GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

**Bachelor's Thesis**

# Untersuchung von Tau Lepton Identifikation mithilfe von Transformer basierenden graphischen Neural Netzwerken für den ATLAS Trigger

# Investigation of Tau Lepton Identification using transformer based graphical neural networks for the ATLAS trigger

prepared by

**Fabio Potrykus**

from Wolfenbüttel

at the II. Physikalischen Institut

| | |
|---|---|
| **Thesis number:** | II.Physik-UniGö-BSc-2026/01 |
| **Thesis period:** | 10th November 2025 until 23rd February 2026 |
| **First referee:** | Prof. Dr. Stan Lai |
| **Second referee:** | Priv.-Dozent Dr. Ralf Bernhard |

# Abstract

This thesis investigates the performance of transformer-based graph neural networks for identifying hadronically decaying tau leptons in the context of the ATLAS trigger system. Efficient tau lepton identification at the trigger level is essential for many physics analyses, particularly under increasing data rates and detector complexity. The primary objective of this work is to evaluate whether an inclusive training approach outperforms prong-exclusive models, in which separate networks are trained for different tau-decay prong multiplicities. The results show that the inclusive model consistently outperforms prong-exclusive models across all evaluated metrics. Additionally, increasing the number of input constituents did not yield a significant improvement in classification performance. These findings suggest that inclusive transformer-based graph neural networks offer a more robust and efficient approach to hadronic tau lepton identification at the trigger level within the new data acquisition framework.

**Keywords:** ATLAS detector, Trigger system, Tau lepton, Transformer, Graphical neural network

# Contents

# 1 Introduction

From the earliest attempts to explain natural phenomena to today's high-energy experiments, humanity has continuously strived to uncover the fundamental laws of nature. Particle physics represents the most advanced stage of this pursuit, aiming to describe matter and its interactions at the smallest accessible scales, as well as the forces that govern them. The discovery of the Higgs boson [1] in 2012 at the Large Hadron Collider [2] (**LHC**) confirmed a key pillar of the Standard Model [3–5], yet many questions remain unanswered, such as the nature of dark matter [6], matter-antimatter asymmetry [7], and physics beyond the Standard Model. Experiments at the **LHC**, like ATLAS [8], probe these questions by colliding protons at unprecedented energies, producing a wealth of data to search for rare signals amid overwhelming backgrounds.

Information is crucial, but there is a limit to how much data can be stored and processed. At the same time, maximizing the use of available data is essential. To address this, the ATLAS Trigger system was developed [8]. The high collision rate of the LHC produces an unprecedented volume of detector data. Since only a small fraction of this data can be recorded, the experiment relies on a highly selective data-acquisition system to identify and retain potentially interesting events. This selection is essential for studies of both Standard Model processes and phenomena beyond the Standard Model, including rare signatures such as Higgs boson production and other processes that occur at extremely low rates.

Among the physics objects of interest at the LHC, hadronically decaying tau leptons are of particular importance, as they appear in a wide range of Standard Model measurements and are prominent signatures in many searches for new physics. An example of one such important signature is the production of Higgs bosons which decay into two tau leptons. However, their identification presents a significant challenge for event selection. Tau leptons have a short lifetime and decay before reaching the detector, producing narrow jets composed of a few charged particles and neutral hadrons. These signatures closely resemble those of ordinary quark- or gluon-initiated (QCD) jets, which are produced at

much higher rates. As a result, efficient tau identification requires sophisticated reconstruction and discrimination techniques to achieve high signal efficiency while maintaining strong rejection of the overwhelming jet background.

In this thesis, the performance of the GNTau architectures developed for identifying hadronically decaying tau leptons is investigated. The goal is to distinguish genuine tau signatures from fake tau leptons. Different graphical neural networks are systematically compared to assess their impact on classification performance, efficiency, and background rejection in the context of the ATLAS trigger.

The remainder of this thesis is structured as follows. Chapter 2 provides a theoretical background of the Standard Model of particle physics, describing its fundamental components and its role as the prevailing theoretical framework in high-energy physics. Chapter 3 details the Experimental Setup, explaining the ATLAS Experiment and Trigger System. Chapter 4 focuses on Machine Learning, presenting the logic of neural networks up to the Transformer, which are primarily used in this thesis. Chapters 5 and 6 present my results for different transformer-based architectures, comparing them with one another and with the results from the ATLAS Tau Trigger group.

# 2 Theoretical Background

One of the greatest achievements of modern physics is the development of a unified framework that explains nearly all observed phenomena involving particles.

## 2.1 The Standard Model of Particle Physics

The **Standard Model (SM) of particle physics** [3–5], derived from experimental observations, predicts, correlates, and unifies these phenomena within a coherent quantum field theory. It unifies the **fundamental particles** and their **interactions** (excluding gravity) in a self-consistent theoretical framework, whose predictions have been verified with extraordinary precision down to distance scales of about $10^{-18}$ m [9].

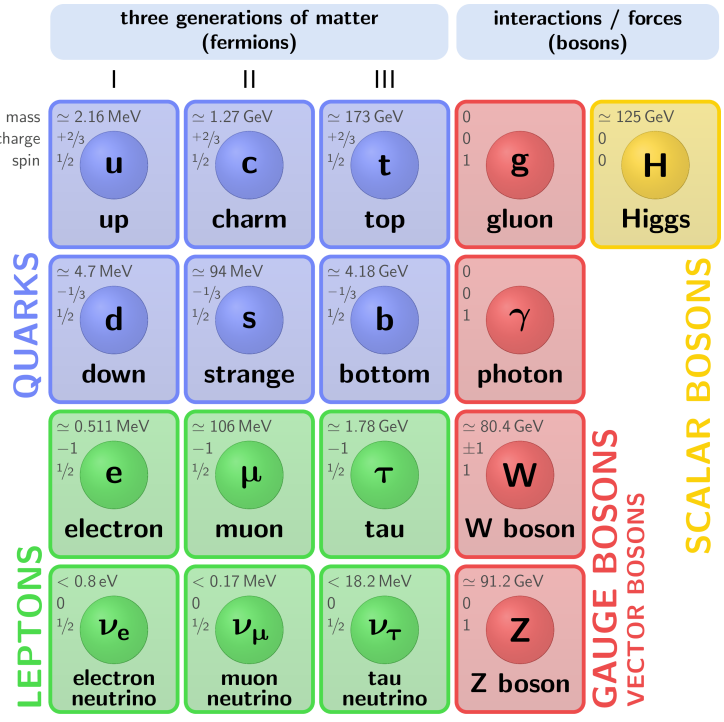



Figure 2.1: Elementary particles of the Standard Model grouped by interaction type and generation [10].

## 2.2 Structure of the Standard Model

The SM of particle physics is a relativistic quantum field theory that provides a unified description of three of the four known fundamental interactions: the electromagnetic, weak, and strong forces. Gravitation is not included in the Standard Model and is instead described by general relativity [11], since gravity is a classical theory of spacetime curvature and cannot be consistently quantized within the framework of perturbative quantum field theory. Within this framework, all known elementary particles are classified according to their intrinsic spin into **fermions**, which constitute matter, and **bosons**, which mediate interactions.

The fermions of the SM consist of three generations, each containing two quarks and two leptons. Quarks participate in the strong interaction and appear in six distinct flavors: up and down (first generation), charm and strange (second generation), and top and bottom (third generation) as shown in Figure 2.1. Quarks are spin-$\frac{1}{2}$ particles and therefore obey Fermi-Dirac statistics [12, 13]. Each quark carries one of three color charges: red, green, or blue [14], and has a fractional electric charge [15]. Specifically, the up, charm, and top quarks each have an electric charge of $+\frac{2}{3}$, whereas the down, strange, and bottom quarks each carry an electric charge of $-\frac{1}{3}$. Due to color confinement [16], quarks do not exist as free particles but instead combine to form color-neutral hadrons, such as baryons and mesons. A notable example is the proton, composed of two up quarks and one down quark ($uud$), carrying a total electric charge of $+1$ [15].

Leptons, in contrast, do not carry color charge and therefore do not participate in the strong interaction. Each generation contains a charged lepton (electron, muon, or tau, with increasing mass) and an associated neutrino, illustrated in Figure 2.1, which is electrically neutral and interacts only via the weak interaction [3].

## 2.3 Bosons and Fundamental Interactions

Interactions in the Standard Model arise from the requirement of **local gauge invariance** under the symmetry group

$$SU(3)_C \times SU(2)_L \times U(1)_Y,$$

[3, 14, 17]. Imposing local gauge symmetry necessitates the introduction of gauge fields, whose quantized excitations correspond to the **gauge bosons** that mediate the funda-

mental forces between fermions.

The **strong interaction**, described by **Quantum Chromodynamics (QCD)**, is governed by the non-Abelian gauge group $SU(3)_C$ [14]. Its gauge bosons are the eight massless **gluons**, which carry color charge and mediate interactions between quarks. A defining feature of QCD is color confinement, which prevents the isolation of individual quarks or gluons, and asymptotic freedom, in which the interaction strength decreases at high energies.

The **weak interaction** [5] is mediated by the massive gauge bosons $W^{\pm}$ **and** $Z^0$, associated with the $SU(2)_L$ component of the electroweak symmetry. Unlike the electromagnetic and strong interactions, the weak interaction violates parity and charge conjugation symmetries and is responsible for processes such as beta decay and neutrino interactions.

The **electromagnetic interaction** is described by a $U(1)_{\text{EM}}$ gauge symmetry and is mediated by the **photon** ($\gamma$), a massless spin-1 boson, shown in Figure 2.1. In the Standard Model, electromagnetism emerges from the electroweak theory through the spontaneous symmetry breaking of the $SU(2)_L \times U(1)_Y$ symmetry down to $U(1)_{\text{EM}}$ [5]. The photon couples to electrically charged particles and is responsible for phenomena such as atomic structure, electromagnetic radiation, and chemical bonding.

## 2.4 Higgs Field and Mass Generation

Within the electroweak sector of the Standard Model, gauge invariance under

$$SU(2)_L \times U(1)_Y$$

forbids explicit mass terms for gauge bosons and fermions. However, experimental observations demonstrate that the weak gauge bosons $W^{\pm}$ and $Z^0$ are massive, while the photon remains massless. This apparent contradiction is resolved by the **Higgs mechanism** [18].

The Higgs mechanism introduces a complex scalar field that acquires a nonzero vacuum expectation value $v \approx 246$ GeV, thereby spontaneously breaking the electroweak symmetry. As a result, three of the four electroweak gauge bosons acquire mass through their interaction with the Higgs field, while the remaining unbroken $U(1)_{\text{EM}}$ symmetry ensures the masslessness of the photon. This excitation of the Higgs Field is quantized as

the scalar Higgs boson. Fermion masses arise through Yukawa couplings [19] to the Higgs field, with the observed mass hierarchy encoded in the corresponding coupling strengths.

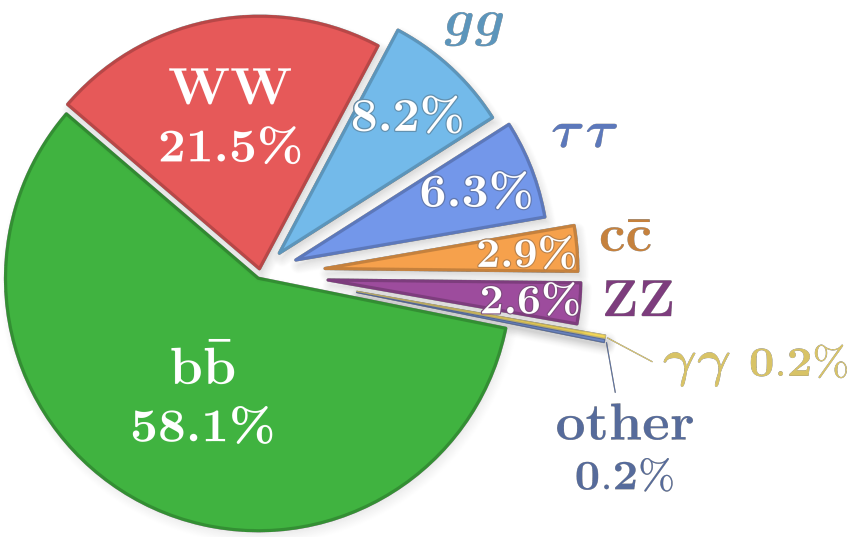



Figure 2.2: The branching ratio of the Higgs boson. Adapted from graphic by © Izaak Neutelings, CC BY 2.5

Figure 2.2 illustrates the branching ratios of the Higgs boson. While most of these decay channels are outside the scope of this thesis, the decay into a pair of tau leptons is of particular importance, as it provides direct access to the Higgs coupling to charged leptons. Although the branching ratio of this decay is only about 6.3%, the $H \to \tau^+\tau^-$ channel is experimentally significant. The presence of neutrinos in tau decays leads to characteristic event topologies and missing transverse momentum, making reliable tau identification essential. This is discussed in Section 3.4.

# 3 Experimental Background

Experimental tests of the Standard Model and searches for physics beyond it require particle collisions at very high energies. According to the principles of relativistic quantum field theory [20], higher collision energies allow access to shorter distance scales and enable the production of heavy particles that cannot be observed at lower energies. In addition, precise measurements of rare processes require large data samples, which can only be obtained at high collision rates. These requirements motivate the construction of high-energy particle accelerators capable of delivering both extreme energies and luminosities.

## 3.1 The Large Hadron Collider

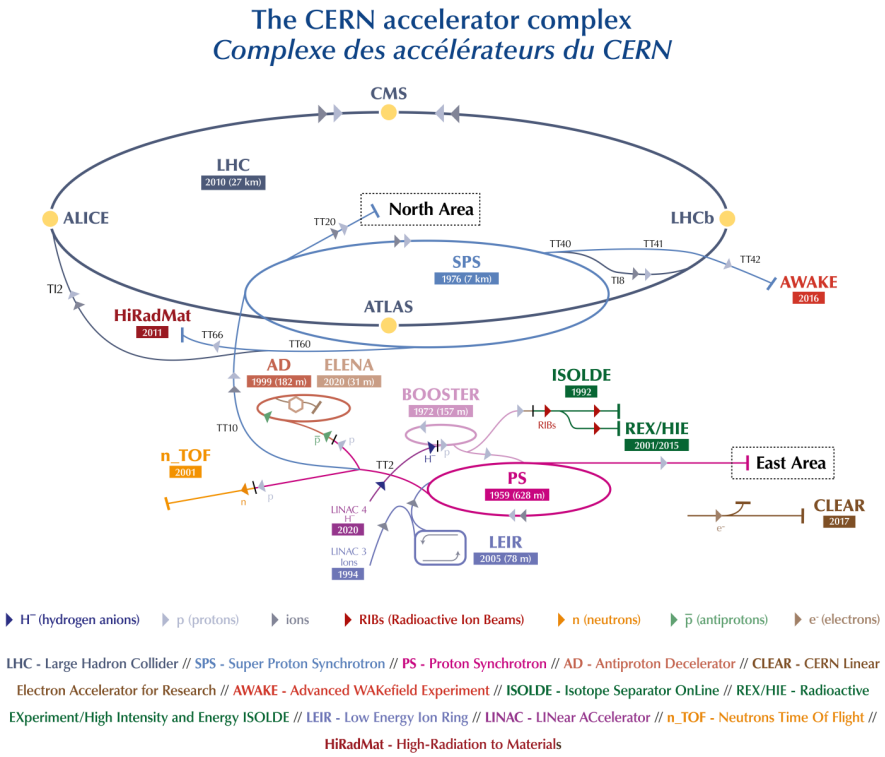



Figure 3.1: The CERN accelerator complex in 2019 [21].

The Large Hadron Collider (LHC) is the world's most powerful particle accelerator. Located at the European Organization for Nuclear Research (CERN) near Geneva, Switzerland, it constitutes the final stage of the CERN accelerator complex, as illustrated in Figure 3.1. The LHC is installed in a circular underground tunnel with a circumference of approximately 27 km and is designed to accelerate and collide beams of protons and heavy ions at unprecedented energies.

During proton-proton operation, two counter-rotating beams are accelerated at the **LHC**. Inside the collider, superconducting dipole magnets guide the beams around the ring. The protons reach energies of up to 6,8 TeV per beam, giving a maximum center-of-mass energy of 13,6 TeV [22]. Collisions occur at four dedicated interaction points, each hosting one of the major experiments: ATLAS [8], CMS [23], ALICE [24], and LHCb [25]. While ATLAS and CMS are general-purpose detectors designed to study a wide range of physics phenomena, ALICE and LHCb are specialized experiments, optimized for heavy-ion physics and precision studies of b-quark decays, respectively. These high-energy collisions reproduce conditions similar to those shortly after the Big Bang, allowing physicists to study rare processes and produce heavy particles.

The LHC operates at a nominal bunch-crossing rate of 40 MHz and up to $1.8 \cdot 10^{11}$ protons per bunch in Run 3. Each crossing can contain multiple proton-proton interactions, known as pile-up. These extreme data rates and radiation levels place stringent demands on the detector technologies, trigger systems, and data-acquisition architectures used by the experiments.

## 3.2 The ATLAS Experiment

The ATLAS experiment (A TOROIDAL LHC APPARATUS) [8] is a large, general-purpose particle detector situated at one of the four main interaction points of the Large Hadron Collider. Measuring approximately 46 m in length and 25 m in diameter, ATLAS is among the largest and most complex scientific instruments ever built. Its primary goal is the precise reconstruction of a wide range of physics signatures, including electrons, photons, muons, jets, missing transverse momentum, and hadronically decaying tau leptons, which underpins a broad physics program spanning precision measurements and searches for new phenomena.

ATLAS employs a right-handed Cartesian coordinate system with the origin at the in-

teraction point. The z-axis aligns with the beamline, the x-y plane lies transverse to it (positive x toward the LHC ring center, positive y upward). The detector is designed with a nearly cylindrical geometry around the beamline as shown in Figure 3.2, allowing for measurements in a coordinate system defined by the azimuthal angle $\phi$ around the beam axis and the pseudorapidity $\eta$, which describes the particle's angle relative to the beam direction. Pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$ where $\theta$ is the polar angle measured from the beam axis. This coordinate is particularly useful in high-energy collisions because differences in $\eta$ are approximately invariant under Lorentz boosts along the beam direction. These coordinates provide a convenient way to describe particle trajectories and detector coverage and will be used extensively in subsequent discussions of ATLAS measurements.

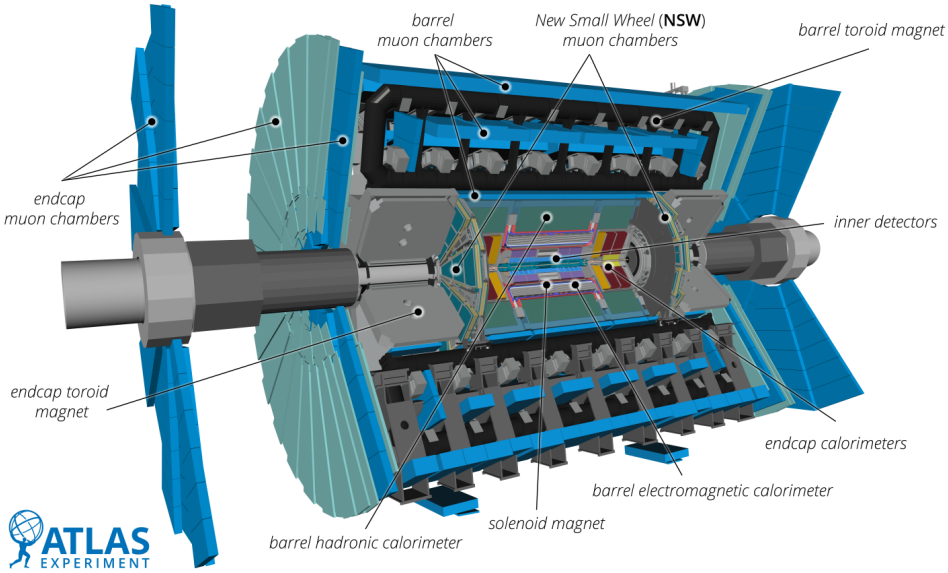



Figure 3.2: Schematic view of the ATLAS detector and its main subsystems [26].

As shown in Figure 3.2, the ATLAS detector is composed of several concentric subsystems arranged around the interaction point. Each subsystem is optimized for a specific task in particle detection, and together they provide nearly full solid-angle coverage.

### 3.2.1 Inner Detector

The innermost component of ATLAS is the Inner Detector (ID), which precisely tracks charged particles produced in proton-proton collisions. It operates within a 2 T axial magnetic field generated by a superconducting solenoid, allowing particle momenta to be measured from the curvature of their trajectories.

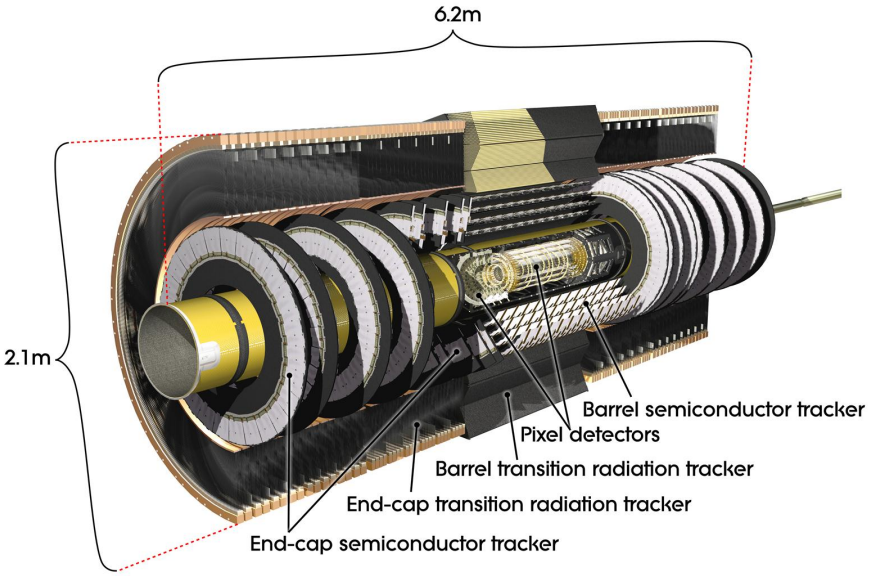



Figure 3.3: Cut-away view of the ATLAS inner detector [27].

