GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Fakultät für Physik

Bachelor's Thesis

Identifikation hadronischer Tau-Zerfälle mittels Transformer-basierter neuronaler Netze in Proton-Proton-Kollisionen am ATLAS-Detektor bei $\sqrt{s} = 13.6$ TeV.

Studies of identification of tau leptons decaying into hadrons and a neutrino with modern transformer-based neural network architectures using data recorded in proton-proton collisions with the ATLAS detector at $\sqrt{s} = 13.6$ TeV.

prepared by

Jerrit Rouven Hamester

from Reinbek

at the II. Physikalischen Institut

Thesis number:	II.Physik-UniGö-BSc-2026/08
Thesis period:	13th April 2026 until 17th July 2026
First referee:	Prof. Dr. Arnulf Quadt
Second referee:	Prof. Dr. Jörn Große-Knetter

Zusammenfassung

Die Identifikation hadronisch zerfallender Tau-Leptonen bei ATLAS wird durch zwei Untergrundprozesse erschwert: QCD-Jets, die die schmale Jetstruktur eines Tau-Zerfalls nachahmen können, und Elektronen, deren elektromagnetische Schauer insbesondere 1-Prong-Zerfällen ähneln. Bisher werden diese getrennt unterdrückt: durch das Graph-Neuronale-Netzwerk GNTau für Jets und ein unabhängiges rekurrentes neuronales Netzwerk (RNN) für Elektronen.

Diese Arbeit stellt GNTau_eVeto vor, ein einzelnes Graph-Neuronales-Netzwerk, das QCD-Jets und Elektronen gleichzeitig unterdrückt und beide Klassifikatoren durch ein vereinheitlichtes Mehrklassen-Modell ersetzt. Das Netzwerk wird auf simulierten ATLAS-Ereignissen mit Spur-, Cluster- und Particle-Flow-Informationen des Tau-Kandidaten trainiert.

Bei fester Signaleffizienz erreicht GNTau_eVeto eine höhere Unterdrückung von QCD-Jets und Elektronen als die Kombination der beiden separaten RNN-Klassifikatoren. Gegenüber dem spezialisierten GNTau ist die Jet-Unterdrückung leicht geringer – ein Kompromiss zwischen Einfachheit und Spezialisierung. GNTau_eVeto kann so den Tau-Identifikationsworkflow vereinfachen, ohne gegenüber dem heutigen Ansatz an Leistung einzubüßen; Strategien zur Schließung der verbleibenden Lücke werden in Kapitel 7 diskutiert.

Abstract

Identifying hadronically decaying tau leptons at ATLAS is complicated by two background sources: QCD jets, which can mimic the narrow jet structure of a tau decay, and electrons, whose electromagnetic showers particularly resemble one-prong decays. These are currently rejected separately: by the graph neural network GNTau for jets, and an independent recurrent neural network (RNN) for electrons.

This thesis presents GNTau_eVeto, a single graph neural network trained to simultaneously reject both QCD jets and electrons, replacing the two separate classifiers with a unified multi-class model. The network is trained on simulated ATLAS events using track, cluster, and particle-flow information of the tau candidate.

At fixed signal efficiency, GNTau_eVeto achieves higher rejection of both QCD jets and electrons than the combination of the two separate RNN-based classifiers. Compared to the specialized GNTau, jet rejection is slightly lower – a trade-off between simplicity and specialization. GNTau_eVeto can thus simplify the tau-identification workflow without a loss in performance relative to today's approach; strategies for closing the remaining gap are discussed in Chapter 7.

Contents

1. Introduction

As the Large Hadron Collider (LHC) continues to deliver increasingly large datasets, the identification of physics objects requires ever more sophisticated reconstruction and classification techniques. One particularly challenging task is the identification of hadronically decaying tau leptons. Owing to its large mass, the tau lepton is the only lepton capable of decaying into hadrons, with approximately 65% of all tau decays proceeding through hadronic final states [1]. These decays typically produce one or three charged hadrons accompanied by neutral pions, resulting in narrow jets with low track multiplicity.

The experimental signature of hadronic tau decays closely resembles that of certain background processes. In particular, fluctuations in QCD jets can produce similarly narrow jets, while electrons generate compact electromagnetic showers that can mimic one-prong tau decays [2]. Because a genuine electron leaves only a single reconstructed track, this confusion is essentially confined to the 1-prong category; electrons are not expected to fake genuine 3-prong tau decays, which is reflected in the very small electron-fake samples available for the 3-prong studies of Chapter 6. Efficiently distinguishing genuine hadronic tau leptons from these backgrounds is therefore essential for many ATLAS analyses, including Higgs boson measurements and searches for physics beyond the Standard Model.

During Run 2 of the LHC, ATLAS employed two dedicated recurrent neural networks (RNN s) for tau identification: one to distinguish hadronically decaying tau leptons from QCD jets [2], and a second, independent network to reject electrons that mimic hadronic tau decays [2]. With the start of Run 3, the jet-discrimination algorithm was replaced by the Graph Neural Tau (GNTau) algorithm, which uses graph neural networks (GNNs) to exploit the relationships between detector constituents and improve the separation of hadronic tau decays from QCD jets [3]. Electron rejection, however, continued to rely on the separate RNN-based classifier.

The objective of this thesis is to extend the GNTau framework by developing and training a graph neural network, GNTau_eVeto, capable of simultaneously distinguishing hadronically decaying tau leptons from both QCD jets and electrons. By combining jet and electron rejection within a single GNN architecture, this work investigates whether a unified model can provide competitive or improved performance while simplifying the tau-identification workflow.

2. The Standard Model of Particle Physics

The Standard Model (SM) of particle physics is the theoretical framework that describes the fundamental particles of matter and the interactions between them [4–7]. Developed throughout the second half of the twentieth century, it combines the electromagnetic, weak, and strong interactions into a single quantum field theory and has been remarkably successful in explaining a wide range of experimental observations. Predictions of the Standard Model have been confirmed with extraordinary precision, making it one of the most thoroughly tested theories in modern physics [1]. The Standard Model describes all known elementary particles and their interactions, with the exception of gravity. It accurately predicts the production and decay of particles over a broad range of energies and culminated in the discovery of the Higgs boson at the Large Hadron Collider in 2012 [8], confirming the mechanism responsible for the generation of particle masses.

Despite its success, the Standard Model is known to be incomplete. It does not incorporate gravity, provide an explanation for dark matter or dark energy, or account for the observed matter-antimatter asymmetry in the Universe. Nevertheless, it forms the theoretical foundation of essentially all analyses performed at the LHC, including the studies presented in this thesis.

2.1. Fundamental Particles

The Standard Model classifies elementary particles into two groups: **fermions**, which constitute matter, and **bosons**, which mediate the fundamental interactions [7].

Fermions are spin-$\frac{1}{2}$ particles and are organized into three generations. Each generation contains two quarks and two leptons. Quarks carry colour charge and therefore participate in the strong interaction, while leptons do not. The first generation consists of the up and down quarks together with the electron and the electron neutrino, which form the stable matter observed in everyday life. The second and third generations contain heavier particles with identical quantum numbers that decay into lighter generations through the

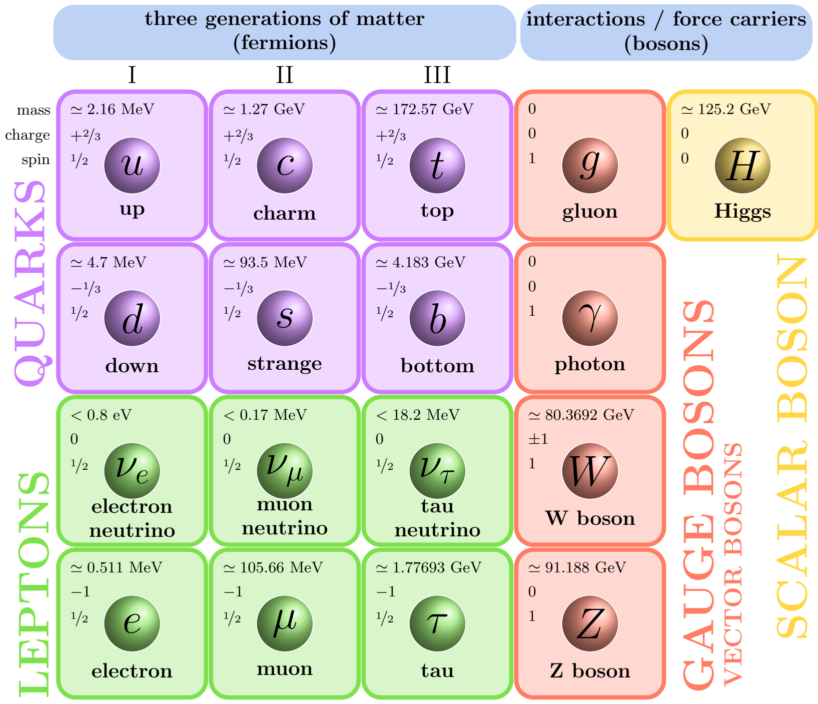


Figure 2.1.: Overview of the Standard Model of Particle Physics

weak interaction. The tau lepton, which is the focus of this thesis, belongs to the third generation of leptons. Figure 2.1 gives an overview of the Standard Model including the particles electric charge and approximate mass [1].

Each fermion is characterized by a set of conserved quantum numbers in addition to its electric charge. Quarks carry one of three **colour charges** (conventionally labelled red, green, and blue) under the $SU(3)_C$ gauge symmetry of the strong interaction, while leptons are colour-neutral **colour singlets** and do not participate in strong-interaction processes. Left-handed fermions additionally carry weak isospin under the $SU(2)_L$ symmetry of the weak interaction, while right-handed fermions do not, a distinction that is responsible for the maximal parity violation observed in weak decays. The full Standard Model gauge symmetry is the direct product $SU(3)_C \times SU(2)_L \times U(1)_Y$, where the strong interaction is mediated by the $SU(3)_C$ colour symmetry and the electromagnetic and weak interactions emerge from the spontaneously broken $SU(2)_L \times U(1)_Y$ electroweak symmetry [4–6].

The force-carrying particles are known as **gauge bosons** and are spin-1 particles arising from the local gauge symmetries of the theory. The photon (γ) mediates the electromagnetic interaction, the $W^{\pm}$ and Z^0 bosons mediate the weak interaction, and eight gluons (g) mediate the strong interaction between quarks. Unlike the photon, gluons themselves carry colour charge, which allows them to self-interact and is ultimately responsible for the confinement of quarks inside colour-neutral hadrons [7].

