONI, F. Primary vertex reconstruction with the ATLAS detector. *Journal of Instrumentation*. 2016, vol. 11, no. 12, pp. C12060–C12060. ISSN 1748-0221. Available from DOI: 10.1088/1748-0221/11/12/C12060.

18. *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*. Geneva, 2016. Tech. rep. CERN. Available also from: https://cds.cern.ch/record/2160731.

19. SOUTH, David; GINGRICH, Doug. *ATLAS MC Production Workflow: Input to the ATLAS Sites Jamboree*. Salle Dirac (CERN): Alessandro Di Girolamo (CERN), 2018. Available also from: https://indico.cern.ch/event/692124/contributions/2898693/attachments/1610808/2557701/mccoord_050318.pdf.

20. HARDT, Moritz; RECHT, Benjamin; SINGER, Yoram. *Train faster, generalize better: Stability of stochastic gradient descent*. 2016. Available from arXiv: 1509.01240 [cs.LG].

21. HOFMANN, Thomas; SCHÖLKOPF, Bernhard; SMOLA, Alexander J. Kernel methods in machine learning. *The Annals of Statistics*. 2008, vol. 36, no. 3. Available from DOI: 10.1214/009053607000000677.

22. EVGENIOU, Theodoros; PONTIL, Massimiliano. Support Vector Machines: Theory and Applications. In: 2001, vol. 2049, pp. 249–257. ISBN 978-3-540-42490-1. Available from DOI: 10.1007/3-540-44673-7_12.

23. PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011, vol. 12, pp. 2825–2830.

24. PLATT, John. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Adv. Large Margin Classif.* 2000, vol. 10.

25. DADO, Tomas; PINAMONTI, Michele; HELD, Alexander. *TRExFitter documentation*. Genève: CERN Organisation Européenne pour la Recherche Nucléaire, 2020. Available also from: https://trexfitter-docs.web.cern.ch/trexfitter-docs/.

26. CRANMER, Kyle. *Practical Statistics for the LHC*. 2015. Available from arXiv: 1503.07622 [physics.data-an].

27. DEVELOPERS, TensorFlow. *TensorFlow*. Zenodo, 2022. Version v2.8.0. Available from DOI: 10.5281/zenodo.5949125.

28. REBACK, Jeff; MCKINNEY, Wes; JBROCKMENDEL; BOSSCHE, Joris Van den; AUGSPURGER, Tom; CLOUD, Phillip; GFYOUNG; HAWKINS, Simon; SINHRKS; ROESCHKE, Matthew; KLEIN, Adam; PETERSEN, Terji; TRATNER, Jeff; SHE, Chang; AYD, William; NAVEH, Shahar; GARCIA, Marc; SCHENDEL, Jeremy; PATRICK; HAYDEN, Andy; SAXTON, Daniel; JANCAUSKAS, Vytautas; MCMASTER, Ali; GORELLI, Marco; BATTISTON, Pietro; SEABOLD, Skipper; DONG, Kaiqi; CHRIS-B1; H-VETINARI; HOYER, Stephan. *pandas-dev/pandas: Pandas 1.2.2*. Zenodo, 2021. Version v1.2.2. Available from DOI: 10.5281/zenodo.4524629.

29. *Search for a charged Higgs boson decaying into a heavy neutral Higgs boson and a W boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV*. Geneva, 2022. Tech. rep. CERN. Available from arXiv: 2207.01046.

Contents of Enclosed CD

```
readme.txt ............................ the file with CD contents description
application ............................ the directory with the application-related files
    preprocessing.py ................... data preprocessing
text ................................... the thesis text directory
    src ................................ the directory of LaTeX source codes of the thesis
    thesis.pdf ......................... the thesis text in PDF format
```

List of abbreviations

2DHM	2 Doublet Higgs Model
2lSS1tau	2 leptons with same sign and 1 tau
CL	Confidence Level
MLP	Multilayer Perceptron
ReLU	Rectified Linear Unit
SGD	Stochastic Gradient Descent
SM	Standard Model