As shown in Figure 3.3, the ID consists of three subsystems. Closest to the beamline is the **Pixel Detector**, composed of multiple layers of silicon sensors with typical pixel sizes of $50 \times 400$ $\mu$m$^2$, covering the pseudorapidity range $|\eta| < 2.5$. The Pixel Detector provides excellent spatial resolution, with intrinsic accuracies of about 10 $\mu$m in $R$-$\phi$ and 115 $\mu$m along the $z$ axis, enabling precise reconstruction of primary and secondary vertices. In Run 2, it was upgraded by adding the Insertable B-Layer (IBL) to cope with higher luminosity, which increased pile-up and radiation levels, and to improve tracking precision and vertexing performance.

Surrounding it is the **Silicon Microstrip Tracker (SCT)**, consisting of eight strip layers of silicon microstrip sensors with a strip pitch of approximately 80 $\mu$m, also covering $|\eta| < 2.5$. The SCT improves the momentum resolution and contributes to precise vertexing when combined with the Pixel Detector.

At larger radii is the **Transition Radiation Tracker (TRT)**, a straw-tube detector composed of $\sim 4$ mm-diameter straws arranged in barrel and end-cap modules covering $|\eta| < 2.0$. The TRT provides only $R - \phi$ tracking, with an intrinsic resolution of about 130 $\mu$m per straw. Beyond continuous tracking, it also identifies electrons by detecting transition-radiation photons produced when charged particles traverse the xenon-based gas mixtures of the straw tubes, thereby enabling efficient electron-hadron separation. The combination of all three subsystems results in robust pattern recognition and high

precision in both $R - \phi$ and z-coordinates.

The inner Detector design achieves a momentum resolution of

$$\frac{\sigma(p_T)}{p_T} = \frac{0.05\%}{p_T[GeV]} \oplus 1\%.$$

### 3.2.2 Calorimeter System

Surrounding the Inner Detector, the ATLAS calorimeter system measures the energies of electrons, photons, and hadrons by absorbing their energy in dense materials. As shown in Figure 3.4, the calorimeters provide nearly hermetic coverage up to $|\eta| < 4.9$, which is essential for jet reconstruction and the measurement of missing transverse momentum.

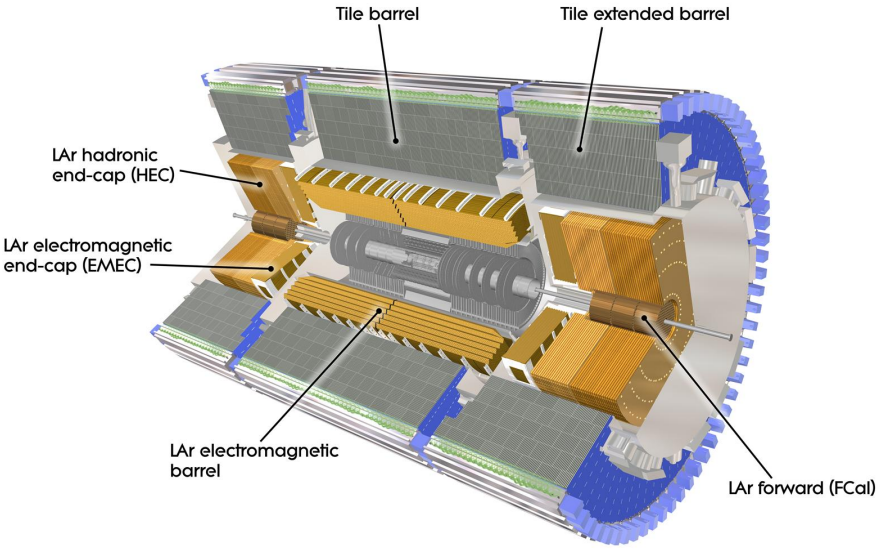



Figure 3.4: Cut-away view of the ATLAS calorimeter system [27].

**Electromagnetic Calorimeter** The electromagnetic (EM) calorimeter is a lead–liquid argon sampling calorimeter with an accordion geometry, providing full azimuthal coverage and fine granularity. It is designed to measure electrons and photons with pseudorapidity $|\eta| < 3.2$. The calorimeter has a thickness of about 22 radiation lengths ($X_0$) in the barrel and up to $24\,X_0$ in the endcap regions, ensuring efficient containment of electromagnetic showers.

The typical energy resolution for electrons and photons is

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E/\text{GeV}}} \oplus 0.7\%,$$

allowing accurate reconstruction of electromagnetic objects and playing a key role in precision measurements and Higgs boson analyses.

**Hadronic Calorimeter** The hadronic calorimeter surrounds the EM calorimeter and measures the energies of hadrons and jets. In the central region ($|\eta| < 1.7$), this is achieved with the Tile Calorimeter, a steel–scintillator sampling detector with a thickness of approximately 7 interaction lengths ($\lambda_I$). In the endcap and forward regions, liquid argon calorimeters with copper or tungsten absorbers are used to withstand the high radiation environment, extending the coverage to $|\eta| < 4.9$.

The combined hadronic calorimeter system provides a typical jet energy resolution of

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E/\text{GeV}}} \oplus 3\%,$$

depending on pseudorapidity and jet properties. Together with the EM calorimeter, the system ensures efficient energy measurement and containment of particle showers across the full detector acceptance.

### 3.2.3 Muon Spectrometer

The outermost subsystem of ATLAS is the Muon Spectrometer (MS), which is designed to identify muons and measure their momentum with high precision. As illustrated in Figure 3.5, it surrounds the calorimeter system and operates in a 0.5-1T magnetic field generated by large air-core toroidal magnets.

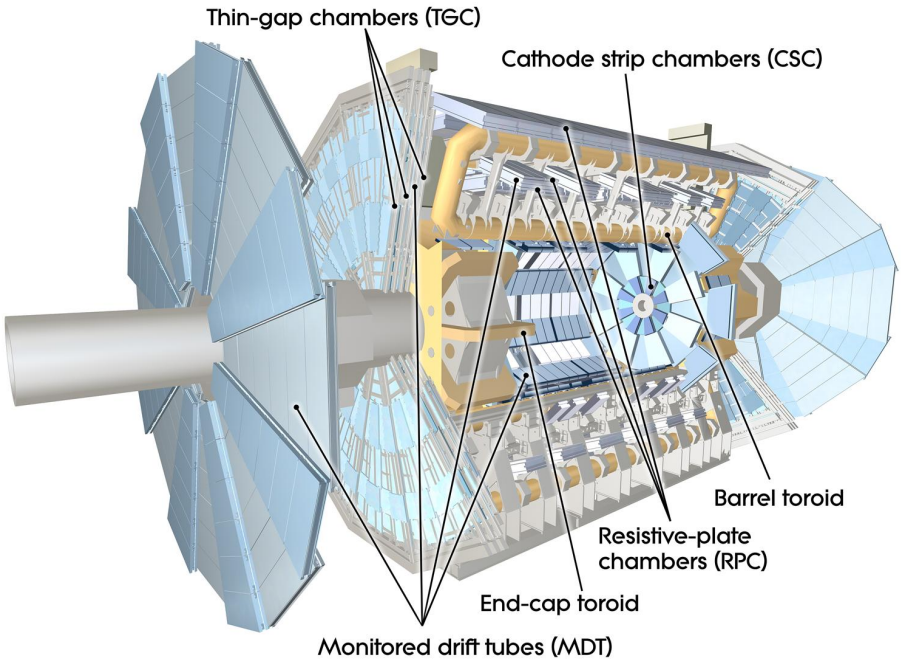



Figure 3.5: Cut-away view of the ATLAS muon detector system [27].

The muon spectrometer provides coverage up to $|\eta| < 2.7$ for precision momentum measurements and up to $|\eta| < 2.4$ for triggering. It consists of three layers of high-precision tracking chambers arranged in the barrel and endcap regions. In the barrel region, Monitored Drift Tubes (MDTs) are used, while Cathode Strip Chambers (CSCs) are employed in the innermost endcap regions, where particle rates are highest.

Muon triggering is achieved using Resistive Plate Chambers (RPCs) in the barrel and Thin Gap Chambers (TGCs) in the endcaps, providing fast timing information for the Level-1 trigger system. The combination of precision tracking and trigger chambers allows for efficient muon identification and momentum reconstruction.

The standalone momentum resolution of the muon spectrometer is approximately 3% for muon transverse momenta up to about 100 GeV and degrades to around 10% at the TeV scale.

## 3.3 Trigger System

The proton-proton collision rate at the LHC far exceeds the ATLAS experiment's data storage and processing capabilities, necessitating an efficient and selective trigger system. At a bunch-crossing frequency of 40 MHz [2], only a tiny fraction of events can be recorded

for offline analysis. Given an average event size of roughly 2 Mb per collision, the combined output of all detector components can reach 80 Tb/s, which far exceeds the available 8 Gb/s of storage bandwidth [28]. The ATLAS trigger system is therefore designed to rapidly identify events of potential physics interest while rejecting background-dominated collisions.

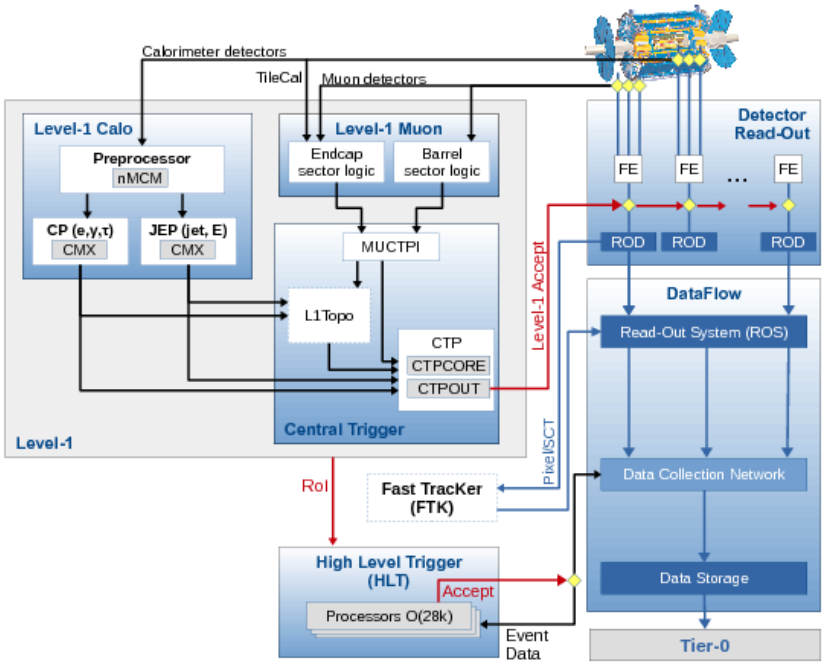



Figure 3.6: The ATLAS Trigger System pipeline [29].

The trigger system operates in multiple stages, as shown in Figure 3.6. The **first-level trigger (Level-1, L1)** is a hardware-based system that uses information from the Level-1 Calorimeters and Level-1 Muon detectors and passes the impact information to the Central Trigger to make a decision within a fixed latency of a few microseconds. It reduces the event rate to approximately 100 kHz [28] by selecting events containing high-transverse-momentum objects, such as electrons, photons, muons, jets, or large missing transverse energy. Additionally it uses topological information from the L1Topo to successfully utilies all available information, to create a clear picture of the decaying products. Level-1 identifies **Regions of Interest (RoIs)**, localized detector regions with significant activity, which are passed to the next trigger stage to focus the processing.

Following a Level-1 Accept, the Detector Read-Out system transfers the full-granularity data from the Front-End (FE) electronics to the Read-Out Drivers (RODs). From here, the DataFlow system takes over, buffering the event fragments in the Read-Out System (ROS).

Simultaneously, events accepted by Level-1 are then processed by the **High-Level Trigger (HLT)**, a software-based system running on a large computing farm. The HLT performs refined reconstruction using full or partial detector readout within the RoIs, further reducing the event rate to about 3 kHz in Run 3.

Trigger selections are organized into **trigger chains**, which define a sequence of reconstruction and selection steps targeting a specific physics signature. For example, a chain to identify Higgs boson decays to tau lepton pairs may involve the following stages:

- **L1 seed:** Require a high-$p_T$ tau candidate in the electromagnetic calorimeter RoI.
- **HLT reconstruction:** Reconstruct tau candidates in the RoI, applying tau identification algorithms based on track and calorimeter cluster shapes.
- **HLT selection:** Require one or more tau candidates passing transverse momentum and isolation criteria to remove QCD background jets.

A set of chains active during a data-taking period constitutes the **trigger menu**, optimized to maintain sensitivity to key physics processes while respecting computing and storage limits. By combining multi-stage selection, RoI-guided reconstruction, and chain-based organization, the ATLAS trigger system efficiently reduces the event rate from 40 MHz to a few kHz while preserving the most physics-relevant collisions.

## 3.4 Tau Identification

Tau leptons [30] play a crucial role in many SM measurements and searches for new physics, particularly in processes involving the Higgs boson and electroweak interactions. The tau lepton has a mass of approximately $m_\tau \approx 1.777$ GeV/$c^2$ and a short lifetime of about $10^{-13}$ s, causing it to decay before reaching the detector. As a result, tau leptons are identified through their decay products. Figure 3.7 shows the branching ratio for tau leptons. Approximately 65% of tau decays proceed hadronically, in green, producing narrow, collimated jets with low charged-particle multiplicity, while the remaining 35% decay leptonically.

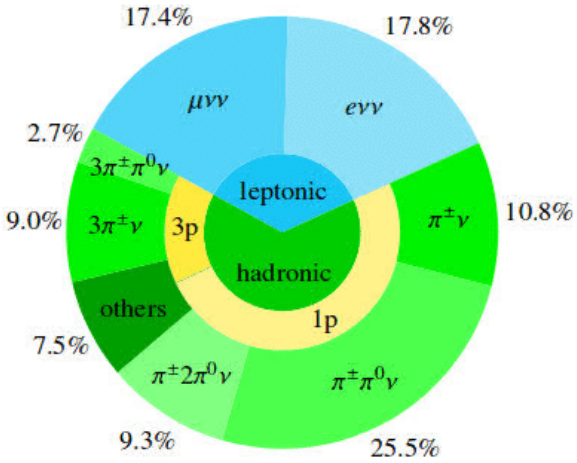



Figure 3.7: Hadronic and leptonic branching ratio for tau leptons. [31].

Hadronically decaying tau leptons are reconstructed in ATLAS using jets seeded by localized energy deposits in the calorimeters and associated tracks in the inner detector. A key concept in tau identification is the prongness of the decay, defined by the number of reconstructed charged particle tracks (prongs) originating from the tau decay vertex. The majority of hadronic tau decays result in one-prong final states, such as $\tau^{\pm} \to \pi^{\pm}\nu_{\tau}$, $\tau^{\pm} \to \pi^{\pm}\pi^{0}\nu_{\tau}$, or $\tau^{\pm} \to \pi^{\pm}\pi^{0}\pi^{0}\nu_{\tau}$, account for approximately 45,6% of all tau decays. Three-prong decays, such as $\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu_{\tau}$, make up about 11,7% of tau decays, as illustrated in Figure 3.7. These decay modes account for the dominant fraction of hadronic tau decays and form the primary signal categories used in tau identification algorithms.

In addition to the one- and three-prong categories, a zero-prong category is retained in the reconstruction and classification chain. In the tau identification algorithm, the leading tau candidate is first selected using calorimeter energy clusters and tracking information. Around this seed, a small cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.1$ defines the tau core region, which is used as the region of interest for associating tracks and neutral energy deposits. Zero-prong tau candidates are reconstructed when no tracks are found within this core region, even though the decay originates from a genuine hadronic tau. Such cases can result from tracking inefficiencies or detector acceptance gaps. Including the zero-prong category in the classification framework improves the robustness of the tau identification against detector-related effects and allows such candidates to be treated consistently within the analysis.

As already mentioned, another key ingredient in tau identification is the use of a signal and isolation cone, as shown in Figure 3.8. The signal cone ($\Delta R < 0.1$) defines a narrow region around the tau axis that contains the decay products of the tau lepton, including charged tracks and neutral energy deposits. Its small angular size reflects the highly collimated nature of hadronic tau decays.

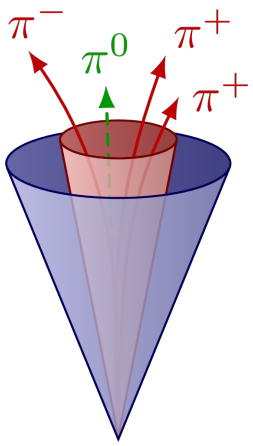



Figure 3.8: A hadronic jet initiated by a $\tau$-lepton that decays to three charged pions ("multi-prong") plus one neutral pion, showing signal and isolation cone. Adapted from graphic by © Izaak Neutelings, CC BY 2.5

Surrounding the signal cone, an isolation region ($\Delta R < 0.2$) is defined to quantify additional activity in the vicinity of the tau candidate. Genuine tau decays are typically isolated, exhibiting little additional track or calorimeter activity outside the signal cone, whereas QCD jets tend to populate a broader area with higher particle multiplicity. Variables derived from the isolation region, such as track and energy sums, therefore provide strong discriminating power and are widely used as inputs to multivariate tau identification algorithms.

In addition to prong-based classification, tau identification exploits vertex-related information in a manner conceptually similar to bottom-quark jet ($b$-jet) tagging. As illustrated in Figure 3.9, long-lived particles produce a secondary vertex that is displaced from the primary interaction point. While the tau lifetime is significantly shorter than that of $b$-hadrons, hadronically decaying tau leptons can still give rise to a measurable secondary vertex due to their finite decay length. The tau lepton's mean lifetime is about $2.9 \cdot 10^{-13}$s, corresponding to a proper decay length $c\tau \approx 87\mu m$. Boosted tau leptons from high-energy collisions travel millimeters before decaying.

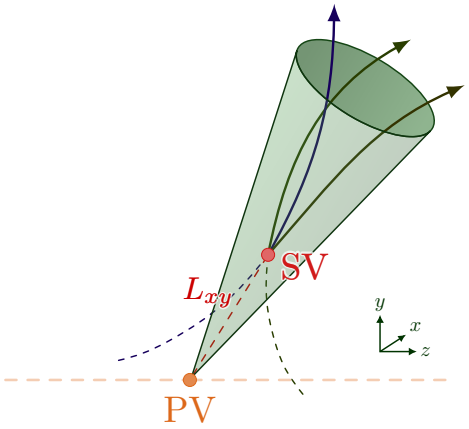



Figure 3.9: Characteristics of a jet initiated by a tau lepton: secondary vertex. Adapted from graphic by © Izaak Neutelings, CC BY 2.5

In ATLAS, tracks associated with the tau candidate are used to reconstruct a tau decay vertex when sufficient prong multiplicity is present, particularly for three-prong decays. Vertex-related observables such as the transverse and longitudinal impact parameters ($d_0$, $z_0$). These variables provide additional discrimination against prompt QCD jets, whose tracks typically originate directly from the primary vertex. By reconstructing and exploiting the tau secondary vertex, the identification performance is significantly improved, especially for multi-prong tau candidates.

# 4 Machine Learning

In the ATLAS trigger system, machine learning models have become an essential component for object identification and event selection. Compared with previously used approaches, such as boosted decision trees, modern neural network architectures have demonstrated improved performance and robustness, particularly in complex environments.

For tau lepton identification at trigger level, ATLAS employs a transformer-based neural network that exploits correlations between reconstructed jet- and track-level features. By processing detailed information about the internal structure of tau candidates, the model captures characteristic signatures of hadronic tau decays and improves discrimination against background jets. Throughout the thesis, the used network is denoted as **GNTau** as it exploits features of a graphical neural network. Before the development of transformer-based deep neural networks, early models were built from simpler components called *perceptrons*, which perform linear transformations followed by nonlinear activation functions.

## 4.1 Perceptrons

The perceptron [32] is one of the earliest and most fundamental models in supervised machine learning, originally introduced as a simplified mathematical abstraction of a biological neuron. It operates by computing a weighted linear combination of input features,

where $x_i$ denote the input variables, $w_i$ the corresponding weights, and $b$ a bias term. The result is passed through an activation function, typically a step or sign function, to produce a binary output as shown in 4.1.

Training a perceptron involves iteratively adjusting its weights based on labeled training data to minimize classification errors. The update rule modifies the weights in proportion to the input features and the discrepancy between the predicted and true labels. Despite its simplicity, the perceptron is guaranteed to converge for linearly separable datasets.

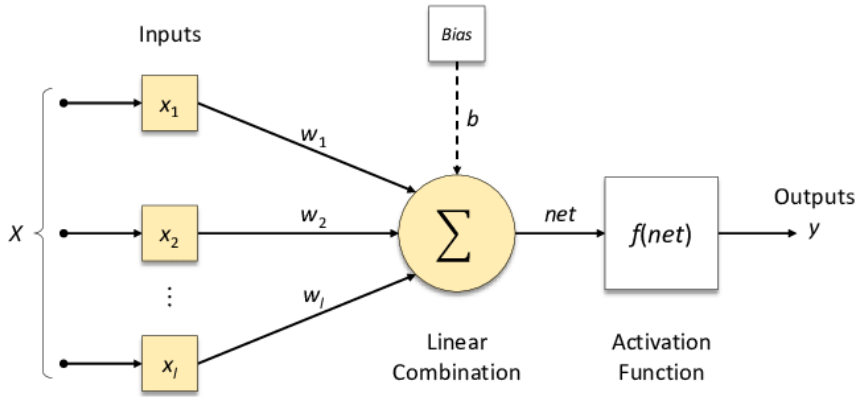



Figure 4.1: The structure of a Single Perceptron [33].

However, the perceptron's reliance on a linear decision boundary severely limits its expressive power, as it cannot solve problems that are not linearly separable. Nonetheless, the perceptron provides the conceptual foundation for multilayer neural networks and modern deep learning architectures.