In addition to the gauge bosons, the Standard Model contains a single scalar (spin-0) particle, the **Higgs boson**. The Higgs field is associated with a spontaneously broken symmetry of the electroweak sector: the field acquires a non-zero vacuum expectation value even in the lowest-energy state of the theory, and this spontaneous symmetry breaking is what allows the $W^{\pm}$ and Z^0 bosons, as well as the charged leptons and quarks, to acquire mass through their coupling to the Higgs field, while leaving the photon massless [9–11]. Without this mechanism, gauge invariance would forbid explicit mass terms for the gauge bosons and fermions in the Standard Model Lagrangian. The Higgs boson itself was discovered by the ATLAS and CMS collaborations at the LHC in 2012 [8], providing direct experimental confirmation of the electroweak symmetry-breaking mechanism.

2.2. Fundamental Interactions

The Standard Model describes three of the four known fundamental interactions: the electromagnetic, weak, and strong interactions; gravity is not incorporated and is negligible at the energy and distance scales relevant to particle physics [7]. The relative strength, range, and mediating boson of each interaction determine its phenomenological role at collider experiments and are summarised in Table 2.1.

Interaction	Mediator	Mediator mass	Relative strength
Strong	gluon (g)	0	1
Electromagnetic	photon (γ)	0	$\sim 10^{-2}$
Weak	$W^{\pm}$, Z^0	$\sim$ 80–91 GeV	$\sim 10^{-6}$

Table 2.1.: Approximate relative strengths and mediator properties of the three interactions described by the Standard Model, evaluated at typical hadronic energy scales [1]. The strength of the weak interaction appears suppressed at low energies because of the large mass of its mediating bosons, while at energies far above m_W it becomes comparable to the electromagnetic interaction, reflecting the underlying electroweak unification.

The electromagnetic interaction acts between electrically charged particles and is mediated by the massless, chargeless photon [6]. Because the photon is massless, the electromagnetic interaction has infinite range, following an inverse-square law familiar from classical electrodynamics.

The strong interaction, mediated by gluons, binds quarks together to form colour-neutral hadrons such as protons, neutrons, and mesons [12]. Its coupling strength grows at large distances (or, equivalently, low energies), a property known as **confinement** that prevents free quarks or gluons from being observed in isolation; at short distances, conversely, the coupling becomes weak, a behaviour referred to as **asymptotic freedom**. This is the reason quarks produced in high-energy collisions cannot be observed directly, but instead fragment and hadronize into the collimated sprays of hadrons known as **jets**, which are extensively discussed in later chapters of this thesis.

The weak interaction is unique in that it can change the flavour of quarks and leptons, allowing unstable particles to decay [6]. It is mediated by the massive $W^{\pm}$ and Z^0 bosons, and its short range of approximately 10^{-18} m is a direct consequence of these large mediator masses. Unlike the electromagnetic and strong interactions, the weak interaction couples only to left-handed fermions (and right-handed antifermions), leading to the maximal parity violation famously observed in weak decays. The weak interaction is responsible for the decay of all unstable fermions, including the tau lepton, which is discussed in detail in Section 2.3.

2.3. The Tau Lepton

The tau lepton $\tau^{\pm}$ is the heaviest of the three charged leptons in the Standard Model, with a mass of $m_\tau = 1.777\,\text{GeV}$ [1]. Owing to its comparatively large mass, it is the only lepton capable of decaying into hadrons via the weak interaction. Its lifetime is approximately 2.9×10^{-13} s, corresponding to a decay length of only a few millimetres at typical LHC energies [1]. Although the tau lepton decays before reaching the detector, this small displacement from the primary interaction vertex can be reconstructed by the Inner Detector and provides valuable information for its identification.

2.3.1. Tau Decays

The tau lepton decays exclusively through the weak interaction and can be classified into two main categories. Approximately **35%** of all decays are **leptonic**, producing either an electron or a muon together with the corresponding neutrinos,

$$\tau^- \to e^- \bar{\nu}_e \nu_\tau, \qquad \tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau\,,$$

while the remaining **65%** are **hadronic** decays [1]. In hadronic decays, the tau lepton produces a tau neutrino together with one or more hadrons, predominantly charged and neutral pions. Since the tau lepton carries a charge of ± 1, charge conservation requires the visible hadronic decay products to contain an odd number of charged particles. Consequently, hadronic tau-lepton decays are dominated by **1-prong** and **3-prong** topologies, corresponding to one or three reconstructed charged particle tracks, respectively. Ap-

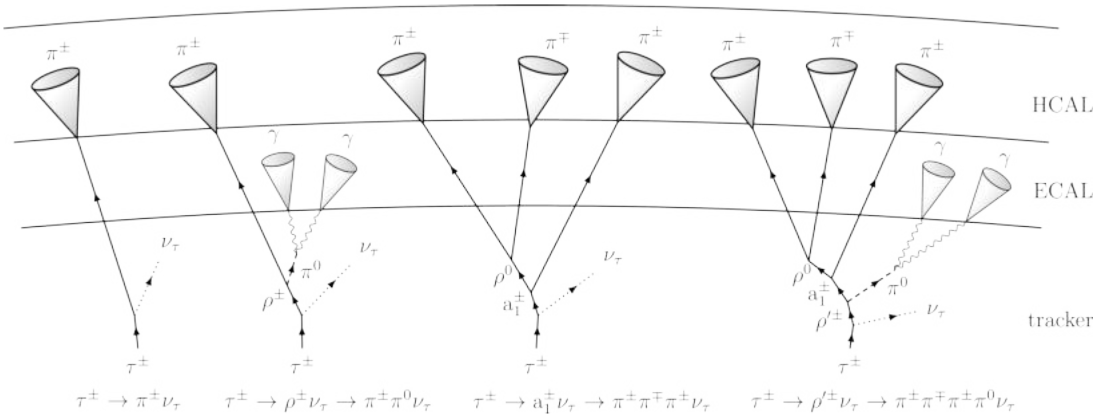


Figure 2.2.: Illustration of the characteristic detector signatures of different jet types, showing the reconstructed tracks and calorimeter clusters for hadronic tau-lepton decays, electrons, and QCD jets [13].

proximately **77%** of hadronic decays are 1-prong decays, while about **23%** are 3-prong decays [1].

The most common 1-prong decays proceed either directly via $\tau^\pm \to \pi^\pm \nu_\tau$, or through intermediate resonances such as the ρ meson, $\tau^\pm \to \rho^\pm \nu_\tau \to \pi^\pm \pi^0 \nu_\tau$. The neutral pion decays almost instantaneously into two photons, $\pi^0 \to \gamma\gamma$, which deposit their energy in the electromagnetic calorimeter. The dominant 3-prong decay proceeds via the a_1 resonance, $\tau^\pm \to a_1^\pm \nu_\tau \to \pi^\pm \pi^\mp \pi^\pm \nu_\tau$, and may likewise be accompanied by one or more neutral pions.

The characteristic detector signature of a hadronically decaying tau lepton therefore consists of one or three charged tracks, a narrow calorimeter energy deposit, and missing energy carried away by the undetected tau neutrino. While these features distinguish tau leptons from many background processes, they also make hadronic tau-lepton decays difficult to identify. In particular, 1-prong decays containing neutral pions closely resemble the detector signature of an electron, whereas both 1-prong and 3-prong decays can mimic the narrow jets produced by quarks and gluons. This effect is visualised in Figure 2.2. Efficient discrimination between hadronic tau leptons, electrons, and QCD jets is therefore one of the central challenges in ATLAS event reconstruction and the primary motivation for the machine learning model developed in this thesis.

Because a genuine electron produces only a single charged-particle track in the Inner Detector, its detector signature can only be confused with the 1-prong hadronic tau topology. Electrons are therefore **not expected** to fake genuine 3-prong tau decays: a 3-prong candidate would require an electron to be accompanied by two additional, unrelated tracks accidentally falling within the tau candidate cone, which is a rare combinatorial occurrence rather than a genuine physics background. This asymmetry between the 1-prong and 3-prong electron-fake rates is a recurring theme in this thesis and directly explains why the electron-veto working points and the available electron-fake test samples differ so strongly between the two prong categories in the performance studies of Chapter 6.

2.4. Physics Motivation for Tau-Lepton Identification

The tau lepton plays a central role in the physics programme of the Large Hadron Collider. As the heaviest lepton in the Standard Model, it provides a unique probe of both Standard Model processes and potential new physics. Precise reconstruction and identification of tau leptons are therefore essential for a wide range of measurements and searches performed by the ATLAS experiment.

One of the most important applications is the study of the Higgs boson. Following its discovery in 2012, measuring its properties and couplings has become a major objective of the LHC physics programme. The decay $H \to \tau^+\tau^-$, with a branching ratio of approximately 6.3% [1], provides direct access to the Yukawa coupling of the Higgs boson to charged leptons and therefore constitutes an important test of the Higgs mechanism. In addition, tau leptons play a significant role in precision measurements of electroweak processes, such as $Z \to \tau^+\tau^-$ and $W \to \tau\nu$, which are used to validate Standard Model predictions with increasing precision.

Beyond the Standard Model, many proposed theories predict an enhanced production of tau leptons in the final state. These include searches for additional Higgs bosons, supersymmetric particles, and other heavy resonances that preferentially couple to third-generation fermions. Consequently, efficient tau-lepton reconstruction is crucial for maximizing the discovery potential of the ATLAS experiment.

As discussed in Section 2.3.1, the dominant hadronic decay channels offer the largest statistical sensitivity for these measurements, but their similarity to QCD-jet and electron signatures makes background rejection experimentally challenging. The development of increasingly sophisticated reconstruction and identification algorithms is therefore essential to achieve high signal efficiency while maintaining strong background rejection; this is the central topic of Chapters 4–6 of this thesis.

3. Experimental Setup

The fundamental goal of particle physics is to understand the basic constituents of matter and the forces governing their interactions. While many particles exist only under extreme conditions, such as those present shortly after the Big Bang, they can be recreated in laboratory environments by accelerating particles to very high energies and bringing them into collision. According to the principle of mass-energy equivalence, the kinetic energy of the colliding particles can be converted into the production of new, more massive particles, allowing physicists to study phenomena that cannot be observed under ordinary conditions [14].

Particle accelerators therefore serve as powerful tools for probing the structure of matter at increasingly smaller distance scales. Higher collision energies correspond to shorter wavelengths, enabling the investigation of the substructure of elementary particles and the search for previously unknown particles and interactions. In addition to precision measurements of the Standard Model, modern accelerators provide the opportunity to explore physics beyond the Standard Model, including the search for new particles and rare processes.

