UNIVERSITÉ Clermont Auvergne ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA technische universität dortmund

A thesis submitted in partial fulfillment of the requirements for the degree Master of Science

Optimizing the Search for Charged Higgs Boson Decaying into a Neutral Higgs Boson Using the ATLAS Detector at CERN

Adwait Meppurath
born in Kerala, India

September 3, 2025

Fakultät Physik
Technische Universität Dortmund

Dipartimento di Fisica e Astronomia
Università di Bologna

École Universitaire de Physique et d'Ingénierie
Université Clermont Auvergne

IMAPP

This thesis summarizes the scientific work carried out at the Chair of Experimental Physics of Technische Universität Dortmund.

Supervisor:	Associate Prof. Dr. André Sopczak
First reviewer:	Dr. Carsten Burgard
Date of submission:	September 2025

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor, Associate Prof. Dr. André Sopczak, for introducing me to the fascinating topic covered in this thesis. His guidance, coupled with the intellectual freedom he provided, was what truly enabled me to explore this subject to the extent that I have. Secondly, I am thankful to IMAPP for giving me this opportunity to work in a research group for my Master thesis and for providing me with three semesters' worth of theoretical background to appreciate the work that I have done.

This work would not have been possible without the support of my colleagues and friends. I am particularly grateful to my friend and colleague, Ishaan Utkarsh. The numerous fruitful (and often late-night) discussions we had on analysis strategies and untangling "Git spaghetti"—frequently over games of ping-pong at the dormitory—were instrumental in advancing this research. I extend my sincere thanks to Ing. Viktoriia Lysenko for her prompt and invaluable assistance in navigating the complexities of ATLAS resources, which saved me countless hours. I am grateful to have been welcomed by the ATLAS group at the Institute of Experimental and Applied Physics, CTU Prague; it was a privilege to work here tackling current issues in HEP.

My time in Prague was made unforgettable by the friends I made in *Hlávkova kolej*. Life here would not have been the same without all of them. I am thankful for the novel perspectives Ms. Tetiana Velehura, as a computer science student, brought to the problems I was tackling in experimental particle physics. I also extend my thanks to my friends from back home—Rohan, Austin, and Patil—for the many weekends spent on call, providing a much-needed respite from writing lines of code.

Finally, to my family, I remain eternally grateful for their unwavering support throughout all these years.

I am able to do all that I have because of the incredible people I have surrounded myself with, and for that, I am profoundly thankful.

Adwait

To,
Amma, Acchan and *Ammu*

Abstract

The search for physics Beyond the Standard Model (BSM) is a primary focus of particle physics. A key target in these searches is the charged Higgs boson ($H^{\pm}$), a particle predicted by well-motivated theoretical frameworks such as the Two-Higgs-Doublet Model (2HDM). This thesis presents an optimized search for the $H^{\pm}$ produced in association with a top and a bottom quark ($tbH^{\pm}$), focusing on the decay channel $H^{\pm} \to W^{\pm}h$ in the complex $2\ell_{\text{SS}}1\tau_{\text{had}}$ final state. The primary challenge of this analysis is the discrimination of the rare signal signature from large and complex Standard Model backgrounds at the Large Hadron Collider.

A full-chain analysis was performed using 140 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV recorded by the ATLAS detector. To enhance the signal-background separation, a novel, scale-based approach was developed, employing two distinct XGBoost models optimized for low-mass and high-mass signal hypotheses. The performance of these models was extensively compared against the attention-based TabNet deep learning architecture. Data-driven template fits were implemented to provide a robust estimate of the significant non-prompt lepton background.

The analysis is expected to set a 95% confidence level upper limit on the charged Higgs boson production cross-section, excluding masses below 800 GeV for the targeted 2HDM Type-II scenario with tan $\beta = 20$. The optimized methods developed in this work yield a substantial gain in sensitivity, resulting in expected limits that are more than twice as stringent as those from the previous analysis using the same dataset. This demonstrates the powerful impact of the tailored machine learning strategy and establishes a new benchmark for sensitivity in this search channel.

Contents

Chapter 1: Introduction

Since the discovery of the Higgs boson in 2012, the particle physics community has intensified its search for phenomena beyond the Standard Model (SM). A key target in this effort is the charged Higgs boson ($H^{\pm}$), whose existence is predicted by several theoretical frameworks, including the Two-Higgs-Doublet Model (2HDM) and Minimal Supersymmetric Model (MSSM) [1]. The observation of such a particle would provide definitive evidence for a more complex Higgs sector and offer new insights into the mechanism of electroweak symmetry breaking. However, identifying a charged Higgs signature amidst the vast number of background events produced at the Large Hadron Collider (LHC) requires sophisticated analysis techniques.

The thesis will begin with a review of the underlying physics, followed by a description of the data and methods. It will then present the model development and performance, culminating in the statistical interpretation of the results and a final conclusion.

1.1 2lSS1tau Final State

The prime focus of this analysis is the tbH^+ production of H^+ in the $2l_{SS}1\tau_{had}$ final state, which can be understood from Fig. 1.1. This is a complex, multi-jet final state, with a hadronically decaying τ lepton, b-jets and missing p_T due to existence of multiple neutrinos in the final state. This makes it a complicated signature to identify among other prominent known background SM processes such as ttH (Fig. 1.2), ttW (Fig. 1.3), $t\bar{t}$, etc.

This thesis addresses this challenge by leveraging machine learning to improve the signal-to-background discrimination. The work is performed on a large sample of simulated proton-proton collision events generated with the latest ATLAS software, benefiting from improved calibrations and high statistics, while also using the data-driven template-fitting method to estimate the fake lepton presence in the simulated data.

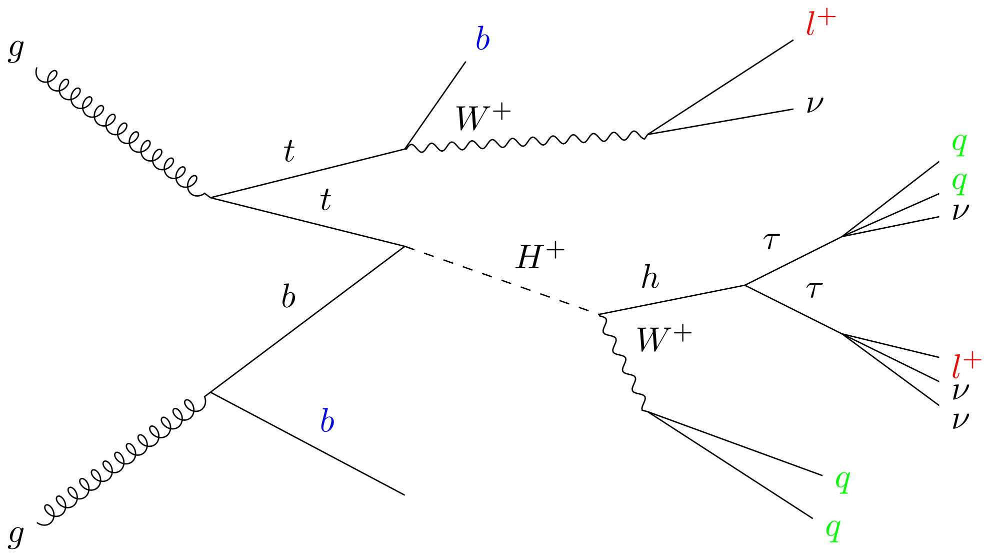


Figure 1.1: *Associative H^+ production with t and b and subsequent decay to the $2l_{SS}1\tau_{had}$ final state*

The primary objectives of this study are:

1. To produce 11 new simulated datasets to extend previous analyses
2. To develop, compare and optimize machine learning models capable of distinguishing charged Higgs boson signal events from dominant SM background processes.
3. To evaluate the performances of the above-mentioned models using established metrics and graphical representations that highlight the separation between event classes.
4. To apply this model on simulated datasets to project the analysis's sensitivity.
5. To estimate the Fakes in this analysis using *Template Fitting*
6. To quantify this sensitivity by deriving the expected upper limits on the charged Higgs production cross-section.

The thesis aims to cover the full analysis chain starting from initial steps of producing MC simulated data to the final steps of setting upper limits on the cross-section of the signal hypothesis.

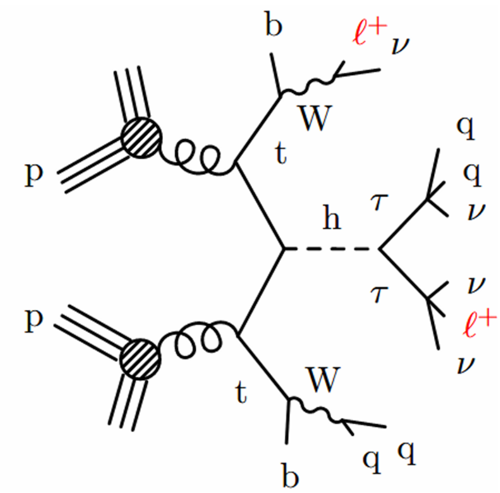


Figure 1.2: *SM ttH background process decaying to the $2l_{SS}1\tau$ final state*

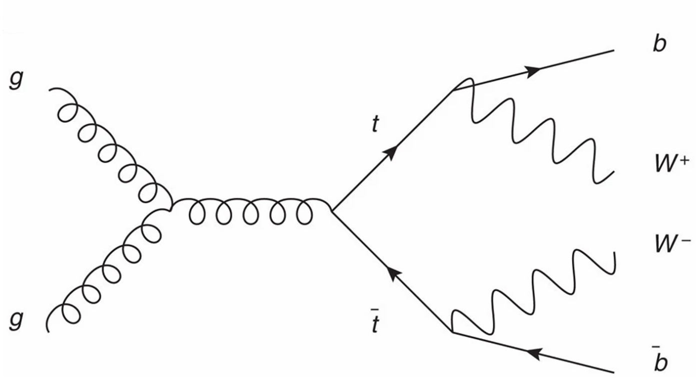


Figure 1.3: *SM ttW background process decaying to $2l_{SS}1\tau$ final state through fake leptons*

Chapter 2: Theory

2.1 The Standard Model

The Standard Model (SM) of particle physics is the theoretical framework that describes the fundamental particles and their interactions, providing our most profound understanding of the subatomic world. Developed through the latter half of the 20th century, it has been rigorously tested and validated by numerous experiments, achieving unprecedented predictive power. The model successfully describes three of the four known fundamental forces—the strong, weak, and electromagnetic forces—and classifies the elementary particles that constitute all visible matter [2].

Fundamental Constituents: Fermions

The matter particles in the Standard Model are fermions, characterized by their half-integer spin. They are organized into two classes: quarks and leptons. Both are further grouped into three successive generations of increasing mass.

- **Quarks** are the building blocks of composite particles called hadrons (e.g., protons and neutrons). They experience the strong force due to their "color charge" and carry fractional electric charges. The six types, or "flavors," of quarks are the up (u), down (d), charm (c), strange (s), top (t), and bottom (b). Due to the phenomenon of color confinement, free quarks are not observed in nature.
- **Leptons** do not carry color charge and are therefore immune to the strong force. This group includes the three charged leptons—the electron (e^-), muon (μ^-), and tau (τ^-)—and their corresponding neutral partners, the neutrinos (ν_e, ν_μ, ν_τ).

Fundamental Interactions: Gauge Bosons

The interactions between fermions are mediated by the exchange of force-carrying particles known as gauge bosons, which have integer spin.

- The **electromagnetic force**, which governs interactions between electrically charged particles, is mediated by the massless **photon** (γ).

- The **strong force**, described by Quantum Chromodynamics (QCD), binds quarks together. It is mediated by eight massless **gluons** (g).
- The **weak force** is responsible for radioactive decay and is unique in its ability to change particle flavor. It is mediated by the massive W^+, W^-, and Z **bosons**.

The electromagnetic and weak forces are unified into a single theoretical structure known as the electroweak theory.

The Higgs Boson and Mass Generation

A crucial component of the Standard Model is the Higgs field, a scalar field that permeates all of space. The massive W and Z bosons, as well as the fundamental fermions, acquire their mass through their interaction with this field via the Brout-Englert-Higgs mechanism. A quantum excitation of this field gives rise to a physical particle: the Higgs boson (H). Its discovery by the ATLAS and CMS experiments at CERN in 2012 was the final experimental confirmation of the SM's particle content.

The complete grouping of Standard Model particles are given in Fig. 2.1.

Limitations and Motivation for New Physics

Despite its remarkable successes, the Standard Model **is acknowledged to be incomplete**. It fails to provide answers to several profound questions in physics [2]:

1. **Gravity:** The SM does not incorporate the fourth fundamental force, gravity.
2. **Cosmological Observations:** It cannot account for the existence of dark matter or the accelerating expansion of our Universe attributed to dark energy, which together constitute approximately 95% of the universe's mass-energy budget.
3. **Neutrino Mass:** In its original formulation, the SM assumes neutrinos are massless. However, the observation of neutrino oscillations has definitively shown that they do have mass.
4. **Matter-Antimatter Asymmetry:** The model does not sufficiently explain why the observable universe is composed almost entirely of matter, with very little antimatter.

These open questions provide **a powerful impetus to search for physics Beyond the Standard Model (BSM)**. Many theoretical extensions, such as the **Two-Higgs-Doublet Model (2HDM)**, address these limitations by postulating a richer particle spectrum. A common feature of such models is an **extended Higgs sector**, which predicts the existence of new scalar particles, including the charged Higgs boson ($H^{\pm}$) that is the subject of this work.

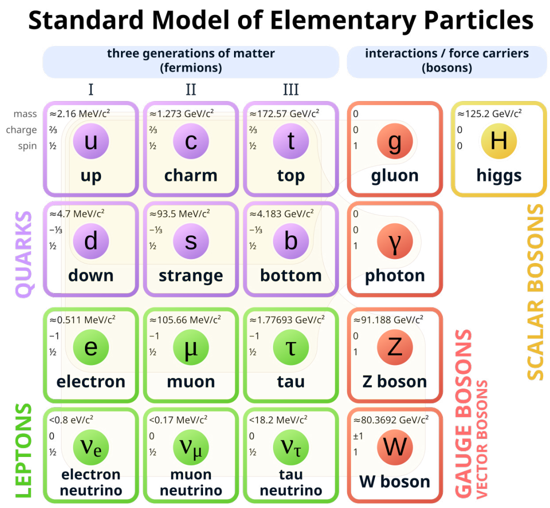


Figure 2.1: *The Standard Model of Particle Physics [3]*

2.2 BSM Extensions

While the Standard Model (SM) has been phenomenally successful, its inability to address several fundamental questions in physics serves as a powerful motivation to explore theories Beyond the Standard Model (BSM). These unresolved issues, often referred to as the theoretical and observational evidence for new physics, guide the development of new models that aim to provide a more complete picture of the universe [4].

From a theoretical perspective, any BSM physics must manifest **as additional terms in the Lagrangian**. This can be approached in two ways [5]. The first is to construct a *UV-complete* model by introducing new particles (fields) and renormalizable interactions, as is done in frameworks like Supersymmetry. The second approach, useful when the new physics scale Λ is much larger than the experimental energy scale, is to use an *Effective Field Theory* (EFT). Here, the effects of heavy, undiscovered particles are parameterized by adding a series of higher-dimensional, non-renormalizable operators ($\mathcal{O}_i$) to the SM Lagrangian:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_{d>4} \frac{c_i}{\Lambda^{d-4}} \mathcal{O}_i^{(d)} \tag{2.1}$$

where d is the mass dimension of the operator. A classic and profoundly important example is the Weinberg operator, which is the unique dimension-five operator that can be constructed from SM fields while respecting SM gauge symmetries. This operator violates lepton number– an "accidental" symmetry of the renormalizable SM–and, after EWSB, naturally generates a small Majorana mass for neutrinos. This elegantly explains the observed neutrino masses by suppressing them with a very high energy scale Λ, providing a direct link between low-energy observables and high-scale BSM physics.

To address the key shortcomings of the SM, a wide range of BSM theories have been proposed. Many of these new frameworks predict the existence of new particles and interactions at the TeV energy scale, making them accessible to experimental verification at the Large Hadron Collider (LHC). The search for charged Higgs bosons is a particularly powerful probe, as an extended Higgs sector is a cornerstone of the SUSY/2HDM framework and serves as a crucial benchmark for physics beyond the Standard Model.

2.2.1 Two-Higgs-Doublet Model (2HDM)

The motivations behind studying BSM phenomena in general can be understood from the sections above and the implications of a potential extended Higgs sector are quite significant, let us now look into a brief overview of what the Two-Higgs-Doublet Model (2HDM) predicts.

Motivation

Among the many proposed theories for BSM physics, 2HDM represents one of the simplest and most compelling extensions of the SM Higgs sector. 2HDM type II (the focus of this study) is the Higgs sector of the MSSM [6]. 2HDM also could provide additional sources of CP Violation in the weak interaction and potentially contribute to the baryon asymmetry in the universe. [7], [8]. A variant of the 2HDMs may play an important role in a Peccei-Quinn theory, predicting Axions, a promising Dark Matter candidate [9]. All of these are strong motivations to search for H^+ under the 2HDM type II model.

Theoretical Structure

The fundamental postulate of the 2HDM is the **introduction of a second complex scalar doublet**, with the same quantum numbers as the SM Higgs doublet, to the electroweak Lagrangian. Where the SM contains one doublet Φ_{SM}, the 2HDM contains two, denoted as Φ_1 and Φ_2. The resulting potential is given in Equation 2.2.

$$V(\Phi_1,\Phi_2) = m_{11}^2\Phi_1^\dagger\Phi_1 + m_{22}^2\Phi_2^\dagger\Phi_2 - [m_{12}^2\Phi_1^\dagger\Phi_2 + h.c] + \frac{1}{2}\lambda_1(\Phi_1^\dagger\Phi_1)^2 + \frac{1}{2}\lambda_2(\Phi_2^\dagger\Phi_2)^2 + \\ + \lambda_3(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_2) + \lambda_4(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) + [\frac{1}{2}\lambda_5(\Phi_1^\dagger\Phi_2)^2 + h.c.] \quad (2.2)$$

Analogously to the SM, the scalar fields will acquire vacuum expectation values (VEVs), v_1 and v_2. The expansion of fields around this VEV is given by

$$\Phi_{1,2} = \begin{pmatrix} \Phi_{1,2}^+ \\ (v_{1,2} + h_{1,2} + i\eta_{1,2})/\sqrt{2} \end{pmatrix} \quad (2.3)$$

This gives rise to 8 degrees of freedom, 3 of which will be absorbed into the longitudinal polarisation states of the weak gauge bosons ($W^\pm, Z$). This leaves **five physical scalar particles**:

- Two neutral, CP-even Higgs bosons: a light scalar (h) and a heavy scalar (H). By convention, h is typically identified with the 125 GeV boson discovered at the LHC.
- Two electrically **charged Higgs bosons:** H^+ and its antiparticle H^-.
- One neutral, CP-odd Higgs boson: a pseudoscalar (A).

The existence of a charged scalar particle, the $H^\pm$, is therefore a generic and robust prediction of any 2HDM. Its discovery would be **unambiguous proof of a non-minimal Higgs sector** and, by extension, of new physics.

By definition, H is chosen to be heavier than h. An important parameter in 2HDM is $\tan\beta$, given by

$$\tan\beta = \frac{v_2}{v_1}. \quad (2.4)$$

The relation of the two VEVs to the SM VEV is

$$v = \sqrt{v_1^2 + v_2^2} = 246 \text{ GeV}. \quad (2.5)$$

The mixing of the two neutral CP-even Higgs (h, H) can be represented by

$$\begin{aligned} h &= h_1\sin(\alpha) - h_2\cos(\alpha), \\ H &= -h_1\sin(\alpha) - h_2\cos(\alpha), \end{aligned} \quad (2.6)$$

which defines the mixing angle α, another important parameter in 2HDM. Since we know that either h or H has to be the observed, SM-like 125GeV Higgs boson, which constrains the parameter-space to the *alignment-limit*. In this limit, **one of the CP-even uncharged scalar has to behave like the SM Higgs boson**. Which then defines the alignment limit according to,

$$h \rightarrow h_{SM} : \sin(\beta - \alpha) \rightarrow 1 \tag{2.7}$$

$$H \rightarrow h_{SM} : \cos(\beta - \alpha) \rightarrow 1 \tag{2.8}$$

Which also relates the parameters α and β.
A critical challenge in constructing a viable 2HDM is to **avoid large, unobserved Flavor-Changing Neutral Currents (FCNCs)** at the tree level. The most common solution is to impose a discrete Z_2 symmetry on the Lagrangian, which restricts the couplings of the fermions. Depending on how the quark and lepton families are assigned to couple to Φ_1 and Φ_2, several distinct "types" of 2HDM emerge (e.g., Type-I, Type-II, Lepton-Specific, Flipped), as given in Table 2.1. The phenomenology of these models is largely governed by the masses of the new particles and by the parameter $\tan\beta$. To phenomenologically study the charged Higgs, we have to look at the Yukawa sector of the Lagrangian, which depends on the 2HDM type. Since this context focuses on Type-II, the the Lagrangian for 2HDM type-II is given in Equation 2.9.

	Type 1	Type 2	Lepton-specific	Flipped
Φ_1		d^i_R, l^i_R	l^i_R	d^i_R
Φ_2	u^i_R, d^i_R, l^i_R	u^i_R	u^i_R, d^i_R	u^i_R, l^i_R

Table 2.1: *The four different 2HDM types, distinguished by the coupling of Φ_1 and Φ_2 to the quarks and leptons. u are the up-type quarks, d the down-type quarks and l the leptons, i is the generation index.*

$$\mathcal{L}_{H^\pm} = -H^+\left(\frac{\sqrt{2}V_{ud}}{v}\bar{u}(m_u(\cot\beta)P_L - m_d(\tan\beta)P_R)d - \frac{\sqrt{2}m_l}{v}(\tan\beta)\bar{\nu}l_R\right) + h.c, \tag{2.9}$$

where V_{ud} correspond to the CKM-matrix element, $P_{L/R}$ the left/right handed projector. The couplings are proportional to the fermion masses, such that decays into the heaviest fermions, i.e top and bottom quark or a tau lepton and a neutrino, depending on $\tan\beta$, dominate. Consequently, searches for charged Higgs bosons have been focused on these decays [10–13].