### 4.1.1 Loss Functions: Cross-Entropy Loss

For classification tasks, neural networks are commonly trained using the cross-entropy loss function. In the binary case, the loss is defined as

$$\mathcal{L}_{\text{CE}} = -\left[y \log(p) + (1-y) \log(1-p)\right], \tag{4.1}$$

where $y$ denotes the true label and $p$ the predicted probability for the positive class [34].

Cross-entropy loss strongly penalizes confident but incorrect predictions and provides well-behaved gradients for probabilistic outputs. In the GNTau model, cross-entropy loss is used to optimize the separation between genuine tau leptons and background jets.

## 4.2 Neural Networks

Artificial neural networks extend the perceptron by organizing multiple neurons into layered architectures [35], enabling the modeling of complex nonlinear relationships in data. The simplest and most widely used architecture is the *Multi-Layer Perceptron* (MLP), a fully connected, feed-forward neural network consisting of an input layer, one or more hidden layers, and an output layer. Figure 4.2 shows a simple example of a Multi-Layer

Perceptron, which is also called a neural network. Each neuron applies an affine transformation followed by a non-linear activation function, such as the rectified linear unit (ReLU) or sigmoid linear unit (SiLU).

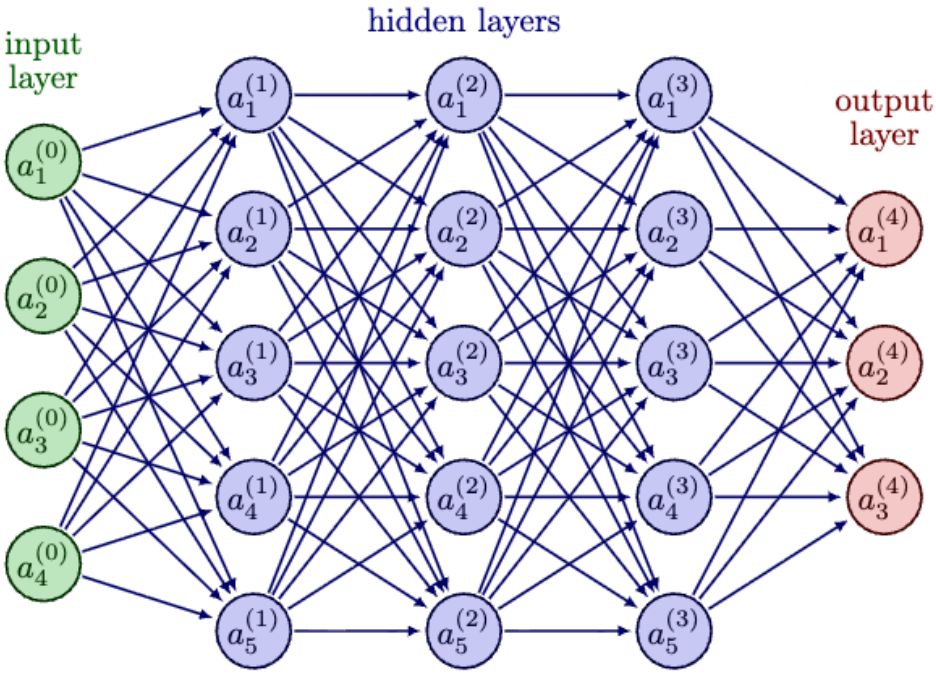



Figure 4.2: Illustration of a neural network with four input nodes, three hidden layers of size five, and an output layer with three nodes. © Izaak Neutelings, CC BY 2.5

Neural network training is performed using gradient-based optimization methods, most commonly stochastic gradient descent and its variants. The **backpropagation** [36] algorithm efficiently computes gradients of a loss function with respect to all network parameters by applying the chain rule of calculus. This allows the network to iteratively minimize a task-specific loss, such as mean squared error for regression or cross-entropy for classification. The loss function here plays a crucial role, as it has a minimum that can be found through optimization, and optimizing it is considered learning.

Deep neural networks, characterized by many hidden layers, can learn hierarchical feature representations directly from raw data. Lower layers typically capture simple, local patterns, while higher layers learn increasingly abstract and task-specific features. This property has led to state-of-the-art performance across a wide range of domains, including computer vision, speech recognition, and high-energy physics data analysis.

### 4.2.1 Sigmoid Linear Unit (SiLU)

Activation functions introduce nonlinearity into neural networks, enabling them to model complex, nonlinear relationships in the input data that cannot be captured by purely linear transformations. They are typically applied after the linear transformation of each layer in the network.

Unlike piecewise-linear activations such as ReLU, SiLU is smooth and non-monotonic, allowing for small negative outputs. This property has been shown to improve optimization stability and performance in deep networks. The GNTau architecture employs SiLU activations to enhance gradient flow and improve representational capacity, particularly in deep and attention-based layers.

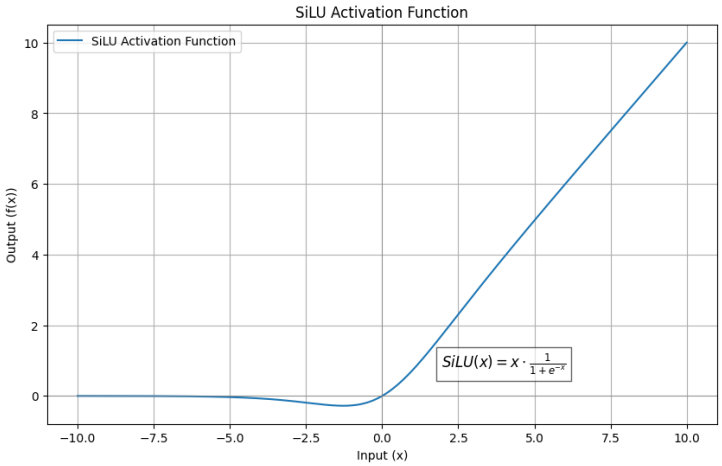



Figure 4.3: The Sigmoid Linear Unit [37].

The Sigmoid Linear Unit (SiLU) as shown in Figure 4.3, also known as the Swish activation function, is defined as

$$\mathrm{SiLU}(x) = x \cdot \frac{1}{1 + e^{-x}}, \tag{4.2}$$

[38].

### 4.2.2 Dropout

Dropout is a regularization technique widely used in deep neural networks to mitigate overfitting [39]. As illustrated in Figure 4.4, during training, a fraction of neuron activations is randomly set to zero with a fixed probability, effectively preventing the network from relying too strongly on individual features or paths through the model.

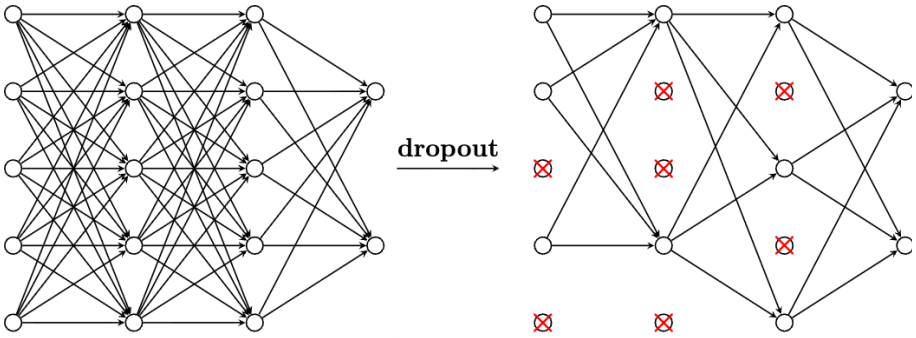



Figure 4.4: Schematic representation of dropout during neural network training, where a random subset of neurons is temporarily deactivated to reduce overfitting. © Janosh Riebesell, CC BY 2.5

By randomly deactivating neurons, dropout encourages the network to learn more robust and redundant feature representations. At inference time, all neurons are active and their outputs are appropriately rescaled to account for the missing activations during training. In the GNTau model, dropout is applied extensively to improve generalization performance, particularly given the high model capacity and the complex correlations present in detector-level inputs.

## 4.3 Transformers

The Transformer architecture, introduced in Ref. [40], is a sequence-to-sequence model that relies entirely on attention mechanisms, dispensing with recurrence and convolution altogether. Figure 4.5 illustrates the schematic of a transformer. It employs an encoder-decoder architecture in which both the encoder and decoder are built from stacks of identical layers. Each encoder layer comprises two main sublayers: a multi-head self-attention mechanism and a position-wise fully connected feed-forward network. Residual connections are applied around each sublayer, followed by layer normalization to facilitate stable and efficient training.

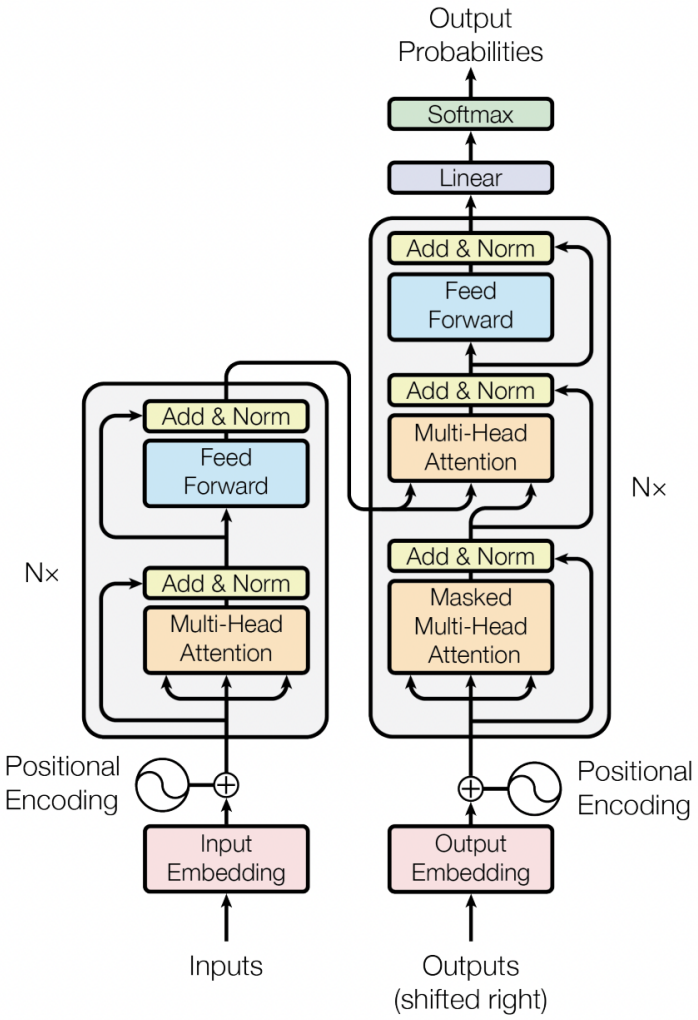



Figure 4.5: Schematic overview of the Transformer architecture [40].

In the encoder, self-attention allows each input token to attend to all other tokens in the sequence, enabling the model to capture global dependencies regardless of distance. The self-attention mechanism operates on learned linear projections of the input embeddings, producing queries, keys, and values. Attention weights are computed using scaled dot-product attention.

The decoder mirrors the encoder structure but includes an additional encoder-decoder attention sublayer. This mechanism enables the decoder to attend to the encoder's output representations and to integrate information from the input sequence during generation. To preserve the autoregressive property during training, masked self-attention is employed in the decoder, preventing positions from attending to subsequent tokens.

Since the Transformer lacks an inherent notion of sequence order, positional encodings are added to the input embeddings to inject information about token positions. In the original formulation, these encodings are deterministic sinusoidal functions with varying frequencies, enabling the model to generalize to sequence lengths unseen during training.

By combining parallelizable attention mechanisms with deep stacking and residual learning, the Transformer achieves superior performance and scalability compared to recurrent architectures. Its ability to model long-range dependencies efficiently has made it the foundation of many state-of-the-art models across natural language processing and beyond.

### 4.3.1 Scaled Dot-Product Attention

Scaled dot-product attention is the fundamental building block of attention-based architectures. Given a set of queries $Q$, keys $K$, and values $V$, attention is computed as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{4.3}$$

where $d_k$ denotes the dimensionality of the key vectors [40]. The **softmax** function converts the scaled dot products into a normalized probability distribution, ensuring that attention weights are positive and emphasize the most relevant features while suppressing relatively less informative ones. The query, key, and value matrices are obtained via learned linear projections of the input representations, i.e. $Q = XW_Q$, $K = XW_K$, and $V = XW_V$, where $W_Q$, $W_K$, and $W_V$ are trainable weight matrices.

The scaling factor stabilizes gradients for high-dimensional features. This mechanism allows the model to dynamically weight interactions between elements, enabling it to focus on the most relevant components of the input. In GNTau, scaled dot-product attention is used to model complex relationships between tracks, clusters, and other detector-level features.

### 4.3.2 Multi-Head Attention

Multi-head attention extends scaled dot-product attention by applying multiple attention operations in parallel. Each attention head operates on a distinct learned projection of the input, enabling the model to capture complementary relationships.

The outputs of all attention heads are concatenated and linearly transformed to produce the final representation. This approach improves expressiveness and stability compared to a single attention mechanism. In the GNTau model, multi-head attention enables the network to simultaneously exploit different correlation patterns among particle-level inputs.

## 4.4 Graph Neural Networks

Graph Neural Networks (GNNs) extend neural network methods to data represented as graphs, consisting of nodes connected by edges, as illustrated in Figure 4.6. Such structures naturally arise in many scientific and engineering domains, including detector data in high-energy physics. Unlike grid-based data, graph-structured data lacks a fixed ordering and spatial locality, and thus requires specialized architectures that respect permutation invariance.

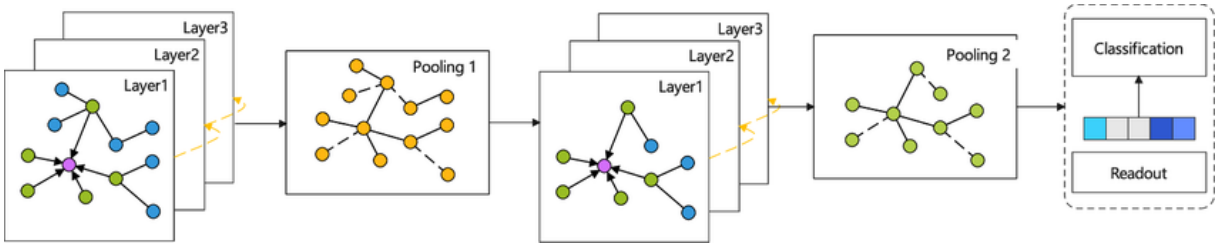



Figure 4.6: Schematic diagram of a graph neural network [41].

For graph-level classification tasks, node-level representations must be aggregated into a fixed-size global feature vector. **Global attention pooling** [42] performs this aggregation by assigning learnable importance weights to individual nodes.

# 5 Data Preprocessing

Data preprocessing is a crucial step in any particle physics analysis. The raw information produced by the detector and simulated Monte Carlo (MC) events is often complex, high-dimensional, and subject to various sources of noise, inefficiencies, and detector effects. Preprocessing transforms raw datasets into a clean, consistent, and physically meaningful format suitable for analysis and machine learning. By performing careful preprocessing, we ensure that subsequent analysis accurately reflects the underlying physics, improves signal-to-background separation, and reduces biases and systematic uncertainties. Preprocessing is a critical bottleneck for GNTau, as it must accurately map the decay topology into a graph structure that prioritizes the most relevant features for high-performance identification.

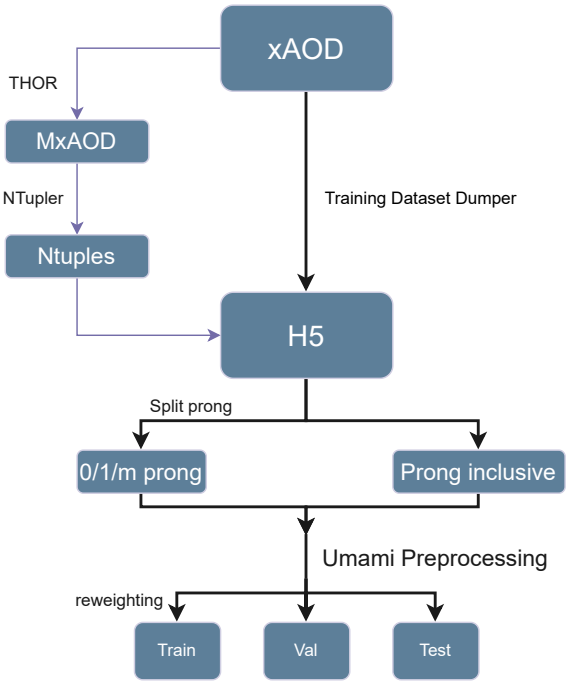



Figure 5.1: Schematic overview of the GNTau pipeline, showing the data flow from xAOD retrieval through preprocessing to train, validation, and test sets.

The workflow begins with the xAOD (**eXperimental Analysis Object Data**) format, the standard analysis data format used within the ATLAS collaboration. The xAOD files contain reconstructed high-level physics objects such as tracks, calorimeter clusters, jets, and tau candidates, together with their associated metadata and auxiliary variables. Compared with lower-level formats, xAOD provides a structured, analysis-ready representation of events while retaining sufficient detail for advanced reconstruction and machine-learning applications. Its widespread use within ATLAS ensures consistency, reproducibility, and compatibility with centrally produced datasets.

For machine learning applications, however, xAOD files are not directly suitable for large-scale training due to their complex structure. Therefore, the relevant physics objects and features must be extracted and converted into a more efficient format for processing. In this workflow, the data are converted to HDF5(Hierarchical Data Format version 5).

Previously, the ATLAS Tau Trigger group relied on the legacy THOR framework for dataset production. THOR implemented its own xAOD retrieval and conversion process to generate training samples for the GNTau model. In the present work, this workflow has been replaced by the newly developed **Training Dataset Dumper (TDD)**. TDD was originally developed and extensively used for $b$-jet tagging, but is now actively maintained and extended for tau trigger applications.

After conversion to the HDF5 format, two different main training strategies are considered: **prong-exclusive** and **prong-inclusive** training. The classification is based on the reconstructed number of associated tracks, $N^{Tracks}$. In the prong-inclusive approach, no selection is applied to $N^{Tracks}$. As illustrated in Figure 5.2, the distribution of $N^{Tracks}$ reveals that the majority of events are classified as 1-prong, with most containing three or fewer tracks. This behavior is expected, as hadronic tau decays are predominantly 1-prong and 3-prong topologies. In contrast, QCD multijet background events typically exhibit much higher particle multiplicities, which explains the long tail in the background distribution extending to 30 or more tracks. The full dataset is used for training, combining all prong multiplicities into a single model. This strategy provides a unified classifier that learns across all decay topologies simultaneously. In contrast, in the prong-exclusive approach, the dataset is divided into categories according to the reconstructed charged-decay multiplicity of the tau candidate. Specifically, events are split into 0-prong, 1-prong, and multi-prong selections, and independent training is performed for each category. This allows the model to specialize in the distinct topological and kinematic characteristics of

each decay mode.

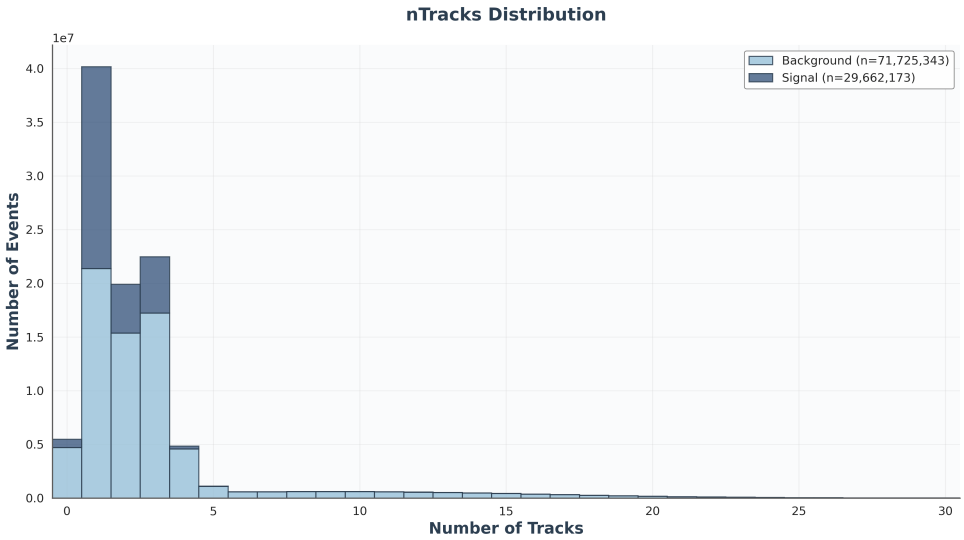



Figure 5.2: Distribution of the Number of Tracks for each Event.

Following the prong selection, the datasets undergo the Umami preprocessing step. This stage performs feature normalization, scaling, masking of invalid entries, and organization of inputs into the structured format required by the graph neural network architecture. The processed datasets are then used for training and validation with the SALT framework [43]. SALT provides the training infrastructure for graph-based taggers, including model configuration, optimization, checkpointing, and monitoring of performance metrics such as loss and background rejection at fixed signal efficiency.

Finally, the trained models are evaluated using the PUMA framework, which standardizes performance assessment. PUMA produces ROC curves, discriminant distributions, and working-point comparisons, enabling a consistent evaluation of signal efficiency and background rejection across different preprocessing strategies and training configurations.

## 5.1 Monte Carlo Simulation Datasets

This analysis relies on a variety of Monte Carlo (MC) simulated datasets to model both signal and background processes. These datasets are produced with different event generators, parton shower configurations, and detector simulation conditions to ensure a realistic description of particle interactions and the detector response. Each dataset is labeled by a unique production name that encodes the generator, physics process, and reconstruction version. The main datasets used in this analysis are shown in Figure 8.1.

Signal samples contain processes that produce genuine hadronically decaying tau leptons, whereas background samples are dominated by QCD dijet production.