3.1. Large Hadron Collider

The **Large Hadron Collider (LHC)** is the world's largest and highest-energy particle accelerator. It is located at the European Organization for Nuclear Research (CERN) on the French-Swiss border near Geneva. The accelerator is installed in a circular tunnel with a circumference of approximately 27 km, originally constructed for the Large Electron-Positron Collider (LEP), and lies between 50 m and 175 m below ground.

The LHC primarily accelerates two counter-rotating beams of protons, although heavy-ion collisions are also performed during dedicated running periods. The particles are accelerated in several stages before being injected into the LHC ring, where more than 1,200 superconducting dipole magnets guide the beams around the accelerator while radio-frequency cavities continuously increase their energy. During Run 3, proton beams are accelerated to an energy of 6.8 TeV per beam, resulting in a center-of-mass energy of

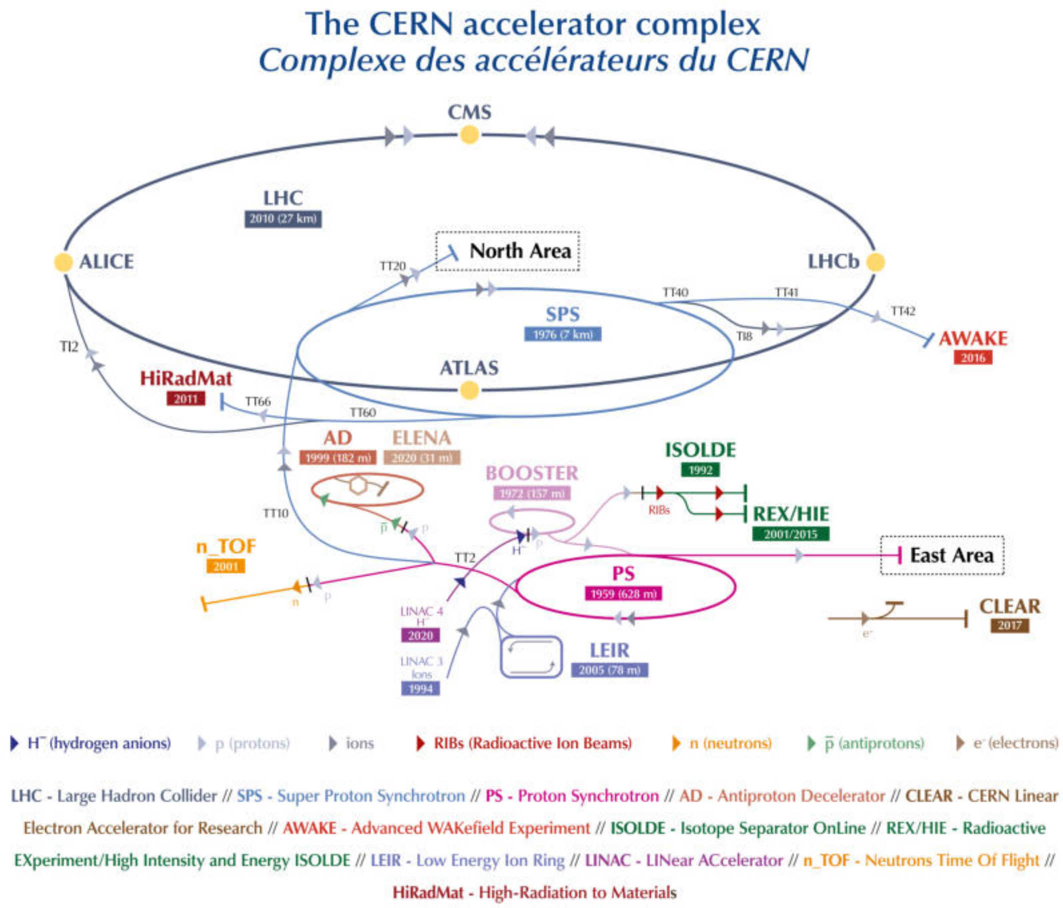


Figure 3.1.: Overview of the CERN accelerator complex and the experiments[16].

13.6 TeV [15]. The beams are grouped into thousands of bunches containing around 10^{11} protons each, which cross every 25 ns at the interaction points of the experiments. Four major experiments are located at these interaction points. **ATLAS** [17] and **CMS** [18] are general-purpose detectors designed to investigate a broad range of physics topics, including precision measurements of Standard Model processes and searches for new phenomena. **ALICE** [19] focuses on heavy-ion collisions to study the properties of the quark-gluon plasma, while **LHCb** [20] investigates the decays of beauty and charm hadrons to improve the understanding of matter-antimatter asymmetries. An Overview of the CERN accelerator complex, including LHC and its experiments can be seen in Figure 3.1.

Since the start of operation in 2009, the LHC has produced an unprecedented amount of collision data, enabling numerous advances in particle physics. Its most significant achievement to date was the discovery of the **Higgs boson** by the ATLAS and CMS collaborations in 2012, confirming the final missing particle predicted by the Standard

Model [8]. Beyond this milestone, the LHC has enabled increasingly precise measurements of Standard Model parameters and has placed stringent limits on many theories of physics beyond the Standard Model.

The high instantaneous luminosity delivered by the LHC also presents significant experimental challenges. Every bunch crossing can contain dozens of simultaneous proton-proton interactions, referred to as **pile-up**, resulting in large amounts of detector data that must be processed within microseconds. Efficient online event selection is therefore essential, making advanced trigger systems and machine learning techniques a key component of modern experiments such as ATLAS.

3.2. ATLAS Detector

The **ATLAS detector** (A Toroidal LHC ApparatuS) [17] is one of the two general-purpose detectors at the Large Hadron Collider. It is a multi-layered, cylindrical detector surrounding one of the proton–proton interaction points, with an overall length of approximately 46 m, a diameter of about 25 m, and a total weight of roughly 7,000 tonnes. Its nearly hermetic design allows the reconstruction and identification of a wide range of particles produced in high-energy collisions. In particular, the identification of hadronically decaying tau leptons requires information from multiple detector subsystems to be combined in order to distinguish them from background objects such as QCD jets and electrons.

A right-handed coordinate system is used throughout the ATLAS experiment. The origin is defined at the nominal interaction point in the center of the detector. The z-axis is aligned with the beam direction, while the x-axis points from the interaction point towards the center of the LHC ring. The y-axis points vertically upwards. In the plane transverse to the beam direction, particle trajectories are described by the azimuthal angle ϕ, measured around the beam axis.

The polar angle θ is measured with respect to the positive z-axis. Instead of using θ directly, particle directions are commonly expressed in terms of the **pseudorapidity**

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right).$$

Pseudorapidity is preferred because particle production in proton–proton collisions is approximately uniform as a function of η, and differences in pseudorapidity remain invariant under Lorentz boosts along the beam axis for highly relativistic particles. Consequently, detector acceptance and physics object reconstruction are naturally described in terms of η rather than the polar angle. A related quantity used throughout this thesis is the angular distance

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}\,,$$

which is used, for example, to associate tracks and calorimeter deposits with a reconstructed tau candidate (Section 3.3).

Figure 3.2 shows a schematic overview of the ATLAS detector and its major subsystems.

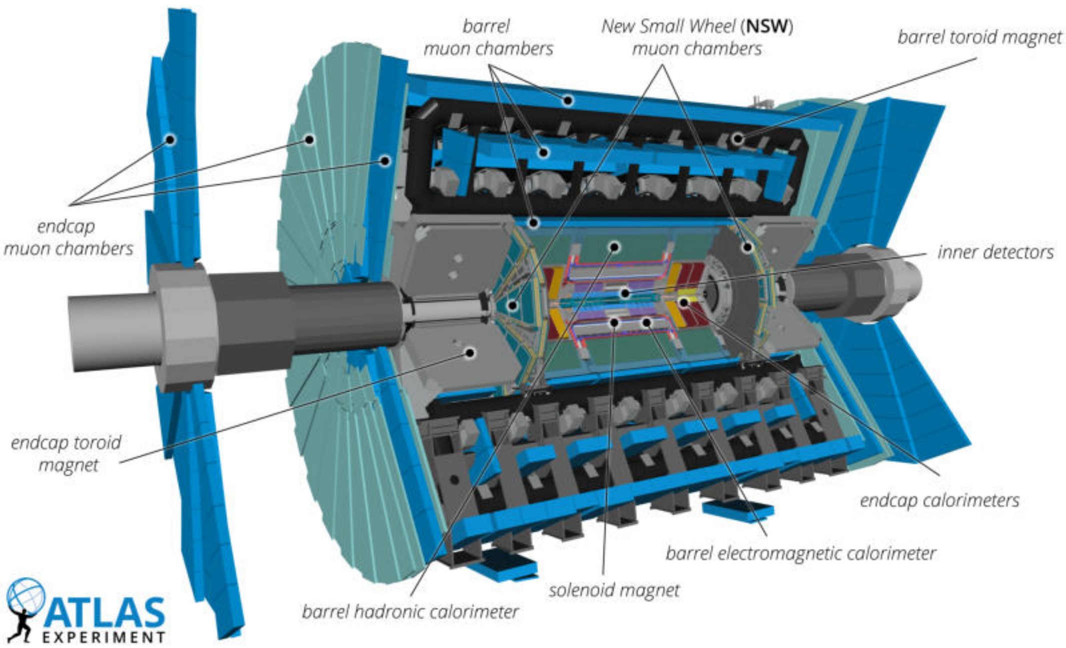


Figure 3.2.: Schematic overview of the ATLAS detector showing its major subsystems [21].