$H^\pm$ Production and Cross-section

The cross-sections for various values of $\tan\beta$ across different H^+ masses are shown in Fig. 2.2. The branching fractions for the $H^\pm \rightarrow Wh$ and $H^\pm \rightarrow tb$ decays are

presented as a function of $\cos(\beta - \alpha)$ and for different $\tan\beta$ and $m_{H^\pm}$ values in Fig. 2.3 for the Type-II 2HDM scenario. The branching fractions for the $H^\pm \to Wh$ decay are more significant **outside of the alignment limit**. The relative importance of the $H^\pm \to Wh$ decay increases also as a function of $m_{H^\pm}$. The branching fractions have been calculated with 2HDMC [14]. The maximum of the production cross section times branching fractions are obtained at higher $\tan\beta$ values (in the case of $H^+ \to W^+h$, the focus of this study).
For this analysis, we target a specific and well-motivated region of the 2HDM parameter space. A high value of $\tan\beta = 20$ is chosen, as this significantly enhances the production cross-section for the charged Higgs boson, as shown in Fig. 2.2, which shows the Santander-matched cross-sections [15]. Furthermore, the search for the $H^+ \to W^+h$ decay channel is a powerful probe of the region slightly away from the perfect alignment limit ($\cos(\beta-\alpha) \neq 1$). As demonstrated in Fig. 2.3, this is precisely the regime where the $H^+ \to tb$ decay is suppressed and the $H^+ \to W^+h$ branching ratio becomes significant. This **makes the chosen parameter space an ideal benchmark to search for new physics**, as it combines a large production rate with a distinct decay signature.

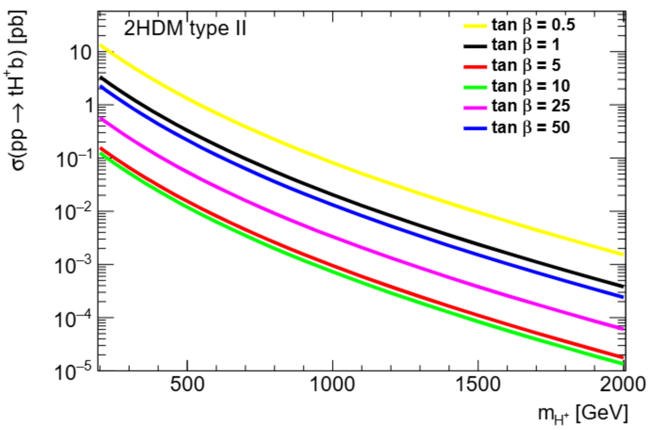


Figure 2.2: Santander-matched cross sections for the process $gg \to tH^+ + X$ for various $\tan\beta$ and $m_{H^\pm}$ values at $\sqrt{s} = 13$ TeV [15]

2.3 The ATLAS Detector

This study uses experimental and simulated data from the ATLAS detector at CERN. The ATLAS (A Toroidal LHC ApparatuS) experiment is one of the four major particle detectors located on the Large Hadron Collider (LHC) ring at CERN. It is a multi-purpose particle detector designed with a forward-backward symmetric, cylindrical

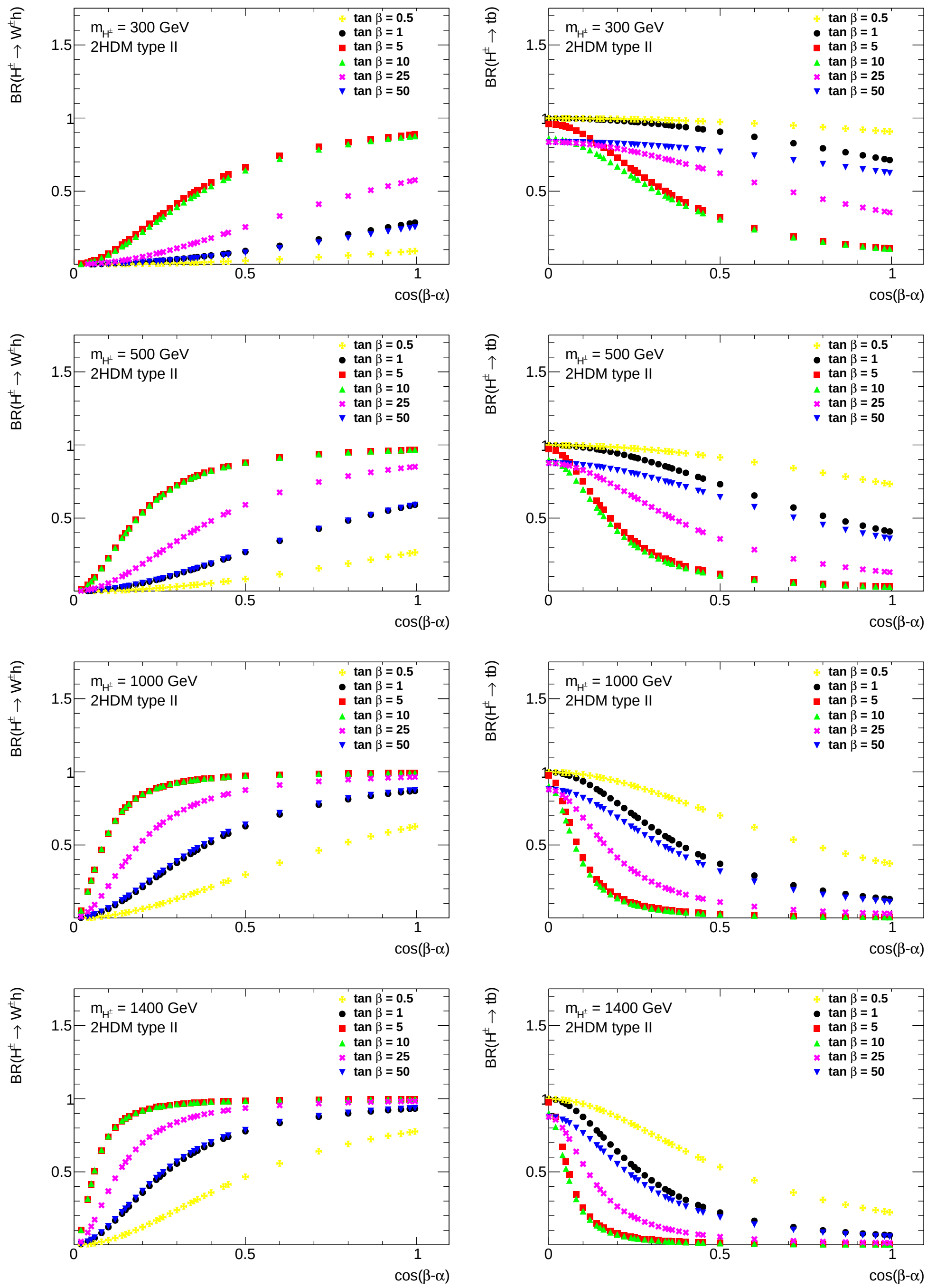


Figure 2.3: *Branching ratios for charged Higgs bosons decaying via $H^\pm \to W^\pm h$ (left) and $H^\pm \to tb$ (right) presented as a function of $\cos(\beta-\alpha)$ for different $\tan\beta$ and $m_{H^\pm}$ values and a type II 2HDM. The branching ratios have been calculated via 2HDMC [14]*

geometry covering nearly the entire solid angle around the proton-proton interaction point [16]. Its primary scientific goals are to perform precision measurements of the Standard Model (SM), including the properties of the Higgs boson, and to search for evidence of physics beyond the Standard Model, such as the charged Higgs bosons that are the focus of this analysis. The overall detector is 46 metres long, 25 metres in diameter, and weighs approximately 7,000 tonnes.

The ATLAS coordinate system is a right-handed system with its origin at the nominal interaction point. The beam direction defines the z-axis, the x-axis points towards the centre of the LHC ring, and the y-axis points upwards. The azimuthal angle ϕ is measured around the beam axis in the transverse ($x-y$) plane, and the polar angle θ is measured from the z-axis. Particle trajectories are more commonly described using pseudorapidity, defined as $\eta = -\ln[\tan(\theta/2)]$.

To achieve its physics goals, ATLAS is composed of several concentric **sub-detector systems**, which, starting from the interaction point and moving outwards, are the Inner Detector, the Calorimeters, and the Muon Spectrometer.

Magnet System The detector's momentum measurement capabilities rely on a powerful magnet system. A thin, superconducting central solenoid surrounds the Inner Detector, providing a 2 T axial magnetic field that allows for the momentum measurement of charged particles. An extensive system of large, air-core toroidal magnets is located outside the calorimeters, generating a strong magnetic field for the independent momentum measurement of muons [16].

Inner Detector (ID) The ID is responsible for precision tracking of charged particles and the reconstruction of primary and secondary vertices. It is immersed in the 2 T field of the central solenoid and consists of three sub-systems:

- The **Pixel Detector**, the innermost component, provides high-granularity spatial measurements, crucial for vertex identification and tracking particles close to the beam pipe.
- The **Semiconductor Tracker (SCT)** surrounds the Pixel Detector and provides further precise spatial points for track reconstruction at a larger radius.
- The **Transition Radiation Tracker (TRT)** is the outermost part of the ID. In addition to providing continuous tracking information, it aids in the identification of electrons by detecting transition radiation photons.

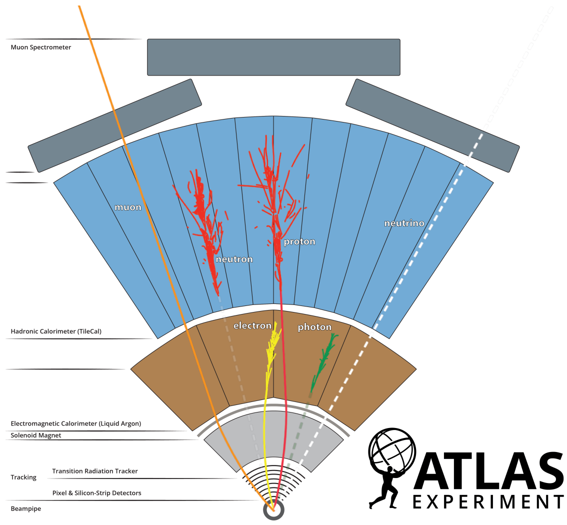


Figure 2.4: *Schematic of the ATLAS detector*

Calorimeters The ATLAS calorimetry system is designed to measure the energy of particles by absorbing them. It consists of two main parts:

- The **Electromagnetic (EM) Calorimeter** uses lead as an absorber and liquid argon (LAr) as the active medium. It is designed with a distinctive "accordion" geometry to ensure complete ϕ coverage and is optimized to measure the energy of electrons and photons with high precision.
- The **Hadronic Calorimeter** surrounds the EM calorimeter and is designed to measure the energy of hadrons (like protons and pions). It uses steel and scintillating tiles in the central barrel region and LAr technology in the end-cap and forward regions.

Together, these systems absorb all known particles except for muons and neutrinos. The energy of neutrinos is inferred indirectly through the measurement of missing transverse energy.

Muon Spectrometer (MS) As the outermost part of ATLAS, the MS is designed to identify muons and measure their momentum with high precision. Since muons penetrate the calorimeters, any particle track observed in the MS is identified as a muon. The spectrometer operates in the magnetic field generated by the large air-core toroids and consists of chambers for precision tracking (Monitored Drift Tubes and Cathode Strip Chambers) and fast triggering (Resistive Plate Chambers and Thin Gap Chambers).

Trigger and Data Acquisition (DAQ) The LHC produces proton-proton collisions at a rate of 40 MHz, far exceeding the capacity for permanent storage. ATLAS employs a two-level trigger system to select interesting events in real-time. A hardware-based Level-1 trigger reduces the rate to approximately 100 kHz, followed by a software-based High-Level Trigger (HLT) that further reduces the event rate to about 1 kHz for offline analysis [16].

2.4 Machine Learning and Neural Networks

2.4.1 Theory

At its core, the supervised machine learning task in this analysis is to learn a mapping from a set of input features to a target label. For each physics event, we define an input vector $\mathbf{x} \in \mathbb{R}^d$, where d is the number of kinematic variables used. The goal is to predict the corresponding true label $y \in \{0, 1\}$, where $y = 1$ denotes a signal event ($tbH^{\pm}$) and $y = 0$ denotes a background event ($ttH, ttW, t\bar{t}$, etc.).

This mapping is approximated by a model, which is a parameterized function $f(\mathbf{x}; \theta)$ that outputs a prediction, $\hat{y}$. The parameters θ (often called weights and biases) are learned during a process called **training**. To guide this process, we define a **loss function**, $\mathcal{L}(\hat{y}, y)$, which quantifies the discrepancy between the model's prediction and the true label for a single event. For binary classification problems, a standard choice is the **Binary Cross-Entropy (BCE)** loss

$$\mathcal{L}(\hat{y}, y) = -\left[y \log(\hat{y}) + (1 - y)\log(1 - \hat{y})\right] \tag{2.10}$$

where $\hat{y}$ is the predicted probability of the event being signal.

The training process aims to find the optimal set of parameters θ^* that minimizes the average loss over the entire training dataset of N events. This is expressed as minimizing an **objective function** $\mathcal{J}(\theta)$

$$\mathcal{J}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(f(\mathbf{x}_i; \theta), y_i) \tag{2.11}$$

This minimization is typically performed iteratively using a gradient-based optimization algorithm, such as stochastic gradient descent (SGD) or its variants (e.g., Adam). In each iteration, the parameters are updated by taking a step in the direction opposite to the gradient of the loss function

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \nabla_{\theta} \mathcal{J}(\theta), \tag{2.12}$$

where η is the **learning rate**, a hyperparameter that controls the step size.

To prevent the model from learning the training data too precisely and failing to generalize to new, unseen data (a phenomenon known as **overfitting**), a **regularization term** $\mathcal{R}(\theta)$ is often added to the objective function. This term penalizes model complexity

$$\mathcal{J}_{\text{reg}}(\theta) = \mathcal{J}(\theta) + \lambda\mathcal{R}(\theta), \tag{2.13}$$

where λ is the **regularization strength**. Common choices for $\mathcal{R}(\theta)$ include the L1-norm ($\sum |\theta_i|$) and the L2-norm ($\sum \theta_i^2$) of the parameters. The specific forms of the function $f(\mathbf{x};\theta)$ and the optimization strategies define the different models, such as TabNet and XGBoost, evaluated in this work.

2.4.2 Models

Google TabNet

While deep learning has achieved state-of-the-art performance in domains like computer vision and natural language processing, its application to tabular data has often been surpassed by traditional methods like gradient-boosted decision trees. To bridge this gap, Google Research introduced TabNet, a novel deep learning architecture specifically designed for tabular data [17]. TabNet aims to combine the representation learning power of deep neural networks with the interpretability and performance that make tree-based models so effective. Given that high-energy physics event data is inherently tabular—where each event is a row and kinematic variables are columns—TabNet presents a compelling model for this analysis.

Architectural Innovation: Sequential Attention The core innovation of TabNet lies in its use of a sequential **attention mechanism** to select the most salient features at each stage of the decision-making process. Unlike a standard Multilayer Perceptron (MLP) which processes all input features simultaneously in dense, fully-connected layers, TabNet processes information in sequential steps, mimicking the decision-making process of a tree-based model but in a soft, differentiable manner.

The architecture is built around a series of decision steps. At each step, two key components are at play:

1. **Attentive Transformer:** This block is responsible for learning which features to focus on. It analyzes the processed information from the previous step and outputs a "mask" that assigns an attention weight to each input feature. This mask is multiplicative, effectively selecting a sparse subset of features for the current step. Crucially, this feature selection is **instance-wise**: the model can learn to focus on different features for different events, a powerful capability for complex datasets.

2. **Feature Transformer:** This block processes the features selected by the attentive transformer. The masked features are fed into a series of shared and step-dependent layers (typically fully connected blocks with batch normalization and Gated Linear Unit (GLU) activations) to extract meaningful representations. The output of the feature transformer is then split: one part contributes to the final overall prediction, and the other part is passed as input to the attentive transformer of the *next* decision step.

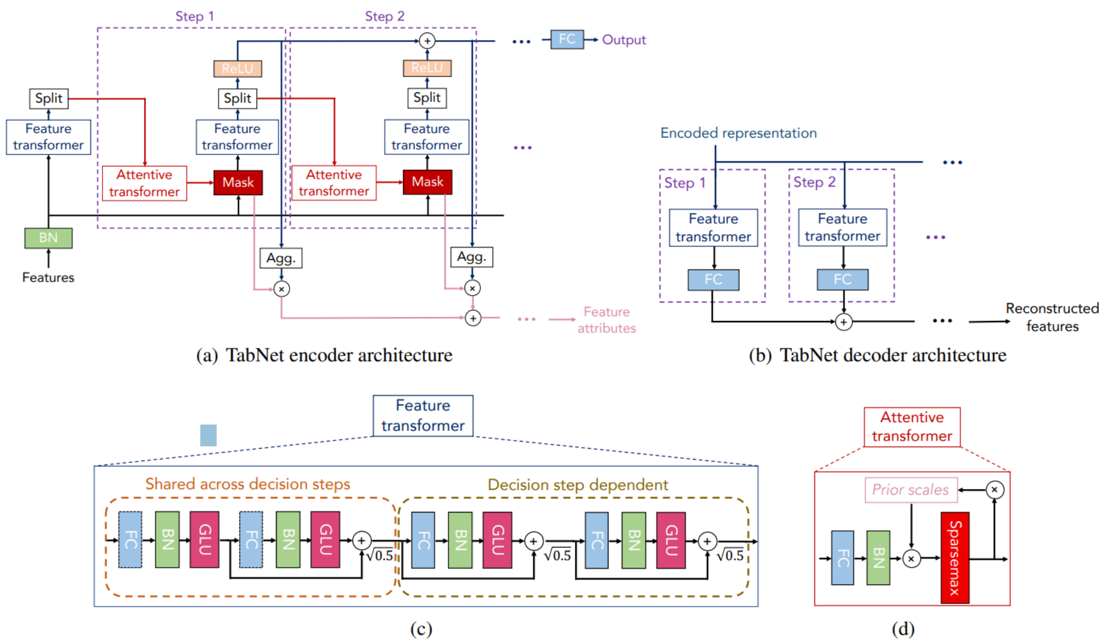


Figure 2.5: *TabNet architecture [17]*

This sequential process allows TabNet to build a complex, non-linear mapping by attending to the most relevant features at each stage. This gives a sort of modularity in the way features are used in the model.

Key Advantages for this Analysis The TabNet architecture provides two significant advantages for the search for a charged Higgs boson:

- **Interpretability:** A major challenge with many deep learning models is their "black box" nature. TabNet offers a high degree of local and global interpretability. The attention masks generated by the attentive transformer can be aggregated across the decision steps and the dataset. This allows for a direct visualization of **feature importance**: one can quantify which variables (e.g., missing transverse energy, lepton p_T, invariant masses) the model finds most discriminative for separating signal ($H^{\pm}$) from background. This is invaluable not only for understanding the model's behavior but also for gaining physical insights and validating that the model has learned physically meaningful correlations.

- **Efficient Learning and Performance:** By selecting only the most necessary features at each step, TabNet allocates its learning capacity more efficiently than a standard MLP. This concept, known as soft feature selection, avoids overwhelming the model with redundant or irrelevant features. This often leads to high performance that is competitive with, and sometimes superior to, state-of-the-art tree-based models like XGBoost, particularly on datasets with a high number of features or intricate dependencies that are not easily captured by axis-aligned splits. Furthermore, since the entire architecture is differentiable, it can be trained end-to-end with standard gradient-based optimization.

Consequently, TabNet was selected as one of the primary models for this analysis. Its ability to learn from raw kinematic variables without extensive manual feature engineering, combined with its built-in interpretability, makes it an ideal candidate for optimizing the search for new physics in a complex, high-dimensional feature space.

XGBoost

eXtreme Gradient Boosting, or XGBoost, is a highly optimized and scalable implementation of the gradient boosting algorithm [18]. Since its introduction, it has become a dominant force in machine learning competitions involving structured or tabular data and is widely regarded as a **state-of-the-art benchmark for classification and regression tasks**. Given its proven performance and robustness, XGBoost serves as an essential reference model in this analysis for discriminating between the charged Higgs boson signal and Standard Model backgrounds.

Core Concept: Gradient Boosted Trees XGBoost is an **ensemble learning method** that builds a strong predictive model by sequentially combining a series of simpler models, known as "weak learners." In the case of XGBoost, these weak learners are **decision trees**. The core principle of boosting is iterative improvement:

1. An initial simple model (a tree) is trained on the data.
2. The errors, or residuals, of this model are calculated.

3. A second tree is then trained not on the original target labels, **but specifically to predict the errors made by the first tree**.

4. The predictions of the first and second trees are combined, yielding a more accurate overall model.

This process is repeated, with each new tree correcting the residual errors of the current ensemble. XGBoost refines this process by using gradient descent. Instead of fitting to the simple residuals, each new tree is trained to fit the *gradient* of the loss function with respect to the predictions of the existing ensemble. This allows the model to optimize any differentiable loss function, making it a highly flexible and powerful framework.

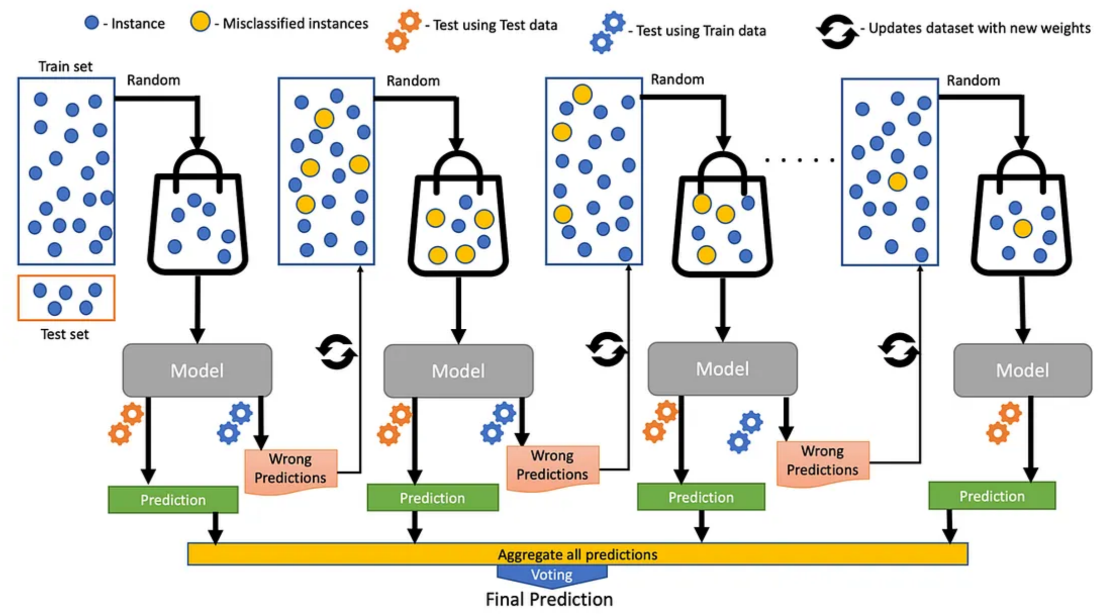


Figure 2.6: XGBoost architecture [19]

Key Advantages and “eXtreme” Features The ”eXtreme” in XGBoost refers to the significant engineering and algorithmic optimizations that set it apart from traditional gradient boosting implementations:

- **Regularization:** To combat overfitting, a common pitfall for tree-based models, XGBoost incorporates both L1 (Lasso) and L2 (Ridge) regularization terms into its objective function. This penalizes model complexity by controlling the number of leaves and the magnitude of the weights at the leaves, leading to better generalization on unseen data. This is particularly important in physics analyses where training statistics can be limited.