The $\boldsymbol{\gamma^* \rightarrow \tau\tau}$ dataset corresponds to electroweak tau-lepton pair production via virtual photons and constitutes the dominant signal contribution, owing to its large production cross-section and well-understood experimental signature. Event generation is performed with Pythia 8.212 [44], which is used both for the matrix-element calculation and for the simulation of parton showering, hadronisation, the underlying event, and $\tau$-lepton decays. The A14 [45] tune and the NNPDF2.3 LO [46] parton distribution functions are used, with the strong coupling constant set to $\alpha_s = 0.13$. To maximise the available training statistics, all $\tau$ leptons in this sample are forced to decay hadronically. A biased phase-space sampling is applied to enhance the population at high invariant masses of the $\tau\tau$ system ($m_{\gamma^*}$), thereby ensuring sufficient coverage of high-$p_{\mathrm{T}}$ hadronic $\tau$ candidates.

The dominant background contribution arises from QCD dijet production, simulated with Pythia 8.186. These samples model jets produced in hard scatterings that can be misidentified as hadronic tau decays. To improve statistical coverage at high transverse momentum, a biased phase-space sampling is employed, resulting in events that are reweighted with continuously decreasing event weights to preserve the physical jet $p_{\mathrm{T}}$ spectrum.

Proton-proton collisions are generated at a center-of-mass energy of $\sqrt{s} = 13$ TeV with a bunch spacing of 25 ns. The decays of bottom- and charm-hadrons in all samples are simulated using EvtGen v1.2.0 [47]. Multiple interactions per bunch crossing (pile-up) are modelled by overlaying simulated minimum-bias events generated with Pythia 8.186, using the A3 tune and the NNPDF2.3 LO parton distribution function set. The average number of interactions per bunch crossing reaches up to $\mu \simeq 70$, representative of Lhc Run-2 conditions.

All generated events are processed with the full Atlas detector simulation, based on Geant4, to provide a realistic modelling of both the detector response and reconstruction effects.

## 5.2 Input Variables for Machine Learning

The selection of input variables is critical for the performance of machine learning classifiers, as they must capture the relevant kinematic, topological, and substructure information of tau jets and their decay products. In this analysis, features are grouped into three categories: jet-level, track-level, and calorimeter cell variables. Some variables are not employed for 0-prong tau candidates because no associated charged tracks are present.

At the jet level, several observables are used to characterise the energy scale, shape, and internal structure of the tau candidate. The transverse momentum ($p_T^{\tau_{had-vis}}$) provides a measure of the overall energy of the reconstructed jet, while the central energy fraction ($f_{cent}^{\Delta R<0.1}$) quantifies how strongly the jet energy is concentrated within $\Delta R < 0.1$ with respect to all energy deposited in $\Delta R < 0.2$ around the $\tau_{had-vis}$ candidate axis.

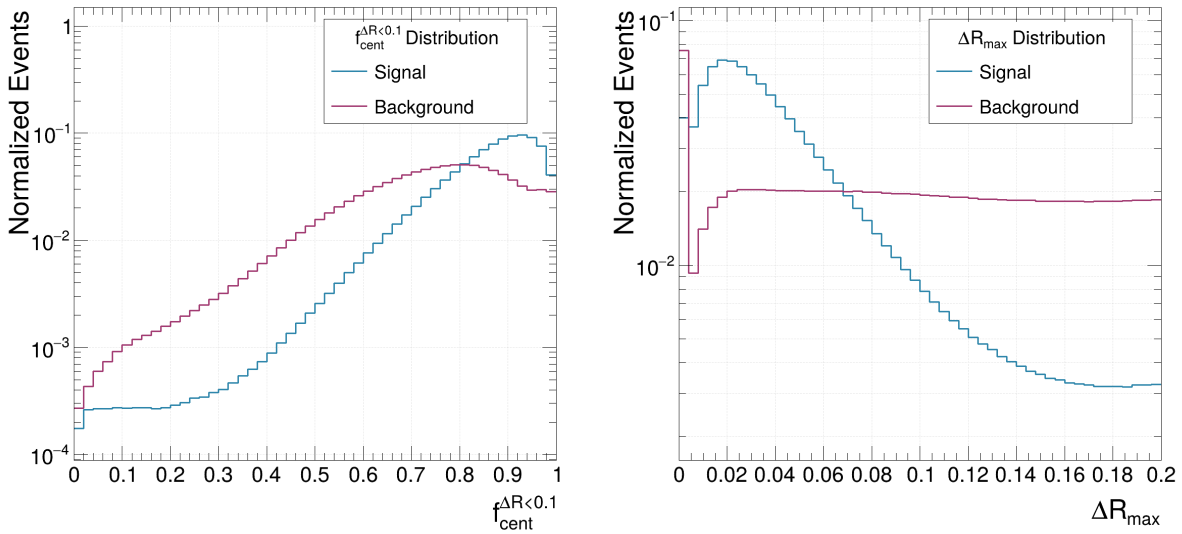



Figure 5.3: Normalized event count of the input variables $f_{cent}^{\Delta R<0.1}$ and $\Delta R_{max}$ for Signal ($\gamma^* \to \tau\tau, t\bar{t}, Z \to \tau\tau$) and Background (Dijet).

Hadronically decaying tau leptons produce narrow and collimated energy deposits, resulting in large values of $f_{cent}^{\Delta R<0.1}$. In contrast, QCD jets typically exhibit broader energy distributions due to parton showering and hadronisation, leading to smaller central energy fractions. This behaviour is clearly visible in Figure 5.3 in the left plot, where signal events peak at high values of $f_{cent}^{\Delta R<0.1}$, while background events feature a flatter distribution.

The maximum angular separation provides complementary information, $\Delta R_{\text{max}}$, defined

as the largest distance in $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ between the tau axis and its associated tracks within the reconstruction cone. Genuine hadronic tau decays are typically highly collimated due to the tau lepton's low multiplicity and limited phase space, resulting in small values of $\Delta R_{\text{max}}$. In contrast, QCD jets originate from quark or gluon fragmentation and tend to exhibit a broader radiation pattern, leading to larger values of $\Delta R_{\text{max}}$. As shown in Figure 5.3 in the right plot, signal events peak at small $\Delta R_{\text{max}}$ and fall off rapidly, while the background distribution extends to significantly larger values. This clear separation demonstrates the strong discriminating power of $\Delta R_{\text{max}}$ between genuine tau decays and dijet background.

Additional jet-level variables describe the consistency between tracker and calorimeter measurements, such as the ratio of electromagnetic energy to track-system momentum ($f_{track}^{EM}$) and track-plus-EM momentum over the calorimeter-only momentum ($\frac{p_T^{track+EM}}{p_T^{calo-only}}$), which measures the ratio of the transverse momentum $p_T$ of a system reconstructed from the charged-classified tracks and the 1st and 2nd most energetic TopoClusters in the EM calorimeter, over the calorimeter-only measurement of the $\tau_{had-vis}$ candidate's $p_T$. The internal jet substructure is further characterised using the track-plus-EM system mass ($m_{track+EM}$), the momentum fraction of isolation tracks ($\frac{\sum p_T^{\text{iso tracks}}}{\sum p_T^{\text{all tracks}}}$), the ratio of transverse energy to the leading-track momentum ($\frac{\sum E_T^{cells}}{p_T^{\text{lead track}}}$), the invariant mass of the track system ($m^{track}$), and the average distance between inner tracks($\langle \Delta R_{trk} \rangle$). As shown in the table, some observables are undefined for zero-prong candidates because no tracks are observed.

| **Input Features missing in the 0-prong category** |
|---|
| $f_{track}^{EM}$ |
| $\Delta R_{\text{max}}$ |
| $\frac{\sum E_T^{cells}}{p_T^{\text{lead track}}}$ |
| $\langle \Delta R_{trk} \rangle$ |

Table 5.1: The not included input variables for the 0-prong model.

Besides kinematic and substructure observables, the number of associated tracks ($N^{tracks}$) plays a central role in the characterisation of hadronic tau decays. This variable corresponds to the reconstructed prong multiplicity of the tau candidate and reflects the decay modes of hadronic tau leptons, which predominantly occur as one-prong or three-prong decays. It also contains a possible 0 value due to detector errors or reconstruction failure. In contrast, $N_{truth}^{tracks}$ derived from Monte Carlo simulations provides the true prong multiplicity and is free from such detector-induced artifacts, making it a reliable reference

for validation.

At the track level, each track associated with a tau candidate contributes detailed kinematic and quality information. The transverse momentum ($p_T^{tracks}$) and angular displacements relative to the jet axis ($\Delta\eta^{track}$ and $\Delta\phi^{track}$) describe the track geometry within the jet. Track impact parameters are computed relative to a reconstructed interaction vertex. As illustrated in Figure 3.9, displaced decay vertices and the associated transverse and longitudinal impact parameters provide powerful discrimination. In offline reconstruction, the Tau Jet Vertex Association (TJVA) algorithm exploits this concept by associating tracks to a reconstructed tau decay (secondary) vertex and selecting the primary vertex most compatible with the tau candidate. At trigger level, where full vertex reconstruction is not available, the High-Level Trigger (HLT) reconstructs a primary vertex that is used as the reference point for impact parameter measurements. The resulting quantities, $d_0^{track}$ and $z_0^{track}\sin\theta$, together with their significances $d_0^{sig}$ and $z_0^{sig}\sin\theta$, quantify the compatibility of individual tracks with the tau decay topology and enhance discrimination against background jets and pile-up. Track quality indicators, such as the number of hits in the pixel detector ($N_{\text{Pixel Hits}}$), the silicon microstrip tracker ($N_{\text{SCT hits}}$), and the innermost pixel layer ($N_{\text{IBL hits}}$), are included to ensure that only well-measured tracks contribute to the classification.

Finally, calorimeter cell features provide detailed information on the energy deposition within the tau jet. Each cluster is characterized by its transverse energy ($E_T^{cluster}$) and angular displacement from the jet axis ($\Delta\phi^{cluster}$ and $\Delta\eta^{cluster}$). Shower shape variables, including $\langle r^2\rangle$, $\langle\lambda^2\rangle$, and $\langle\lambda_{cluster}\rangle$, describe the spatial distribution of energy within the cluster, and are particularly sensitive to the number of prongs in the decay.

## 5.3 Prong Exclusive Cuts

After assessing the dataset, dedicated selections are applied to partition the tau candidates according to their reconstructed track multiplicity. The Table 5.2 summarizes the kinematic selection criteria applied to all reconstructed tau jet candidates and the conditions used to assign events to the individual prong categories.

| Category | Selection Criteria |
|---|---|
| All | $20\ \text{GeV} < p_t < 600\ \text{GeV} \ \wedge\ \lvert\eta\rvert < 2.5$ |
| 0-prong (0p) | $\text{N}^{tracks} = 0$ |
| 1-prong (1p) | $\text{N}^{tracks} = 1 \ \wedge\ (\text{N}^{tracks}_{truth} = 1 \vee 3)$ |
| Multi-prong (mp) | $(\text{N}^{tracks} = 2 \vee 3) \ \wedge\ (\text{N}^{tracks}_{truth} = 1 \vee 3)$ |

Table 5.2: Selection cuts and prong-based categorization applied by the `cuts_and_prongs` preprocessing script.

Figure 5.4 shows the distribution of reconstructed jets across the different prong categories, separated into signal and background contributions. The majority of jets populate the 1-prong category, with most of these originating from the $\gamma^* \to \tau\tau$ process. This channel constitutes the dominant and most relevant signal contribution in the analysis due to its large production cross-section and well-understood experimental signature.

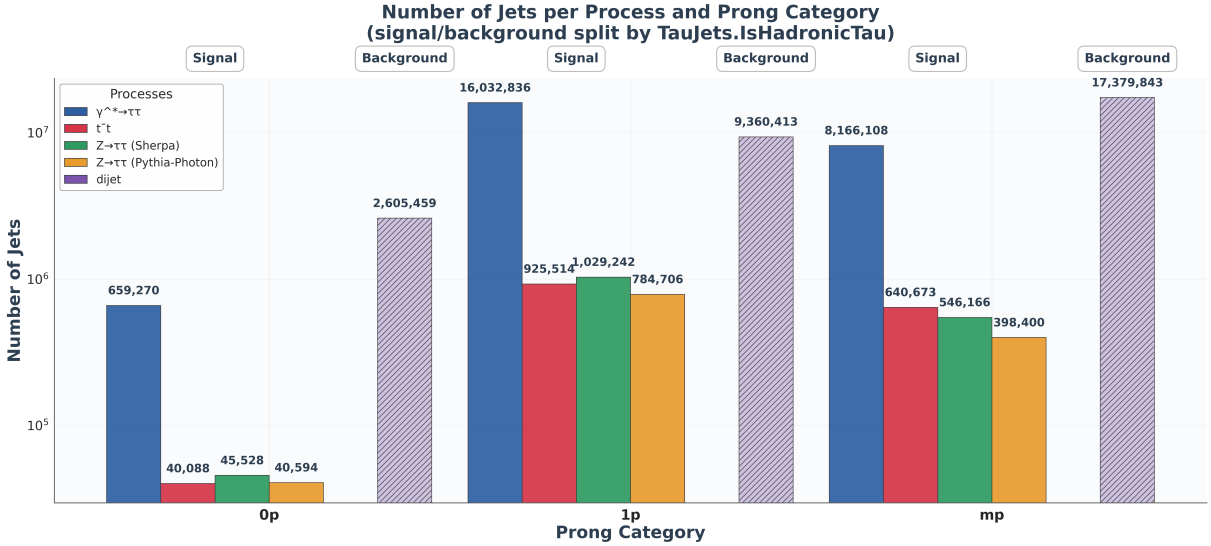



Figure 5.4: Number of reconstructed jets categorized by prongness and signal or background origin.

The 0-prong category contains only a small number of entries, as genuine hadronic tau decays without associated charged tracks are not physically expected. Multi-prong decays, while less abundant than 1-prong decays, provide complementary information and are therefore treated separately due to their distinct topological and kinematic characteristics.

## 5.4 Umami Framework

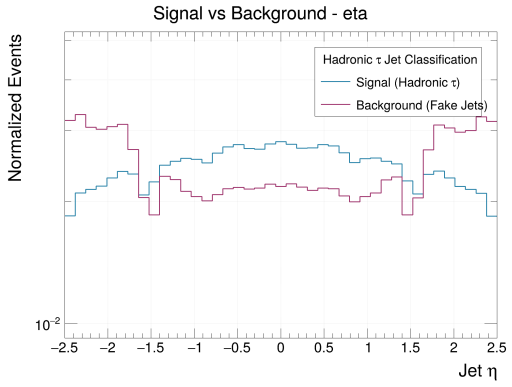
After the dataset has been acquired, a comprehensive preprocessing step is required to ensure its suitability for machine learning applications. This preprocessing is performed using the Umami framework, which provides standardized tools for dataset preparation, event selection, and sample balancing in the context of jet-based analyses. The primary goals of this step are to construct statistically robust training, validation, and test samples, and to reduce potential biases arising from kinematic differences between signal and background processes.

The dataset is first divided into mutually exclusive training (80%), validation (10%), and test (10%) sets. This separation is performed at the event level to prevent information leakage between datasets and to allow for an unbiased evaluation of model performance. The validation set is used for hyperparameter tuning and model selection, while the test set is reserved exclusively for the final performance assessment.

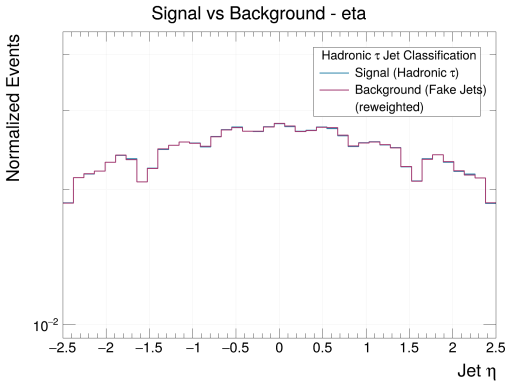
A key feature of the Umami preprocessing pipeline is the normalization of jet kinematics across the transverse momentum ($p_T$) and pseudorapidity ($\eta$) spectra. Since machine learning models are highly sensitive to kinematic differences between signal and background samples, Umami applies a reweighting procedure that equalizes the jet distributions in $p_T$ and $\eta$. This ensures that the classifier does not exploit trivial kinematic features but instead learns meaningful differences in jet substructure and decay topology.

This kinematic normalization improves the generalization of the trained models and enhances their robustness across different physics processes. The procedure is applied consistently across all considered tau decay topologies, namely the 0-prong (0p), 1-prong (1p), multi-prong (mp), and inclusive categories.

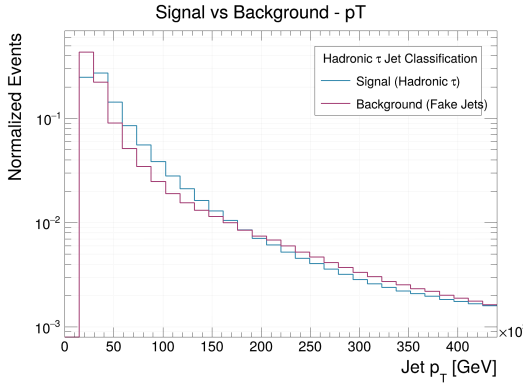
Figures 5.5(a) and 5.5(b) illustrate the effect of the Umami reweighting on the $\eta$ distribution for 1-prong tau candidates before and after preprocessing. A similar comparison for the transverse momentum distribution is shown in Figures 5.5(c) and 5.5(d). In both cases, a clear improvement in the agreement between signal and background distributions is observed after preprocessing, confirming the effectiveness of the normalization procedure. The same observation can be made in Figure 5.6, it shows the $p_T$ and $\eta$ distribution for the multi-prong tau candidates.



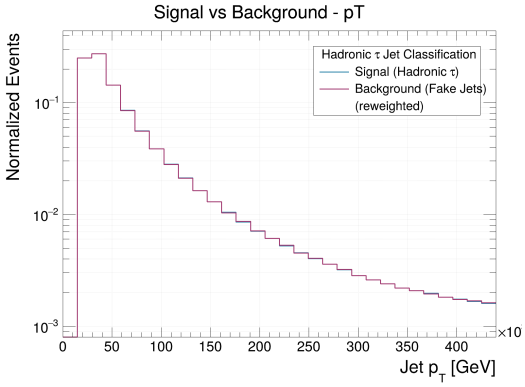

(a): Normalized $\eta$ distribution before Umami preprocessing.




(b): Normalized $\eta$ distribution after Umami preprocessing.

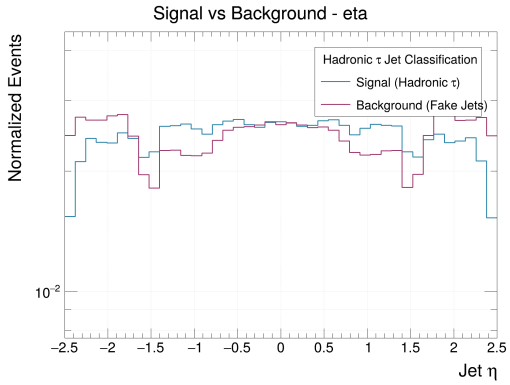



(c): Normalized $p_T$ distribution before Umami preprocessing.

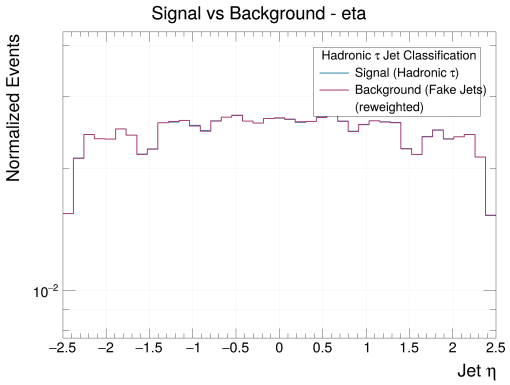



(d): Normalized $p_T$ distribution after Umami preprocessing.

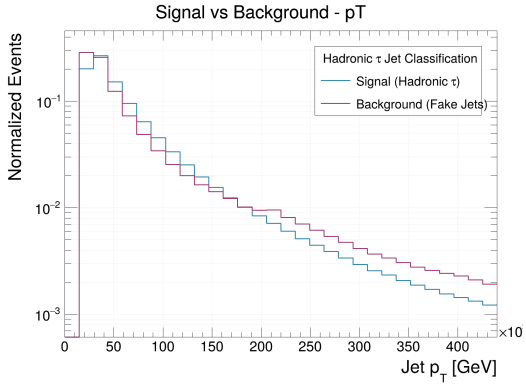
Figure 5.5: Normalized event count of 1-prong tau-decay.



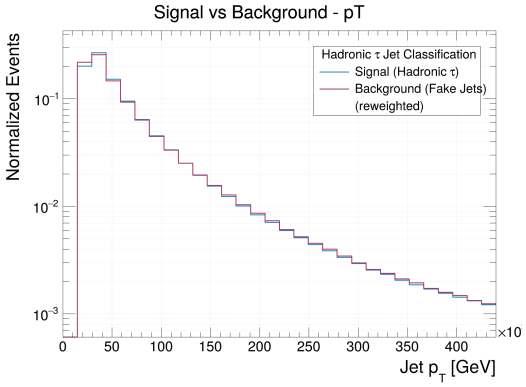

(a): Normalized multi-prong $\eta$ distribution before Umami preprocessing.




(b): Normalized multi-prong $\eta$ distribution after Umami preprocessing.




(c): Normalized multi-prong $p_T$ distribution before Umami preprocessing.




(d): Normalized multi-prong $p_T$ distribution after Umami preprocessing.

Figure 5.6: Normalized event count of multi-prong tau-decay.

# 6 Performance Studies

The performance of tau identification models critically depends on both the quality of the input data and the neural network architecture. In this chapter, we evaluate the performance of the GNTau model across various data-preparation strategies, input configurations, and training approaches. The goal is to assess whether modifications to the data retrieval, such as replacing the legacy THOR workflow with the newly developed TDD, or adopting an inclusive training strategy, affect the model's ability to distinguish genuine hadronic tau decays from background jets.