3.2.1. Inner Detector

The **Inner Detector (ID)** surrounds the beam pipe and is immersed in a uniform axial magnetic field of **2 T**, generated by a superconducting solenoid magnet. It is responsible for reconstructing the trajectories of charged particles produced in proton–proton collisions. As charged particles traverse the magnetic field, they follow curved trajectories whose curvature is inversely proportional to their momentum. By precisely measuring these trajectories, the Inner Detector determines the momentum and charge of particles, reconstructs interaction and decay vertices, and provides the tracking information required for particle identification. It covers the pseudorapidity range of $|\eta| < 2.5$ and consists of three complementary subdetectors: the Pixel Detector, the Semiconductor Tracker

(SCT), and the Transition Radiation Tracker (TRT) and can be seen in Figure 3.3. The

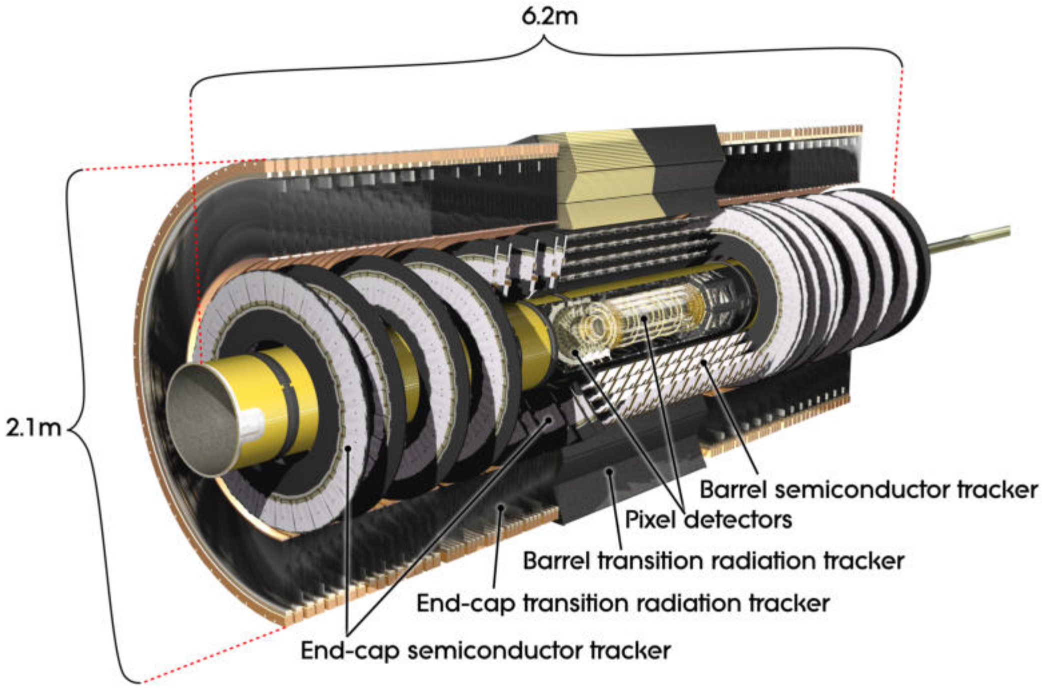


Figure 3.3.: Schematic overview of the Inner Detector [22].

Pixel Detector forms the innermost part of the tracking system and is located closest to the interaction point. It consists of several cylindrical barrel layers together with end-cap disks, providing high-precision three-dimensional position measurements. The innermost layer, the Insertable B-Layer (IBL), is situated only about **33 mm** from the beam axis and significantly improves the precision of vertex reconstruction. The detector employs silicon pixel sensors with a typical pixel size of $50 \times 250\,\mu\text{m}^2$, resulting in an excellent spatial resolution of approximately $10\,\mu$m in the transverse plane. These precise measurements allow the separation of the primary interaction vertex from displaced secondary vertices, making the Pixel Detector particularly important for identifying particles with finite lifetimes, including hadronically decaying tau leptons and heavy-flavour hadrons.

Surrounding the Pixel Detector is the **Semiconductor Tracker (SCT)**, which consists of four barrel layers and nine end-cap disks on each side of the detector. The SCT uses silicon microstrip sensors rather than pixels, providing a larger active area while maintaining excellent position resolution. Each module contains pairs of strips mounted with a small stereo angle, allowing the reconstruction of three-dimensional hit positions. Compared to the Pixel Detector, the SCT offers slightly lower spatial resolution but substantially increases the number of precision tracking measurements. Together with the pixel system, it provides accurate momentum determination and efficient reconstruction of charged-particle trajectories over the full tracking volume.

The outermost component of the Inner Detector is the **Transition Radiation Tracker (TRT)**. It consists of approximately **300,000** thin gaseous straw tubes, each with a diameter of about **4 mm**, arranged in barrel and end-cap sections. Besides contributing additional tracking measurements that improve the overall momentum resolution, the TRT is capable of detecting **transition radiation** emitted by highly relativistic charged particles when crossing boundaries between materials with different dielectric constants. Since electrons emit significantly more transition radiation than heavier charged particles at the same momentum, the TRT provides valuable discrimination between electrons and hadrons. The combination of the Pixel Detector, SCT, and TRT achieves a transverse momentum resolution of approximately

$$\frac{\sigma_{p_T}}{p_T} = \frac{0.05\%}{p_T[\mathrm{GeV}]} \oplus 1\%,$$

where p_T is given in GeV. This precise momentum measurement and the excellent vertex reconstruction provided by the Inner Detector are essential for the identification of hadronically decaying tau leptons and for distinguishing them from background processes.

3.2.2. Calorimeter System

Surrounding the Inner Detector is the ATLAS calorimeter system, which measures the energy of particles by absorbing them within dense detector material and can be seen in Figure 3.4. As particles traverse the calorimeters, they produce cascades of secondary particles, known as showers, whose deposited energy is proportional to the energy of the incident particle. The calorimeter system provides nearly complete coverage up to $|\eta| < 4.9$ and consists of two main components: the **electromagnetic calorimeter (ECAL)** and the **hadronic calorimeter (HCAL)**. Together, they enable the reconstruction of electrons, photons, hadronic jets, and missing transverse momentum. **Electromagnetic Calorimeter.** The electromagnetic calorimeter is designed to measure the energy of electrons and photons, which primarily interact through electromagnetic processes such as bremsstrahlung and pair production. In the ATLAS barrel region $|\eta| < 1.475$ and endcaps $1.375 < |\eta| < 3.2$, the ECAL is a lead–liquid argon (LAr) sampling calorimeter with an accordion-shaped electrode geometry. This design provides continuous azimuthal coverage without inactive regions and allows for fast signal collection.

The thickness of the electromagnetic calorimeter reaches approximately **24 radiation lengths** X_0 in the barrel and up to **26** X_0 in the endcaps, ensuring that electromagnetic

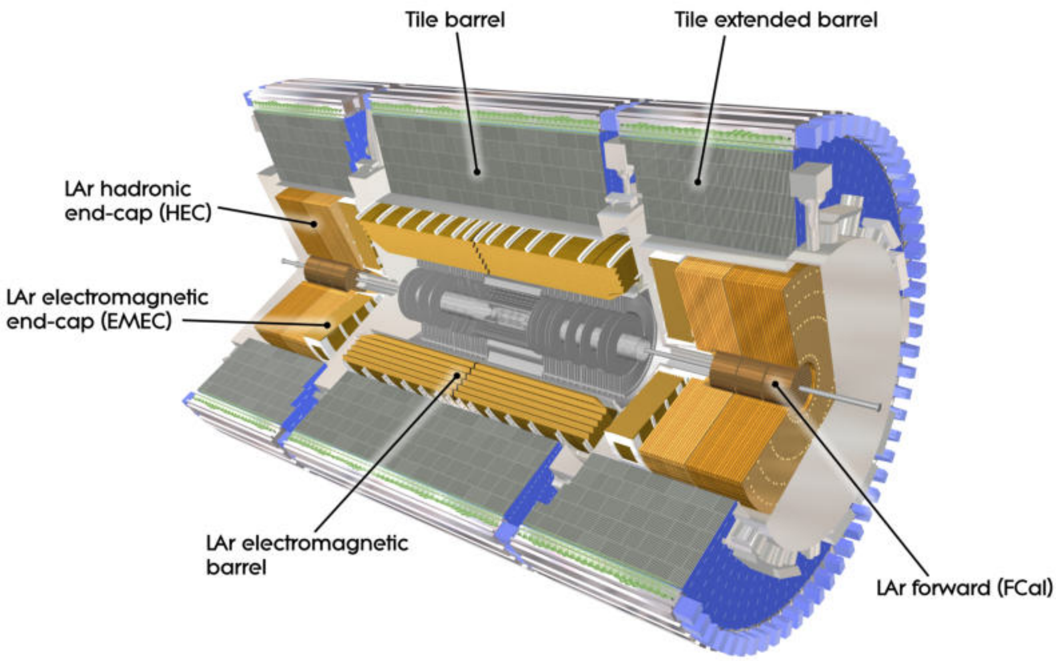


Figure 3.4.: Schematic overview of the ATLAS calorimeter system [23].

showers are almost completely contained within the detector. The energy resolution can be parameterized as

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

where the first term represents stochastic fluctuations in the shower development and the second term accounts for constant detector effects. The fine longitudinal and transverse segmentation of the ECAL allows precise energy measurements and contributes significantly to the identification of electrons and photons.

Hadronic Calorimeter. The hadronic calorimeter surrounds the ECAL and measures the energy of hadrons, including charged and neutral hadrons produced within jets and hadronic tau-lepton decays. In the central detector region $|\eta| < 1.7$, it consists of a steel absorber with scintillating plastic tiles as the active medium, commonly referred to as the Tile Calorimeter. At higher pseudorapidities, liquid argon calorimeters with copper or tungsten absorbers are used to withstand the increased radiation levels.

Since hadronic showers are more extended than electromagnetic showers, the HCAL is designed in terms of **interaction lengths** λ rather than radiation lengths. The combined electromagnetic and hadronic calorimeters provide approximately **10 interaction lengths** of material in the barrel region, ensuring efficient containment of hadronic showers. The hadronic energy resolution is approximately

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

reflecting the larger fluctuations inherent in hadronic shower development.

The calorimeter system plays a crucial role in the reconstruction of hadronically decaying tau leptons. While the Inner Detector measures the trajectories of charged decay products, the calorimeters determine the energy deposited by both charged and neutral particles, allowing the reconstruction of characteristic narrow tau-jet signatures and providing essential inputs for machine learning-based identification algorithms.

3.2.3. Particle Flow Reconstruction

Rather than reconstructing tracks and calorimeter deposits independently, ATLAS combines the information from the Inner Detector and the calorimeter system through a **particle-flow algorithm** [24]. Each track reconstructed in the Inner Detector is extrapolated into the calorimeter and matched to its associated energy deposits, which are then subtracted from the calorimeter image. The energy remaining after this subtraction is attributed to neutral particles, most importantly photons from $\pi^0 \to \gamma\gamma$ decays.

This procedure yields a set of **particle flow objects (PFOs)**: charged PFOs, corresponding to the momentum-based four-vector of a reconstructed track, and neutral PFOs, corresponding to calorimeter energy not associated with any track. A further category, **shot PFOs**, identifies small, localized electromagnetic energy deposits within the core of a tau candidate that are individually too collimated to be reconstructed as ordinary calorimeter clusters, but which provide additional sensitivity to closely spaced photons from π^0 decays. Particle-flow reconstruction therefore provides a more complete and less redundant description of the visible tau decay products than either subdetector alone, and forms the basis of the low-level input features used by the Graph Neural Network described in Chapters 4 and 5.