- **High Predictive Performance:** XGBoost is renowned for its ability to achieve exceptionally high accuracy. The boosting process allows it to model complex, non-linear relationships in the data, making it highly effective at separating signal and background events that are defined by subtle correlations between multiple kinematic variables.

- **Computational Efficiency and Scalability:** The algorithm is designed for performance. It includes novel techniques like a sparsity-aware split-finding algorithm to handle missing values efficiently, and a cache-aware block structure for out-of-core computation. Furthermore, tree construction can be parallelized, significantly reducing training time on large datasets.

- **Feature Importance:** Like other tree-based models, XGBoost provides metrics to rank the importance of input features. These scores (e.g., based on "gain," "cover," or "frequency") quantify how much each variable contributes to the model's predictive power, offering valuable physical insight and a crucial tool for model validation.

In this thesis, XGBoost represents the gold standard for gradient boosted decision trees. Its inclusion allows for a direct comparison between a top-tier classical machine learning approach and the deep learning architectures of TabNet, providing a comprehensive evaluation of the most effective methods for this physics search.

2.4.3 Model Evaluation Metrics

Confusion Matrix

A confusion matrix is a table that provides a detailed summary of a classifier's performance for a specific discrimination threshold. For a binary classification problem, it is a 2x2 matrix that tabulates the four possible outcomes

- **True Positives (TP):** Signal events correctly identified as signal,

- **True Negatives (TN):** Background events correctly identified as background,

- **False Positives (FP):** Background events incorrectly identified as signal (Type I error),

- **False Negatives (FN):** Signal events incorrectly identified as background (Type II error).

The matrix provides a clear view of not only the correct classifications (the main diagonal) but also the types of errors the model is making. It serves as the foundation from which many other performance metrics, including precision and recall, are derived.

ROC

The Receiver Operating Characteristic (ROC) curve is a fundamental tool for visualizing the performance of a **binary classifier**. It is generated by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various discrimination threshold settings. The TPR, also known as signal efficiency or sensitivity, is the fraction of signal events correctly classified as signal

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{2.14}$$

The FPR is the fraction of background events incorrectly classified as signal (i.e., background rejection is $1 - \mathrm{FPR}$):

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}. \tag{2.15}$$

An ideal classifier would yield a point at the top-left corner of the plot (TPR=1, FPR=0). A random classifier corresponds to the diagonal line (TPR = FPR). The overall performance is often summarized by a single scalar value, the Area Under the Curve (AUC). An AUC of 1.0 signifies a perfect classifier, while an AUC of 0.5 indicates no discriminative power.

F1 and Recall

While the ROC curve evaluates performance across all thresholds, it is often necessary to assess a model at a single operating point. For this, metrics like Recall, Precision, and the F1-Score are used.

Recall measures the signal efficiency—the ability of the model to find all the relevant (signal) cases within the dataset.

Precision measures the signal purity—the proportion of events classified as signal that are actually signal. It is defined as

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2.16}$$

There is often a trade-off between precision and recall; improving one may degrade the other.

The **F1-Score** provides a single metric that balances both precision and recall by calculating their harmonic mean

$$\mathrm{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.17}$$

The harmonic mean heavily penalizes extreme values, making the F1-score a robust measure of a classifier's accuracy, particularly when dealing with imbalanced datasets, which are common in particle physics where signal events are far rarer than background.

2.5 Data-Driven Background Estimation and Statistical Model

2.5.1 Fake Lepton Estimation with Template Fits

In many analyses, particularly those with multi-lepton final states, a significant background arises from "fake" or "non-prompt" leptons. These are lepton candidates that do not originate from an electroweak boson decay (W, Z) but from other sources, such as the semi-leptonic decay of heavy-flavor hadrons, photon conversions, or the misidentification of jets. The rate of these processes is notoriously difficult to model accurately with Monte Carlo (MC) simulation alone.

To obtain a reliable estimate, this analysis employs the semi-data-driven **Template Fit** method [20]. The core principle is to use MC simulation to model the kinematic *shapes* ("templates") of all physics processes, but to use collision data to constrain the overall *normalization* of the fake lepton backgrounds. This is achieved by performing a fit in signal-depleted but background-enriched **Control Regions (CRs)**. In this fit, the normalizations of the fake lepton templates are allowed to float as free parameters, known as **Normalization Factors (NFs)**. The process provides a data-corrected estimate of the fake background, and the resulting NFs, along with their uncertainties, are then incorporated as **nuisance parameters** into the final statistical model used for inference.

2.5.2 Exclusion Limits using Profile Likelihood Fits

The final statistical interpretation, which combines all analysis channels and systematic uncertainties to set exclusion limits, is performed using a **profile likelihood fit**. This frequentist method is the standard for searches at the LHC.

The foundation of the fit is the construction of a binned **likelihood function**, $\mathcal{L}(\text{data}|\mu, \theta)$, which models the probability of observing the data in all analysis bins given a set of model parameters. These parameters are:

- The **signal strength**, μ, which is the parameter of interest (POI). It scales the predicted normalization of the signal process. A value of $\mu = 1$ corresponds to the nominal theoretical cross-section, while $\mu = 0$ corresponds to the background-only hypothesis.

- A vector of **nuisance parameters**, θ. These parameters encode all sources of systematic uncertainty in the analysis. This includes uncertainties on detector performance (e.g., jet energy scale, b-tagging efficiency), theoretical cross-sections, and critically, the **data-driven Normalization Factors** for the fake lepton backgrounds derived from the template fits described in Section 2.5.1. Each nuisance parameter is typically constrained by a Gaussian or log-normal probability distribution.

To set a limit on the signal strength μ, we "profile" the likelihood function. For each specific value of μ being tested, the likelihood is maximized with respect to all nuisance parameters, θ. This yields the conditional maximum likelihood estimate, $\hat{\hat{\theta}}(\mu)$.

The primary test statistic, q_μ, is the **profile likelihood ratio**:

$$q_\mu = -2\ln\frac{\mathcal{L}(\text{data}|\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})} \tag{2.18}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the values that give the global maximum of the likelihood function. This statistic measures the compatibility of the data with a signal of strength μ and is used to exclude signal hypotheses, typically at the 95% confidence level (CL), using standardized methodologies like the CL_s procedure.

To prevent any potential experimenter's bias, this analysis follows a **blinded** procedure, meaning the real data in the signal region (SR) is not examined until all analysis choices are finalized. Consequently, to assess the analysis's sensitivity, an **Asimov dataset** is used in place of real data. This dataset represents the median, background-only outcome of the experiment and is generated by `TRExFitter` to estimate the **expected** upper limit on μ. The limits presented in this thesis are based on this Asimov dataset. They do not yet include the systematic uncertainties and thus serve as an initial estimate of the analysis sensitivity.

Chapter 3: Tools and Libraries Used

3.1 ATLAS Athena

Athena is the primary software framework for offline data processing in the ATLAS experiment. It is built upon the Gaudi architecture, a flexible and modular C++ framework designed to handle the immense scale and complexity of data from event simulation, reconstruction, and physics analysis [21]. Athena provides a common environment for all of these tasks, ensuring consistency and reusability of code across the collaboration.

The framework operates on a component-based model, where the main processing units are known as **Algorithms**. These algorithms are executed sequentially on each event, reading data from the event store, performing specific tasks—such as object calibration, event selection, or variable calculation—and writing new information back to the event store. The configuration and orchestration of these components are managed through Python-based **Job Options (JO)** files, which allow physicists to steer the execution flow and customize the analysis chain without modifying the underlying C++ code.

In the context of this thesis, the Athena framework was instrumental during the initial stages of data preparation and dataset production. It was used to produce the primary Monte Carlo simulations to prepare **xAOD** (extended Analysis Object Data). The final output from this step of analysis is a set of `DAOD_PHYS` **(Derived Analysis Object Data)** files. Which can then be processed with `TopCPToolkit` to produce NTuples which can subsequently be processed by FastFrames before ML and statistical analysis. The following configurations were used to produce the MC datasets

- The hard-scatter processes were generated by `MadGraph5AMC@NLO`
- `Pythia8` was used for showering.
- The PDF Base Fragment used was `NNPDF30NLO` in the 4-Flavour Scheme.

As part of this study, JOs were written to configure simulation processes using the `2HDMTypeII` BSM model. Specific parts of the Control Option (CO) are listed below.

3.1.1 MadGraph Configuration

The code snippet in Listing 3.1 shows how the hard-scatter process is defined in MadGraph. The imported BSM model is `2HDMtypeII` from the ATLAS cvmfs library. In the following lines, the particles (protons, jets, leptons, etc.) are defined using the "define" command and finally the the hard-scatter process is generated using the "generate" command where a proton-proton scatter is defined to produce the $\bar{t}bH^+$ **process, the signal channel that is the prime focus of this study.**

```
#Defining the process in MadGraph
process="""
    set group_subprocesses Auto
    set ignore_six_quark_processes False
    set loop_optimized_output True
    set complex_mass_scheme False
    import model 2HDMtypeII
    define p = g u c d s u~ c~ d~ s~
    define j = g u c d s u~ c~ d~ s~
    define l+ = e+ mu+
    define l- = e- mu-
    define vl = ve vm vt
    define vl~ = ve~ vm~ vt~
    generate p p > t~ h+ b [QCD]
    output -f
    """
```

Listing 3.1: *Process definition in the JO for MadGraph*

Parameters for the masses are defined in the block shown in Listing 3.2. The neutral Heavy Higgs is configured to have a mass dependent on the hypothesized charged Higgs mass according to $H = \sqrt{m_{H^+}^2 + m_W^2}$. It should be noted that the free parameter β is not defined in the JO, so the default value from the model is used, $\tan\beta = 20$.

```
mh1 = 125.0               # light neutral Higgs
mhc = hp_mass             # charged Higgs
mw = 80.399               # W boson mass
mh2 = math.sqrt(mhc**2 + mw**2) #Heavy Neutral Higgs
mh3 = mh2

masses = {'25': mh1,
          '35': mh2,
          '36': mh3,
          '37': mhc
          }
```

Listing 3.2: *Setting masses for particles in 2HDMtypeII*

3.1.2 Pythia Showering

Once the hard-scatter process is generated with MG, the LHE file thus produced is interfaced with Pythia to shower the particles and generate subsequent decays to stable final states, the configurations can be seen in Listing 3.3, where the allowed decay channels are restricted to $H^+ \to Wh$ and $h \to \tau^+\tau^-$. Note that **no further restrictions** are applied to W or τ.

```
evgenConfig.description = 'aMcAtNlo High mass charged Higgs NLO4FS'
evgenConfig.keywords+=['Higgs','MSSM','BSMHiggs','chargedHiggs']
evgenConfig.contact = ['Adwait Meppurath : adwait.meppurath@cern.ch']
runArgs.inputGeneratorFile=outputDS

genSeq.Pythia8.Commands += ["Higgs:useBSM = on",
                            "37:onMode = off",# turn off all mhc decays
                            "37:onIfMatch = 24 25"]# switch on H+ to Wh
evgenConfig.keywords+=['W','Higgs']
genSeq.Pythia8.Commands += ["Higgs:useBSM = on",
                            "25:onMode = off",# turn off all Higgs(h)
    decay
                            "25:onIfMatch = 15 -15"]#Turn on h to tau tau
     decay
evgenConfig.keywords+=['tau','tau']
```

Listing 3.3: *Setting allowed decay modes for Higgs and charged Higgs in Pythia*

3.1.3 xAOD Filter: Lepton pT and Eta Cuts

Finally, the events are filtered according to the p_T of the final-state light-leptons. The filter is imported from the Generator Filters available within the Athena framework, namely the `xAODMultiElecMuTauFitler`. Truth information is used to filter events. Since we apply two separate p_T cuts on the leading and sub-leading leptons the events are filtered through a Di-Lepton filter first with the lower p_T and then a Single-Lepton filter with the higher p_T to achieve the desired cuts on the two leptons. No restrictions are applied on the τ as can be seen in Listing 3.4

```
createxAODSlimmedContainer("TruthGen",prefiltSeq)
prefiltSeq.xAODCnv.AODContainerName = 'GEN_EVENT'

from GeneratorFilters.GeneratorFiltersConf import xAODMultiElecMuTauFilter
filtSeq += xAODMultiElecMuTauFilter("DiLeptonFilter")
filtSeq += xAODMultiElecMuTauFilter("SingleLeptonFilter")

xAODMultiElecMuTauFilter2 = filtSeq.DiLeptonFilter
xAODMultiElecMuTauFilter2.NLeptons = 2
xAODMultiElecMuTauFilter2.MinPt = 8000.
xAODMultiElecMuTauFilter2.MaxEta = 2.8
```

```
xAODMultiElecMuTauFilter2.IncludeHadTaus = False

xAODMultiElecMuTauFilter1 = filtSeq.SingleLeptonFilter
xAODMultiElecMuTauFilter1.NLeptons = 1
xAODMultiElecMuTauFilter1.MinPt = 20000.
xAODMultiElecMuTauFilter1.MaxEta = 2.8
xAODMultiElecMuTauFilter1.IncludeHadTaus = False

filtSeq.Expression = "SingleLeptonFilter and DiLeptonFilter"
```

Listing 3.4: *Setting the Di-Lepton Filter based on p_T and η cuts*

The full CO with the individual JOs for each mass hypothesis can be found in the linked GitLab repo[1]

3.2 TopCPToolkit: TauX Subgroup

The TopCPToolkit is a modern, centralized ntuple production framework developed by members of the ATLAS Top Working Group for widespread use in Run 2 and Run 3 physics analyses [22]. It serves as the successor to the older AnalysisTop framework and is designed to standardize the creation of analysis-level datasets by building upon the centrally managed Common Physics (CP) algorithms. The use of these common tools ensures that physics objects (leptons, jets, etc.) across different analyses are defined, corrected, and selected using a consistent and validated approach.
While Athena provides the foundational environment for low-level data processing, the TopCPToolkit is employed in the subsequent step to convert the large xAOD files into lightweight, "flat" NTuples. This analysis was done under the Tau+X umbrella (part of the **LPX** Subgroup under **Exotics** working group at CERN) , hence the TauX version of TopCPToolkit was used (called `taux_gnt1makeralg`[2].). The toolkit was configured to apply a baseline event and object selection to retain all potentially interesting events for the final analysis.
The primary event selection required events to pass standard data quality checks (GRL, PV, goodCalo flags) and a set of single-lepton or di-lepton triggers. A minimal pre-selection was applied to categorize events based on their lepton content, primarily requiring at least one hadronic tau candidate in conjunction with zero, one, or two light leptons (e, μ).

The baseline physics object definitions used to produce the NTuples are summarized below [23]:

[1]The CO can be found in the DSID folder 567607, the rest of the DSIDs have the COs symlinked to the main one in 567607

[2]The GitLab repo for the TauX version of TopCPToolkit

- **Electrons:** Required to have $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.47$ (excluding the calorimeter crack region), pass the `TightLH` identification working point, and satisfy `Loose_VarRad` isolation criteria, along with standard impact parameter requirements.
- **Muons:** Required to have $p_{\mathrm{T}} > 10$ GeV, $|\eta| < 2.5$, pass the `Loose` identification working point, and satisfy `Loose_VarRad` isolation and standard impact parameter cuts.
- **Jets:** Reconstructed using the Anti-Kt algorithm with a radius of $R = 0.4$. Jets were required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 4.9$. A jet vertex tagger (`NNJVT`) was applied to central jets to suppress pileup.
- **Hadronic Taus:** Required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ (excluding crack regions). Multiple identification working points, based on both the legacy RNN algorithm and the newer GNTau algorithm, were considered and stored to allow for flexibility in the final analysis.

A key feature of the NTuple production was the approach to Overlap Removal (OR). Instead of applying a single, fixed OR procedure, the toolkit was configured to save multiple boolean flags for several OR hypotheses (referred to as `wpSet0`, `wpSet1`, etc.). Each set corresponds to a different combination of lepton identification working points. This sophisticated approach provides maximum flexibility, allowing the final object selection and ambiguity resolution to be optimized in the downstream analysis rather than being fixed at the NTuple level.

By handling this crucial data-slimming stage, the TopCPToolkit produces output files that are highly optimized for rapid processing. This significantly accelerates the later stages of the analysis.

3.3 FastFrames

Once the analysis-specific NTuples are produced by TopCPToolkit, the next crucial step is to add custom-defined variables/features relevant to the ML models as well as trim down the dataset with pre-selections relevant to the analysis. For this task, the FastFrames package was used [24]. FastFrames is a high-performance tool, written in C++ with convenient Python bindings, designed for the rapid processing of flat NTuples.

The central design philosophy of FastFrames is declarative configuration. Instead of requiring the user to write explicit C++ or Python event loops, it allows all aspects of the histogramming process—including variable definitions, event weighting, systematic variations, and final selections—to be defined in a single configuration file (typically in YAML format). Internally, FastFrames leverages the power of ROOT's `RDataFrame`. This modern ROOT interface enables highly optimized, multi-threaded processing of data through a columnar analysis approach, which minimizes I/O operations and efficiently parallelizes the workload across multiple CPU cores. This results in a significant reduction in the time required to produce a full set of analysis histograms compared to traditional methods.

In this analysis, FastFrames was the primary tool used to process the NTuples. It was configured to apply the final event selections for the signal and control regions. The SR NTuples were then used to train the ML/DL models which were then subsequently used to infer signal probability on the SR+CR dataset (also produced with FastFrames) and then analyzed statistically with `TRExFitter`. The most amount of effort in this study went to defining over 70 custom features, mostly specific to this analysis channel using a custom frame, which is briefly discussed in the Results section 4.2.

3.4 TRExFitter

`TRExFitter` [25] is a powerful and widely used statistical fitting framework designed to perform the final statistical interpretation of high-energy physics analysis results. It serves as the bridge between the final histogram outputs and the ultimate physics conclusions, such as setting exclusion limits or measuring cross-sections.

The framework takes as input the histograms of kinematic variables (in this case, produced by FastFrames and/or inferred upon by ML models) for the signal and all relevant background processes. Through a set of human-readable text-based configuration files (called configs), the user defines the complete statistical model. This includes specifying the analysis regions (signal and control regions), the list of samples, and, most importantly, the full set of systematic uncertainties[3] and their correlations across different regions and samples.

Based on this configuration, TRExFitter automatically constructs a binned likelihood function, $\mathcal{L}(\text{data}|\mu, \theta)$. This function models the probability of observing the data given a signal strength parameter, μ, and a set of nuisance parameters, θ, which represent the systematic uncertainties. The fit then maximizes this likelihood function to find the best-fit values for the nuisance parameters, effectively constraining

[3] In this study, however, systematic uncertainties are not included. It falls in the future scope of this analysis

the background model using data from the **Control Regions** (CRs). A list of CR definitions for fake analysis of $m_{H^+} = 400$ GeV is given in Table 3.1 [4]

In this analysis, TRExFitter was used to perform several critical final steps:

- It was used to fit the norm-factors for fake leptons in the CRs to real data, to mitigate over/under-estimation of fakes by the simulated data.
- It set 95% confidence level (CL) upper limits on the production cross-section of the charged Higgs boson using the CL_s method [26], [27].
- It generated pre-fit and post-fit plots showing the agreement between the observed data and the simulation in the fitted CRs.

In essence, `TRExFitter` is the final tool used in this long chain of analysis to finally determine tangible, statistically sound results.

3.5 Python

Beyond its role in configuring the Athena JOs and ML/DL models, the Python programming language served as the primary tool for custom analysis tasks, data manipulation and custom plot creation. Its extensive ecosystem of open-source scientific libraries makes it an ideal environment for modern data analysis. A (non-exhaustive) list of libraries used in this study is given below :

- **Core Data Handling and Numerics:**
  - `NumPy`: The fundamental package for numerical computation in Python, used for handling arrays and performing mathematical operations.
  - `Pandas`: Used for data manipulation and analysis, particularly for managing the structured datasets before training.
  - `Uproot`: The primary tool for efficiently reading and writing data from the `.root` files produced by the analysis frameworks.
- **Machine Learning Frameworks:**
  - `Scikit-learn`: A comprehensive library for various machine learning tasks, including data splitting (`train_test_split`), model evaluation metrics (e.g., ROC, AUC, F1-score), and dimensionality reduction (`PCA`).

[4] This is specific to $m_{H^+} = 400$ GeV, especially in the way `sig_prob` is used to remove signal contamination. The full config used for each of the 14 mass points can be found in the linked repo. The cut on `sig_prob` is specific to each mass point based on the optimal F1 threshold.

Region Name	Key Selection Criteria	Fit Variable
FakeElSS_CR	• Two same-sign light leptons. • ≤ 3 jets, ≥ 0 b-jets. • High-quality tau candidate (quality > 0.5). • Exactly one lepton is an electron that **fails** its quality selection (< 0.5), while the other light lepton **passes** its quality selection (> 0.8).	Leading Lepton p_T
FakeMuSS_CR	• Two same-sign light leptons. • ≤ 3 jets, ≥ 0 b-jets. • High-quality tau candidate (quality > 0.6). • Exactly one lepton is a muon that **fails** its quality selection (≤ 0.5), while the other light lepton **passes** its quality selection (> 0.8). • Signal suppression cut (sig_prob ≤ 0.38).	Leading Lepton p_T
FakeLepTauSS_CR	• Two same-sign light leptons. • ≤ 3 jets, ≥ 0 b-jets. • Both light leptons must be high-quality (≥ 0.8). • The tau candidate must **fail** its quality selection (< 0.45). • Signal suppression cut (sig_prob ≤ 0.38).	Tau p_T
FakeTauOS_CR	• Two **opposite-sign** light leptons. • ≤ 3 jets. • Both light leptons must be high-quality (≥ 0.8). • The tau candidate must **fail** its quality selection (< 0.45). • Signal suppression cut (sig_prob ≤ 0.38).	Tau η

Table 3.1: *Definitions of the Control Regions (CRs) used in the template fit for fake-lepton background estimation. The selection criteria are designed to enrich the sample in a specific fake lepton category while remaining* ***orthogonal*** *to the Signal Region. These selections are specific to $m_{H^+} = 400$ GeV*

  - `XGBoost`: The high-performance implementation of the gradient-boosted decision tree algorithm used as a primary model in this thesis.
  - `PyTorch-TabNet`: The implementation of the attention-based deep learning model, TabNet.
  - `PyTorch`: The underlying deep learning framework that serves as the backend for the TabNet model.

- **Data Visualization:**
  - `Matplotlib`: The foundational library for creating all static plots, figures, and graphs presented in this work.
  - `Seaborn`: Used for creating enhanced statistical visualizations, such as the heatmaps for the confusion matrices.
  - `mplhep`: A specialized library used to apply the official ATLAS collaboration styling to all plots, ensuring they meet publication standards.