Throughout this chapter, the evaluation focuses on both overall discrimination power, quantified by ROC curves and discriminant scores, and operationally relevant metrics, such as background rejection at predefined working points. The discriminant score is a numerical value assigned by each model to quantify the likelihood that a candidate is a tau rather than background: higher scores indicate greater confidence that the candidate is a true tau lepton, while lower scores correspond to a higher probability of being background. Working points, commonly categorized as loose, medium, and tight, define specific thresholds on the discriminant score. Candidates with scores above a given threshold are classified as a signal tau candidate, with loose working points prioritizing higher signal efficiency, tight working points emphasizing stronger background rejection, and medium working points providing an intermediate balance. Comparisons are performed for different tau decays (0-prong, 1-prong, multi-prong) to ensure a comprehensive understanding of model behavior across the full range of tau candidates.

## 6.1 Architecture of the GNTau model

The **GNTau** model is a transformer-based graphical neural network architecture designed for hadronic tau identification using low-level detector information. It follows a modular, multi-input design that processes physics objects, tracks, calorimeter, cluster, and jet-level features, within a unified attention-based framework.

Three categories of inputs represent each tau candidate:

- **Jet-level (global) features**, such as transverse momentum, isolation variables, and track-system observables.
- **Track-level features**, describing charged-particle trajectories associated with the tau candidate.
- **Calorimeter cluster (cell) features**, encoding local energy deposits and shower-shape information.

Tracks and clusters are treated as unordered sets of variable size, while jet-level quantities act as a global context shared across all constituents. After that, separate input embedding networks are defined for tracks and calorimeter clusters. Each embedding network consists of a fully connected multilayer perceptron (MLP) with SiLU activation, mapping the raw input features to a 128-dimensional latent space. Jet-level variables are concatenated as global features and injected into both embedding streams, ensuring that local object representations are conditioned on global jet properties.

The embedded track and cluster representations are processed by a shared Transformer encoder. This encoder consists of two stacked self-attention layers with multi-head attention (two attention heads) and feed-forward sublayers. The Transformer enables information exchange between all objects in the set, allowing the model to learn correlations between tracks, clusters, and the global jet context without imposing an explicit ordering.

After contextualization by the Transformer encoder, the variable-length set of object embeddings is reduced to a single fixed-size representation via global attention pooling. This pooling layer learns attention weights over objects, allowing the model to focus on the most discriminative tracks or clusters for tau identification.

The pooled representation is passed to a task-specific classification head, implemented as a shallow MLP with dropout regularization. The final layer outputs logits for a binary classification task (tau vs. notau), trained using a cross-entropy loss. The model is trained end-to-end with the AdamW optimizer [48] and a one-cycle learning rate schedule, starting with a high learning rate and reducing it as training progresses.

In addition to the baseline configuration described above, two further variants of the GNTau model are considered: an inclusive model and an extended model. Importantly,

the core architecture remains unchanged across all variants. In particular, the embedding networks, Transformer encoder, attention pooling, and classification head are identical, ensuring that performance differences can be attributed solely to input feature content rather than architectural modifications.

For the inclusive model, the main differences lie in the data preprocessing and input feature definitions. The architectural components and their hyperparameters are kept fixed, whereas preprocessing adapts the input to include all types of prong multiplicity. No changes are introduced at the model level.

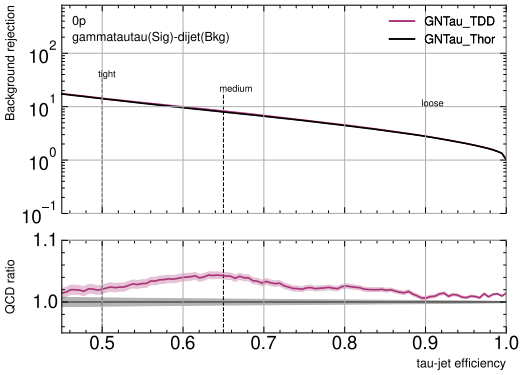
For the extended model, the preprocessing pipeline is likewise unchanged from the inclusive approach, but the dimensionality of the low-level inputs is increased to incorporate additional track- and cluster-level observables

## 6.2 Training Dataset Dumper (TDD) compared to THOR-based models

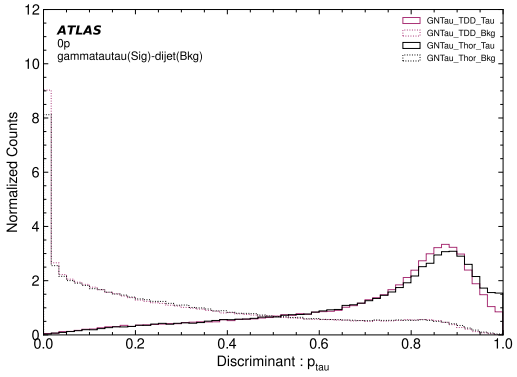
As already mentioned, the previously used THOR-based workflow has been replaced by the actively maintained Training Dataset Dumper (TDD). Since changes in dataset preparation can affect the input feature distributions and, consequently, model performance, it is essential to verify that this transition does not significantly degrade tau identification performance.

The goal of this section is to compare the performance of the GNTau model, trained and evaluated using the TDD workflow, with the corresponding results obtained with the legacy THOR framework. This comparison validates the TDD framework and demonstrates that it delivers performance consistent with the previously established workflow.

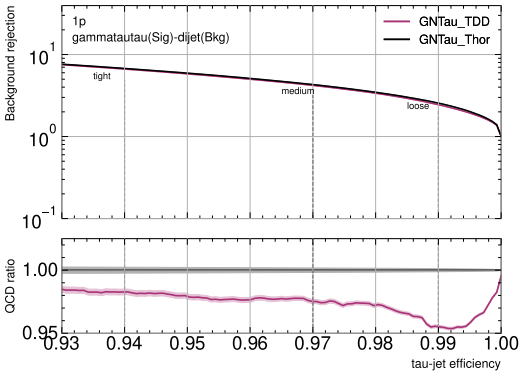
The performance is evaluated using Receiver Operating Characteristic (ROC) curves, which illustrate the trade-off between signal efficiency and background rejection across different discriminator thresholds. In these plots, the x-axis represents the tau-jet signal efficiency ($= \frac{N_{\text{signal, selected}}}{N_{\text{signal, total}}}$), while the y-axis shows the background rejection ($= 1/\frac{N_{\text{background, selected}}}{N_{\text{background, total}}}$). Figures 6.1(a), 6.1(c), and 6.1(e) compare the performance of the GNTau_TDD model trained using TDD with the corresponding results obtained using the legacy THOR workflow. The comparison is performed separately for the 0-prong, 1-prong, and multi-prong tau decay modes for the $\gamma^* \to \tau\tau$ signal against the dijet background.



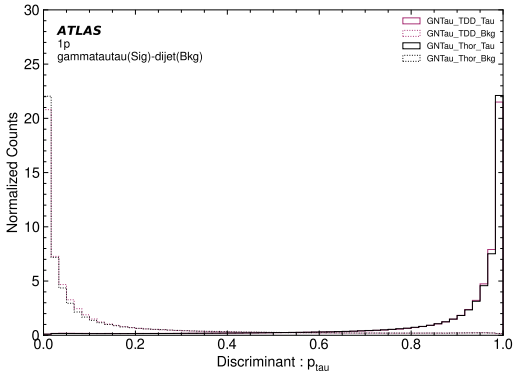

(a): ROC curve of the GNTau using Thor and TDD preprocessing for the 0-prong selection.



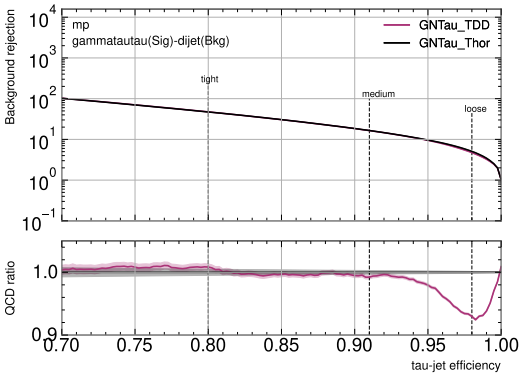

(b): Discriminant distribution of the GNTau using Thor and TDD preprocessing for the 0-prong selection.



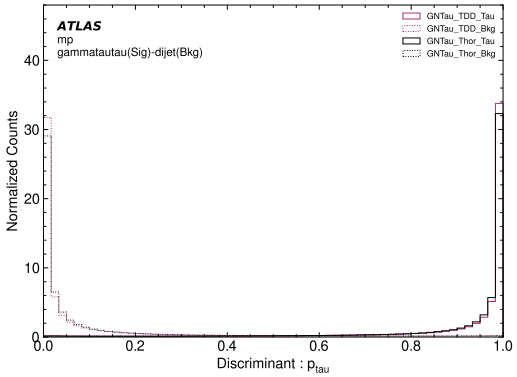

(c): ROC curve of the GNTau using Thor and TDD preprocessing for the 1-prong selection.




(d): Discriminant distribution of the GNTau using Thor and TDD preprocessing for the 1-prong selection.




(e): ROC curve of the GNTau using Thor and TDD preprocessing for the multi-prong selection.




(f): Discriminant distribution of the GNTau using Thor and TDD preprocessing for the multi-prong selection.

Figure 6.1: Performance and Discriminant score of the GNTau, evaluated on $\gamma^* \rightarrow \tau\tau$ signal and dijet background. The THOR model is shown as the black line, while the TDD workflow is shown in purple.

Across all prong categories, the ROC curves obtained with the TDD workflow closely match those from the THOR-based approach, exhibiting very similar shapes and background rejection levels over the full range of tau identification efficiencies. This observation is supported by the comparison of the discriminant scores of each model, shown in Figures 6.1(b), 6.1(d), and 6.1(f). The discriminant score. In each case, the separation between tau signal and background is very similar across models. By comparing these scores across models, one can directly assess how well each model separates tau signals from background events.

For trigger applications, the medium working point is particularly relevant. At the corresponding tau identification efficiencies of approximately 65% for 0-prong, 97% for 1-prong, and 91% for multi-prong candidates, the background rejection achieved with TDD is comparable to that obtained with THOR. While the largest differences between the models are observed in the 1-prong category, the deviation in background rejection at the medium working point is limited to at most a 3% performance loss for the inclusive model, which is considered acceptable. Minor differences in the results may arise from small variations in the number of signal and background jets introduced by the different implementation of the THOR-based and TDD-based framework. This demonstrates that the classifier performance at the operationally most relevant working point is preserved.

Furthermore, the relative performance ordering among prong types is preserved, indicating that the model's physics sensitivity is largely unaffected by the change in the data-preparation workflow. These results demonstrate that TDD delivers performance comparable to established THOR-based training, thereby validating its suitability as a replacement for future tau trigger studies.

## 6.3 Inclusive vs. Exclusive model performance

Building upon the prong-specific baseline, we move toward the **prong-inclusive** modeling strategy. This evaluation is done in two ways. Figure 6.2 visualizes these two different evaluation steps, in order to achieve an unbiased comparison of the prong-inclusive and prong-exclusive models.

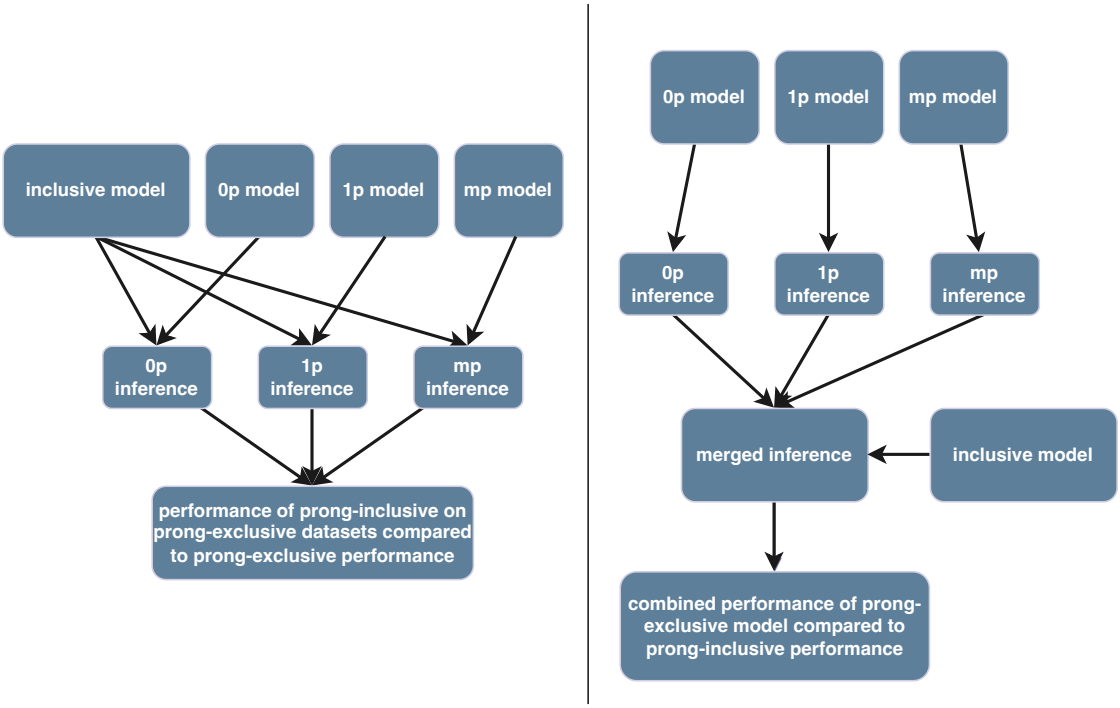
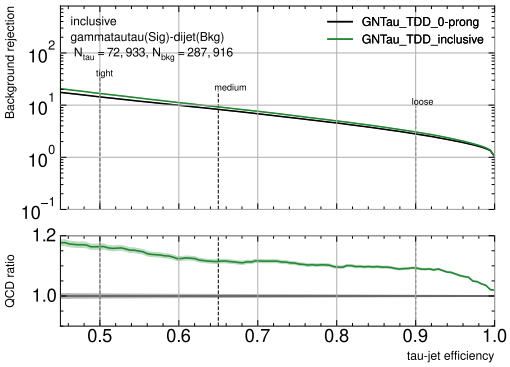
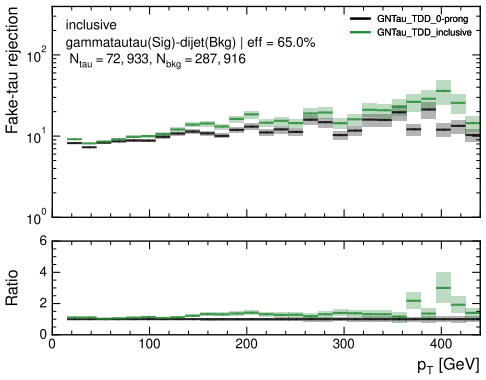



Figure 6.2: Flowchart of the 2 different approaches to compare the performance of the different architectures.

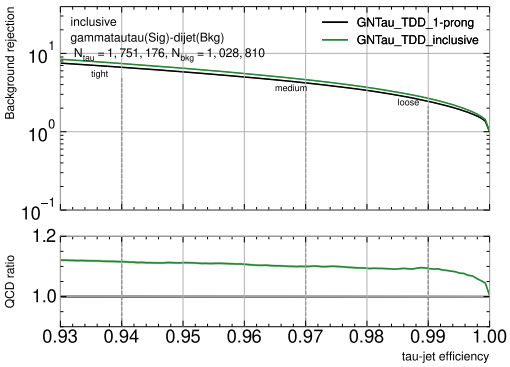
First, the inclusive model is tested on the same datasets as the prong-exclusive models. For instance, the 1p model is evaluated on the 1p dataset and compared with the inclusive model's predictions on that same dataset, and similarly for the other prongs. Second, we evaluate each prong model on its respective prong test dataset and then merge the 0p, 1p, and mp outputs into a single combined dataset. The inclusive model is then evaluated on this **merged prong-exclusive** dataset. This approach allows for a direct comparison of how tau classification performance differs across the inclusive and exclusive architecture.



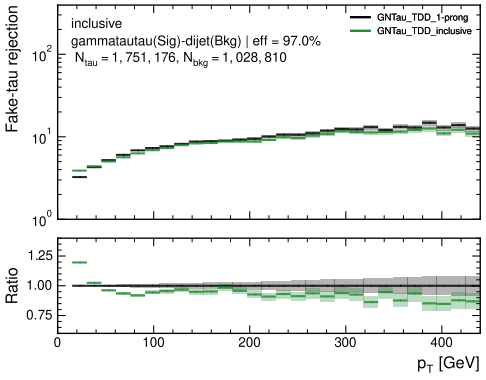

(a): Receiver operating characteristic (ROC) curves for the 0-prong and inclusive cases.



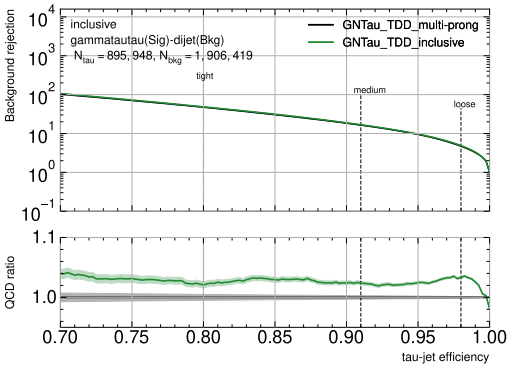

(b): Background rejection as a function of $p_T$ for the 0-prong medium working point.



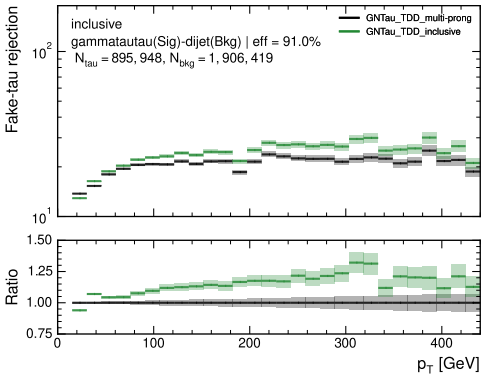

(c): Receiver operating characteristic (ROC) curves for the 1-prong and inclusive cases.




(d): Background rejection as a function of $p_T$ for the 1-prong medium working point.




(e): Receiver operating characteristic (ROC) curves for the multi-prong and inclusive cases.




(f): Background rejection as a function of $p_T$ for the multi-prong medium working point.

Figure 6.3: Receiver operating characteristic (ROC) and background rejection as a function of $p_T$ using $\gamma^* \rightarrow \tau\tau$ as signal and dijet as background. The prong-exclusive model is shown in black, while the inclusive model is shown in green.

Figures 6.3(a), 6.3(c), and 6.3(e) show the performance of the inclusive model evaluated on the individual prong-specific test datasets and compared to the corresponding prong-trained models. In terms of overall ROC performance, the inclusive model achieves at least as good discrimination as the dedicated 0-prong, 1-prong, and multi-prong models on their respective datasets.

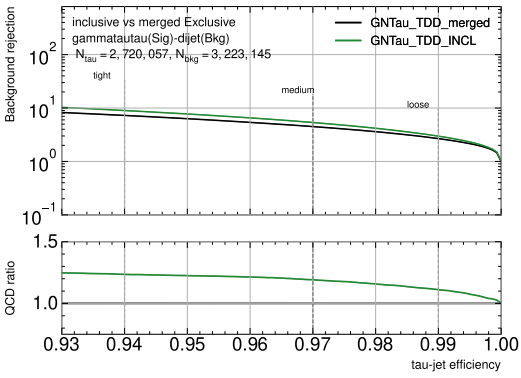
To assess the performance at an operationally relevant working point, the $p_T$-dependent background rejection at the medium working point is shown in Figures 6.3(b), 6.3(d), and 6.3(f). For the 0-prong and multi-prong categories, the inclusive model consistently provides superior background rejection across the full $p_T$ range, indicating improved discrimination performance without relying on explicit prong-specific training.

In contrast, for the 1-prong category, the prong-specific model exhibits improved background rejection in the high-$p_T$ region compared to the inclusive model, as shown in Figure 6.3(d). This indicates that, while the inclusive model performs well overall, the dedicated 1-prong model retains an advantage for high-$p_T$ tau candidates, which is particularly important for the tau trigger, as high-$p_T$ tau leptons are typically associated with the most relevant and potentially rare physics processes.

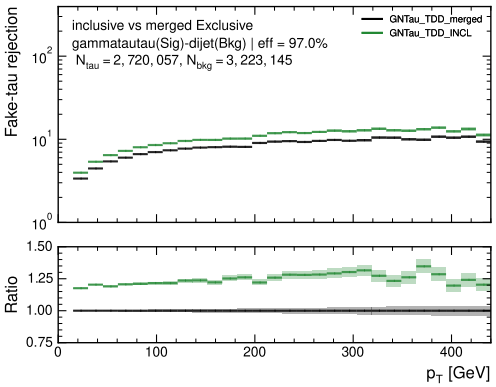
This behaviour is consistent with the $p_T$ distribution observed after the Umami selection shown in Figure 5.5(d), where the majority of background jets populate the low-$p_T$ region. As a result, the inclusive model benefits from improved discrimination at low $p_T$, which dominates the overall performance metrics, while the prong-specific model better exploits prong-dependent features relevant at higher transverse momenta.

Overall, these results demonstrate that the inclusive training approach yields robust and competitive performance across all prong categories, while highlighting a potential benefit of prong-specific training for high-$p_T$ 1-prong tau leptons.

Figure 6.4(a) compares the performance of the inclusive model with that of the merged prong-exclusive models. For the latter, the individually trained 0-prong, 1-prong, and multi-prong models are each evaluated on their respective test sets, and the resulting predictions are combined into a single aggregated dataset. This merged sample reflects the overall performance across all decay modes and enables a direct comparison with the inclusive model, which is trained and evaluated on the full mixed dataset from the outset.




(a): Receiver operating characteristic (ROC) curves for the merged prong exclusive and inclusive cases.




(b): Background rejection as a function of $p_T$ for the merged prong exclusive and inclusive at medium working point.