3.2.4. Muon Spectrometer

The **Muon Spectrometer (MS)** forms the outermost subsystem of the ATLAS detector and is designed to identify and measure the momentum of muons. Since muons interact only weakly with the calorimeters, they typically traverse the entire detector with only a small loss of energy before reaching the spectrometer. The spectrometer is embedded in a magnetic field generated by a system of large air-core toroidal magnets, consisting of

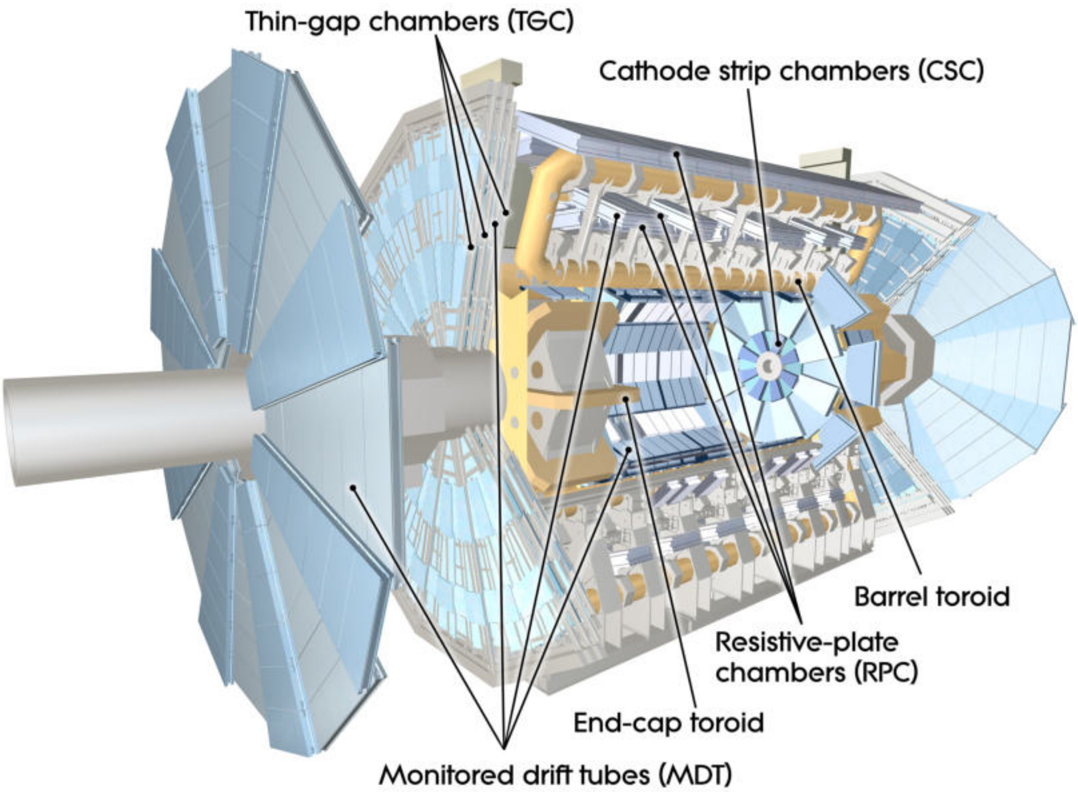


Figure 3.5.: Schematic overview of the ATLAS Muon Spectrometer [25].

one barrel toroid and two end-cap toroids and can be seen in Figure 3.5. The trajectories of muons are measured using high-precision tracking chambers, while dedicated trigger chambers provide fast timing information for the ATLAS trigger system. The Muon Spectrometer covers the pseudorapidity range of approximately $|\eta| < 2.7$ for precision tracking and up to $|\eta| < 2.4$ for triggering.

By combining measurements from the Inner Detector and the Muon Spectrometer, ATLAS achieves excellent momentum resolution over a wide range of muon momenta. Although the Muon Spectrometer is not directly used for the hadronic tau-lepton identification studied in this thesis, it plays a crucial role in many ATLAS physics analyses involving muons in the final state.

3.3. Tau Candidate Reconstruction

Hadronically decaying tau leptons are not reconstructed as a single object type but are built from the standard detector-level objects described above. Reconstruction proceeds in two stages: the identification of a **tau candidate** from calorimeter and tracking information, followed by a classification step that estimates how likely the candidate is to originate from a genuine hadronic tau decay rather than a background process [26].

Tau candidates are seeded by jets reconstructed with the anti-k_t algorithm using calorimeter energy deposits, typically with a radius parameter of $R = 0.4$ [26]. For each jet seed satisfying minimal kinematic requirements, a dedicated **tau vertex** is identified among the reconstructed primary vertices using the tracks compatible with the jet direction, improving the association of tracks and calorimeter energy to the correct proton–proton interaction in the presence of pile-up. Tracks within $\Delta R < 0.2$ of the tau candidate axis are classified as **core tracks** and used to determine the track multiplicity (1-prong or 3-prong) of the candidate, while tracks in the surrounding region $0.2 < \Delta R < 0.4$, referred to as the **isolation** region, provide additional discriminating information without contributing to the track multiplicity decision. This is visualised in Figure 3.6. Once

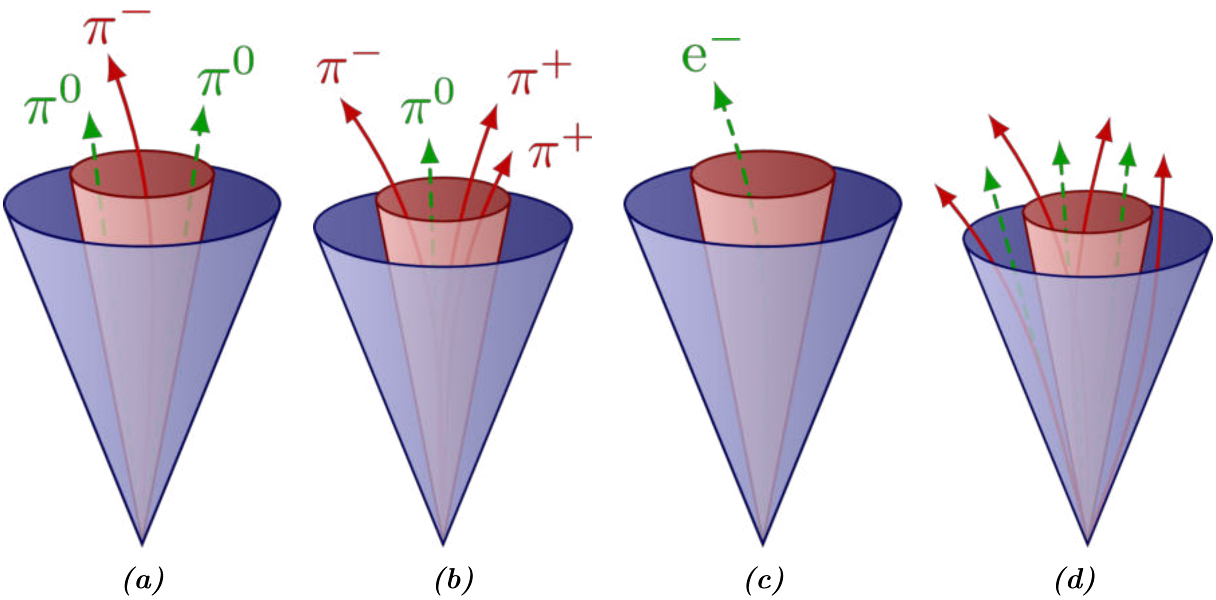


Figure 3.6.: Comparison of the reconstructed tracks and clusters in the detector for different jet types. Showing a 1-prong tau decay in (a), a 3-prong tau decay in (b), a single electron in (c) and a general QCD jet topology in (d) [27–30].

a tau candidate has been built, its four-momentum, substructure, and the associated tracks, calorimeter clusters, and particle-flow objects (Section 3.2.3) are used as inputs to a dedicated identification algorithm that computes a continuous **tau-identification score**. This score is used to define working points (commonly *loose*, *medium*, and *tight*); the precise signal-efficiency definitions used in this thesis are given per prong multiplicity in Chapter 6 (Section 6.1 of the Performance Studies chapter), since they differ between the jet-rejection and electron-veto tasks and between 1-prong and 3-prong candidates.

4. Machine Learning

Machine learning has become an essential component of ATLAS reconstruction and identification algorithms, enabling efficient particle identification by exploiting complex correlations between reconstructed detector observables. For the identification of hadronically decaying tau leptons, ATLAS has evolved from recurrent neural networks (RNN s) to the Graph Neural Network (GNN) GNTau for jet rejection [2, 3], while electron rejection is still handled by a separate RNN-based veto. The objective of this thesis is to develop GNTau_eVeto, a unified GNN that simultaneously distinguishes between hadronic tau leptons, QCD jets, and electrons, replacing the two previously separate classifiers with a single multi-class network.

4.1. Perceptrons

One of the earliest machine learning models was inspired by the biological neuron. This idea led to the development of the perceptron, introduced by Frank Rosenblatt in 1958 [31], which represents one of the first computational models capable of learning from data.

A perceptron receives a set of input features (x_i), each associated with a weight (w_i), and computes their weighted sum together with a bias term (b):

$$z = \sum_i w_i x_i + b.$$

The resulting value is passed through an activation function, originally a binary step function, to produce the final output. In this way, the perceptron performs a binary classification by deciding on which side of a linear decision boundary a sample lies. Figure 4.1 shows an overview of the perceptron. During training, the predicted output is compared to the true label, and the model parameters are updated iteratively to reduce classification errors. The perceptron learning rule adjusts the weights in the direction that improves future predictions, allowing the model to learn directly from labeled examples without requiring explicitly programmed decision rules.