- **Hyperparameter Optimization and Data Processing:**
  - `Optuna`: The framework used for the efficient, automated hyperparameter optimization studies.
  - `Imbalanced-learn`: Utilized to address class imbalance in the training dataset via the SMOTE (Synthetic Minority Over-sampling Technique) algorithm.
  - `joblib`: Used for saving the final, trained machine learning models to disk and for loading them during the inference stage.

- **Utilities and Workflow Management:**
  - `os`, `sys`, `argparse`, `json`: Standard libraries used for interacting with the operating system, parsing command-line arguments to steer the analysis scripts, and managing configuration files.

Chapter 4: Results

4.1 ATLAS Athena Dataset Production

The new MC datasets to extend upon Azad Afandizada's analysis, which used 3 mass points (512185-512187 in Table 4.1) of the tbH^+ channel [28], were generated using the ATLAS Athena framework. This involves writing Job Options (JOs), which control what BSM model is to be used for the generation of the hard-scatter processes and the subsequent parton-level information stored in the Les Houches Event File (LHEF) format [29]. More information about the MC generation is given in Sec. 3.1.

11 new NTuples were successfully produced after processing the `DAOD_PHYS` files from Athena with `TopCPToolkit`. The DSIDs of the datasets used in this analysis can be found in Tables 4.1 and 4.2.

DSID	H^+ **Mass (GeV)**	**DSID**	H^+ **Mass (GeV)**
512185	250	567613	900
512186	3000	567614	1000
512187	800	567615	1200
567607	300	567616	1400
567608	350	567617	1600
567609	400	567618	1800
567610	500	567619	2000
567611	600	567620	2500
567612	700		

Table 4.1: *DSIDs and hypothesized H^+ mass points in Release 22*

Sample Name	**DSIDs**	**Sample Name**	**DSIDs**
$t\bar{t}$	410470	tZ	410560
threeTop	304014	$ttWW$	410081
Z-jets	700322, 700324, 700325, 700335,	*fourTop*	412043
	700336, 700337	WtZ	410408
		$ttWZ$	500463
W-jets	700338-700340	$ttHH$	500460
		$ttWH$	500461
		$tt\gamma$	500800, 504554
VV	700589-700605,	$V\gamma$	700398-700404,
SingleTop	410659, 410644, 410645		
VH	346645, 346646	$ttZZ$	500462
ttZ	410276, 410277, 410278	tH	545796
ttW	700168		
ttH	346343, 346344, 346345		

Table 4.2: *Full list of DSIDs of background SM processes used in this analysis*

4.2 Tau + X FastFrames Custom Frame

The `TauXFastFrame` custom frame was used in FastFrames to define custom features with basic objects. An example of a feature definition in the custom frame is shown in code snippet 4.1. In similar fashion **more than 70 features were added** to the custom frame. The linked custom frame[1] is a clone of the original `TauXFastFrame`[2] repo with more than 70 custom features added to it, mostly specific to the tbH^+ signal in the $2l_{SS}1\tau$ final state, which could be useful for future analyses. The list of features are given in Table A.1.

```
//Inv. mass of the 2l1tau4j system
  auto get_total_invariant_mass_2l1t4j = [](const std::vector<TauXParticles::
    Lepton>& leptons, const std::vector<TauXParticles::Jet>& jets, const std
    ::vector<TauXParticles::Tau>& taus) {
    if (leptons.size() < 2 || jets.size() < 4 || taus.empty()) return
    C_INVALID_VAL;

    ROOT::Math::PtEtaPhiMVector total_system_p4(0.0, 0.0, 0.0, 0.0);

    // Add 2 leptons, 1 tau, 4 jets
    total_system_p4 += ROOT::Math::PtEtaPhiMVector(leptons[0].pt(), leptons
    [0].eta(), leptons[0].phi(), leptons[0].mass());
```

[1]The full custom frame with all custom defined features can be found here
[2]The original TauXFastFrame repo, maintained by the Tau+X umbrella

```
    total_system_p4 += ROOT::Math::PtEtaPhiMVector(leptons[1].pt(), leptons[1].eta(), leptons[1].phi(), leptons[1].mass());
    total_system_p4 += ROOT::Math::PtEtaPhiMVector(taus[0].pt(), taus[0].eta(), taus[0].phi(), taus[0].mass());
    for (size_t i = 0; i < 4; ++i) {
        total_system_p4 += ROOT::Math::PtEtaPhiMVector(jets[i].pt(), jets[i].eta(), jets[i].phi(), jets[i].mass());
    }

    return total_system_p4.M() * 0.001; // Convert to GeV
};
mainNode = MainFrame::systematicDefine(mainNode, "tot_invMass_2l1t4j_NOSYS", get_total_invariant_mass_2l1t4j, {"Leptons_NOSYS", "Jets_NOSYS", "Taus_NOSYS"});
```

Listing 4.1: *Feature definition for the total invariant mass of the 2l1tau4j system in the custom frame in FastFrames*

4.3 ATLAS AF3 vs. FullSim Comparison

The existing three datasets were produced with ATLAS FullSim configuration. The usage of ATLAS AF3 [30] was considered for the new 11 new datasets, but to ensure consistency between the two simulation configurations, comparisons between AF3 and FS were made for the existing three mass points. The final limit upper estimates were found to be within 5% of each other for all three mass points. Further, bin by bin comparisons were made for features, shown in Fig. 4.1. Comparions for other mass points are given in the Appendix (Fig. A.4 and A.5)

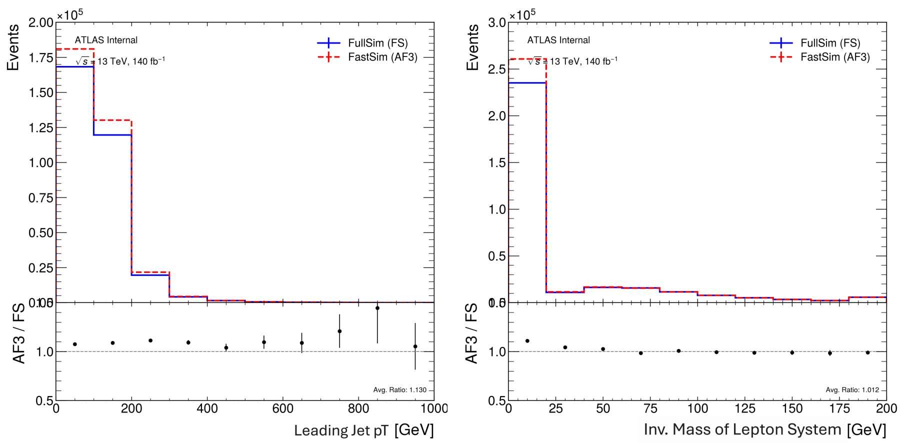


Figure 4.1: *Bin-by-bin comparison of the leading jet p_T and the invariant mass of the leading and sub-leading leptons (M_{ll}^{01}) for $m_{H^+} = 250$ GeV*

Similar comparisons were made for other features as well, each showing an average bin-by-bin ratio close to 1.

4.4 Model Training and Comparison of XGB vs. TabNet

Two models were identified to be used in this analyis - Google TabNet and XGBoost[3]. Both models were trained on the same set of labeled signal-background dataset. The full analysis chain was conducted using the signal probability output of both models and the final limit estimates as well the $S/\sqrt{B}$ ratios were compared.

The configuration for both models were iteratively tested and tuned with trial-and-error and finalized to a set of hyper-parameters shown in Tables 4.3 and 4.4.

Table 4.3: *TabNet Hyperparameters*

Parameter	Value
`n_d`	24
`n_a`	24
`n_steps`	5
`gamma`	1.1
`lambda_sparse`	`1e-5`
`optimizer_params`	`dict(lr=2e-3)`
`mask_type`	`'entmax'`

Table 4.4: *XGBoost Hyperparameters*

Parameter	Value
`n_estimators`	5000
`learning_rate`	0.05
`max_depth`	7
`subsample`	0.8
`colsample_bytree`	0.8
`gamma`	0.1
`tree_method`	*tree_method*

4.4.1 Scale-based Approach

In preliminary tests, it was found that there was a **severe lack of performance for lower hypothesized masses** using a single model, possibly due the model being biased towards the higher masses. So, to maximize the discrimination power of the models, a two-model approach was adopted - namely, low and high-scale models. Where the lower hypothesized mass points are used to train the low-scale model and the high-scale model was trained on the rest of the masses.

- **Low-scale masses :** 250, 350, 400 GeV
- **High-scale masses :** 500, 600, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 3000 GeV

[3]The training codes for both models, which specifies the usage of SMOTE in the low-scale model and the (unused) option to use PCA to reduce dimensionality : TabNet and XGB

Low-scale Model Training

The training of both models were evaluated with the analysis of the Loss, ROC, AUC and Validation accuracy curves.
Looking at Fig. 4.2 and 4.3, we can already make a few inferences about the performances of the two models. The validation LogLoss curve (Fig. 2.4.2 , Left) shows a classic, desirable shape. It starts high (around 0.70) and decreases rapidly, indicating that the model is quickly learning the patterns in the data. The curve then flattens out after approximately 1000 boosting rounds, settling at a low value (around 0.29). This plateau signifies that the model has reached its optimal performance, and adding more trees yields diminishing returns. Crucially, the loss does not start to increase, which indicates that the model is **not overfitting**. On the other hand looking at the validation accuracy curve of TabNet in Fig. 4.2, the accuracy generally increases over the 25 epochs, rising from around 0.37 to a peak of 0.64. However, the curve is quite jagged and volatile, with significant fluctuations from one epoch to the next. This volatility can suggest that the learning rate is slightly too high for the dataset or that the model is sensitive to the specific batches of data it sees in each epoch.

The validation AUC curve for XGB (Fig. 4.3, Right) rises sharply and smoothly, mirroring the drop in LogLoss. It reaches a very high final value of approximately 0.92. An AUC score this high indicates a strong classifier. The plot also explicitly shows that early stopping (to prevent overfitting) was used, with the "Best Iteration" found at round 1741. For TabNet this curve rises quickly to a high value of approximately 0.89 - 0.90. This is an excellent score, indicating that the model is very effective at ranking signal events higher than background events. However, unlike the XGBoost curve which was still slightly climbing, the TabNet AUC curve appears to plateau and even slightly decrease after its peak around epoch 18. This suggests that continuing to train beyond this point is not beneficial and might be leading to the model memorizing noise in the training data, thus slightly hurting its generalization performance.

So far XGB-low has been proving to be better than TabNet-low. But we see a different story when analyzing each of their ROC curves in Fig. 4.4. Overall, both models achieve an AUC of 0.895 (orange line). But when we look at the individual ROCs for the three mass hypotheses chosen for the low-scale models to be trained with, we see that TabNet has a tight clustering of the individual ROCs as compared to XGB, where the 250 GeV ROC (violet dash) is slightly lower than the overall curve as compared to TabNet-low. More importantly, in the low FPR (False-Positive Rate) region (towards the left of the plot), where there is a high background rejection rate, TabNet achieves a much higher signal efficiency.

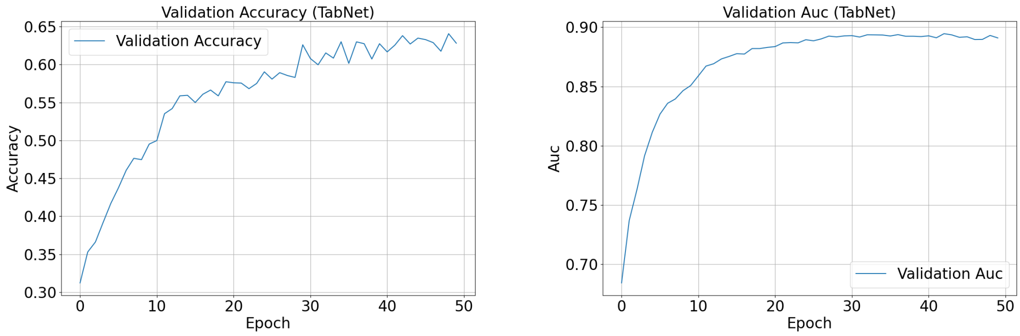


Figure 4.2: *Validation accuracy and Validation AUC values for TabNet-low*

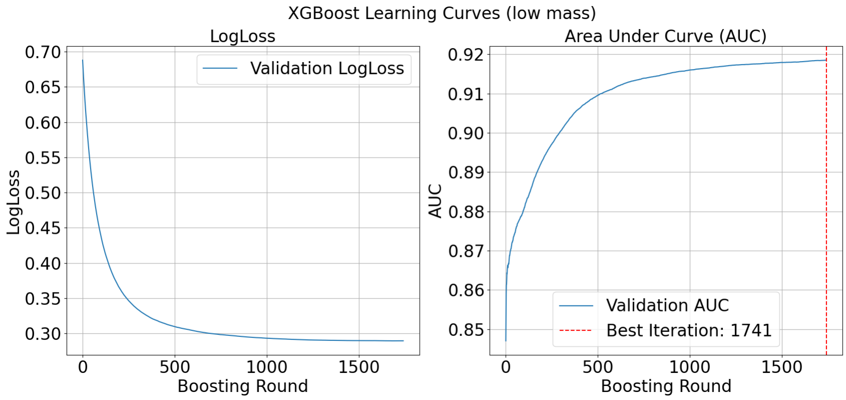


Figure 4.3: *Log loss over epochs and Validation AUC values for XGB-low*

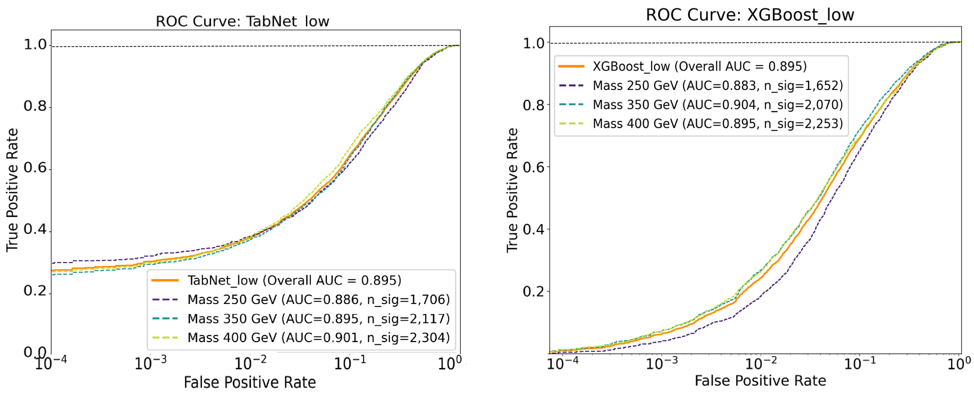


Figure 4.4: *Comparison of overall and individual ROC curves for TabNet-low and XGB-low*

To summarize, for the lower mass hypotheses, overall both models achieve the same AUC for the ROCs but a nuanced look at the clustering of individual ROCs imply that TabNet generalizes well without biases towards the kinematics of a particular mass point and could *potentially* provide a purer signal background separation because of the higher signal efficiency in the low FPR region, even though other training metrics (loss, val. accuracy) favor XGBoost.

High-scale Model Training

For the high-scale models, both XGB-high and TabNet-high achieve a very high AUC value of ≈ 0.97. However their training dynamics indicate that XGB is supremely stable while TabNet is very volatile in it's training. From Figures 4.5 and 4.6, this is evident. XGB is more reliable and more likely to give reproducible results. In the same vein as the low-scale model, we need to also look at the individual ROCs to have a complete picture.

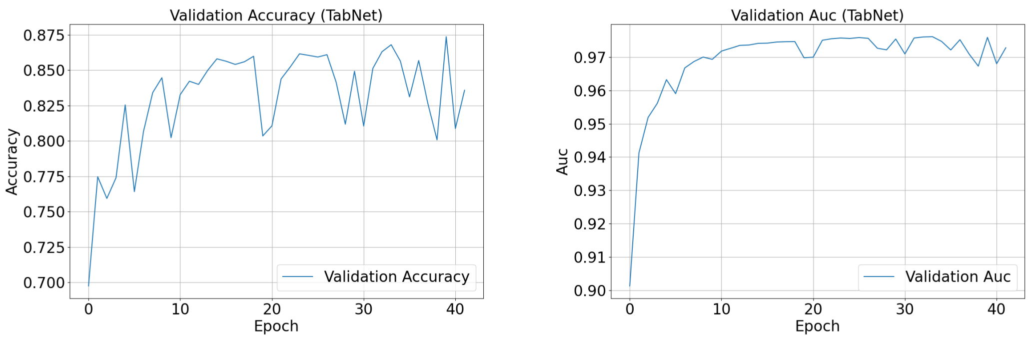


Figure 4.5: *Validation accuracy and Validation AUC values for TabNet-high*

Similar to the low-scale models, looking at the ROC curves in Figs. 4.7 and 4.8, show that overall, both models achieve an AUC of about 0.97 but TabNet is able to deliver better signal efficiencies in the low FPR region to the left of the plot. Another thing that can be observed is that in both cases, the models are able to deliver higher signal efficiencies for higher mass points (Cyan to light green dashes) in the low FPR region, possibly due to the kinematics of the higher mass signals being distinct from the background and thus making them easier to identify.

4.4.2 Precision, Recall and F1

To have a complete understanding of how the models compare with each other, it is also helpful to look at other evaluation metrics. At first glance, looking at the overall confusion matrices for both models, the result points towards XGB being

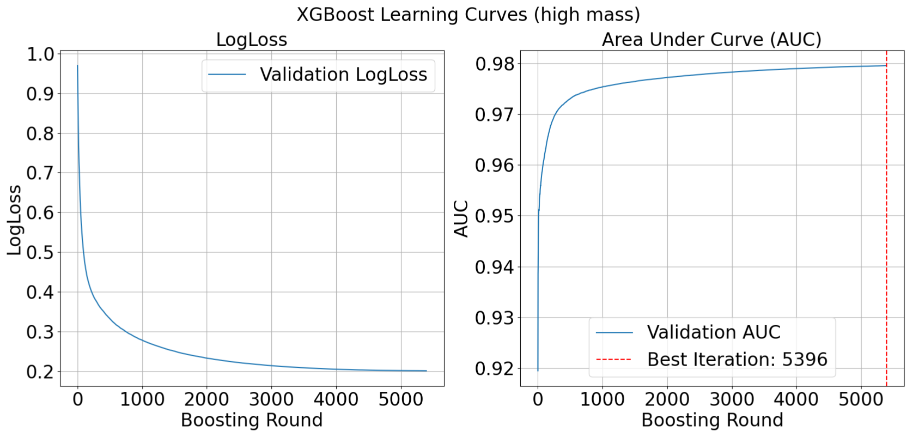


Figure 4.6: *Log loss over epochs and Validation AUC values for XGB-high*

better, shown in Fig. 4.9.

However, upon closer inspection, we understand the finer details. Consider the F1, Recall and Precision curves for both the models, shown in Fig. 4.10. We see that, although XGB achieves a higher $S/\sqrt{B}$ ratio overall (at the optimal threshold) in both mass-scales, **it has an optimal threshold at a significantly higher value than XGB**, especially in the low mass region (0.91 vs 0.36), implying that TabNet is able to learn the complexities between the features well enough to assign a high signal probability for **True Signal** events, which is consistent with what was seen in the low FPR region on the ROC curves (Fig. 4.4). However, it still fails to achieve a higher significance due to also assigning a high signal probability for a proportionally larger amount of background events. Nevertheless, it is interesting to note that TabNet has an optimal threshold almost three times that of XGB. For a full understanding of how the optimal thresholds for individual signal mass points are different between XGB and TabNet (especially for low masses), it is helpful to look at Fig. A.1 and A.2 in the Appendix.