Figure 6.4: Receiver operating characteristic (ROC) and background rejection as a function of $p_T$ using $\gamma^* \to \tau\tau$ as signal and dijet as background for the merged-exclusive and inclusive models. The merged-exclusive model is shown in black, and the inclusive model in green.

At the medium working point, the inclusive model demonstrates a clear improvement in discrimination performance relative to the merged prong models. In the $\gamma^* \to \tau\tau$ sample, which provides the highest available signal statistics, the inclusive model achieves approximately a 20% higher background rejection at a signal efficiency of 97% compared to the merged prong approach. This indicates a significantly improved separation of genuine $\tau$ candidates from the dijet background at the operating point.

The $p_T$-dependent background rejection at the medium working point, shown in Figure 6.4(b), further confirms this behaviour. Across the full transverse-momentum range, the inclusive model consistently outperforms the combined prong models, no region exhibits superior performance for the prong-based approach. This suggests that the inclusive training strategy effectively leverages information from all prong categories and yields more robust, uniform discrimination performance.

Although the results appear promising, there are still unexplained effects that require further validation to ensure the correctness of the testing procedure. For the prong-exclusive models, as shown in Figure 5.4, the 1-prong configuration in the $\gamma^* \to \tau\tau$ process is the most prevalent. Consequently, a performance improvement of approximately 10% is expected to persist in the second testing scenario, where prong-exclusive classifications are combined. However, when evaluating the merged prong-exclusive test set using the

inclusive model, the observed performance gain increases to approximately 20%. This discrepancy is unexpected and not readily explained by the composition of the test set alone, raising the question of why the inclusive model outperforms the merged prong-exclusive models by such a large margin.

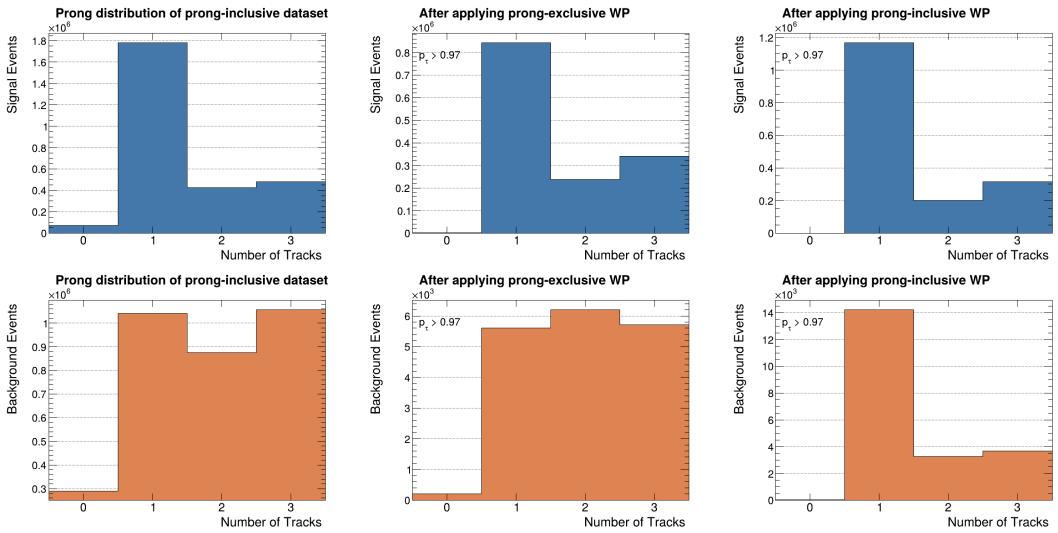



Figure 6.5: Distribution of the number of tracks ($N^{Tracks}$) for signal (blue) and background (orange) events. The left panels show the inclusive dataset without applying the medium working point identification. The middle panels correspond to the merged prong-exclusive models, while the right panels show the results for the inclusive model.

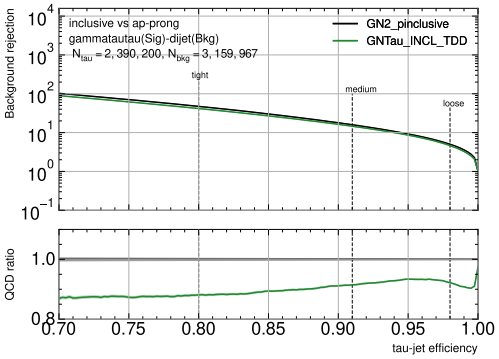
Figure 6.5 provides additional insight into the origin of this behaviour. The left panels show the inclusive dataset before applying the medium working point, while the middle and right panels present the post-selection distributions for the merged prong-exclusive models and the inclusive model, respectively. In the inclusive model, the 1-prong signal component is particularly well preserved, indicating superior discrimination performance in this dominant topology. This enhanced signal retention, however, comes with a higher residual background contribution, suggesting that the inclusive model operates with a slightly higher background misclassification rate in this region.

It is important to note that the background sample is biased towards 1-prong candidates. Consequently, the inclusive model's strong signal preservation in the 1-prong topology is accompanied by a comparatively larger residual background contribution, as even modest background misclassification rates in this dominant prong category translate into a noticeable excess after selection.

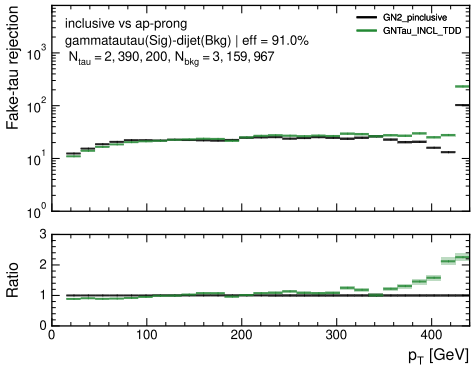
In contrast, the merged prong-exclusive models show reduced background contamination after selection, but at the expense of signal efficiency, especially in the 1-prong category. This behaviour suggests a more conservative classification boundary, resulting in stronger background suppression but weaker overall signal retention.

Therefore, further validation using an independent model is required to ensure that the testing procedure is sound and to determine whether the observed performance gain remains consistent across different modelling approaches. To validate the observed performance gain of the inclusive architecture, we perform an independent cross-check using a model developed by another member of the ATLAS Tau Trigger group [49]. Rather than directly benchmarking absolute performance, this comparison aims to verify whether the relative improvement achieved with the inclusive strategy is reproducible across different model implementations.

The model used by the Tau Trigger group differs from ours in both its hyperparameter choices and overall capacity. Its larger size and more optimized configuration lead to higher absolute performance. However, in the context of this study, the key quantity is the *relative change* in performance observed when moving from prong-exclusive to inclusive inference.




(a): Receiver operating characteristic (ROC) curves for the Tau Trigger group inclusive and TDD inclusive model.

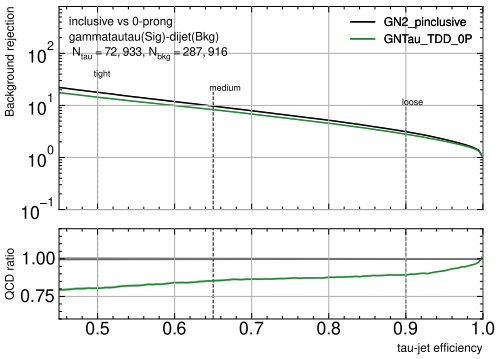



(b): Background rejection as a function of $p_T$ for the Tau Trigger group inclusive and TDD inclusive model at medium working point.

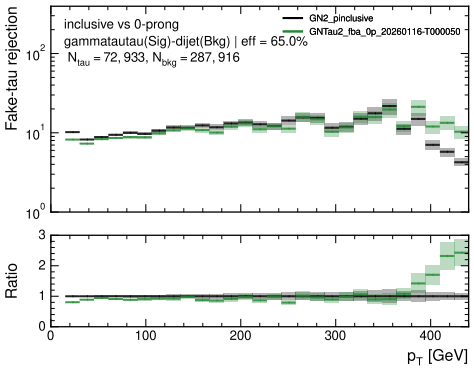
Figure 6.6: Evaluation, using $\gamma^* \to \tau\tau$ as signal and dijet as background. The inclusive model is shown in black, while the Tau Trigger group inclusive model is shown in green.

As expected, Figure 6.6(a) shows that the Tau Trigger group's inclusive model achieves an average performance improvement of approximately 10% compared to the inclusive

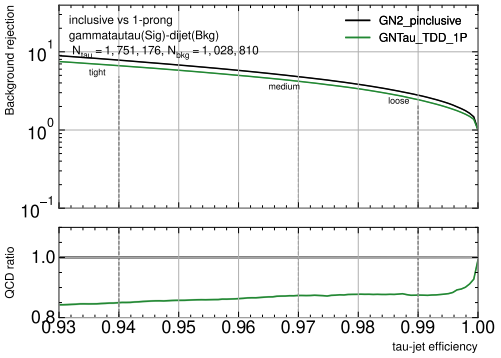
model. Using this model as a baseline, we anticipate an additional 10% improvement with the Tau Trigger group inclusive architecture.



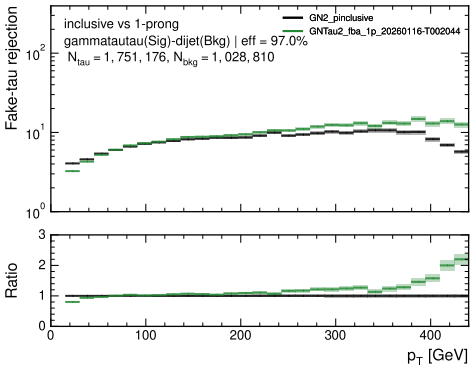

(a): ROC curve of the GNTau 0-prong model compared to the Tau Trigger group inclusive.



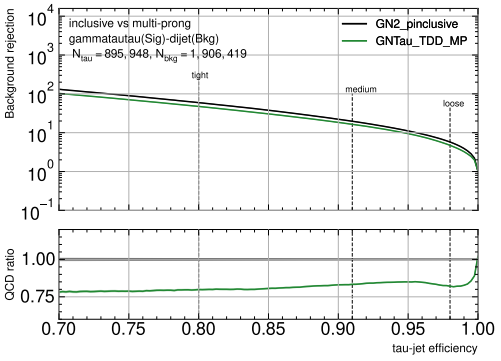

(b): Background rejection as a function of $p_T$ for the GNTau 0-prong model at medium working point.



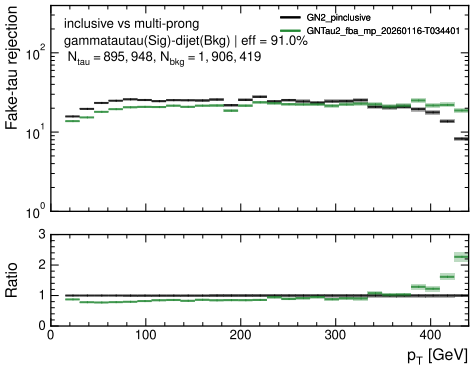

(c): ROC curve of the GNTau 1-prong model compared to the Tau Trigger group inclusive.




(d): Background rejection as a function of $p_T$ for the GNTau 1-prong model at medium working point.




(e): ROC curve of the GNTau multi-prong model compared to the Tau Trigger group inclusive.

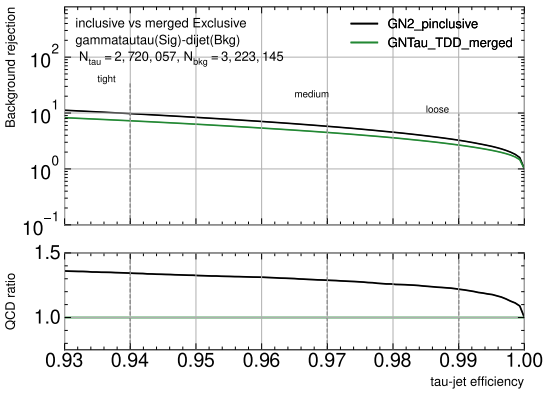



(f): Background rejection as a function of $p_T$ for the GNTau multi-prong model at medium working point.

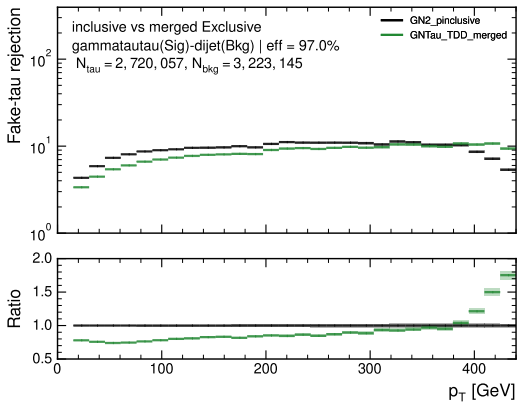
Figure 6.7: Evaluated respectively on the prong exclusive $\gamma^* \to \tau\tau$ signal and dijet background dataset. The Tau Trigger group inclusive model is shown in black, while the prong-exclusive model is shown in green.

We therefore repeat the prong-dependent and merged prong-exclusive procedures using the inclusive model. The prong-separated ROC evaluations for the 0p, 1p, and multi-prong categories (Figures 6.7(a), 6.7(c), and 6.7(e)) demonstrate that the inclusive model consistently outperforms the baseline across the full ROC curve in all prong categories. In particular, for any fixed signal efficiency, the inclusive inference achieves a higher background rejection, indicating a systematic improvement in discrimination performance over the entire operating range.

At the medium working point, shown in Figures 6.7(b), 6.7(d), and 6.7(f), the behaviour is more nuanced. The inclusive model provides improved performance in the low-$p_T$ region across all prong categories, while at high $p_T$ the prong-exclusive model exhibits slightly better background rejection. This indicates that the inclusive training for the Tau Trigger group model particularly enhances discrimination in the low-$p_T$ regime, whereas at high transverse momentum the gains are reversed.




(a): ROC curve of the merged GNTau exclusive models and the Tau Trigger group inclusive model.




(b): Background rejection as a function of $p_T$ for the merged GNTau exclusive and the Tau trigger group inclusive model at medium working point.

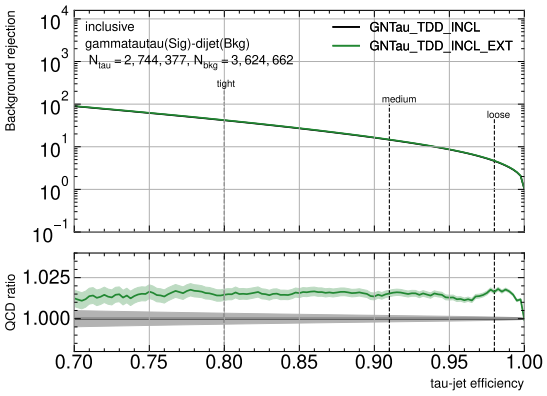
Figure 6.8: Performance and Background rejection as a function of $p_T$ for the merged GNTau exclusive models and the Tau Trigger group inclusive model, evaluated on the inclusive $\gamma^* \to \tau\tau$ signal and dijet background dataset. The Tau Trigger group inclusive model is shown in black, while the merged-exclusive model is shown in green.

The merged prong-exclusive evaluation, shown in Figure 6.8(a), further confirms this behaviour. At the medium working point, the improvement reaches approximately 30%. While this gain is numerically larger than the ∼20% observed in our study, the relative enhancement compared to the respective baseline is similar.

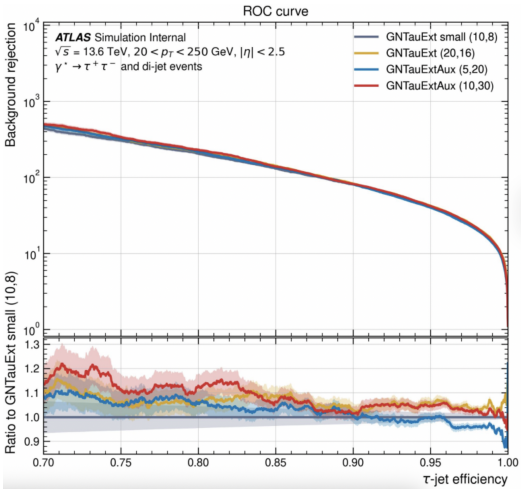
The consistent results across two independently developed models show that the observed performance gain is not tied to a particular architecture, hyperparameter configuration, or training setup. Rather, they indicate that the inclusive architecture offers a systematic advantage over prong-exclusive approaches. The aligned rate of improvement further supports the conclusion that the inclusive strategy yields more robust and generally superior performance.

## 6.4 Extending the number of constituents

The GNTau model architecture restricts the maximum number of track and cluster constituents used for tau identification. In the standard configuration, up to 10 tracks and 8 clusters are provided to the network. An extended configuration is also considered, in which these limits are increased to 20 tracks and 16 clusters, to evaluate whether additional low-level information improves discrimination performance.




(a): Receiver operating characteristic (ROC) curves for the inclusive and extended inclusive model.




(b): Receiver operating characteristic (ROC) curves for the inclusive and extended inclusive model from the tau trigger group.

Figure 6.9: Comparison of extended models on $\gamma^* \rightarrow \tau\tau$ signal and dijet background. The inclusive model is shown in black, and the extended inclusive model in green. On the right, results from the Tau Trigger group are shown: grey for the inclusve model and yellow for the extended-inclusive model.

Figure 6.9(a) shows the ROC curve and the $p_T$-dependent background rejection at the medium working point for the standard and extended input configurations, evaluated on the $\gamma \rightarrow \tau\tau$ signal against dijet background. The extended model exhibits a marginal improvement of around 1% in overall performance relative to the standard configuration.

However, the observed difference is small and falls within the expected statistical fluctuations associated with training variability, and there is no clear or systematic gain across the full efficiency or transverse-momentum ranges.

Before tau candidates are passed to the GNTau model, a track-counting preselection is applied. This selection evaluates the number of well-reconstructed charged tracks associated with each candidate, distinguishing between those in the core and isolation regions. Candidates are rejected if the number of core tracks exceeds three for transverse momenta below 430 GeV, or if more than one isolation track is present for transverse momenta below 200 GeV. Candidates passing either of these criteria are accepted. This procedure preferentially suppresses QCD jet backgrounds while preserving the characteristic one- and three-prong topologies of hadronic tau decays.

As a consequence of this preselection, the effective number of tracks and clusters available per tau candidate is already significantly limited before the data reach the network. In most events, the constituent multiplicities fall well below the imposed maximum input sizes. Therefore, increasing the maximum number of allowed track and cluster inputs from (10,8) to (20,16) rarely yields additional informative inputs to the model and does not substantially increase the available information content.

A similar comparison performed independently by the Tau Trigger group is shown in Figure 6.9(b). In that study, the standard model (GNTauExt small) with input sizes (10,8) is compared to an extended configuration (GNTauExt) using (20,16). Although a modest performance preference for the extended model is reported, corresponding to an approximately 4% improvement, this effect is not reproduced in the present study, despite the use of the same datasets and evaluation procedures.

Overall, these results indicate that, under the current dataset preparation and track setup, counting preselection, which effectively limits the available low-level information, does not yield a significant performance gain by increasing the maximum number of track and cluster inputs. The standard input configuration, therefore, represents an efficient and sufficiently performant choice for trigger-level tau identification within the present framework.

# 7 Conclusion and Outlook

This thesis presents a detailed study of transformer-based graph neural networks for hadronic tau lepton identification at the trigger level in the ATLAS experiment. The analysis focuses on how variations in data preparation, training strategy, and input configuration influence the discrimination between genuine hadronic tau decays and background jets. The investigations focused on simulated $\gamma^* \to \tau\tau$ signal and dijet background samples, since these represents the dominant processes.

A key result of this work is the successful validation of the Training Dataset Dumper (TDD) as a replacement for the legacy THOR-based workflow. By comparing receiver operating characteristic curves and discriminant score distributions across various tau-decay modes, we show that models trained and evaluated with TDD achieve performance similar to that obtained using THOR. In particular, background rejection at the medium working point, crucial for trigger applications, is preserved within a small and acceptable deviation. These results demonstrate that the transition to TDD does not degrade tau-identification performance and support its adoption for future trigger studies.

The second major outcome of this thesis is a systematic comparison between prong-exclusive and prong-inclusive training strategies. For each prong category, the inclusive model achieves discrimination performance that exceeds that of the corresponding prong-specific models. When evaluated on an inclusive dataset that combines all prong categories, the inclusive model delivers a significant improvement in background rejection, reaching approximately 20% at the medium working point. This gain is observed consistently across the full transverse-momentum range and is independently confirmed by a cross-check using a separately developed model from the Tau Trigger group. The agreement between the two studies shows that the observed performance improvement is robust and not tied to a specific architecture or hyperparameter configuration. Overall, these results indicate that inclusive training offers a more effective and robust strategy for trigger-level tau identification than prong-exclusive approaches. In the high-pile-up environment of Run 3 and the future High-Luminosity LHC, such an improvement in

trigger-level background rejection is highly valuable. It directly translates to reduced trigger rates, affording the experiment the flexibility to maintain or lower transverse-momentum thresholds, thereby increasing the acceptance for key physics processes such as $H \to \tau\tau$ measurements and searches for Beyond Standard Model phenomena involving final-state taus.

Finally, we investigated the impact of increasing the number of track and cluster constituents provided to the network. Raising the maximum number of input tracks and clusters does not produce a significant or systematic performance gain: the observed differences are small and compatible with expected training fluctuations. This behaviour is explained by the dataset preselection, which already restricts the number of relevant constituents available to the model. Therefore, the standard input configuration adequately captures the information required for effective tau discrimination and is the preferred choice for efficient trigger deployment.

In summary, this work demonstrates that the inclusive GNTau architecture, when trained using the TDD workflow, provides a robust, high-performance, and operationally well-suited solution for hadronic tau lepton identification at the trigger level. By using an inclusive strategy, it outperforms prong-exclusive models and eliminates the added complexity of maintaining multiple prong-specific networks. From an operational standpoint, this reduction in complexity ensures a more stable and maintainable trigger menu. Ultimately, the deployment of this optimized GNTau architecture ensures that the ATLAS trigger system can efficiently navigate the overwhelming QCD background, safeguarding the physics potential of the tau sector in an increasingly demanding collision environment.