A major advantage of the perceptron is its mathematical simplicity. It can be proven

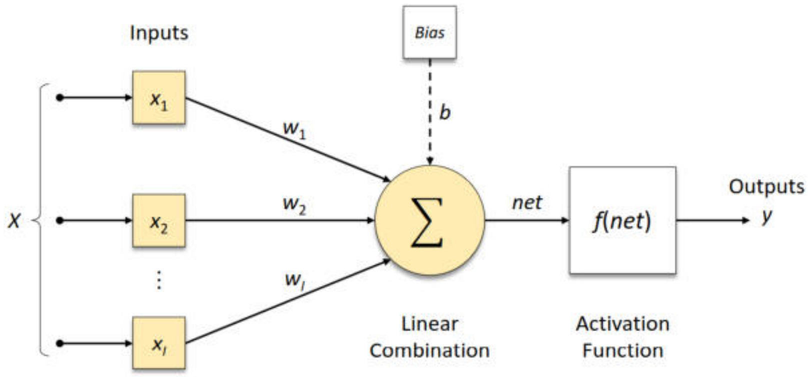


Figure 4.1.: Schematic overview of a single perceptron [32].

that the learning algorithm converges to a solution whenever the training data are linearly separable, meaning that the two classes can be divided by a single hyperplane. However, this property also highlights its main limitation. Problems that are not linearly separable, such as the well-known XOR problem, cannot be solved by a single perceptron regardless of the amount of training.

Nevertheless, the perceptron established the fundamental concepts that remain central to modern machine learning: representing data through weighted connections, learning by optimizing model parameters from examples, and performing inference using activation functions. These ideas ultimately led to the development of multilayer neural networks, which overcome the linear limitations of a single perceptron and form the foundation of today's deep learning methods.

4.1.1. Loss Functions: Cross-Entropy Loss

To train a neural network, a loss function is required to quantify the difference between the predicted output and the true class labels. For the multi-class classification task in this thesis, the **cross-entropy loss** is used, as it is the standard choice for classification problems.

The network produces a probability distribution over the three target classes, QCD jets, hadronically decaying tau leptons, and electrons, using a softmax activation function in the final output layer. The cross-entropy loss measures how well this predicted probability distribution matches the true class label. For a single training sample, it is defined as

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

where C is the number of classes, y_i is the one-hot encoded true label, and $\hat{y}_i$ is the predicted probability for class i [33]. Since only one element of the target vector is equal to one, the loss simplifies to the negative logarithm of the predicted probability assigned to the correct class.

Minimizing the cross-entropy loss encourages the network to assign high probabilities to the correct class while suppressing the probabilities of the remaining classes. In this work, equal class weights are used during training, such that all three classes contribute equally to the optimization.

4.2. Multi-Layer Perceptrons

The inability of a single perceptron to solve non-linearly separable problems motivated the development of **multi-layer perceptrons (MLPs)** [34]. By combining multiple perceptrons into a network of interconnected neurons, MLPs can learn increasingly complex relationships between input features and target classes.

A neural network consists of an **input layer**, one or more **hidden layers**, and an **output layer**. The input layer receives the feature vector describing each sample, while the output layer produces the final prediction. Between them, the hidden layers successively transform the input into increasingly abstract representations. Each neuron receives the outputs of the previous layer, computes a weighted sum together with a bias term, and applies an activation function to produce its output,

$$a = f\left(\sum_i w_i x_i + b\right) ,$$

where f denotes the activation function. The activation function introduces non-linearity into the network, enabling it to model complex relationships that cannot be represented by a single linear transformation. Common activation functions include the sigmoid function, the hyperbolic tangent, and the Rectified Linear Unit (ReLU), which is widely used in modern deep learning due to its computational efficiency and favourable optimization properties. Activation functions are discussed in more detail in Section 4.2.1. The trainable parameters of the network consist of the weights and biases connecting the neurons. During training, these parameters are adjusted such that the network minimizes a predefined loss function. Starting from randomly initialized values, the network repeatedly processes batches of training data, computes predictions, and evaluates the corresponding loss. The parameters are then updated using gradient descent, an optimization algorithm that iteratively moves the network parameters in the direction that reduces the loss. A schematic overview of neural network can be seen in Figure 4.2.

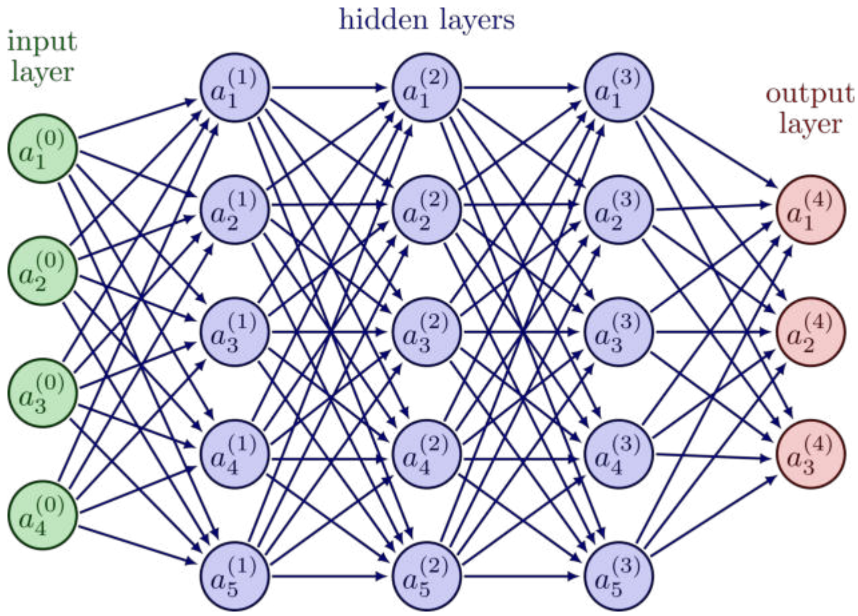


Figure 4.2.: Schematic representation of a neural network with four nodes in the input layer, several hidden layers and three nodes in the output layer [35].

To determine how each parameter contributes to the prediction error, the gradients of the loss function with respect to all weights and biases are computed using **back-propagation** [36]. Applying the chain rule of differential calculus, the prediction error is propagated backwards through the network from the output layer to the input layer. These gradients are subsequently used by the optimizer to update the parameters, gradually improving the network's performance over successive training iterations. Training continues until the loss converges or another stopping criterion is reached.

4.2.1. Activation Functions

Activation functions are a fundamental component of neural networks, as they introduce non-linearity into the transformations performed by each layer. Without activation functions, a network consisting of multiple layers would still only represent a single linear transformation, limiting its ability to learn complex relationships between input features. By applying a non-linear function after each weighted combination, neural networks can approximate more complex functions and identify patterns within high-dimensional data.

A commonly used activation function in modern deep learning is the **Rectified Linear Unit (ReLU)** [37], defined as

$$f(x) = \max(0, x)\,.$$

The ReLU function outputs zero for negative inputs and retains positive inputs unchanged. Compared to older activation functions such as the sigmoid function, ReLU is computationally simple and reduces the problem of vanishing gradients, allowing deeper networks to be trained more effectively. Due to these advantages, ReLU has become the standard choice for many hidden layers in modern neural networks.

In the network developed in this thesis, ReLU activation functions are applied after the dense layers in the input embedding networks and the classification head (Section 4.5).

4.2.2. Dropout

As neural networks become deeper and contain an increasing number of trainable parameters, they are prone to **overfitting**, where the model learns patterns specific to the training data rather than general features. To reduce this effect, a regularization technique known as **dropout** [38] is commonly employed and is visualised in Figure 4.3. During training, dropout randomly deactivates a fraction of the neurons in a layer with a

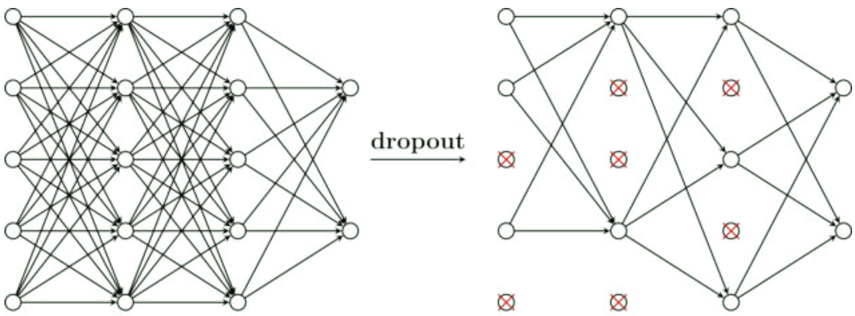


Figure 4.3.: Schematic representation of dropout during the training of a neural network, by temporarily deactivating a number of random neurons to reduce overfitting [39].

predefined probability. The remaining neurons must therefore learn representations that are not dependent on specific connections, improving the robustness and generalization of the network. During inference, dropout is disabled and all neurons are used, while their activations are appropriately scaled to account for the training procedure.

In the classification head of the model presented in this thesis, a dropout rate of 0.1 is applied between the fully connected layers to reduce overfitting while maintaining the network's learning capacity.

4.3. Transformers

Transformers are a class of neural network architectures built around the attention mechanism [40]. Originally developed for natural language processing, they have since been successfully applied to many other fields, including computer vision and particle physics. Their main advantage is the ability to model relationships between all input elements simultaneously, without relying on sequential processing as required by recurrent neural networks. A visual representation can be seen in Figure 4.4. A Transformer processes a

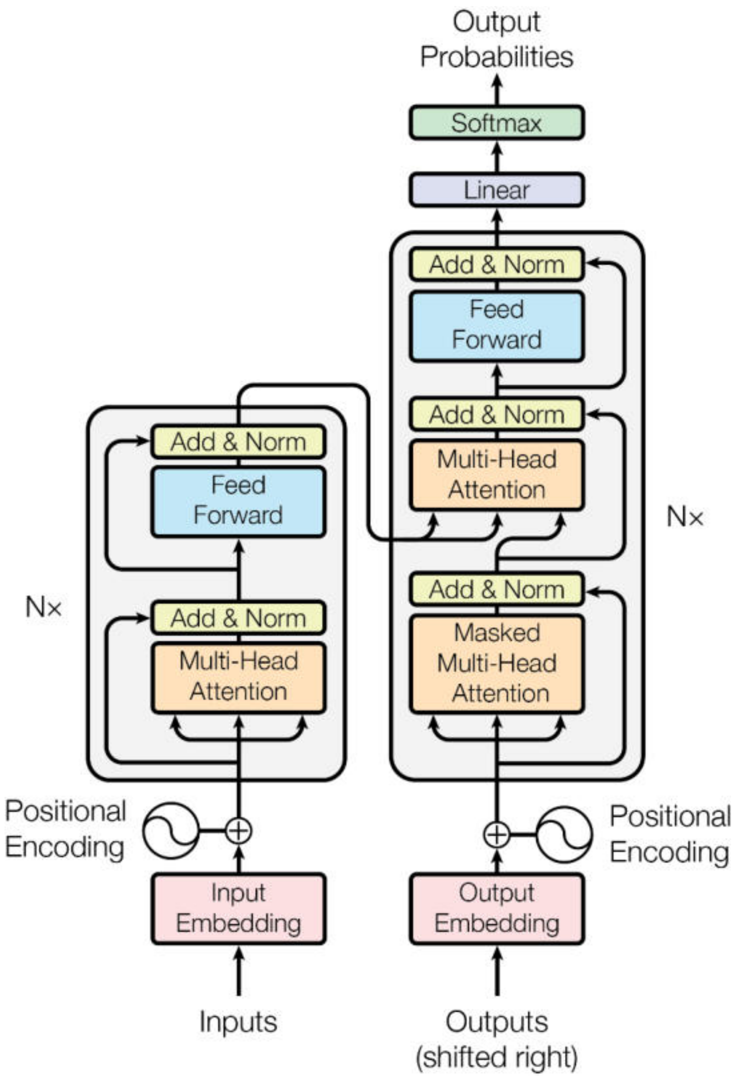


Figure 4.4.: Schematic Overwiew of the Transformer architercture [40].

set of input elements by first embedding them into a common feature space. These embeddings are then passed through multiple Transformer blocks, which typically consist of multi-head self-attention layers followed by fully connected feed-forward networks. Residual connections and normalization layers are commonly used to improve the stability of training. Through repeated application of these blocks, the network learns increasingly complex representations of the input data.