Taking a look at the assigned probability distributions for one of the lower mass points, $m_{H^+} = 250$ GeV is helpful in understanding why XGB performs better, even though it has a drastically lower optimal threshold. Figure 4.11 shows why the optimal threshold for TabNet is high – because the signal events are mostly assigned high probabilities and so it is **concentrated to the right-side of the plot** (note that the y-axis is in log-scale), while for XGB the signal events are **scattered across the range**, hence the threshold has to be lower to capture all of those events. However it is clear that XGB is more successful in classifying backgrounds correctly, indicated by the **grey peak on the left-side of the plot**. Resulting in a higher *relative* number

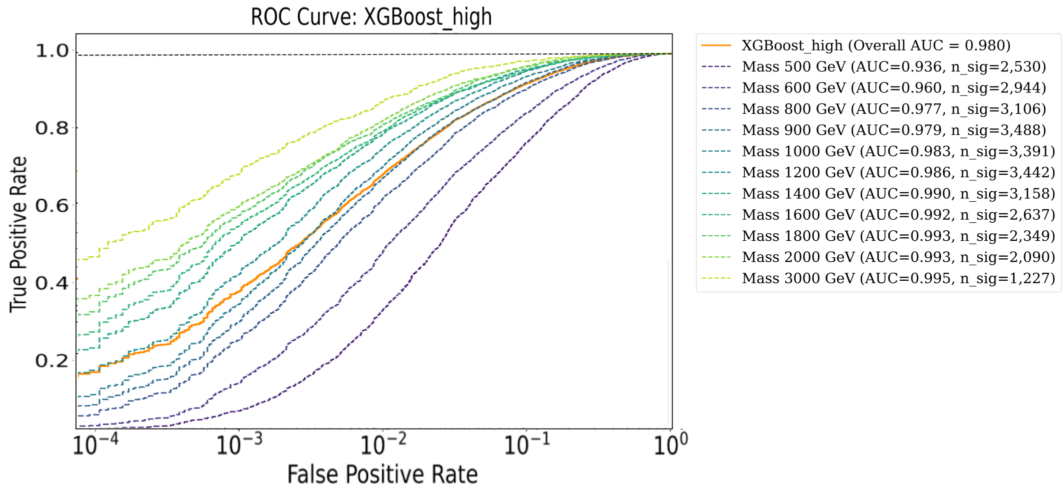


Figure 4.7: *Overall and individual ROC curves for XGB-high*

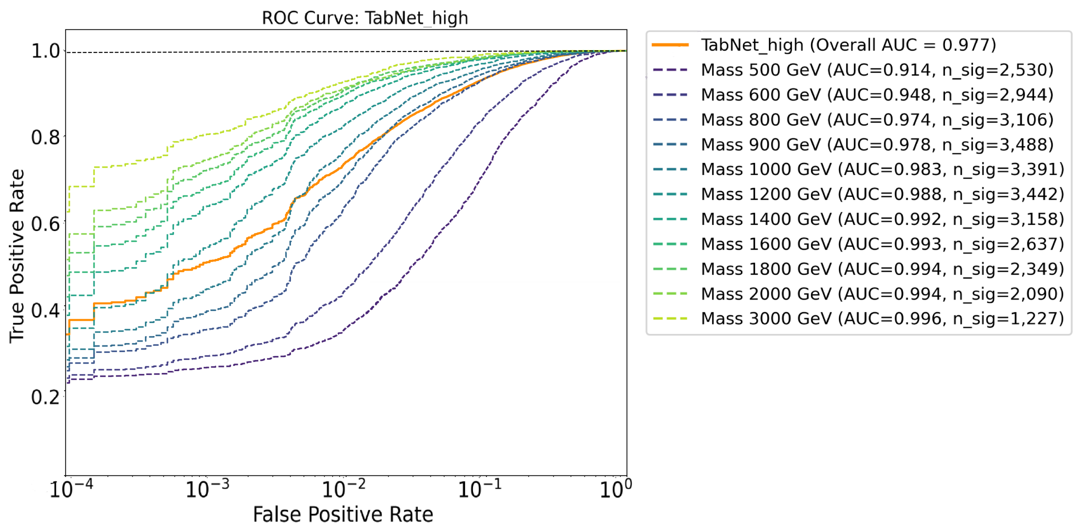


Figure 4.8: *Overall and individual ROC curves for TabNet-high*

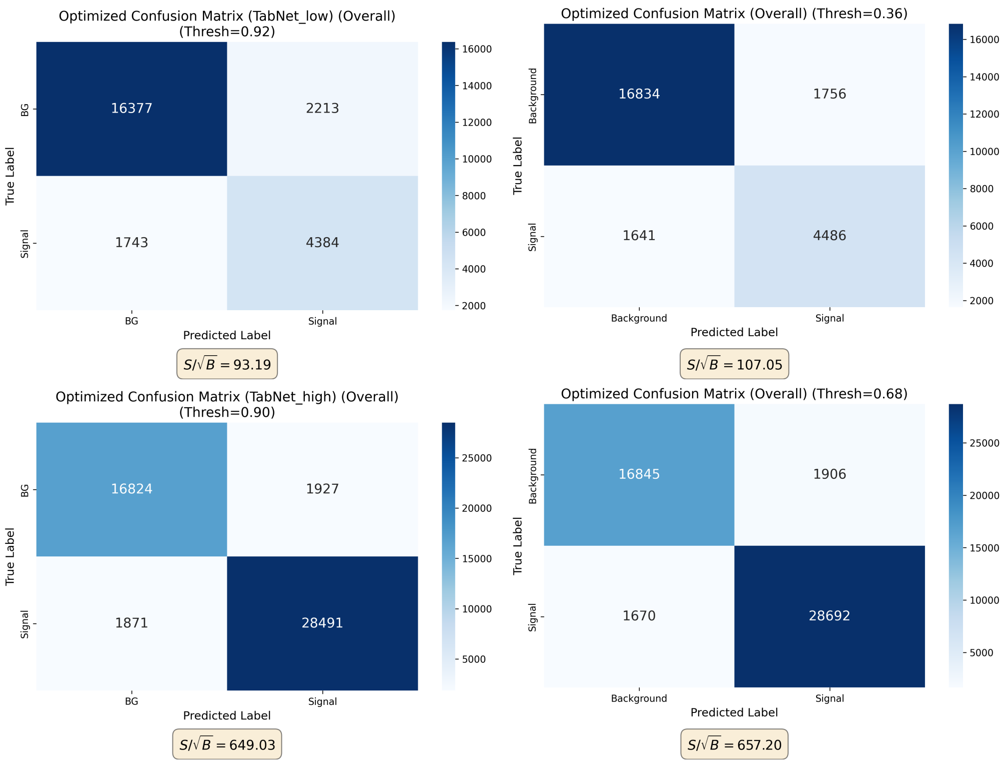


Figure 4.9: *Confusion matrices for low (top) and high(bottom) scale models for TabNet (left) and XGB (right) at the optimal overall threshold*

of signal events as compared to the background.

Another interesting observation can be made if we look at the **individual** $S/\sqrt{B}$ ratio at the optimized threshold for each mass point which can be evaluated by calculating the F1 scores for individual mass points and the Confusion Matrix at the optimal threshold for each mass point. The comparison of the significances for each mass point is shown in Fig. 4.12. Overall, XGB scores consistently higher than TabNet, except a few particular mass points where TabNet outperforms XGB by a significant margin (eg. $m_{H^+} = 1400$ GeV).

After having this detailed understanding of how both models classify signal and background, in order to conclude what model we should proceed with, the expected upper limit was calculated on the signal probability output feature (`sig_prob`) for each model, using a profile-likelihood fit in `TRExFitter`. The estimated limits with both models are presented in Fig. 4.13. Fits on the XGB signal probability consistently produces stronger limits, across the board than TabNet, implying that having an

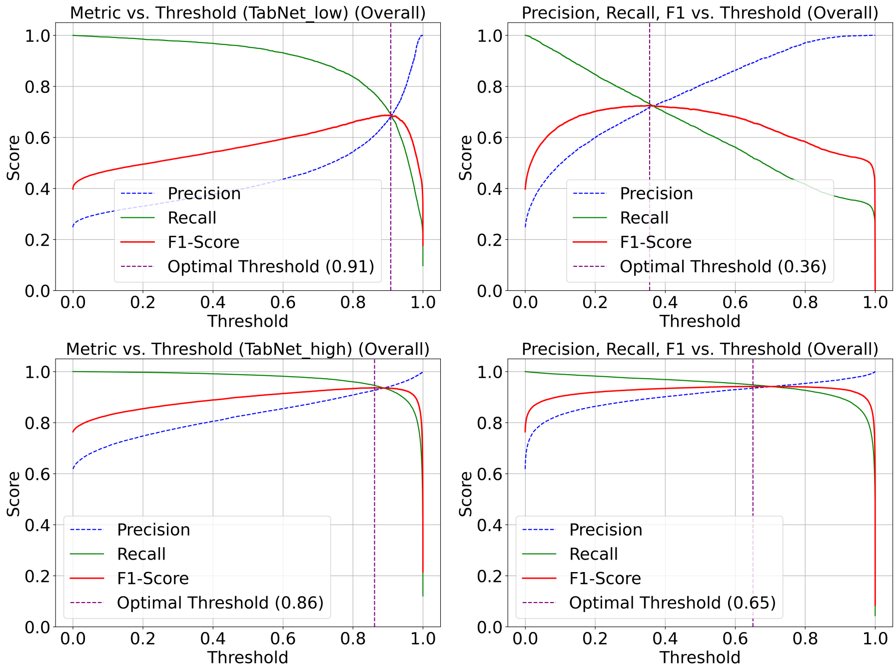


Figure 4.10: *Precision,Recall and F1 curves for low (top) and high(bottom) scale models for TabNet (left) and XGB (right)*

overall higher $S/\sqrt{B}$, regardless of the threshold (as seen with XGB), aides in stronger limits rather than having high F1-based optimal thresholds (as seen in TabNet)

To conclude, from Fig. 4.4, 4.8 and 4.10, TabNet shows great potential in its classification power, but as far as the goal of this study is concerned, XGB performs the task better, as of now. The above reasons are the motivations behind **proceeding with a decision-tree based model** than an attention-based neural network. The reasons as to why TabNet performs better at some mass points, how it is able to have a drastically higher optimal threshold although providing an overall worse performance than XGB require further analysis and are beyond the scope of this study.

4.4.3 Optuna Studies

To maximally optimize the XGB model, both XGB-low and XGB-high were iteratively tested with different hyper-parameters on Optuna[4]. The results are presented in Table 4.5. To see what effect this has had on the analysis, the limits were re-estimated with the optimized models, shown in Fig. 4.14. The plots have the limit estimates

[4]The Optuna studies were conducted with this code

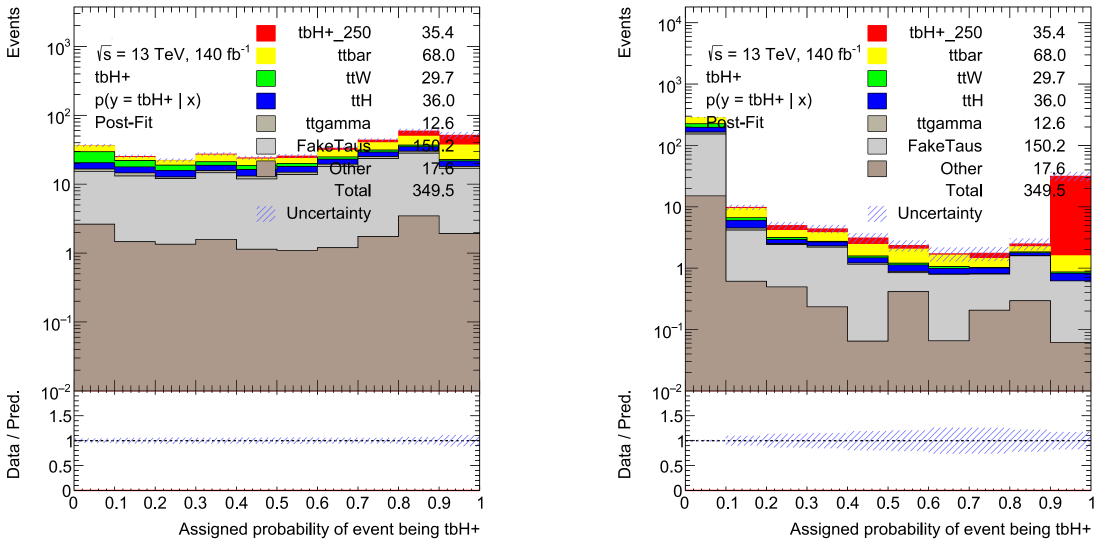


Figure 4.11: ***Left :*** *TabNet assigned probabilities ,* ***Right :*** *XGB assigned probabilities*

based on the signal probability from *both* the models on the entire mass-range. We make note of two main observations from these plots

- As expected, the low-scale model provides stronger limits on the lower mass points of H^+ and the high-scale model provides stronger limits on the higher mass points.
- With Optuna optimized hyper-parameters, the limits on the higher masses don't seem to be affected as much as the lower masses, which has achieved noticeably stronger limits than before.

4.4.4 Feature Importance

It is also interesting to look at the way the same set of features are used by both models by looking at the *Feature-Importance* ranking (Figs. 4.15 and 4.16). The ranking functions come with their respective packages (`model.feature_importances_`). For the low-scale models, both XGB and TabNet give great importance to `sumPsbtag`. However on a relative scale, TabNet seems to depend on a lot of features than XGB which depends greatly on the top 5 features and then the importance sharply drops

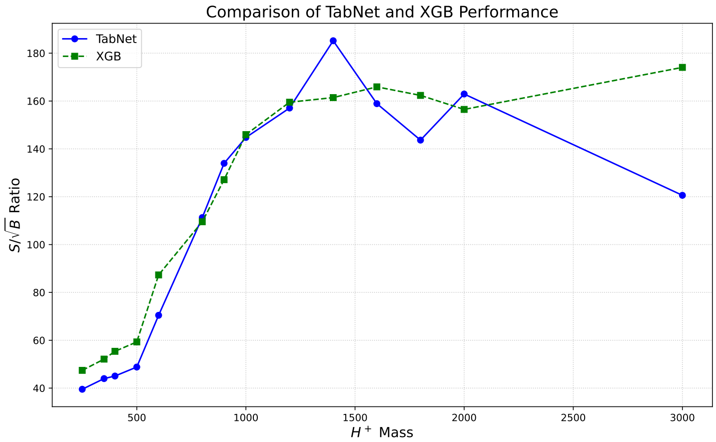


Figure 4.12: *$S/\sqrt{B}$ ratios at individually optimized thresholds for XGB and TabNet*

for the subsequent features, as opposed to the steady decline in case of TabNet. There are also some obvious glaring differences between feature usage in XGB and TabNet such as how the top 10 (except *sumPsbtag, taus_RNNJetScoreSigTrans_0*) features are fairly different for XGB-low and TabNet-low

Similar inferences is made for the high-scale models as well, however here, both models show a stark difference in their feature usage. TabNet-high gives relatively high importance to way more features than XGB-high, for which importance drops after the first 3 features (Fig. 4.15). Between the two models there are some disparities in feature-ranking. For example, TabNet-high rates `lep_pt_0` as the **3rd most important feature** which is **ranked 20th for XGB-high**.

Even between the same models for different scales there are similarities as well as glaring differences. For example **both XGB-low and XGB-high have the same top 2 important features** (`sumPsbtag, Ml1l01`) but in the case of `jet_phi_0` which is ranked 5 for XGB-high, it is **nowhere in the top 40 for XGB-low**. This disparity of feature usage is even more prominent between TabNet-low and TabNet-high, where the relative importance and order of important features are both strikingly different. This indicates that there is some **correlation between feature-importance and scale of the signal mass hypothesis**, as we see that for the same model, **different sets of features are given higher importance**, indicating that a scale-based model approach could be beneficial in better signal-background discrimination.

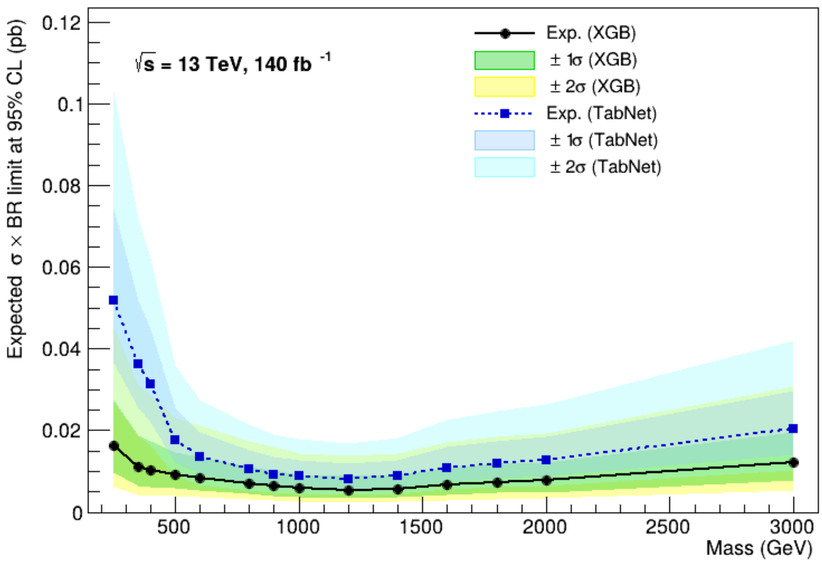


Figure 4.13: *Initial estimates (no fake norms applied) of upper limits by profile-likelihood fit on the respective signal probability variables of XGB and TabNet*

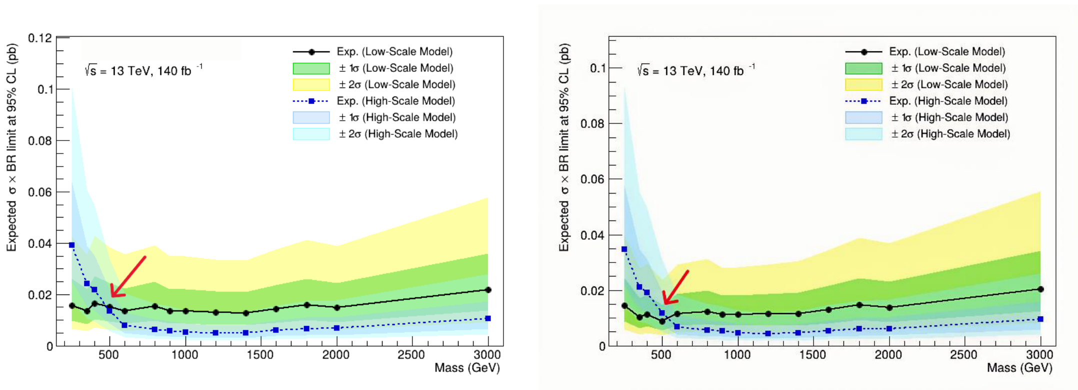


Figure 4.14: *Limit estimates by profile likelihood fit on the signal probability of XGB-high and XGB-low models on the the whole mass range – Left: Pre-Optuna; Right: Post-Optuna*

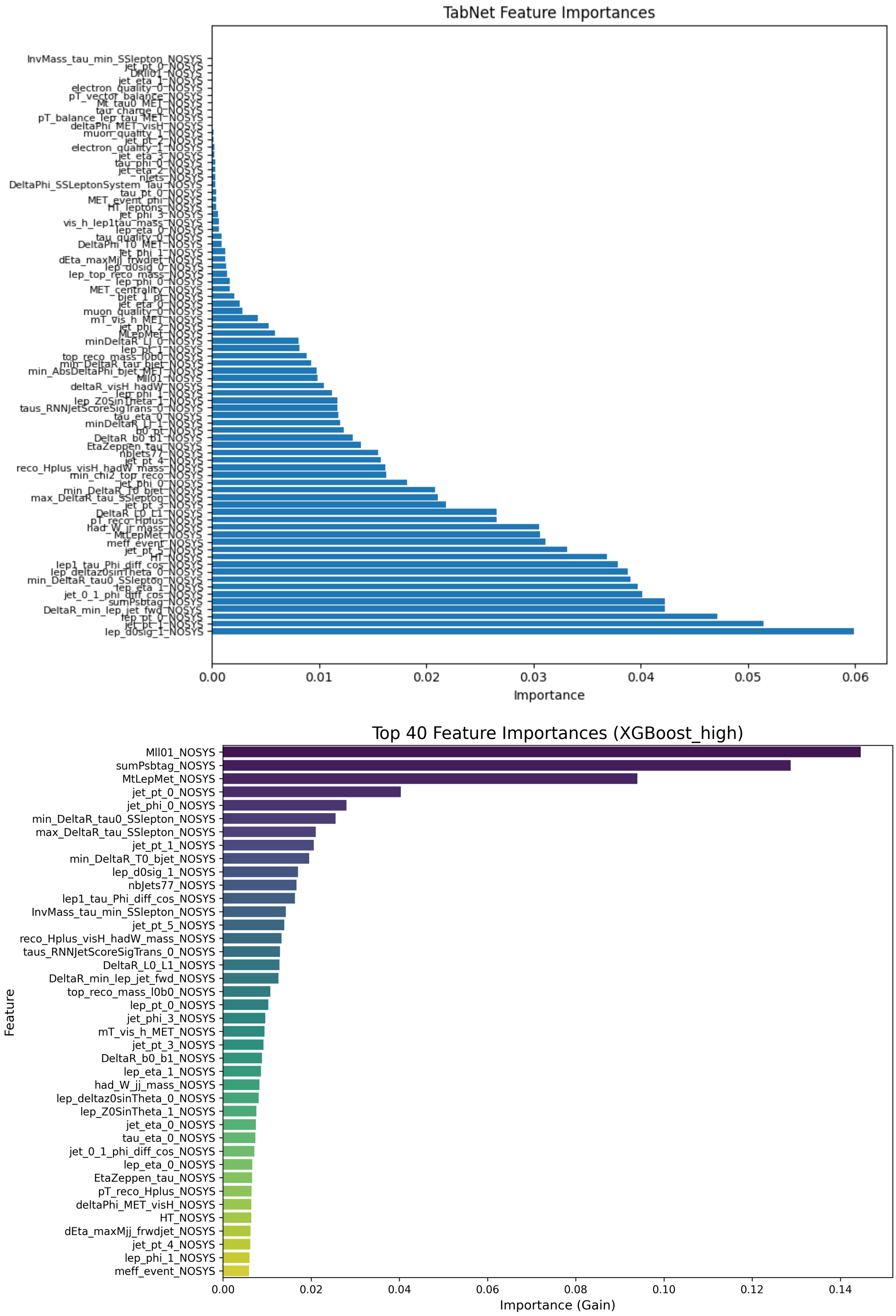


Figure 4.15: *Feature importance rankings for XGB and TabNet with high scale models*

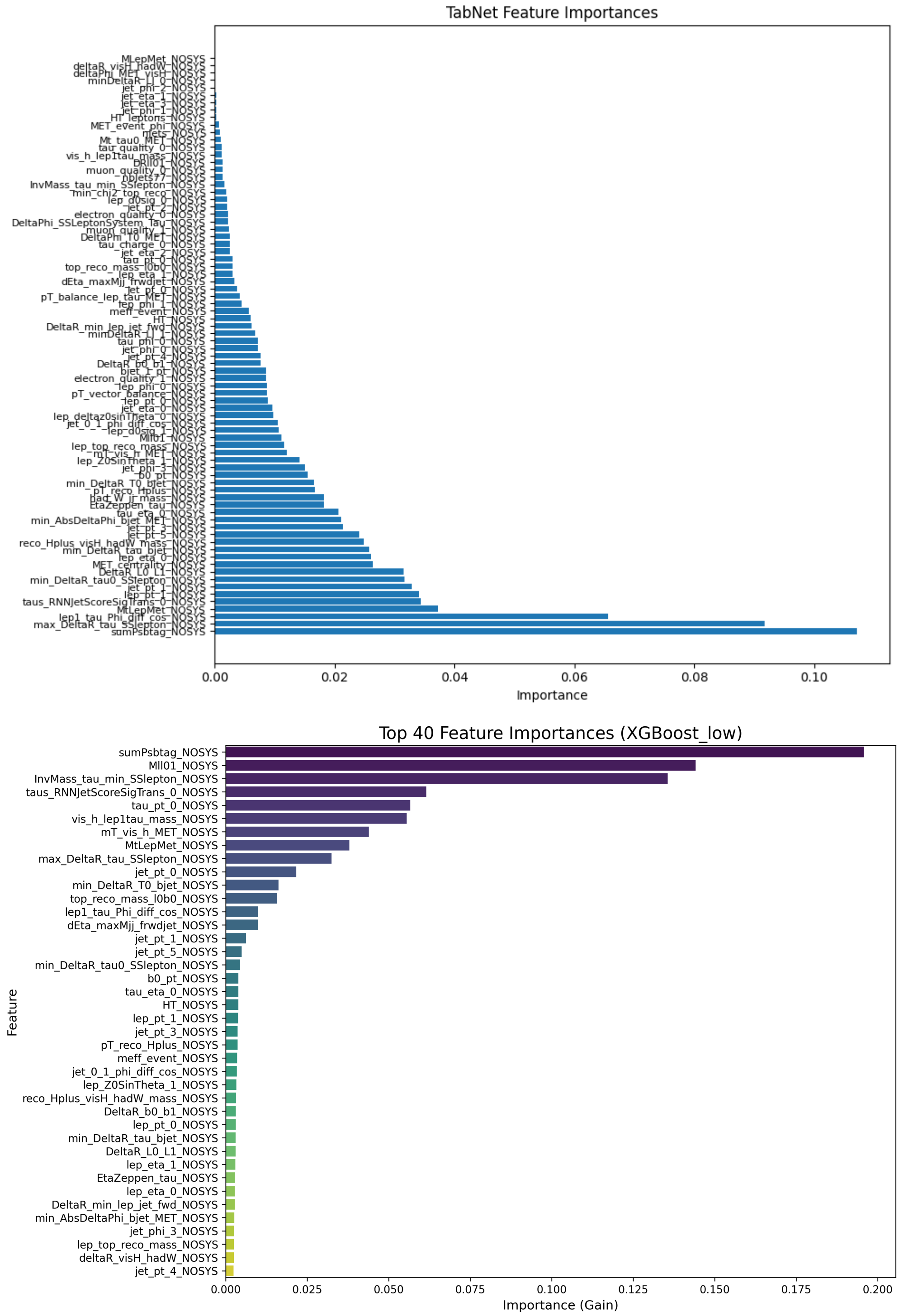


Figure 4.16: *Feature importance rankings for XGB and TabNet with low scale models*

Low-scale Model

Parameter	Value
n_estimators	6000
learning_rate	0.0109
max_depth	10
subsample	0.6656
colsample_bytree	0.8735
gamma	0.0070
reg_lambda	0.0095
alpha	0.2164
early_stopping_rounds	25

High-scale Model

Parameter	Value
n_estimators	6000
learning_rate	0.0157
max_depth	8
subsample	0.8577
colsample_bytree	0.6873
gamma	3.43e-05
reg_lambda	3.95e-07
alpha	0.0003
early_stopping_rounds	25

Table 4.5: *Optimized Hyper-parameters for XGB-low and XGB-high after Optuna study*

4.4.5 Effect of Number of Features on Performance

As part of this study, more than 70 custom features were defined in `FastFrames` and used to train and test ML/DL models. However, Figs. 4.15 and 4.16 show that many of these features are not nearly as important as the top few. So, to quantify the effect of feature importance on model performance, the optimized XGB model was trained and tested on the top 70, 40, 30 and 20 features. Their respective $S/\sqrt{B}$ at **F1 optimized thresholds for each hypothesized mass point** is shown in Fig.4.17.