Several avenues for future work arise naturally from the results of this thesis. The inclusive training strategy shows clear performance advantages but implicitly relies on the network's ability to learn prong-dependent features without explicit supervision. A promising extension would be to introduce an auxiliary learning task in which the model is trained not only to distinguish hadronic tau decays from background jets, but also to explicitly classify the tau decay prongness into 0-prong, 1-prong, and multi-prong categories. This multi-task learning setup could encourage the network to learn more structured internal representations and improve both discrimination performance and model interpretability. As the field transitions into the High-Luminosity LHC era, the traditional boundary between real-time trigger algorithms and precision offline reconstruction will increasingly vanish. The robust, inclusive architecture validated in this work serves as a crucial step-

ping stone toward that future, proving that sophisticated deep learning pipelines can be operated efficiently and reliably at the very front lines of data acquisition to safeguard the discovery potential of the energy frontier.

# Bibliography


[1] The ATLAS Collaboration, et al., *Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716.1 pages 1–29 (2012)

[2] L. Evans, P. Bryant, *LHC machine*, JZNST **3(08)**, S08001 (2008)

[3] S. L. Glashow, *Partial-symmetries of weak interactions*, Phys. Rev. Lett. Nucl. Phys. **22(4)**, 579 (1961)

[4] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. Nucl. Phys. **19(21)**, 1264 (1967)

[5] A. Salam, J. C. Ward, *Weak and electromagnetic interactions*, Il Nuovo Cimento (1955-1965) **11(4)**, 568 (1959)

[6] V. C. Rubin, W. K. Ford Jr, *Rotation of the Andromeda nebula from a spectroscopic survey of emission regions*, Astrophysical Journal, vol. 159, p. 379 **159**, 379 (1970)

[7] A. D. Sakharov, *Violation of CP-invariance, C-asymmetry, and baryon asymmetry of the Universe*, in *In The Intermissionsâ/ Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16*, pages 84–87, World Scientific (1998)

[8] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 **S08003** (2008)

[9] M. K. Gaillard, P. D. Grannis, F. J. Sciulli, *The standard model of particle physics*, Rev. Mod. Phys. **71(2)**, S96 (1999)

[10] *Standard Model of Elementary Particles*, https://tikz.net/wp-content/uploads/2026/01/SM_particles-002.png Accessed: 2026-02-04

[11] A. Einstein, *Die Feldgleichungen der Gravitation*, Sitzungsber. K. Preuss. Akad. Wiss. pages 844–847 (1915)

[12] E. Fermi, *Sulla quantizzazione del gas perfetto monoatomico*, Rend. Lincei **145** (1926)

[13] P. A. M. Dirac, *On the theory of quantum mechanics*, Proc. R. Soc. Lond. A **112(762)**, 661 (1926)

[14] H. Fritzsch, M. Gell-Mann, H. Leutwyler, *Advantages of the color octet gluon picture*, Phys. Lett. B **47(4)**, 365 (1973)

[15] M. Gell-Mann, *A schematic model of baryons and mesons*, Phys. Lett. B **8**, 1 (1964)

[16] K. G. Wilson, *Confinement of quarks*, Phys. Rev. D **10(8)**, 2445 (1974)

[17] C.-N. Yang, R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96(1)**, 191 (1954)

[18] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13(16)**, 508 (1964)

[19] H. Yukawa, *On the interaction of elementary particles. I*, Proc. Phys.-Math. Soc. Japan, 3rd Ser. **17**, 48 (1935)

[20] E. Fermi, *High energy nuclear events*, Prog. Theor. Phys. **5(4)**, 570 (1950)

[21] *The Large Hadron Collider*, https://cds.cern.ch/record/2684277 Accessed: 2026-01-26

[22] O. Brüning, H. Burkhardt, S. Myers, *The Large Hadron Collider*, Prog. Part. Nucl. Phys. **67(3)**, 705 (2012)

[23] The CMS Collaboration, et al., *The CMS experiment at the CERN LHC*, JZNST **3**

[24] The ALICE Collaboration, *Performance of the ALICE experiment at the CERN LHC*, Int. J. Mod. Phys. A **29(24)**, 1430044 (2014)

[25] The LHCb Collaboration, *LHCb detector performance*, Int. J. Mod. Phys. A **30(07)**, 1530022 (2015)

[26] *The ATLAS Experiment*, https://cds.cern.ch/record/2837191 Accessed: 2026-01-26

[27] *ATLAS Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, CERN/LHCC/94-43, LHCC/P2, 1994 (unpublished)

[28] The ATLAS Collaboration, et al., *Operation of the ATLAS trigger system in Run 2*, JZNST **15(10)**, P10004 (2020)

[29] *The ATLAS Trigger System*, https://cds.cern.ch/record/2624583 Accessed: 2026-01-26

[30] M. L. Perl, G. Abrams, A. Boyarski, M. Breidenbach, D. Briggs, F. Bulos, W. Chinowsky, J. Dakin, G. Feldman, C. Friedberg, et al., *Evidence for anomalous lepton production in e+- e- annihilation*, Phys. Rev. Lett. **35(22)**, 1489 (1975)

[31] T. Saleem, *Feynman diagram of the tau leptonic decay*, https://www.researchgate.net/figure/a-Feynman-diagram-of-the-tau-leptonic-decay-by-the-emission-of-an-o-shell-W-boso fig85_333942621 Accessed: 2026-01-30

[32] F. Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain.*, Psychol. Rev. **65(6)**, 386 (1958)

[33] G. E. B. Antonio Rafael Sabino Parmezan, Vinícius Mourão Alves de Souza, *Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model*, Information Sciences (2019)

[34] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, USA (2016), cross-Entropy Loss for Classification

[35] A. G. Ivakhnenko, *Polynomial theory of complex systems*, IEEE transactions on Systems, Man, and Cybernetics **(4)**, 364 (2007)

[36] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Learning representations by back-propagating errors*, Nature **323(6088)**, 533 (1986)

[37] M. Reda, *Activation Functions: All You Need To Know*, https://mlarchive.com/machine-learning/activation-functions-all-you-need-to-know/ Accessed: 2026-02-02

[38] P. Ramachandran, B. Zoph, Q. V. Le, *Searching for Activation Functions*, arXiv preprint (2017), arXiv:1710.05941

[39] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, J. Mach. Learn. Res. **15**, 1929 (2014)

[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Åukasz Kaiser, Illia Polosukhin, *Attention is all you need*, Adv. Neural Inf. Process. Syst. 30 (2017)

[41] W. Xie, S. Zhang, L. Wang, K. Yu, W. Li, *Feature selection of microarray data using multidimensional graph neural network and supernode hierarchical clustering*, Artif. Intell. Rev. **57** (2024)

[42] J. Lee, I. Lee, J. Kang, *Self-Attention Graph Pooling*, in K. Chaudhuri, R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3734–3743, PMLR (2019), URL https://proceedings.mlr.press/v97/lee19c.html

[43] J. Barr, D. Biswas, M. Draguet, P. Gadow, E. Haines, O. Karkout, D. Kobylianskii, W. S. Lai, M. Leigh, N. Luongo, I. Oleksiyuk, N. Pond, S. Rettie, A. Vaitkus, S. V. Stroud, J. Wagner, *Salt: Multimodal Multitask Machine Learning for High Energy Physics*, J. Open Source Softw. **10** (2025), URL https://joss.theoj.org/papers/10.21105/joss.07217

[44] T. Sjöstrand, S. Mrenna, P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178(11)**, 852 (2008)

[45] The ATLAS Collaboration, et al., *ATLAS Pythia 8 tunes to 7 TeV data*, Technical report, LHC/ATLAS Experiment (2014), report: ATL-PHYS-PUB-2014-021

[46] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, et al., *Parton distributions with LHC data*, Nucl. Physics B **867(2)**, 244 (2013)

[47] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Methods Phys. Res., Sect. A **462(1-2)**, 152 (2001)

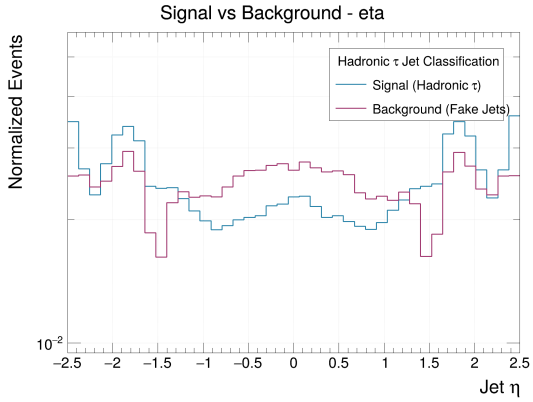
[48] I. Loshchilov, F. Hutter, *Decoupled weight decay regularization*, arXiv preprint arXiv:1711.05101 (2017)

[49] The ATLAS Tau Trigger group, *Advancing GNTau's Trigger Performance* (2025), (only ATLAS internally available), URL https://indico.cern.ch/event/1595957/contributions/6726714/attachments/3149083/5591640/08-10-2025_TGM_Derin.pdf

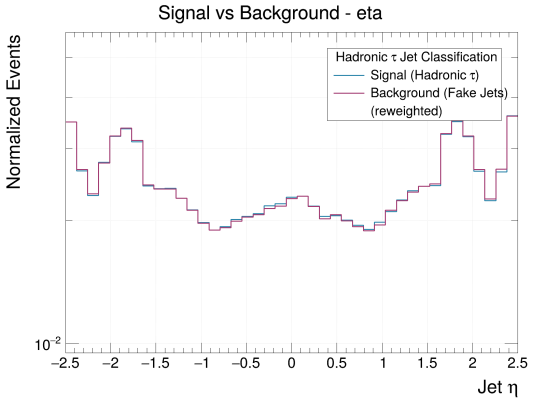
# 8 Appendix

- **Gammatautau:** valid1.801002.Py8EG_A14NNPDF23LO_Gammatautau_MassWeight.recon.AOD.e8514_e8528_s4159_s4114_r15271
- **ttbar:** valid1.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.recon.AOD.e8514_e8528_s4159_s4114_r15271
- **Ztautau (Sherpa):** valid1.700794.Sh_2214_Ztautau_maxHTpTV2_CVetoBVeto.recon.AOD.e8514_e8528_s4159_s4114_r15271
- **Ztautau (Pythia-Photos):** valid1.601191.PhPy8EG_AZNLO_Ztautau.recon.AOD.e8514_e8528_s4159_s4114_r15271
- **Dijet JZ1:** valid1.801166.Py8EG_A14NNPDF23LO_jj_JZ1.recon.AOD.e8514_e8528_s4159_s4114_r15271
- **Dijet JZ2:** valid1.801167.Py8EG_A14NNPDF23LO_jj_JZ2.recon.AOD.e8514_e8528_s4159_s4114_r15271
- **Dijet JZ3:** valid1.801168.Py8EG_A14NNPDF23LO_jj_JZ3.recon.AOD.e8514_e8528_s4159_s4114_r15271
- **Dijet JZ4:** valid1.801169.Py8EG_A14NNPDF23LO_jj_JZ4.recon.AOD.e8514_e8528_s4159_s4114_r15271

Figure 8.1: Monte Carlo simulated ATLAS datasets used in this analysis, showing the generator configuration and reconstruction campaign.

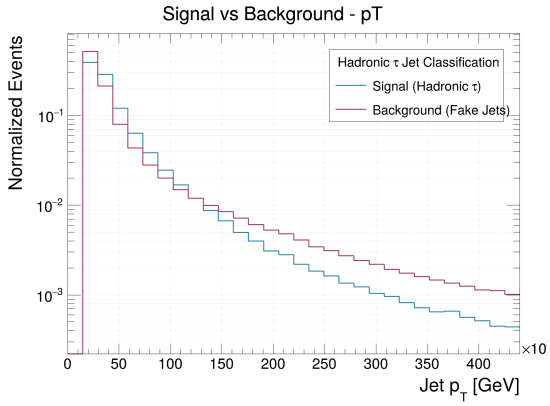



(a): 0-prong $\eta$ distribution before Umami preprocessing.

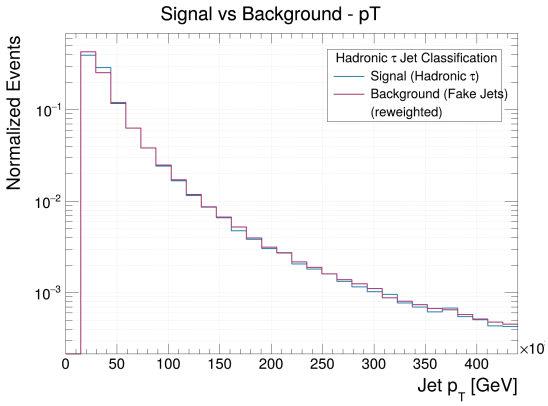



(b): 0-prong $\eta$ distribution after Umami preprocessing.

Figure 8.2: Pseudorapidity ($\eta$) distributions of signal and background events in the 0-prong $\tau_{\text{had}}$ training dataset before (left) and after (right) Umami preprocessing.

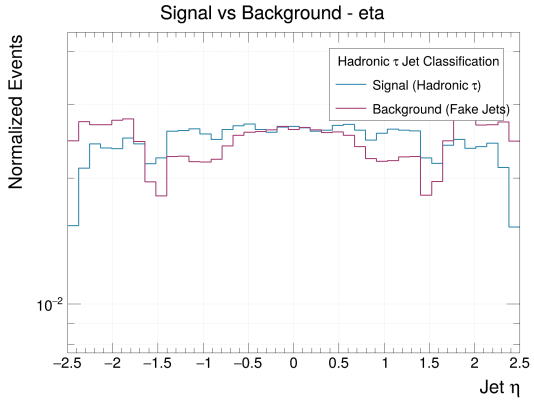



(a): 0-prong $p_T$ distribution before Umami preprocessing.

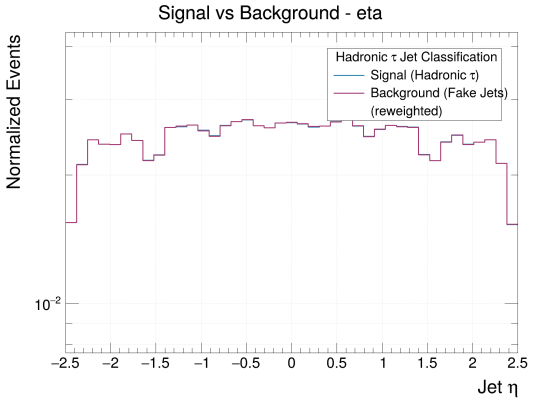



(b): 0-prong $p_T$ distribution after Umami preprocessing.

Figure 8.3: Transverse momentum ($p_T$) distributions of signal and background events in the 0-prong $\tau_{\text{had}}$ training dataset before (left) and after (right) Umami preprocessing.

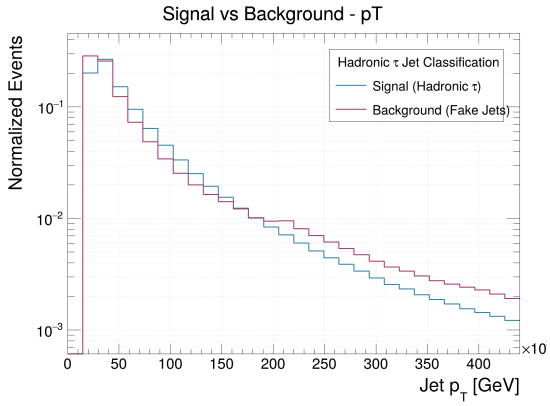



(a): multi-prong $\eta$ distribution before Umami preprocessing.

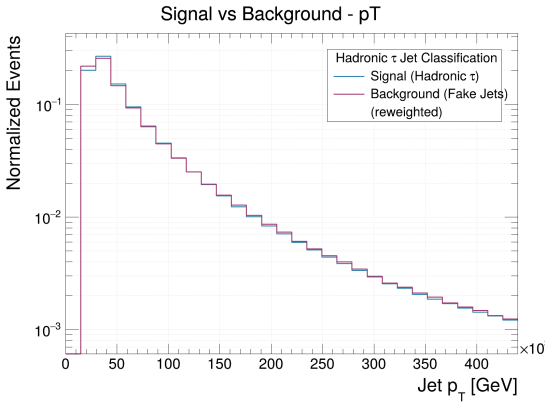



(b): multi-prong $\eta$ distribution after Umami preprocessing.

Figure 8.4: Pseudorapidity ($\eta$) distributions of signal and background events in the multi-prong $\tau_{\text{had}}$ training dataset before (left) and after (right) Umami preprocessing.

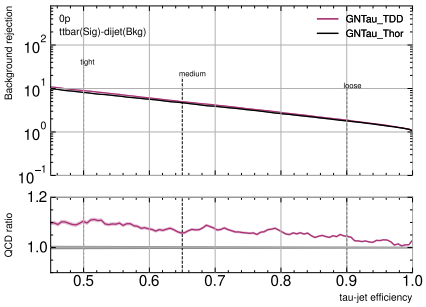



(a): multi-prong $p_T$ distribution before Umami preprocessing.

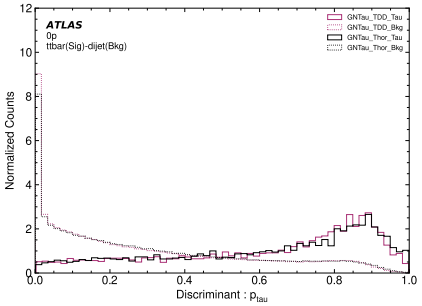



(b): multi-prong $p_T$ distribution after Umami preprocessing.

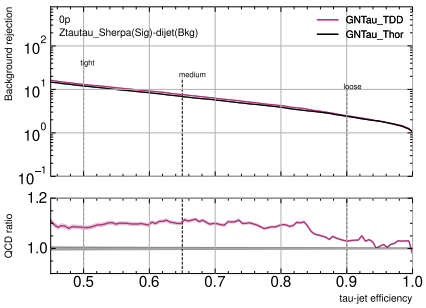
Figure 8.5: Transverse momentum ($p_T$) distributions of signal and background events in the multi-prong $\tau_{\text{had}}$ training dataset before (left) and after (right) Umami preprocessing.



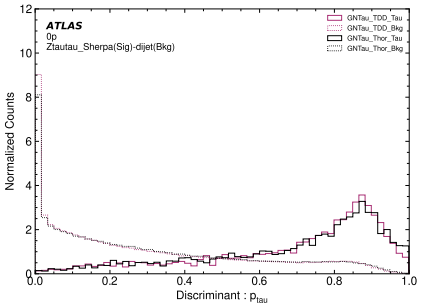

(a): ROC curve of the GNTau using Thor and TDD, tested on $t\bar{t}$-signal and Dijet-background.



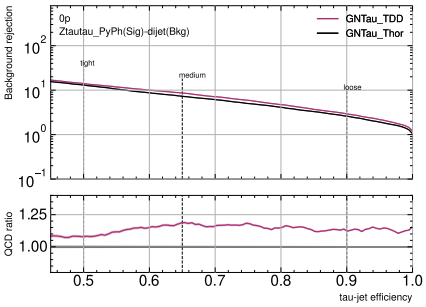

(b): Discriminant distribution of the GNTau using Thor and TDD, tested on $t\bar{t}$-signal and Dijet-background.



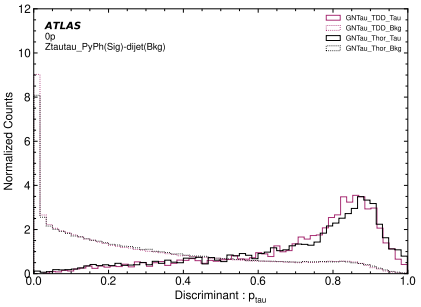

(c): ROC curve of the GNTau using Thor and TDD, tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet-background.




(d): Discriminant distribution of the GNTau using Thor and TDD, tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet-background.

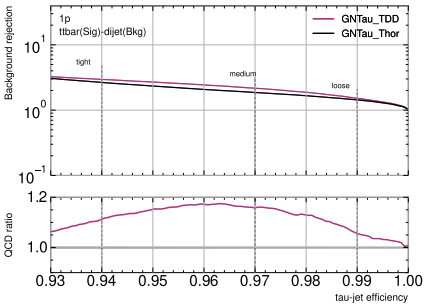



(e): ROC curve of the GNTau using Thor and TDD, tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet-background.

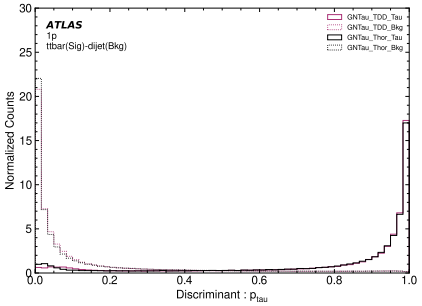



(f): Discriminant distribution of the GNTau using Thor and TDD, tested on $ZZ_{PyPh} \rightarrow \tau\tau$-signal and Dijet-background.

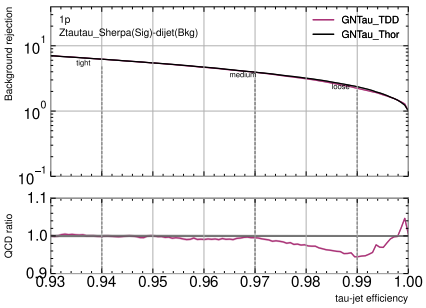
Figure 8.6: Performance and Discriminant score of the GNTau 0-prong, evaluated on different signal and dijet background datasets



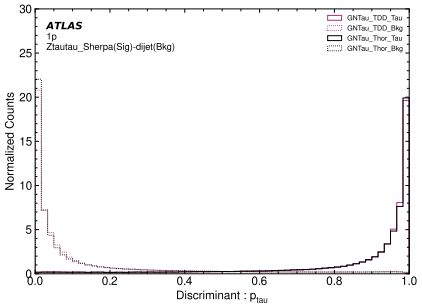

(a): ROC curve of the GNTau using Thor and TDD, tested on $t\bar{t}$-signal and Dijet-background.