The self-attention mechanism is especially relevant for particle physics applications, where a reconstructed object is often composed of many detector-level constituents. Instead of treating these constituents independently, Transformers can learn the relationships between them and identify which interactions are most relevant for the classification task.

In the context of ATLAS tau-lepton identification, Transformer-based architectures have enabled improved performance by exploiting correlations between tracks, calorimeter measurements, and other reconstructed objects. By learning attention patterns between detector constituents, these models can capture characteristic signatures of hadronic tau-lepton decays and improve the separation from background processes such as QCD jets and electrons.

4.3.1. Attention

A central concept in modern deep learning architectures is the **attention mechanism**, which allows a network to dynamically determine the importance of different input features when producing an output. Unlike traditional neural networks, where all inputs are processed with fixed connections, attention enables the model to assign different weights to different parts of the input depending on their relevance for the current task.

The attention mechanism is based on three learned representations: **queries** (Q), **keys** (K), and **values** (V). The similarity between a query and all available keys determines the attention weights, which are then used to compute a weighted combination of the corresponding values. The scaled dot-product attention is defined as

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

where d_k represents the dimensionality of the key vectors [40]. The scaling factor prevents excessively large values in the dot products, improving numerical stability during training.

By using attention, a model can selectively focus on the most relevant information and capture relationships between different elements of the input, even when these elements are far apart. This property makes attention particularly useful for complex data structures where the relationships between individual objects contain important information.

4.4. Graph Neural Networks

Many machine learning architectures are designed for data with a fixed structure, such as images arranged in a grid or sequences with a defined order. However, data from particle physics experiments often consist of sets of reconstructed objects, where the relationships between these objects contain important information. **Graph Neural Networks (GNNs)** [41] are a class of neural networks specifically designed to process such graph-structured data.

A graph consists of a set of **nodes**, **edges**, and optionally **global features**. In the context of particle physics, nodes can represent reconstructed detector objects, such as tracks, calorimeter clusters, or particle-flow objects, while edges describe relationships between these objects. Unlike conventional neural networks, GNNs are not restricted to a fixed number or ordering of inputs, making them well suited for the variable-size and complex structure of particle collision events.

The fundamental principle of GNNs is **message passing**, where information is exchanged between connected nodes. During each message-passing step, every node updates its representation by combining its own features with information received from neighbouring nodes. Through multiple iterations, the network learns increasingly complex representations that include information about the local and global structure of the graph.

A simplified representation of a message-passing operation can be written as

$$h_i^{(k+1)} = f\left(h_i^{(k)}, \sum_{j \in \mathcal{N}(i)} g(h_i^{(k)}, h_j^{(k)})\right),$$

where $h_i^{(k)}$ represents the feature representation of node i after the k-th update step and $\mathcal{N}(i)$ denotes the neighbouring nodes. The functions g and f describe the learned transformations used to exchange and update information.

The ability of GNNs to incorporate relationships between individual detector objects makes them particularly powerful for particle identification tasks. For hadronic tau-lepton identification, the internal structure of the reconstructed tau candidate, including correlations between tracks, calorimeter deposits, and other particle-flow objects, provides valuable information for distinguishing signal from background processes. In ATLAS, GNN-based architectures have therefore been introduced to improve upon earlier RNN-based approaches, with the GNTau architecture used as the starting point for the model developed in this thesis.

4.5. The GNTau_eVeto Architecture

The model developed in this thesis, **GNTau_eVeto**, combines the building blocks introduced above into a single graph neural network that receives the five groups of input features described in Section 5.3.2 (jet, track, cluster, neutral PFO, and shot PFO variables) and outputs a probability for each of the three target classes: QCD jet, hadronic tau lepton, and electron.

- **Input embedding.** Each input feature group is first mapped into a common latent representation by a dedicated feed-forward embedding network, using ReLU activations (Section 4.2.1).
- **Relational processing.** The embedded objects are then processed by self-attention layers that exchange information between the constituents of the tau candidate.
- **Readout.** The updated node representations are aggregated into a single graph-level representation via aggregated into a single graph-level representation via a global attention pooling layer [42].
- **Classification head.** The graph-level representation is passed through a series of fully connected layers with ReLU activations and a dropout rate of 0.1 (Section 4.2.2), followed by a final linear layer and a softmax activation producing the three-class output used by the cross-entropy loss of Section 4.1.1.

While the preprocessed, resampled training data themselves are prepared with the Umami Preprocessing framework (Section 5.2), the GNTau_eVeto model is trained using **Salt** [43], a general-purpose, PyTorch-Lightning-based framework for training multi-modal, multi-task machine-learning models that is maintained within the ATLAS flavour-tagging (FTAG) software ecosystem alongside Umami. Salt handles the model definition, optimizer and learning-rate scheduling, checkpointing, and logging of the training and validation loss shown in Chapter 6 (Section 6.6), and is configured entirely through YAML configuration files rather than bespoke training code, which improves the reproducibility of the training procedure.

4.6. Performance Metrics

The performance of a tau-identification classifier is quantified in terms of the **signal efficiency** and the **background rejection**. For a given classifier output threshold, the signal efficiency is defined as the fraction of genuine hadronic tau leptons correctly classified as signal,

$$\varepsilon_{\text{sig}} = \frac{N_{\text{sig}}^{\text{selected}}}{N_{\text{sig}}^{\text{total}}}\,,$$

while the background rejection is defined as the inverse of the corresponding background efficiency,

$$R_{\text{bkg}} = \frac{1}{\varepsilon_{\text{bkg}}} = \frac{N_{\text{bkg}}^{\text{total}}}{N_{\text{bkg}}^{\text{selected}}}\,.$$

Since GNTau_eVeto performs a three-class classification, both background rejections (against QCD jets and against electrons) are evaluated as a function of the same signal efficiency, typically by scanning the classifier threshold and reporting a **receiver operating characteristic (ROC) curve** of background rejection versus signal efficiency for each background class separately. Fixed reference points on this curve, commonly referred to as *loose*, *medium*, and *tight* working points, correspond to representative signal efficiencies (e.g. 60%, 75%, and 85%) and are used to compare GNTau_eVeto to the GNTau and RNN-based baselines in Chapter 6.

5. Data Preparation

The performance of a supervised machine learning model strongly depends on the quality and composition of the training data. In addition to a large number of examples, each event must be assigned a truth label that identifies its underlying physics process. Furthermore, the input data must be preprocessed to remove biases that could otherwise be learned by the model instead of the physical characteristics distinguishing the different particle classes.

For this purpose, simulated proton–proton collision events are used throughout this thesis. This chapter describes the simulated datasets (Section 5.1), the preprocessing chain applied to them using the Umami framework (Section 5.2), and the input features presented to the Graph Neural Network (Section 5.3). It is worth distinguishing at the outset between the two software frameworks used in this thesis: **Umami** [44] performs the preprocessing described in this chapter (event selection, truth matching, feature normalization, and kinematic resampling), whereas the actual training of the GNTau_eVeto model itself is performed using **Salt** [43], a separate, complementary framework described in Section 4.5 of Chapter 5.

5.1. Simulated Event Samples

The datasets used in this work are produced using **Monte Carlo (MC) simulation**, which models both the underlying proton–proton interactions and the response of the ATLAS detector. In contrast to collision data, Monte Carlo simulation provides complete truth information about the generated particles and their decay chains, allowing reconstructed objects to be matched to their true particle origin. This truth information is essential for supervised machine learning, as it provides the class labels required during training.

The simulated samples used in this thesis consist of both signal and background processes. Hadronically decaying tau leptons are primarily obtained from $Z/\gamma^* \to \tau^+\tau^-$ events, while the principal background processes are represented by $Z \to e^+e^-$ and other Drell–Yan electron production, together with QCD dijet events. Together, these samples provide representative examples of the three target classes considered by the classifier: hadronic tau leptons, electrons, and QCD jets.

Process	**Target class**	N_{events}
$Z/\gamma^* \to \tau^+\tau^-$	hadronic τ	55 million
$Z \to e^+e^-$ / Drell–Yan	electron	12 million
QCD dijet	jet	770 million

Table 5.1.: Simulated samples used for training, validation, and testing.

The generated events are reconstructed using the standard ATLAS reconstruction software and stored as **ntuples**, which contain both reconstructed detector quantities and the corresponding truth information. Since these ntuples are not directly suitable as input for neural network training, they are converted into HDF5 (`.h5`) format, merged into common datasets, and assigned truth labels according to the generated particle origin. The resulting labelled HDF5 files are then processed using the **Umami** preprocessing toolkit, as described in the following section.

5.2. Data Preprocessing

Before the labelled HDF5 datasets introduced in Section 5.1 can be used for machine learning, they must undergo several further preprocessing steps. The objective of this preprocessing is to transform the simulated detector data into a format suitable for neural network training while preserving the relevant physical information, and to ensure that the model learns the characteristic properties of the different particle classes rather than artificial biases introduced by the simulation. The remaining preprocessing described in this section is performed using the **Umami Preprocessing (UPP)** framework [44].

5.2.1. Umami Preprocessing Framework

The Umami Preprocessing framework provides a standardized workflow for preparing ATLAS datasets for machine learning applications. Its primary objective is to ensure that the neural network learns genuine physical differences between the considered particle classes instead of exploiting biases originating from the underlying Monte Carlo samples. Besides converting the data into the format expected by the network, UPP performs event selection, truth matching, feature normalization, resampling of the input kinematic distributions, and the creation of independent training, validation, and test datasets.