Two main inferences can be made from this plot:

- There is a general trend of higher mass points achieving significantly higher $S/\sqrt{B}$ values which can also be seen in Fig. 4.12. This implies that signal-background discrimination for lower mass points is challenging, possibly because of similar kinematics of the signal with the background. This further solidifies the observations made from the ROC curves seen in Figs. 4.4, 4.7 and 4.8.
- There are some mass points that benefit from a reduction in features, such as $m_{H^+} = 600, 2000, 3000$ GeV. However, there is a general trend of loss of significance as the number of features decreases, where **the difference is not so drastic from 70 to 40 features**.

In future extensions of this study, we will include systematics and more features mean more systematics to be taken into account. Considering this fact as well as taking into account Fig. 4.17, we chose to go with top 40 features to have a balance of performance and efficiency.

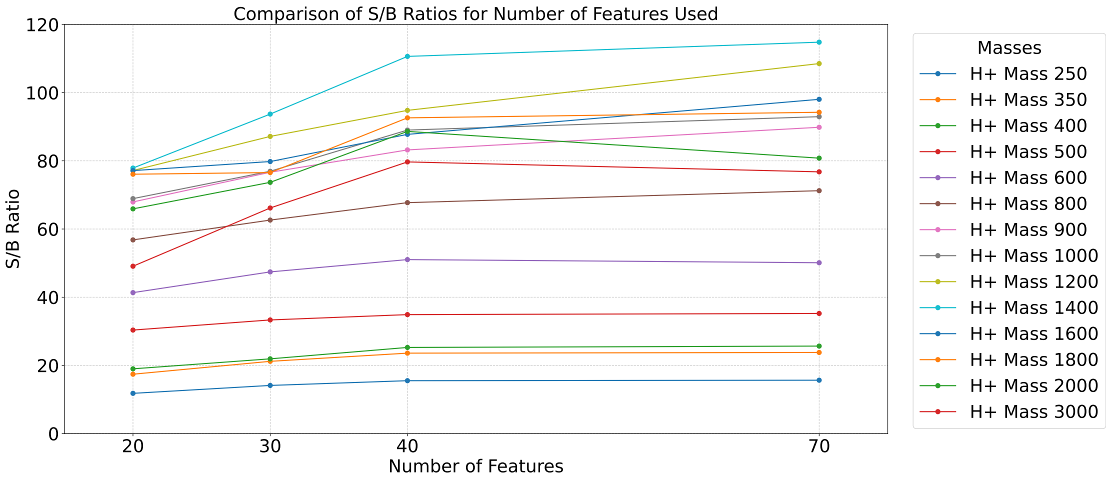


Figure 4.17: *Comparison of achieved $S/\sqrt{B}$ ratio for individual mass points with their respective F1 optimized thresholds for top 70, 40, 30 and 20 features*

4.5 Fake Estimation with Template Fits

Fake Estimation depends on high-quality "Templates" for fake leptons. In the $2l_{SS}1\tau$ final state, fake contributions can arise from either the two light leptons or the tau or any combinations of these three. There are two primary tasks to be done before template fits on real ATLAS data can be made – Sample re-classification and Control-Region definition.

4.5.1 Sample Re-classification

The signal and background samples are now characterized by the process they represent and they contain both prompt and non-prompt leptons. For developing templates, it is essential that we reclassify *all* the samples into prompt and fake samples. This is done with the help of **truth-level variables** from the simulation, in this case the ATLAS Isolation and Fake Forum (IFF) tools [31] was used. The prompt samples were named after the process which they represent (eg. ttH, $t\bar{t}$, etc) while the fake samples are classified as one of the following

1. **Fake** e: EXACTLY one fake electron. Norm : $\lambda(e_f)$
2. **Fake** μ: EXACTLY one fake muon. Norm : $\lambda(\mu_f)$
3. **Fake** $l + \tau$: EXACTLY one fake lepton and fake tau. Norm : $\lambda(multi - fake)$
4. **Double Fake**: Both light leptons are fake. Norm : $\lambda(multi - fake)$

5. **Fake τ**: EXACTLY one fake tauon. Norm : $\lambda(\tau_f)$

6. **Triple Fake**: All three leptons are fake. Norm : $\lambda(\tau_f)$

Note that each event can only be part of *only one* of these samples, they absolutely cannot overlap. In total, there are 4 Norm-Factors assigned to 6 fake samples. The grouping of fake samples for a single Norm-Factor (eg. Double Fake and Fake $l + \tau$ to $\lambda(multi - fake)$) has been done to account for low statistics and overlap of samples in their CRs (Fig. 4.20). The effect of **Re-classification** in the SR (for $m_{H^+} = 400$ GeV) is shown in Fig. 4.18 where the individual sample counts are different, but the total event count remains the same. The criteria used for sample re-classification is given in Table 4.6.

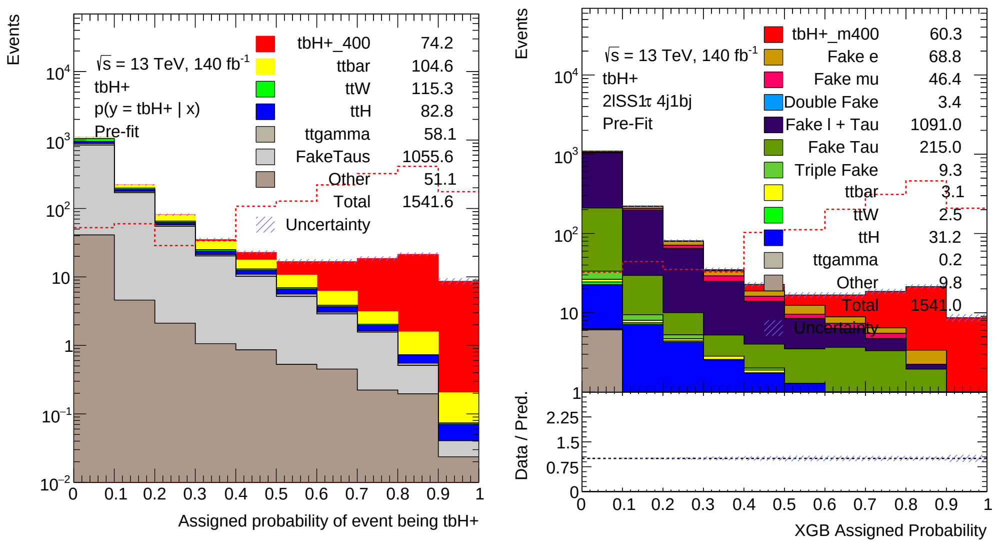


Figure 4.18: Signal-Region re-classification into prompt and non-prompt lepton classes for $m_{H^+} = 400$ GeV

The usage of IFFtools to determine fakes and prompts is given in Table 4.7

4.5.2 Control Regions

For the four assigned Norm-Factors, four dedicated CRs (orthogonal to SR) must be defined. Ideally these 4 regions must be populated purely with the samples associated with one Norm-Factor. **CRs must be defined with reconstructed variables** as the fitting will be performed on recorded ATLAS data. To control sample purity, 4 extra variables were added to all datasets to help with identifying lepton quality, namely `electron_quality_X`, `muon_quality_X`, `tau_quality_X` where X =

Fake Category	Selection Logic
Fake e	`(((lep_flav_0_NOSYS == 0 && XXX_OnlyLeadFake) || (lep_flav_1_NOSYS == 0 && XXX_OnlySubLFake))) && HadTau_truth_NOSYS`
Fake μ	`(((lep_flav_0_NOSYS == 1 && XXX_OnlyLeadFake) || (lep_flav_1_NOSYS == 1 && XXX_OnlySubLFake))) && HadTau_truth_NOSYS`
Double Fake	`!XXX_Prompt_0 && !XXX_Prompt_1 && HadTau_truth_NOSYS`
Fake $l + \tau$	`(XXX_OnlyLeadFake || XXX_OnlySubLFake) && !HadTau_truth_NOSYS`
Fake τ	`(!HadTau_truth_NOSYS && XXX_BothPrompt)`
Triple Fake	`!XXX_Prompt_0 && !XXX_Prompt_1 && !HadTau_truth_NOSYS`

Table 4.6: *Boolean flags used to define different categories of fake leptons and taus*

Flag Name	Boolean Logic Definition
XXX_Prompt_0	`(leps_IFFtype_0_NOSYS == 2 || leps_IFFtype_0_NOSYS == 4 || leps_IFFtype_0_NOSYS == 7)`
XXX_Prompt_1	`(leps_IFFtype_1_NOSYS == 2 || leps_IFFtype_1_NOSYS == 4 || leps_IFFtype_1_NOSYS == 7)`
XXX_BothPrompt	`(XXX_Prompt_0 && XXX_Prompt_1)`
XXX_OnlyLeadFake	`(!XXX_Prompt_0 && XXX_Prompt_1)`
XXX_OnlySubLFake	`(!XXX_Prompt_1 && XXX_Prompt_0)`

Table 4.7: *Prompt recognition using IFFtool-based variables*

0,1 indicating whether the lepton is leading or sub-leading. These quality parameters are based on a simple weighted scoring system that makes use of object definition flags shown in Tables 4.8 and 4.9. The scores are then normalized to 1. A lepton passing more flags will naturally have a higher quality score. This is a primitive but effective way to veto fake and prompt leptons using reconstructed variables for region definitions.
In addition, the signal probability variable from XGB was used to remove signal contamination in the CRs[5]. The exact cut on `sig_prob` was determined by the optimal threshold for each mass point to **exclude the region with maximum** $S/\sqrt{B}$.

The Pearson coefficient of correlation between the new set of features were evaluated, shown in Fig. 4.19. The feature correlation matrix has a few yellow spots (excluding the diagonal) indicating that some features have significant correlation with each other and could potentially be removed to improve model performance. A notable *anti-correlation* is seen between the leading as well as sub-leading muon and electron quality parameters (the dark blue spots for features 4 to 7). This is expected

[5]An example is given in Section 3.4

Electron ID	
Identifier	**Score**
`LooseBLayerLH`	`+= 1.0`
`LooseDNN`	`+= 2.0`
`MediumDNN`	`+= 4.0`
`TightLH_Loose`	`+= 2.0`
`TightDNN`	`+= 8.0`
`TightLH_Tight`	`+= 4.0`
`TightLH_TrackOnly`	`+= 8.0`

Muon ID	
Identifier	**Score**
`Loose_Loose`	`+= 1.0`
`Medium`	`+= 2.0`
`Loose_Tight`	`+= 4.0`
`Tight`	`+= 8.0`
`HighPt`	`+= 16.0`

Table 4.8: *Light lepton quality scoring based on object definition flags*

Tau ID	
Identifier	**Score**
`LooseRNN`	`+= 1.0`
`MediumRNN`	`+= 2.0`
`TightRNN`	`+= 4.0`
`LooseGNTau`	`+= 8.0`
`MediumGNTau`	`+= 16.0`
`TightGNTau`	`+= 32.0`

Table 4.9: *Tau quality scoring based on object definition flags*

because if the leading lepton is detected to be an electron, it is likely to have a very high `electron_quality_0` value while `muon_quality_0` would have a value of zero or close to zero.

Pre-Fit and Post-Fit Plots

The pre-fit and post-fit plots for $m_{H^+} = 400$ GeV are shown in Figs. 4.22, 4.23, 4.21 and 4.20. The norms for the fake leptons were estimated by a simultaneous CR+SR fit (SR blinded) separately for each individual mass points.

Statistics in the Fake μ region is quite limited and that is the case with all mass points, this is seen in the uncertainty margin for $\lambda(\mu_f)$ in Fig. 4.24. Ultimately, it is a trade-off between CR purity and statistics. The correlation factors between the norms are also shown in Fig. 4.24. The correlations between norm factors are minimal except with $\lambda(\mu_f)$ which shows some anti-correlation with $\lambda(multi - fake)$, but only around 40% for all mass points. The correlation matrices and estimated

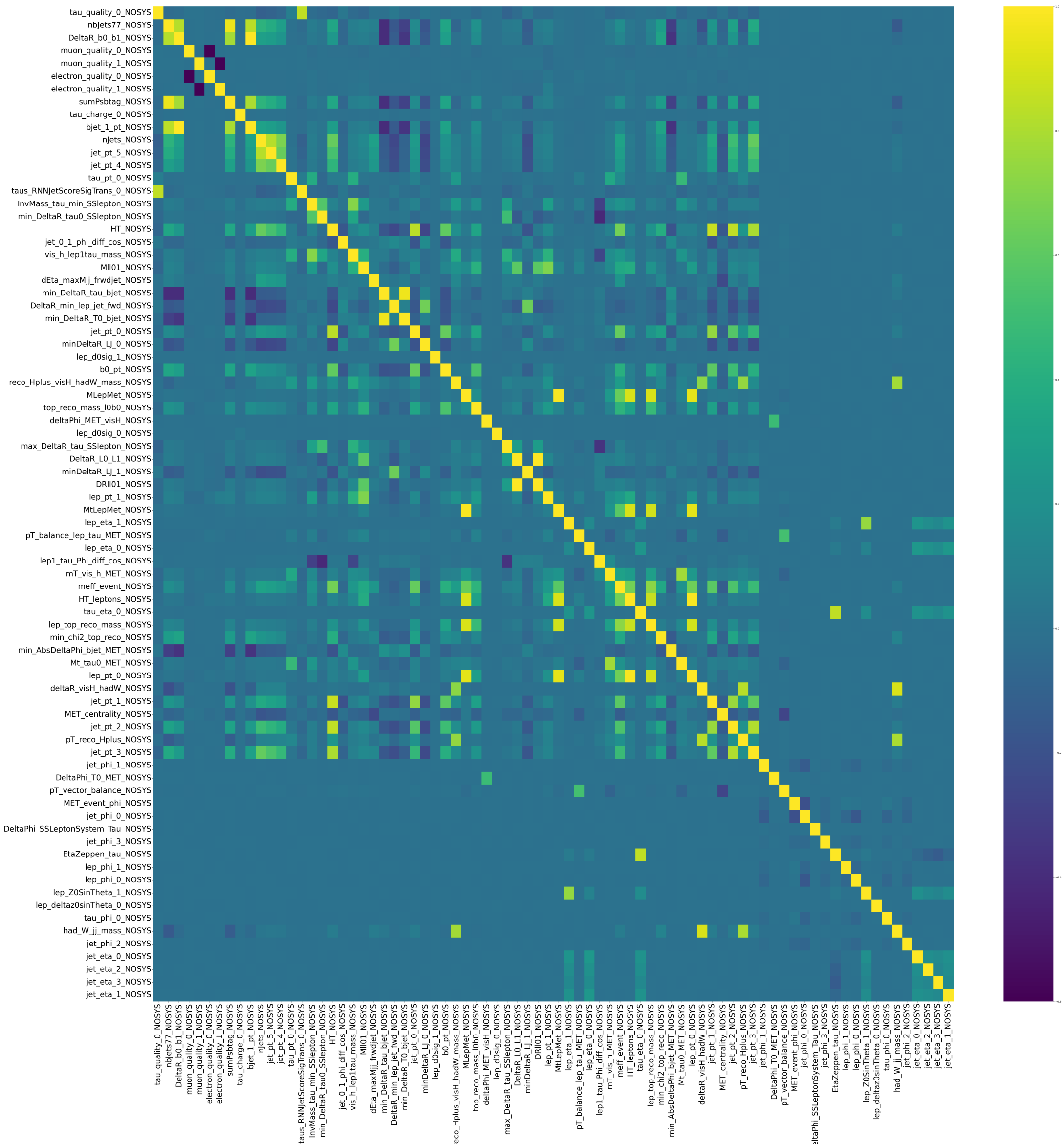


Figure 4.19: *Pearson correlation matrix after the inclusion of lepton quality parameters*

norms for each mass point are shown in the Appendix in Figs. A.6 and A.7
The fitted norm-factor estimates are very stable for $\lambda(multi - fake)$, $\lambda(e_f)$ and $\lambda(\tau_f)$ while $\lambda(\mu_f)$ varies over a range of values for each mass point. This could be improved in future analyses.

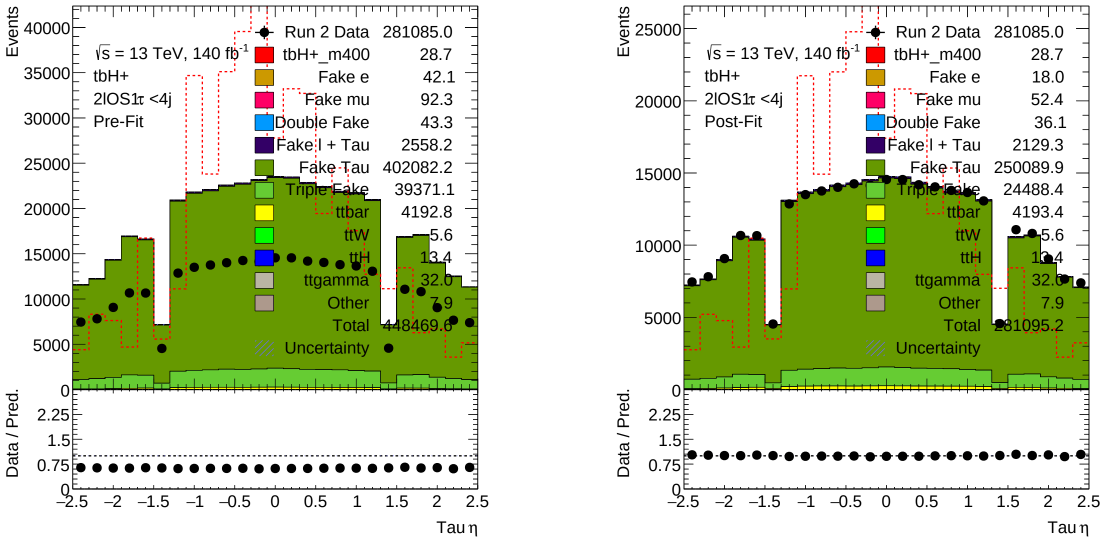


Figure 4.20: *Pre-fit and Post-fit η_τ distribution in the CR for Fake τ and Triple Fake; Norm-Factor $\lambda(\tau_f)$*

In general, all the norm-factors are less than one, indicating that the simulation over-estimates the amount of fakes, especially fake electrons, as it has the lowest estimated norm factor of around 0.4. Simulations for *multi-fakes* and to a certain extent fake taus are reasonably estimated by the simulation. As for fake μ, in most cases it is either overestimated by simulation or close to being correctly estimated ($\lambda(\mu_f) \approx 1$). However, the statistical uncertainty is high.

4.6 Sensitivity of Analysis: Expected Upper Limits

The final sensitivity of this analysis is determined by setting 95% confidence level (CL) upper limits on the signal cross-section. The limits are calculated using a profile-likelihood fit performed with `TRExFitter` on the final XGBoost signal probability distribution (`sig_prob`) in the signal region ($2\ell_{\text{SS}}1\tau_{\text{had}}4j1b$). Since the analysis is blinded, an Asimov dataset is used to calculate the median expected limit.