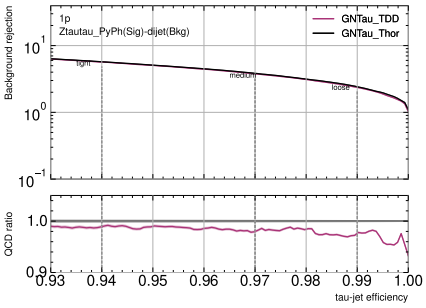

(b): Discriminant distribution of the GNTau using Thor and TDD, tested on $t\bar{t}$-signal and Dijet-background.



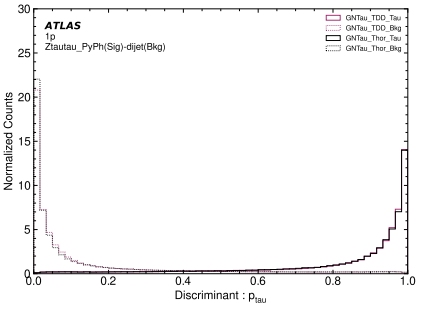

(c): ROC curve of the GNTau using Thor and TDD, tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet-background.




(d): Discriminant distribution of the GNTau using Thor and TDD, tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet-background.

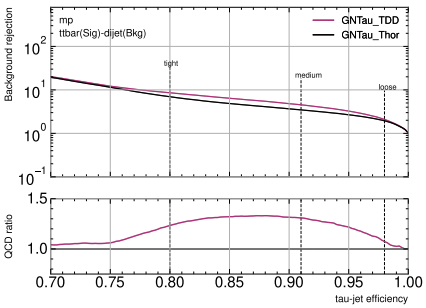



(e): ROC curve of the GNTau using Thor and TDD, tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet-background.

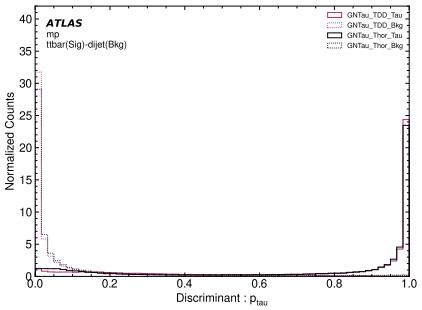



(f): Discriminant distribution of the GNTau using Thor and TDD, tested on $ZZ_{PyPh} \rightarrow \tau\tau$-signal and Dijet-background.

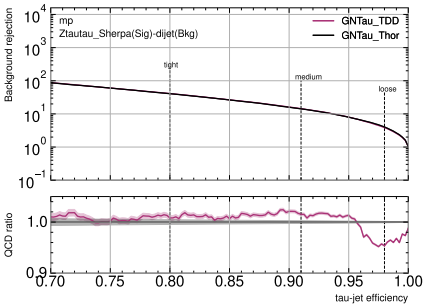
Figure 8.7: Performance and Discriminant score of the GNTau 1-prong, evaluated on different signal and dijet background datasets



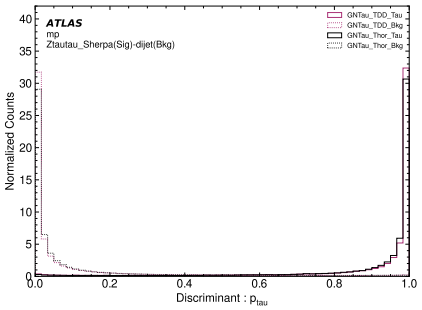

(a): ROC curve of the GNTau using Thor and TDD, tested on $t\bar{t}$-signal and Dijet-background.



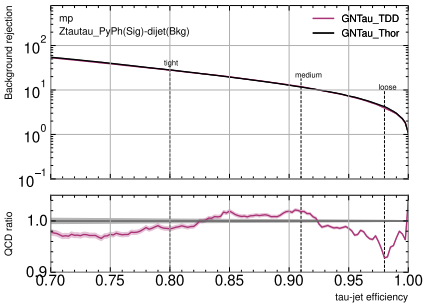

(b): Discriminant distribution of the GNTau using Thor and TDD, tested on $t\bar{t}$-signal and Dijet-background.



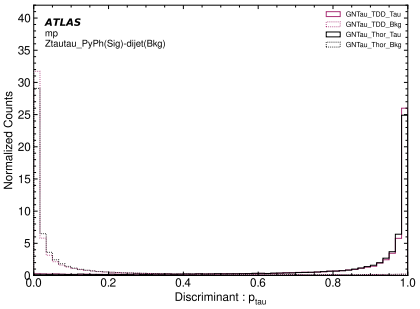

(c): ROC curve of the GNTau using Thor and TDD, tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet-background.




(d): Discriminant distribution of the GNTau using Thor and TDD, tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet-background.

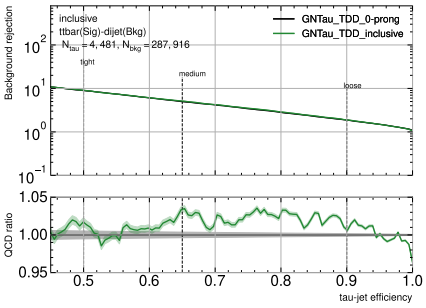



(e): ROC curve of the GNTau using Thor and TDD, tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet-background.

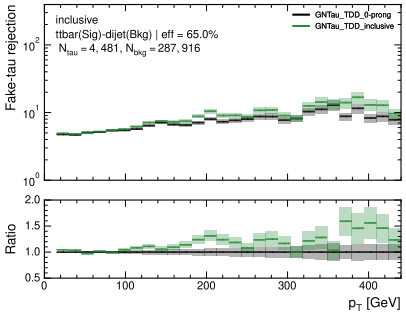



(f): Discriminant distribution of the GNTau using Thor and TDD, tested on $ZZ_{PyPh} \rightarrow \tau\tau$-signal and Dijet-background.

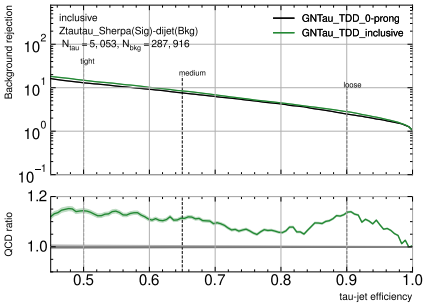
Figure 8.8: Performance and Discriminant score of the GNTau multi-prong, evaluated on different signal and dijet background datasets



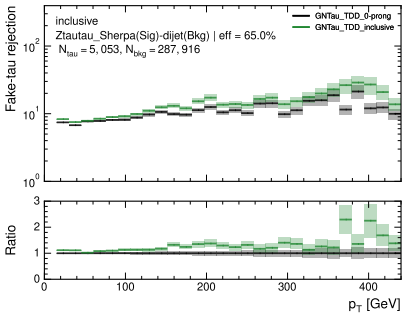

(a): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



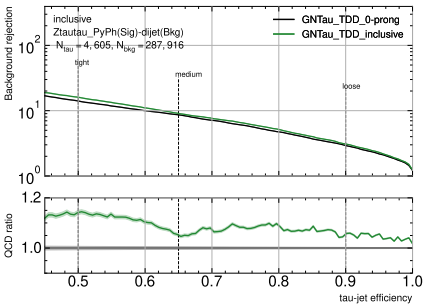

(b): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



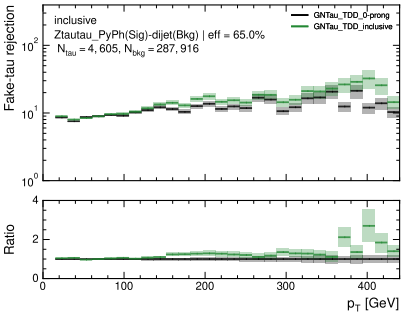

(c): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.




(d): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.

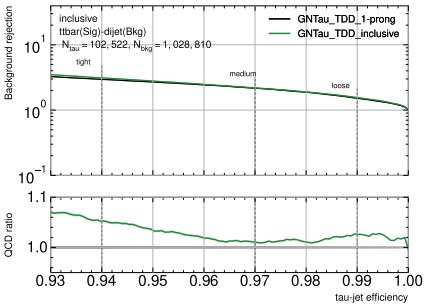



(e): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

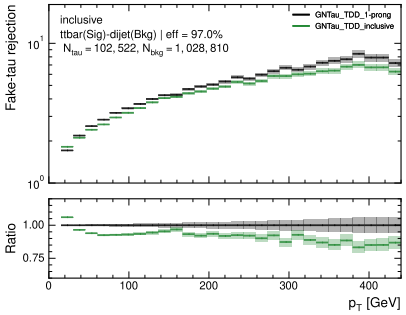



(f): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

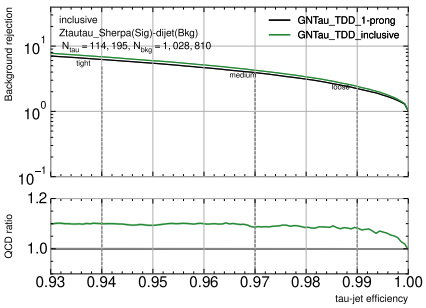
Figure 8.9: Receiver operating characteristic (ROC) and background rejection as a function of $p_T$ at the medium working point for the 0-prong and inclusive model.



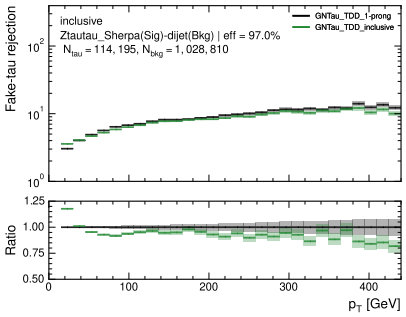

(a): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



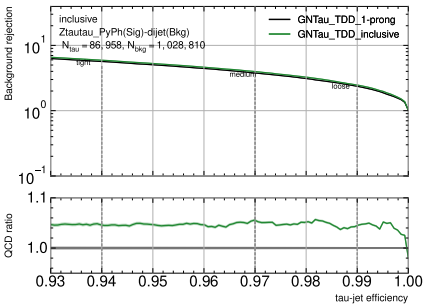

(b): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



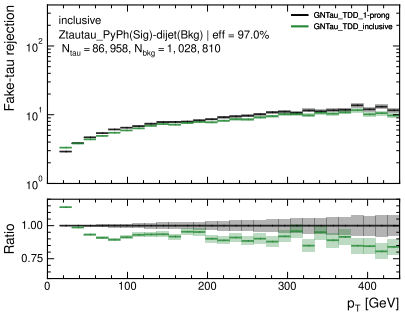

(c): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.




(d): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.

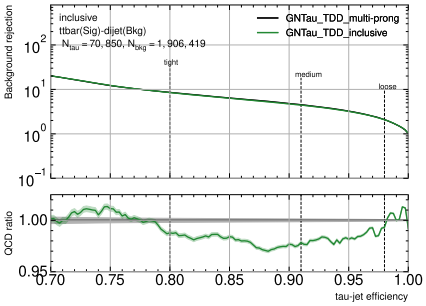



(e): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

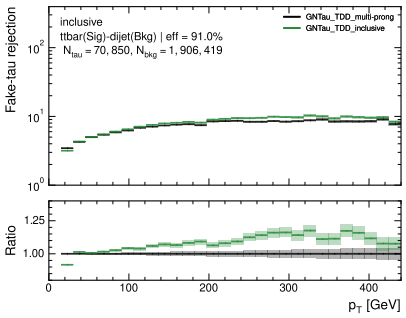



(f): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

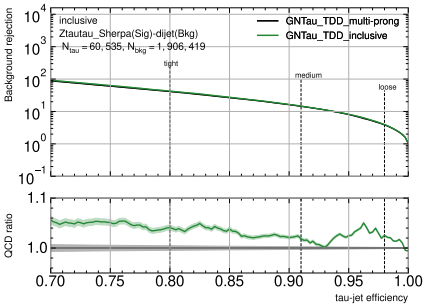
Figure 8.10: Receiver operating characteristic (ROC) and background rejection as a function of $p_T$ at the medium working point for the 1-prong and inclusive model.



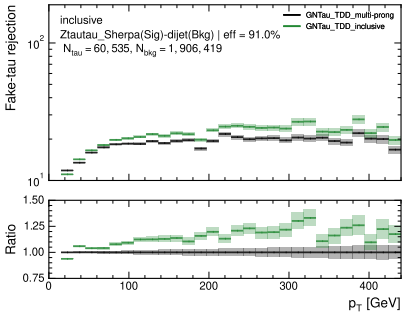

(a): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



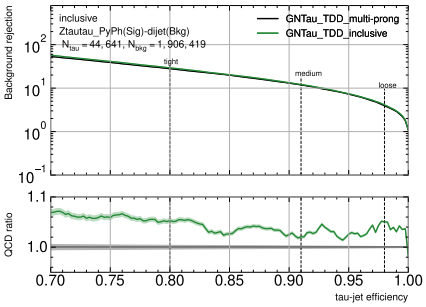

(b): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



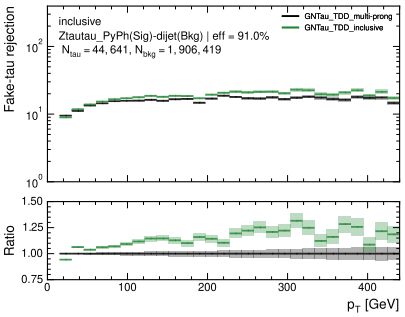

(c): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.




(d): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.

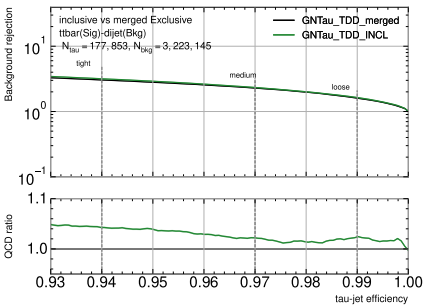



(e): Receiver operating characteristic (ROC) curves of GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

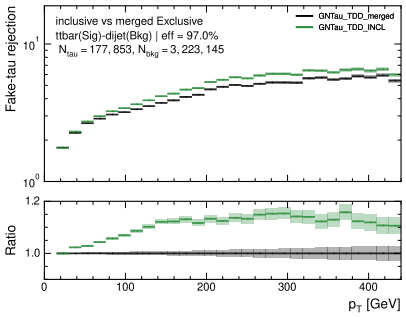



(f): Background rejection as a function of $p_T$ of GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

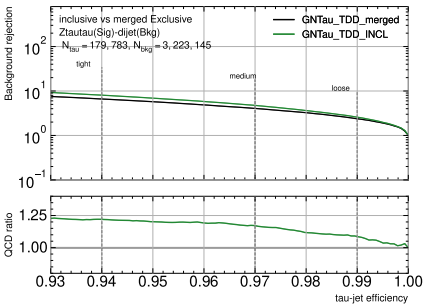
Figure 8.11: Receiver operating characteristic (ROC) and background rejection as a function of $p_T$ at the medium working point for the multi-prong and inclusive model.



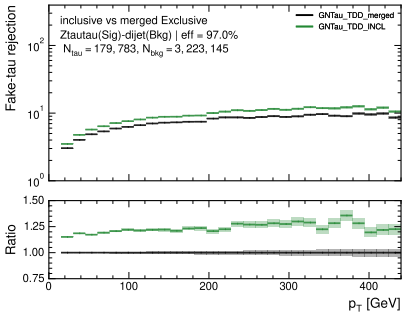

(a): Receiver operating characteristic (ROC) curves of the merged GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



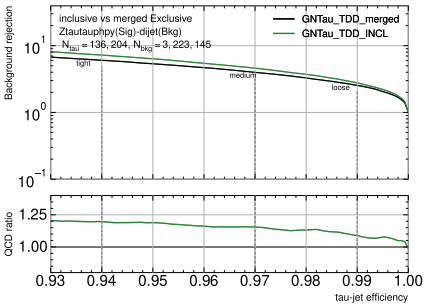

(b): Background rejection as a function of $p_T$ of the merged GNTau prong-exclusive and inclusive tested on $t\bar{t}$-signal and Dijet background dataset.



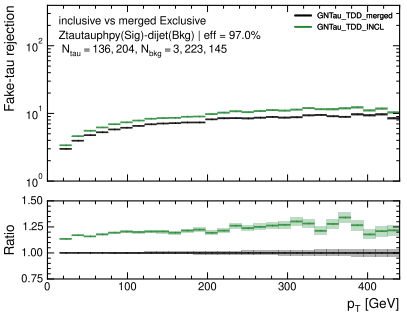

(c): Receiver operating characteristic (ROC) curves of the merged GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.




(d): Background rejection as a function of $p_T$ of the merged GNTau prong-exclusive and inclusive tested on $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet background dataset.

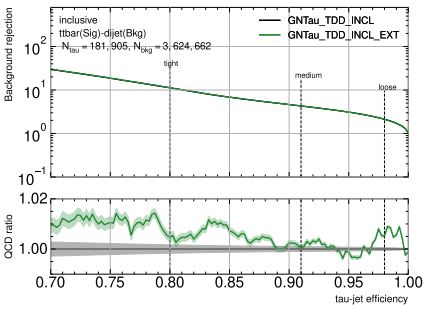



(e): Receiver operating characteristic (ROC) curves of the merged GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

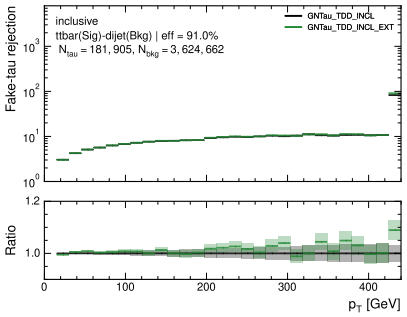



(f): Background rejection as a function of $p_T$ of the merged GNTau prong-exclusive and inclusive tested on $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet background dataset.

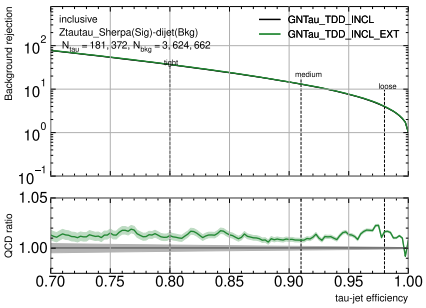
Figure 8.12: Receiver operating characteristic (ROC) and background rejection as a function of $p_T$ at the medium working point for the multi-prong and inclusive model.



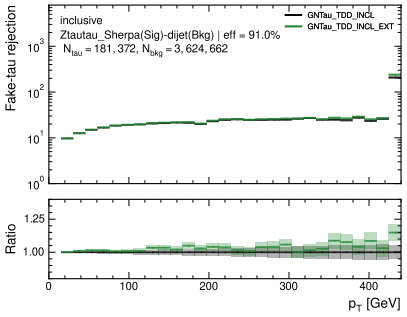

(a): Receiver operating characteristic (ROC) curves for the inclusive and extended inclusive model using $t\bar{t}$ as signal and Dijet as background.



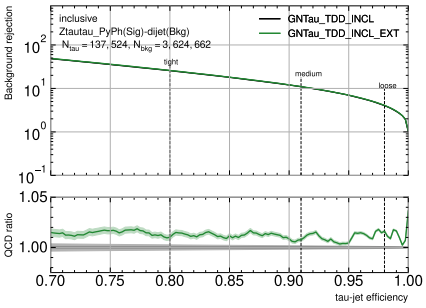

(b): Background rejection as a function of $p_T$ for the inclusive and extended inclusive model at medium working point, using $t\bar{t}$-signal and Dijet-background.



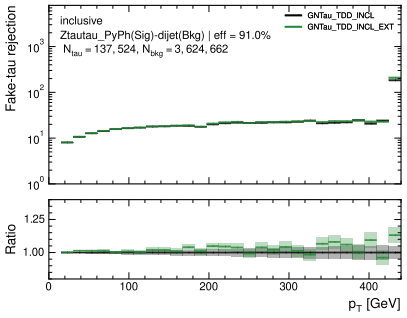

(c): Receiver operating characteristic (ROC) curves for the inclusive and extended inclusive model using $Z_{Sherpa} \rightarrow \tau\tau$ as signal and Dijet as background.




(d): Background rejection as a function of $p_T$ for the inclusive and extended inclusive model at medium working point, using $Z_{Sherpa} \rightarrow \tau\tau$-signal and Dijet-background.




(e): Receiver operating characteristic (ROC) curves for the inclusive and extended inclusive model using $Z_{PyPh} \rightarrow \tau\tau$ as signal and Dijet as background.




(f): Background rejection as a function of $p_T$ for the inclusive and extended inclusive model at medium working point, using $Z_{PyPh} \rightarrow \tau\tau$-signal and Dijet-background.

Figure 8.13: Receiver operating characteristic (ROC) and background rejection as a function of $p_T$ at the medium working point for the inclusive and extended inclusive model.

# Danksagung

First and foremost, I would like to thank my parents. Without their support, I would not have been able to pursue my studies or even be in Göttingen at this moment. Opportunities to engage in academia and to learn are not available to everyone, and my parents gave me this chance. I am deeply grateful for their encouragement and help.

I would also like to express my sincere gratitude to my supervisor, Stan, and the entire AG Lai. My respect for physicists was already immense, but my time in the working group has shown me even more clearly how remarkable they are at understanding the most abstract concepts and finding the proverbial needle in the haystack. Stan's passion for particle physics, coupled with his kindness and mentorship, inspired me to dive deeper into this field.

I am particularly grateful to Athul for patiently answering all of my questions, even when they were repeated multiple times, and for providing constant, timely advice throughout this entire journey.

# Declaration on the use of ChatGPT and comparable tools in the context of examinations

In this paper, I have used ChatGPT or another AI as follows.:
[ ] not at all
[ ] during brainstorming
[x] when creating the outline
[ ] to write individual passages, altogether to the extent of ...% of the entire text
[ ] for the development of software source texts
[ ] for optimising or restructuring software source texts
[x] for proofreading or optimising
[x] further, namely: sentence structure and spell correction
I hereby declare that I have stated all uses completely. Missing or incorrect information will be considered as an attempt to cheat.

**Erklärung** nach §13(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäÃ aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüber hinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den Februar 19, 2026

(Fabio Maxim Potrykus)