The resulting median expected limits, along with the $\pm1\sigma$ (green) and $\pm2\sigma$ (yellow) uncertainty bands, are presented in Fig. 4.25. This analysis is expected to

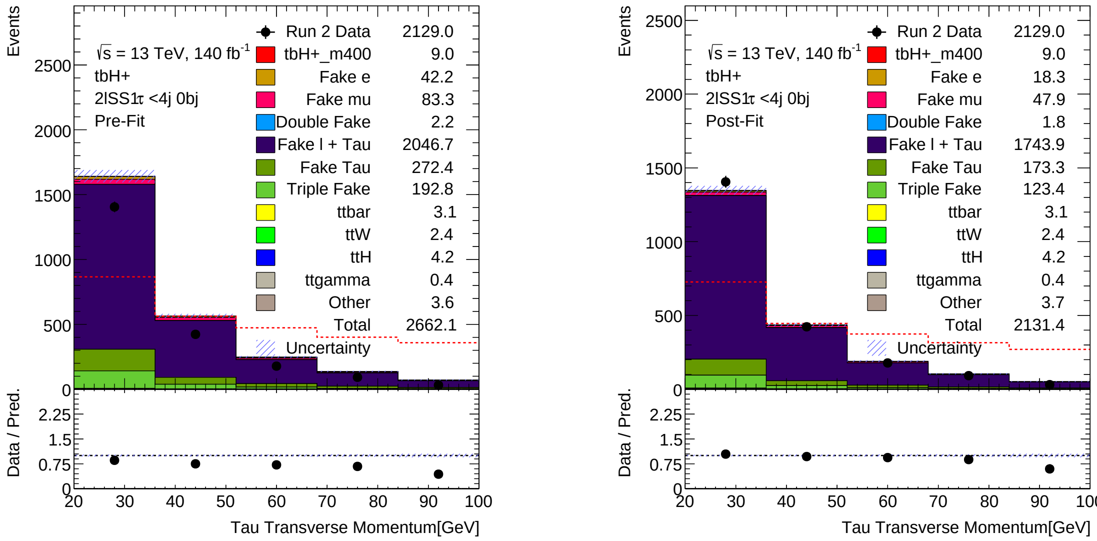


Figure 4.21: *Pre-fit and Post-fit $p_{T,\tau}$ distribution in the CR for Fake $l+\tau$ and Double Fake; Norm-Factor $\lambda(multi-fake)$*

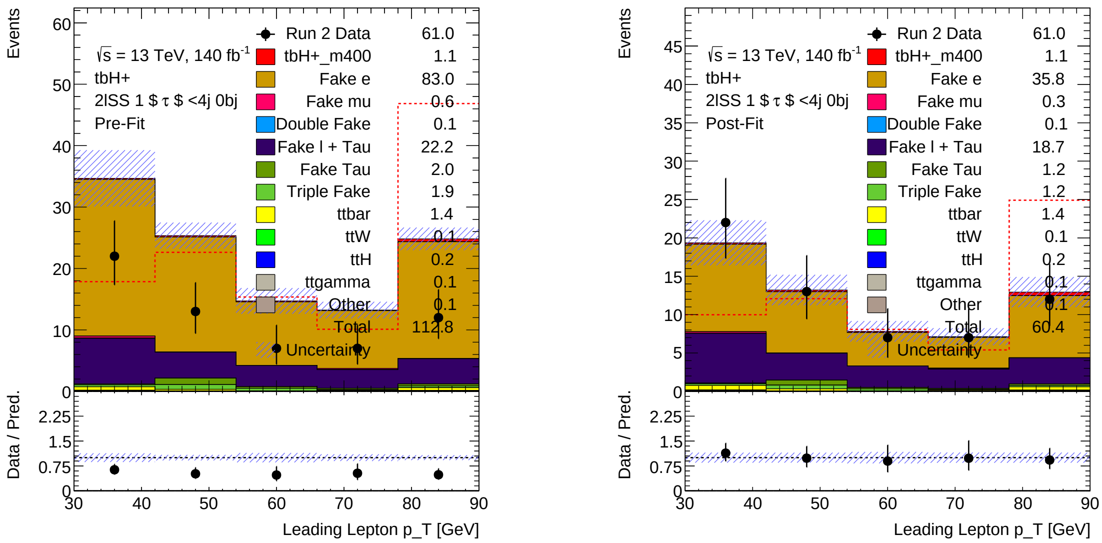


Figure 4.22: *Pre-fit and Post-fit lepton p_T distribution in the CR for Fake e ; Norm-Factor $\lambda(e_f)$*

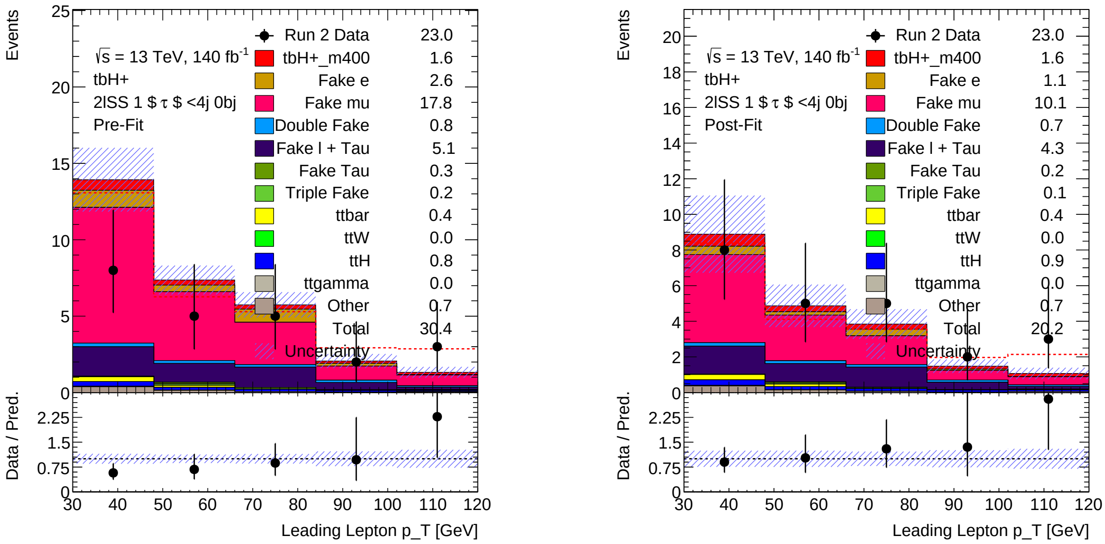


Figure 4.23: *Pre-fit and Post-fit lepton p_T distribution in the CR for Fake μ ; Norm-Factor $\lambda(\mu_f)$*

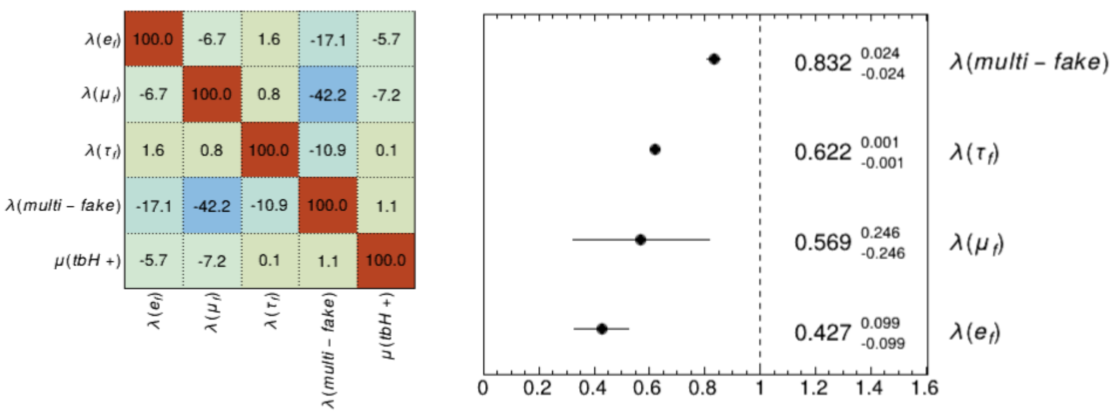


Figure 4.24: *The correlation matrix between norm-factors and the estimated norm-factor after template fit*

exclude the existence of the targeted charged Higgs boson with a mass below approximately 800 GeV, where the expected limit falls below the theoretical prediction. The analysis achieves its best sensitivity in the mass range of 800-1200 GeV.

This result represents a significant improvement over previous work. As shown in Fig. 4.26, the expected limits from this analysis are **more than twice as stringent** as those from the previous analysis using Release 21 software [32]. For example, at a mass of 1000 GeV, the limit improves from 0.02 pb to 0.009 pb. Furthermore, when compared to the published observed limits from CMS [33], this analysis demonstrates superior expected sensitivity across a significantly extended mass range up to 3000 GeV. **This substantial gain in sensitivity can be directly attributed to better statistics in Rel. 22 and the enhanced signal-background discrimination provided by the optimized, scale-based machine learning models developed in this thesis.**

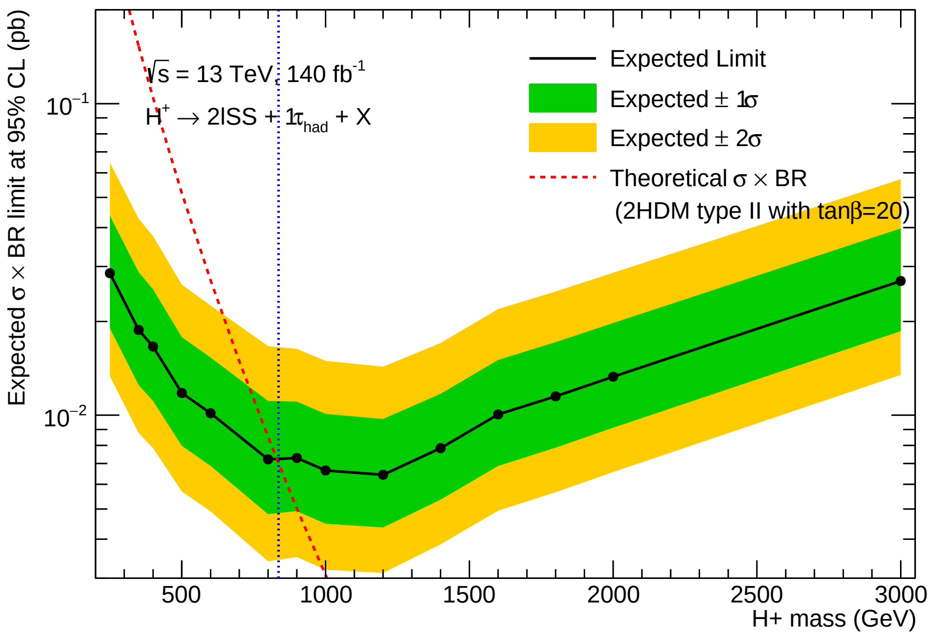
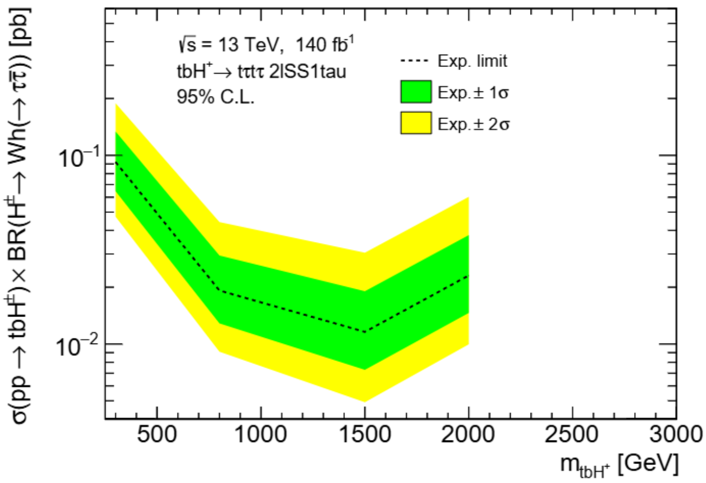
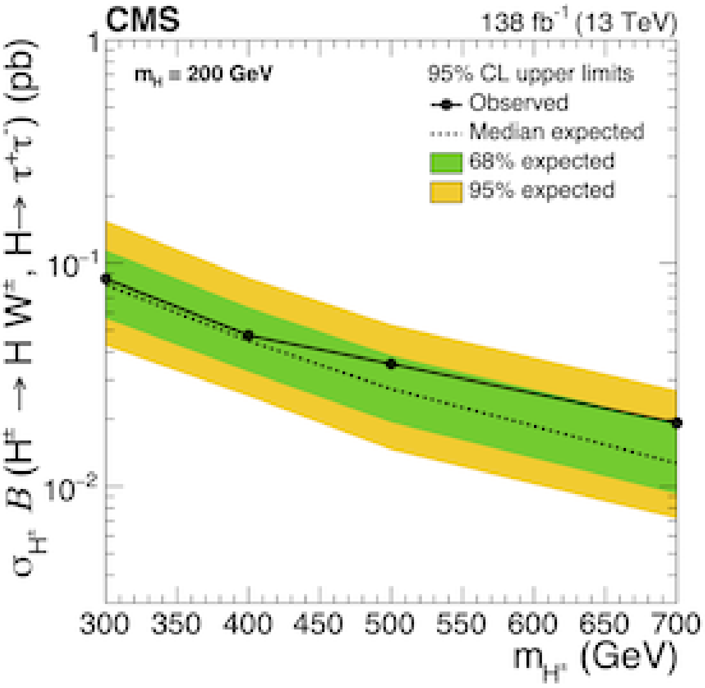


Figure 4.25: *Expected 95% CL upper limit on the cross-section for all mass hypotheses. The solid black line shows the median expected limit, with the green and yellow bands representing the* $\pm 1\sigma$ *and* $\pm 2\sigma$ *uncertainties. The red dotted line is the theoretical prediction for the signal.*



(a) *Previous ATLAS expected limit (R21) [32].*



(b) *Published CMS observed limit [33].*

Figure 4.26: *Comparison of the expected sensitivity from previous results.*

Chapter 5: Conclusion

In this thesis, a comprehensive, full-chain analysis was conducted to optimize the search for a charged Higgs boson ($H^{\pm}$) in the $2\ell_{\text{SS}}1\tau_{\text{had}}$ final state. This work successfully navigated the entire analysis pipeline, from initial Monte Carlo (MC) production to the final statistical interpretation of the results. The foundation of the analysis was the generation of eleven new $H^{\pm}$ signal samples in Release 22, significantly extending the mass range of previous studies. These were processed through a customized workflow using the Tau+X TopCPToolkit and FastFrames to produce analysis-level n-tuples, which were enhanced with over 70 new physics-based features designed to maximize signal-background discrimination.

The central achievement of this work was the development of a novel, scale-based machine learning strategy. Following an extensive comparative study, two optimized XGBoost models—one for low-mass and one for high-mass signal hypotheses—were chosen for their superior performance and stability. To ensure a robust estimate of instrumental backgrounds, a data-driven template fit was performed on dedicated control regions, with lepton quality parameters introduced to improve the purity of fake lepton samples. The final analysis, leveraging the enhanced discrimination power of these optimized models, yielded an expected 95% confidence level upper limit on the signal cross-section. This result represents a significant advancement, demonstrating a sensitivity that is more than twice as stringent as the previous ATLAS analysis and surpasses the published limits from CMS.

Future extensions of this work will focus on the incorporation of the full set of systematic uncertainties, which can be seamlessly integrated using the established FastFrames framework. Once systematics are included, the signal region can be un-blinded, with the ultimate goal of using the techniques developed in this thesis to obtain the best sensitivity on the charged Higgs boson to date.

Chapter A: Appendix

A.1 Complete Feature List

Table A.1: Feature importance ranking for the high scale XGB model, rounded to five decimal places. The features are presented in order of decreasing importance.

Feature	Importance	Feature	Importance
sumPsbtag_NOSYS	0.195 72	jet_eta_0_NOSYS	0.002 40
Ml1l01_NOSYS	0.144 19	MET_centrality_NOSYS	0.002 37
InvMass_tau_min_SSlepton_NOSYS	0.135 55	deltaPhi_MET_visH_NOSYS	0.002 36
taus_RNNJetScoreSigTrans_0_NOSYS	0.061 50	jet_phi_0_NOSYS	0.002 33
tau_pt_0_NOSYS	0.056 64	lep_phi_1_NOSYS	0.002 33
vis_h_lep1tau_mass_NOSYS	0.055 48	bjet_1_pt_NOSYS	0.002 32
mT_vis_h_MET_NOSYS	0.043 94	had_W_jj_mass_NOSYS	0.002 12
MtLepMet_NOSYS	0.037 89	lep_d0sig_1_NOSYS	0.002 11
max_DeltaR_tau_SSlepton_NOSYS	0.032 43	lep_d0sig_0_NOSYS	0.002 11
jet_pt_0_NOSYS	0.021 65	HT_leptons_NOSYS	0.002 10
min_DeltaR_T0_bjet_NOSYS	0.016 20	pT_balance_lep_tau_MET_NOSYS	0.002 07
top_reco_mass_l0b0_NOSYS	0.015 75	nbJets77_NOSYS	0.002 07
lep1_tau_Phi_diff_cos_NOSYS	0.009 86	lep_deltaz0sinTheta_0_NOSYS	0.002 05
dEta_maxMjj_frwdjet_NOSYS	0.009 84	nJets_NOSYS	0.002 02
jet_pt_1_NOSYS	0.006 22	MLepMet_NOSYS	0.001 99
jet_pt_5_NOSYS	0.004 89	minDeltaR_LJ_1_NOSYS	0.001 96
min_DeltaR_tau0_SSlepton_NOSYS	0.004 42	min_chi2_top_reco_NOSYS	0.001 92
b0_pt_NOSYS	0.003 91	pT_vector_balance_NOSYS	0.001 91
tau_eta_0_NOSYS	0.003 91	MET_event_phi_NOSYS	0.001 90
HT_NOSYS	0.003 89	jet_phi_1_NOSYS	0.001 89
lep_pt_1_NOSYS	0.003 82	tau_phi_0_NOSYS	0.001 89
jet_pt_3_NOSYS	0.003 68	jet_eta_1_NOSYS	0.001 89
pT_reco_Hplus_NOSYS	0.003 68	lep_phi_0_NOSYS	0.001 89
meff_event_NOSYS	0.003 49	Mt_tau0_MET_NOSYS	0.001 89
jet_0_1_phi_diff_cos_NOSYS	0.003 44	DeltaPhi_T0_MET_NOSYS	0.001 88
lep_Z0SinTheta_1_NOSYS	0.003 27	jet_phi_2_NOSYS	0.001 88
reco_Hplus_visH_hadW_mass_NOSYS	0.003 16	jet_eta_2_NOSYS	0.001 87
DeltaR_b0_b1_NOSYS	0.003 11	DeltaPhi_SSLeptonSystem_Tau_NOSYS	0.001 87
lep_pt_0_NOSYS	0.003 07	jet_pt_2_NOSYS	0.001 87
min_DeltaR_tau_bjet_NOSYS	0.003 05	tau_charge_0_NOSYS	0.001 84
DeltaR_L0_L1_NOSYS	0.003 03	jet_eta_3_NOSYS	0.001 84
lep_eta_1_NOSYS	0.002 94	minDeltaR_LJ_0_NOSYS	0.001 83
EtaZeppen_tau_NOSYS	0.002 86	DRl1l01_NOSYS	0.001 81
lep_eta_0_NOSYS	0.002 76	electron_quality_0_NOSYS	0.001 80
DeltaR_min_lep_jet_fwd_NOSYS	0.002 75	electron_quality_1_NOSYS	0.001 79
min_AbsDeltaPhi_bjet_MET_NOSYS	0.002 66	muon_quality_0_NOSYS	0.001 76
jet_phi_3_NOSYS	0.002 56	muon_quality_1_NOSYS	0.001 75
lep_top_reco_mass_NOSYS	0.002 51	tau_quality_0_NOSYS	0.001 74
deltaR_visH_hadW_NOSYS	0.002 46		
jet_pt_4_NOSYS	0.002 42		

A.2 Recall, Precision and F1 Curves for Individual Mass Points

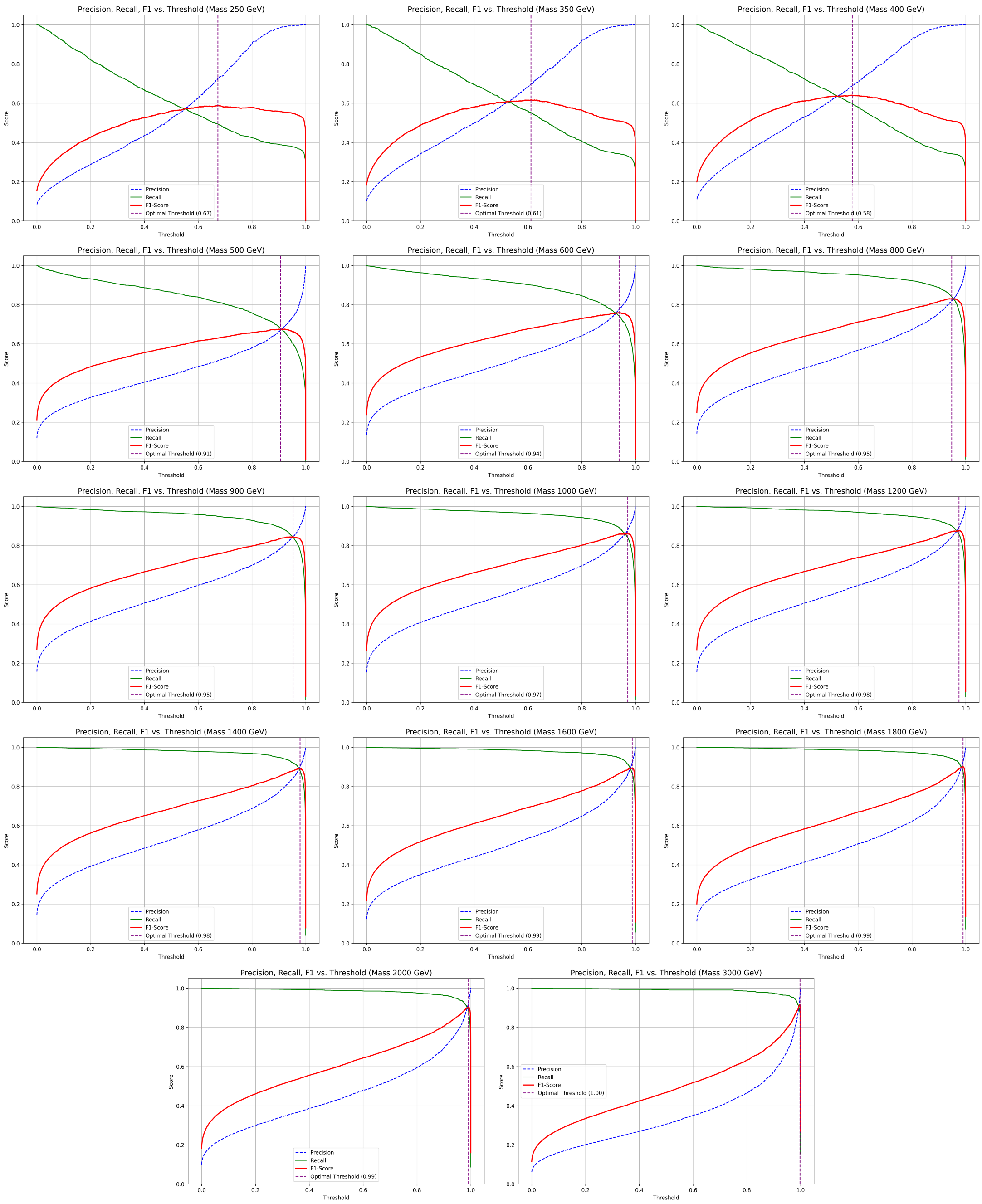


Figure A.1: *Precision, Recall and F1 curves for indvidual mass points with the XGBoost model*

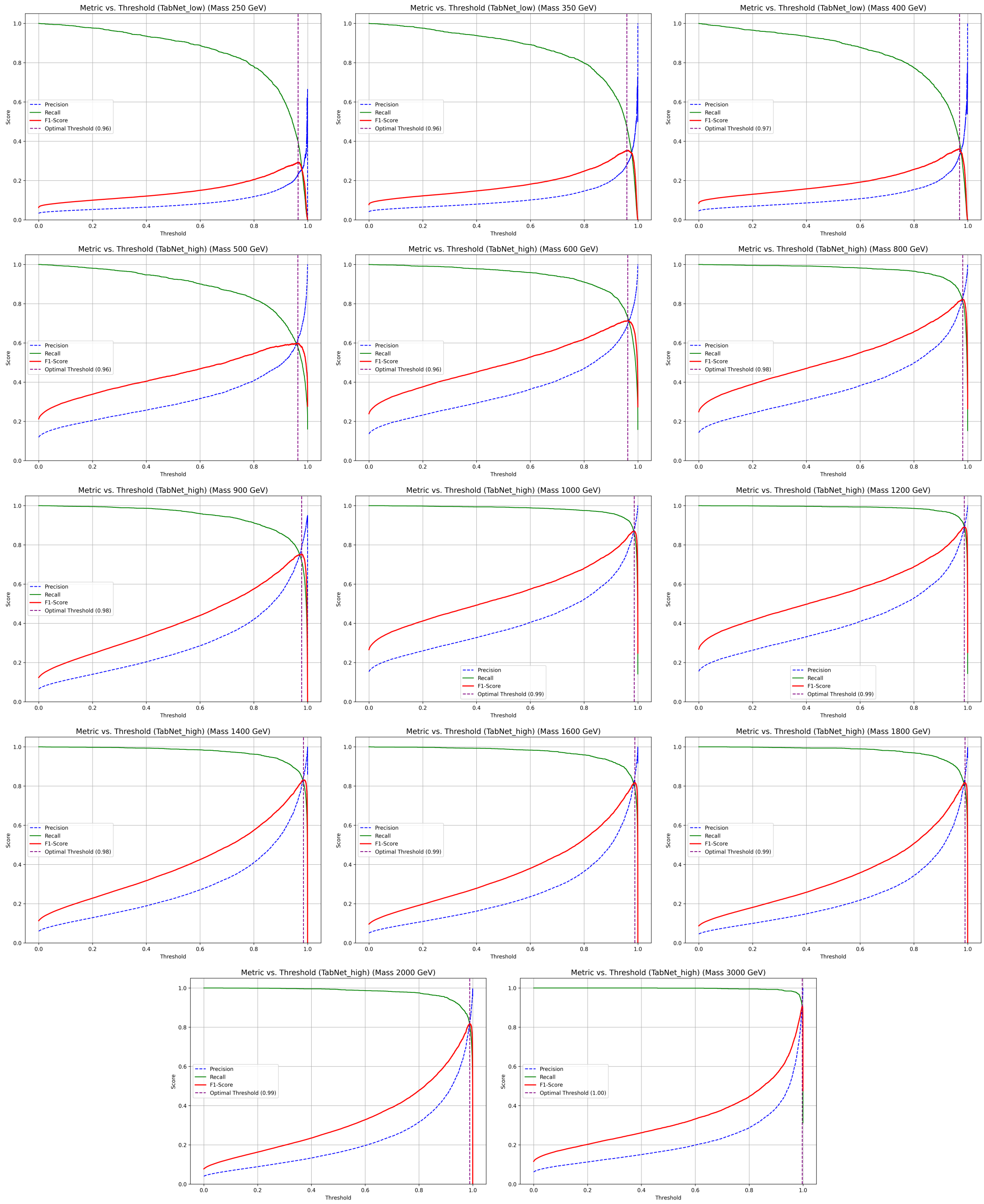


Figure A.2: *Precision, Recall and F1 curves for indvidual mass points with the TabNet model*

A.3 Signal Probability

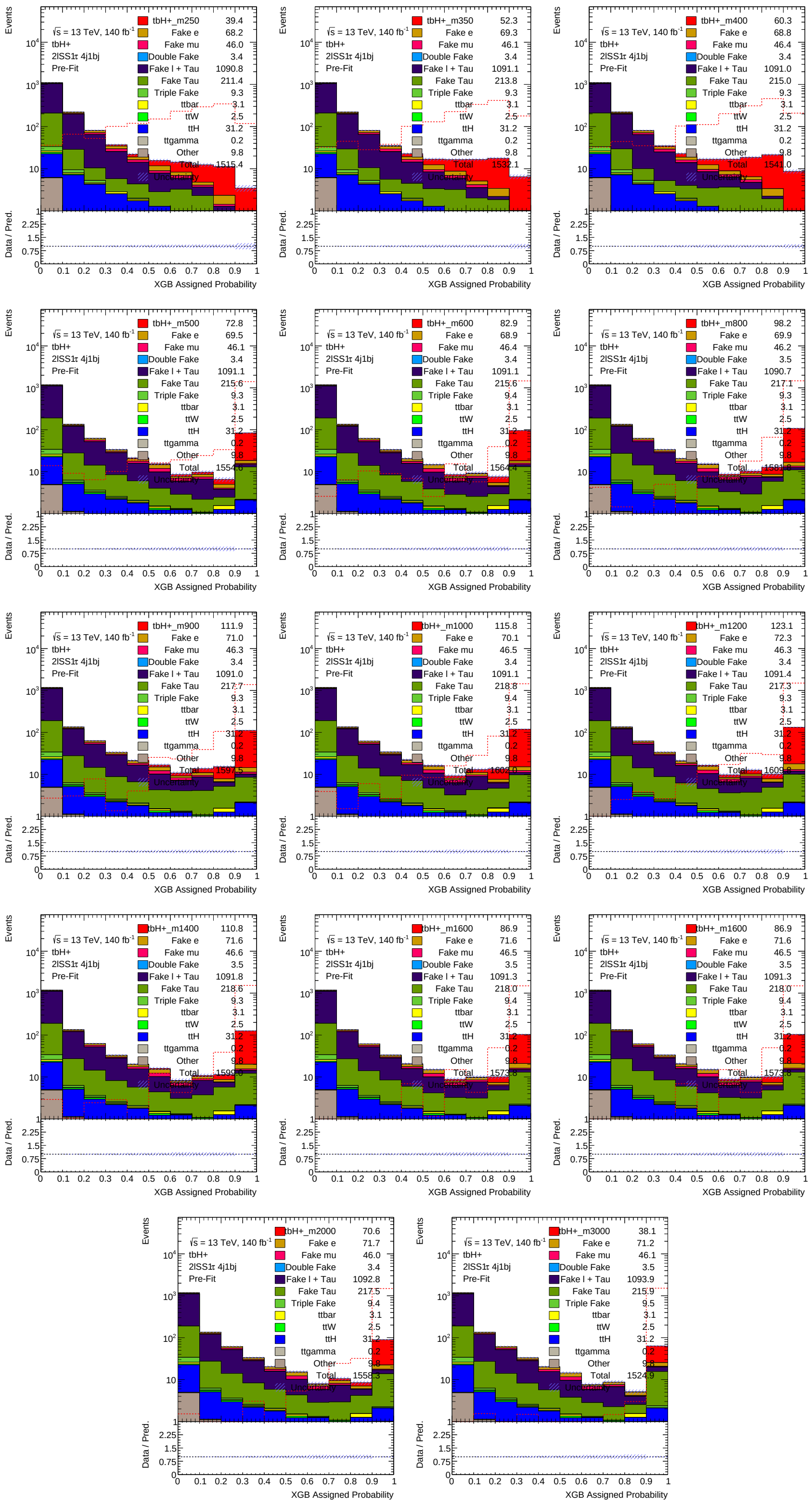


Figure A.3: Signal probability distributions for all mass hypotheses

A.4 AF3 v FullSim Comparison

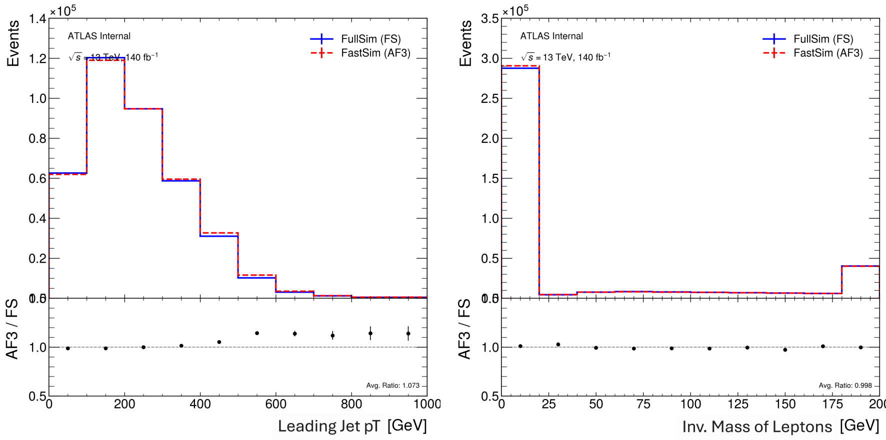


Figure A.4: *Bin-by-bin comparison of the leading jet p_T and the invariant mass of the leading and sub-leading leptons (M_{ll}^{01}) for $m_{H^+} = 800$GeV*

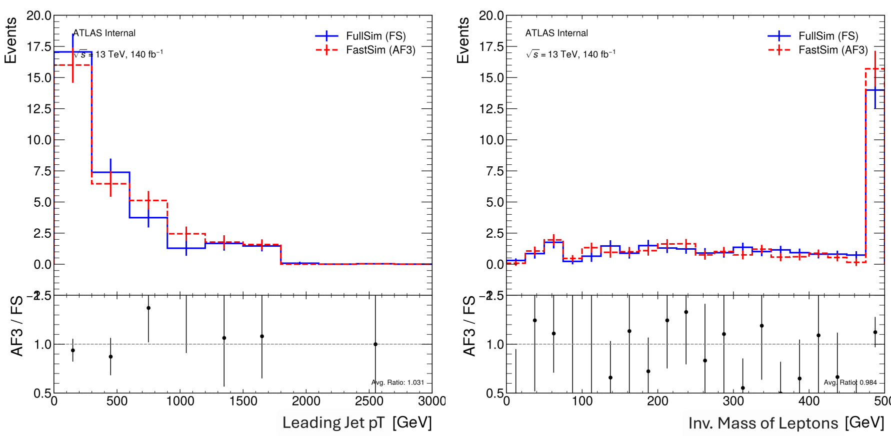


Figure A.5: *Bin-by-bin comparison of the leading jet p_T and the invariant mass of the leading and sub-leading leptons (M_{ll}^{01}) for $m_{H^+} = 3000$GeV*

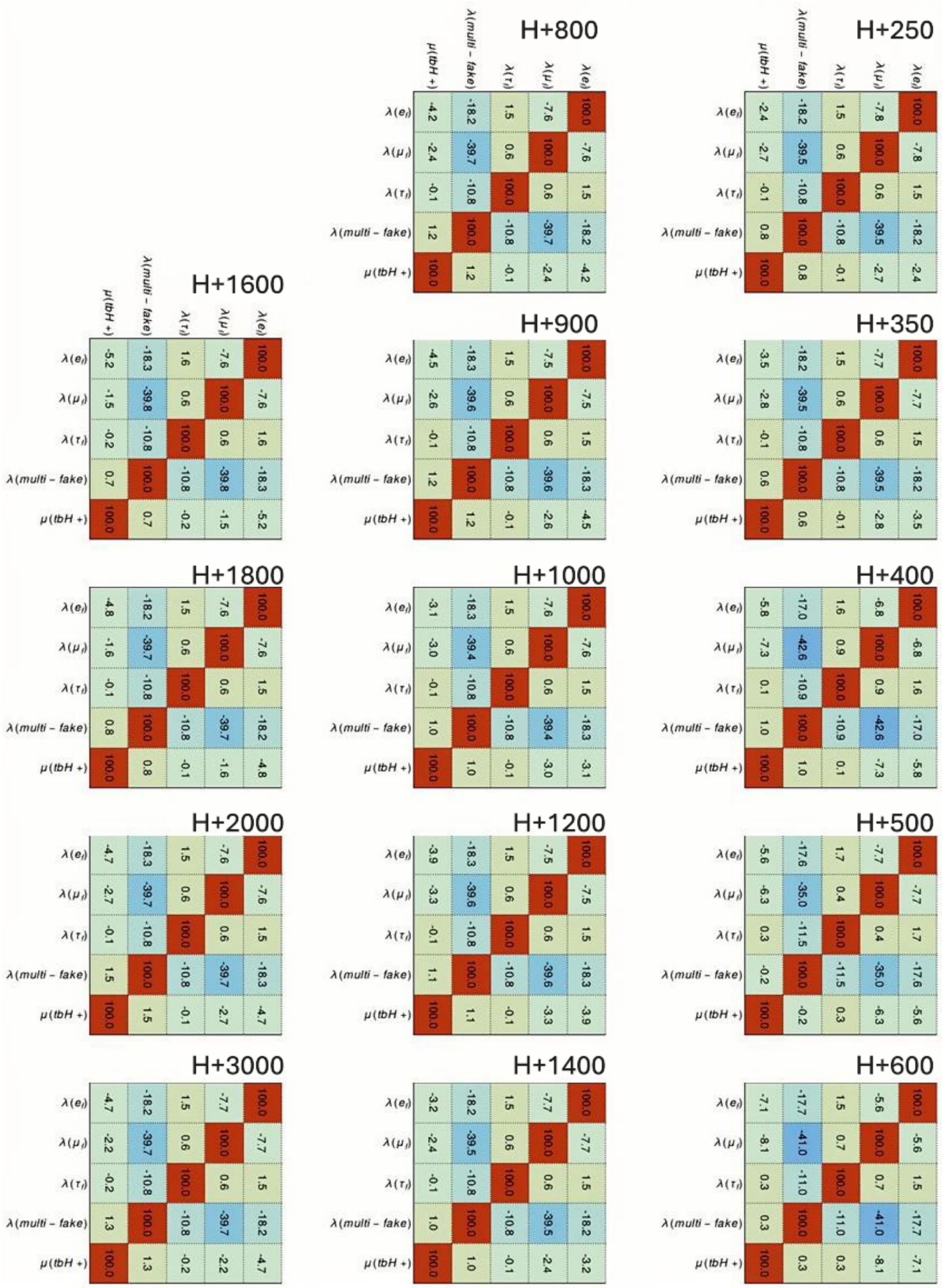


Figure A.6: *Correlation matrices between norm factors for fake analysis done for each mass point*

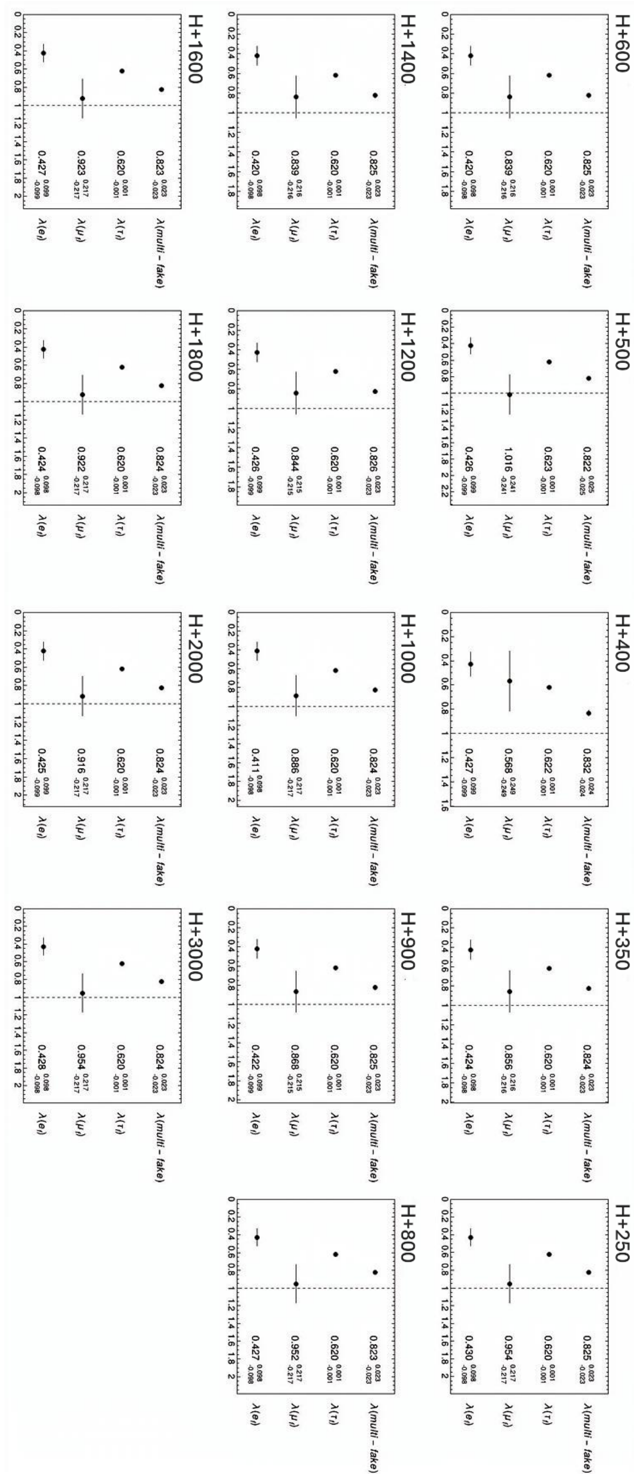


Figure A.7: Estimated fake norm factors for each mass point

Bibliography

[1] A. Sopczak. "Higgs physics at LEP-1." In: *Phys. Rept.* 359 (2002), pp. 169–282. DOI: 10.1016/S0370-1573(01)00062-X.

[2] *The Standard Model*. CERN. 2023. URL: https://home.cern/science/physics/standard-model (visited on 10/27/2023).

[3] MissMJ. *Standard Model of Elementary Particles*. https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. 2006.

[4] J. D. Lykken. "Beyond the Standard Model." In: *Proceedings, 1st International Summer School on Particle Physics (SCSS2009)*. Vol. 88. Can. J. Phys., 2010, pp. 573–581. DOI: 10.1139/P10-040. arXiv: 1005.1676 [hep-ph].

[5] M. Schmaltz and D. Shih. "The Standard Model and Beyond." In: (2022). DOI: 10.1142/9789811263300_0001. arXiv: 2206.05284 [hep-ph].

[6] H. Haber and G. Kane. "The search for supersymmetry: Probing physics beyond the standard model." In: *Physics Reports* 117.2 (1985), pp. 75–263. DOI: https://doi.org/10.1016/0370-1573(85)90051-1.

[7] A. Tranberg and B. Wu. "Cold Electroweak Baryogenesis in the Two Higgs-Doublet Model." In: *JHEP* 07 (2012), p. 087. DOI: 10.1007/JHEP07(2012)087. arXiv: 1203.5012 [hep-ph].

[8] S. Iguro and Y. Omura. "The direct CP violation in a general two Higgs doublet model." In: *JHEP* 08 (2019), p. 098. DOI: 10.1007/JHEP08(2019)098. arXiv: 1905.11778 [hep-ph].

[9] D. Espriu, F. Mescia, and A. Renau. "Axion-Higgs interplay in the two Higgs-doublet model." In: *Phys. Rev. D* 92.9 (2015), p. 095013. DOI: 10.1103/PhysRevD.92.095013. arXiv: 1503.02953 [hep-ph].

[10] M. Aaboud et al., ATLAS Collaboration. "Search for charged Higgs bosons decaying into top and bottom quarks at $\sqrt{s}$ = 13 TeV with the ATLAS detector." In: *JHEP* 11 (2018), p. 085. DOI: 10.1007/JHEP11(2018)085. arXiv: 1808.03599 [hep-ex].

[11] ATLAS Collaboration. "Search for charged Higgs bosons decaying via $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ in the τ+jets and τ+lepton final states with $36\,\mathrm{fb}^{-1}$ of pp collision data recorded at $\sqrt{s} = 13\,$TeV with the ATLAS experiment." In: *JHEP* 09 (2018), p. 139. DOI: 10.1007/JHEP09(2018)139. arXiv: 1807.07915 [hep-ex].

[12] CMS Collaboration. "Search for a charged Higgs boson decaying into top and bottom quarks in events with electrons or muons in proton–proton collisions at $\sqrt{s} = 13$ TeV." In: *JHEP* 01 (2020), p. 096. DOI: 10.1007/JHEP01(2020)096. arXiv: 1908.09206 [hep-ex].

[13] CMS Collaboration. "Search for charged Higgs bosons in the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ decay channel in proton–proton collisions at $\sqrt{s} = 13$ TeV." In: *JHEP* 07 (2019), p. 142. DOI: 10.1007/JHEP07(2019)142. arXiv: 1903.04560 [hep-ex].

[14] D. Eriksson, J. Rathsman, and O. Stal. "2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual." In: *Comput. Phys. Commun.* 181 (2010), pp. 189–205. DOI: 10.1016/j.cpc.2009.09.011. arXiv: 0902.0851 [hep-ph].

[15] C. Degrande et al. "Accurate predictions for charged Higgs production: Closing the $m_{H^{\pm}} \sim m_t$ window." In: *Phys. Lett. B* 772 (2017), pp. 87–92. DOI: 10.1016/j.physletb.2017.06.037. arXiv: 1607.05291 [hep-ph].

[16] *ATLAS Experiment*. CERN. 2024. URL: https://www.home.cern/science/experiments/atlas (visited on 05/21/2024).

[17] S. O. Arik and T. Pfister. "TabNet: Attentive Interpretable Tabular Learning." In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 8. May 2021, pp. 6679–6687.

[18] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System." In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[19] Simplilearn. *XGBoosting Helps Data Scientists Transforming Predictive Analytics*. https://python.plainenglish.io/xgboosting-helps-data-scientists-transforming-predictive-analytics-1b3395a4b503. Sept. 2022.

[20] A. Alkandari and A. Sopczak. *Determination of Fake Electrons, Muons and Taus in the tbH+ Search for the 2lSS1tau Channel Using Template Fit Method - Summer Student Project*. Tech. rep. Geneva: CERN, 2023.

[21] The ATLAS Collaboration. *Introduction to Athena — Athena master documentation*. https://atlas-software.docs.cern.ch/athena/basics/intro/. 2024.

[22] The ATLAS TopCPToolkit Developers. *TopCPToolkit documentation*. https://topcptoolkit.docs.cern.ch/latest/. 2024.

[23] Tau + X Sub-group. *TopCPToolkit N-Tuple Content Documentation*. https://gitlab.cern.ch/groups/atlas-phys/exot/lpx/exot-2022-34/-/wikis/TopCPToolkit-ntuple-content. 2024.

[24] The FastFrames Developers. *FastFrames documentation*. https://atlas-project-topreconstruction.web.cern.ch/fastframesdocumentation/latest/. 2024.

[25] The TRExFitter Developers. *TRExFitter documentation*. https://trexfitter-docs.web.cern.ch/trexfitter-docs/latest/. 2024.

[26] A. L. Read. *Presentation of search results: The CL(s) technique*. Tech. rep. CERN, 2002, pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[27] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics." In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[28] A. Afandizada. *Search for a charged Higgs boson decaying into a neutral Higgs boson and a W-boson Using Machine Learning with ATLAS Data*. Tech. rep. CERN, 2025. DOI: 10.17181/endr0-7m408.

[29] J. Alwall et al. "A standard format for Les Houches event files." In: *Comput. Phys. Commun.* 176 (2007), pp. 300–304. DOI: 10.1016/j.cpc.2006.11.010. arXiv: hep-ph/0609017.

[30] A. Hasib. "ATLAS Fast Simulation - from classical to deep learning." In: (2022).

[31] The ATLAS Collaboration. *TruthClassification Tool Documentation (Athena Release 21.2)*. https://gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/AnalysisCommon/TruthClassification. 2024.

[32] M. Rames. "Search for tb$H^+(\tau\tau)$ with Performance Optimisation for Signal and Background Separation Using Machine Learning with ATLAS Data." MA thesis. CTU, 2023.

[33] CMS Collaboration. "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." In: *Journal of High Energy Physics* 2023.9 (). DOI: 10.1007/jhep09(2023)032.