After applying the required event selections and truth matching, all input variables are normalized using statistics determined from the training sample. This normalization transforms variables with different physical units and numerical ranges into comparable scales, improving the stability and convergence of the neural network during training.

All input variables have been plotted in normalized form, separated by origin — signal (tau), QCD, and electron and can be found in Appendix A.2.

5.2.2. Kinematic Resampling

Since the individual Monte Carlo samples originate from different physics processes, their transverse momentum p_T and pseudorapidity η distributions differ significantly. If used directly, the neural network could exploit these kinematic differences instead of learning the characteristic detector signatures of hadronic τ leptons, electrons, and QCD jets. To avoid this bias, UPP performs a dedicated kinematic resampling in p_T and η, removing the strong correlations between the signal and background samples that would otherwise allow the model to classify events based on their production process rather than their detector signature.

Two different resampling methods are implemented within the Umami framework: *count-up resampling* and *probability density function (PDF) resampling* [44]. In this work, the latter approach is used. Unlike count-up resampling, which operates on discrete histogram bins with fixed boundaries, PDF resampling approximates the underlying event distributions by continuous probability density functions, so that each event contributes continuously to the estimated distribution rather than being assigned exclusively to an individual bin. Consequently, events located close to bin boundaries are treated smoothly, avoiding discontinuities introduced by hard bin edges, and sharp peaks that may arise from fixed-bin resampling are significantly reduced, leading to smoother and more representative training samples.

Figures 5.1–5.3 illustrate the effect of the PDF resampling procedure for the pseudorapidity and transverse momentum distributions. The initial distributions exhibit clear differences between the various Monte Carlo samples, whereas the resampled training and validation datasets show significantly improved agreement. The test dataset is deliberately left unresampled, so that it retains the original kinematic distributions. This ensures that the neural network cannot simply exploit kinematic differences between the underlying physics processes during training and is instead forced to learn discriminating features based on the detector response, while still allowing its performance to be evaluated under realistic, unaltered kinematic conditions.

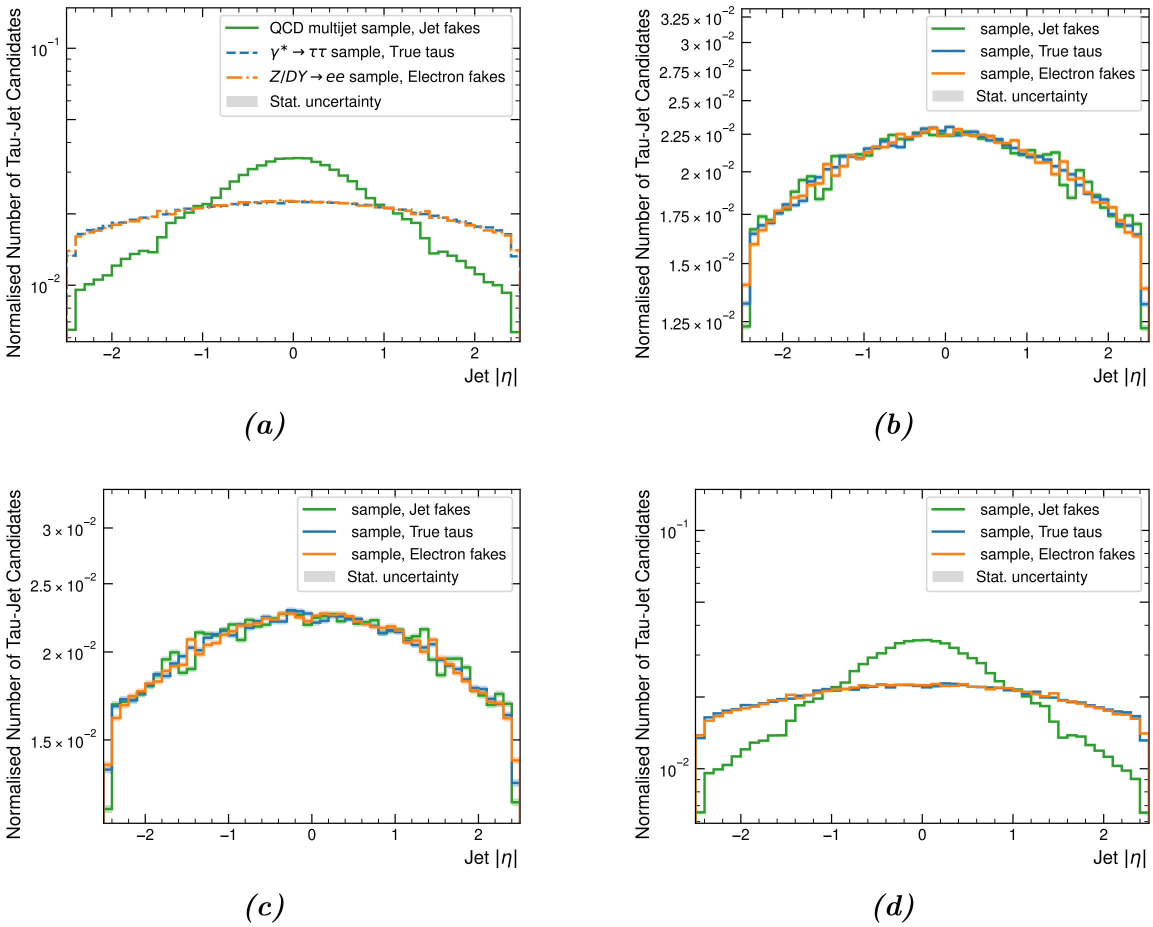


Figure 5.1.: Pseudorapidity (η) distribution of the tau-jet sample before resampling (a), and after PDF resampling in the training (b), validation (c), and test (d) datasets.

5.2.3. Training, Validation and Test Samples

Following the preprocessing and resampling procedure, the complete dataset is divided into three statistically independent subsets. The **training dataset** is used to optimize the network parameters during gradient descent. The **validation dataset** is employed during training to monitor the model performance and detect potential overfitting without influencing the optimization itself. Finally, the completely independent **test dataset** is used only after training has been completed to evaluate the final performance of the network.

Since all three datasets originate from the same preprocessing chain and are resampled using identical procedures, they exhibit nearly identical kinematic distributions while remaining statistically independent. This guarantees an unbiased evaluation of the trained model and allows a reliable assessment of its generalization performance.

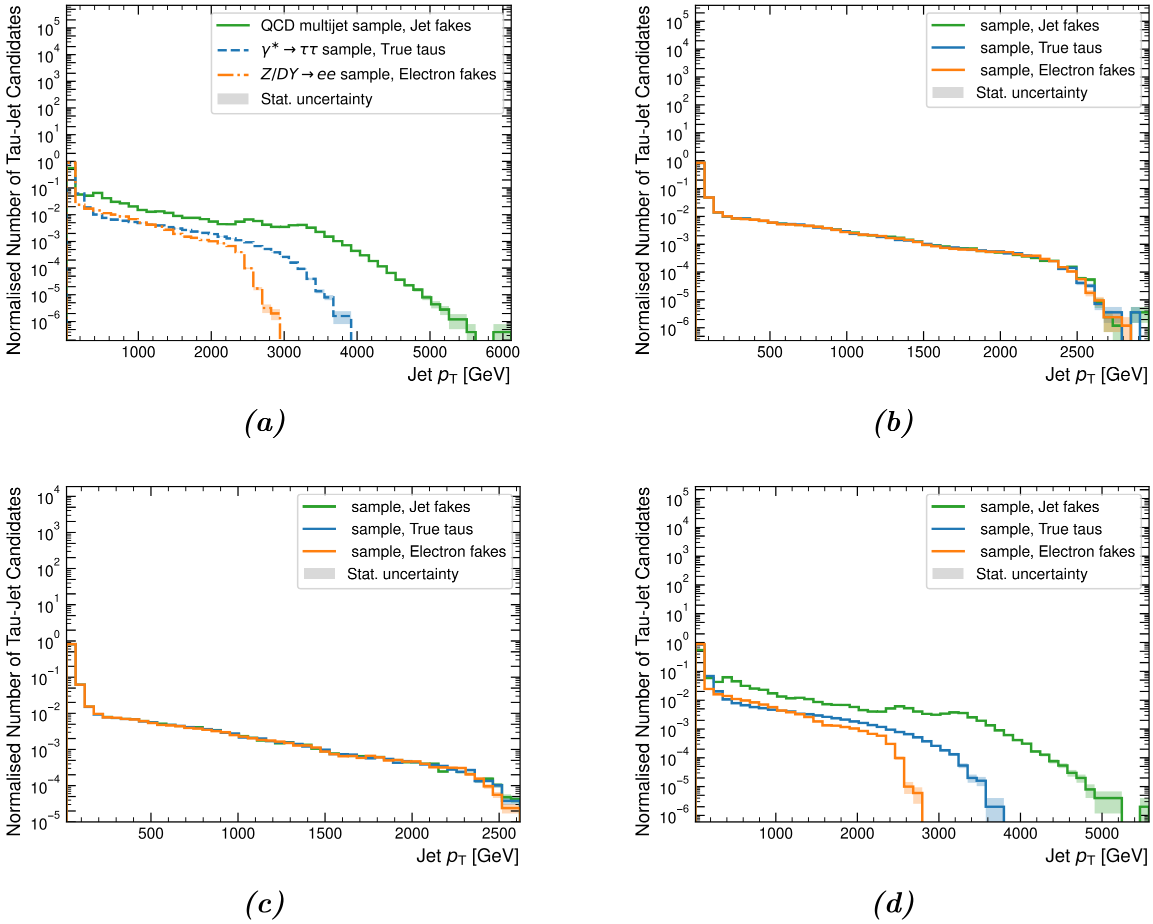


Figure 5.2.: Transverse momentum (p_T) distribution of the tau-jet sample before resampling (a), and after PDF resampling in the training (b), validation (c), and test (d) datasets.

5.3. Input Features

The performance of a machine learning model strongly depends on the information provided as input. The datasets generated by the Umami preprocessing framework contain a large number of reconstructed detector quantities as well as additional variables required for preprocessing and performance evaluation. However, only a subset of these variables is used during the training of the Graph Neural Network.

The variables stored in the preprocessed HDF5 files can therefore be divided into two categories. The first category consists of variables required by the Umami framework for event selection, truth matching, sample weighting, and performance evaluation (Section 5.3.1). The second category comprises the physics features used directly as inputs to the neural network (Section 5.3.2).

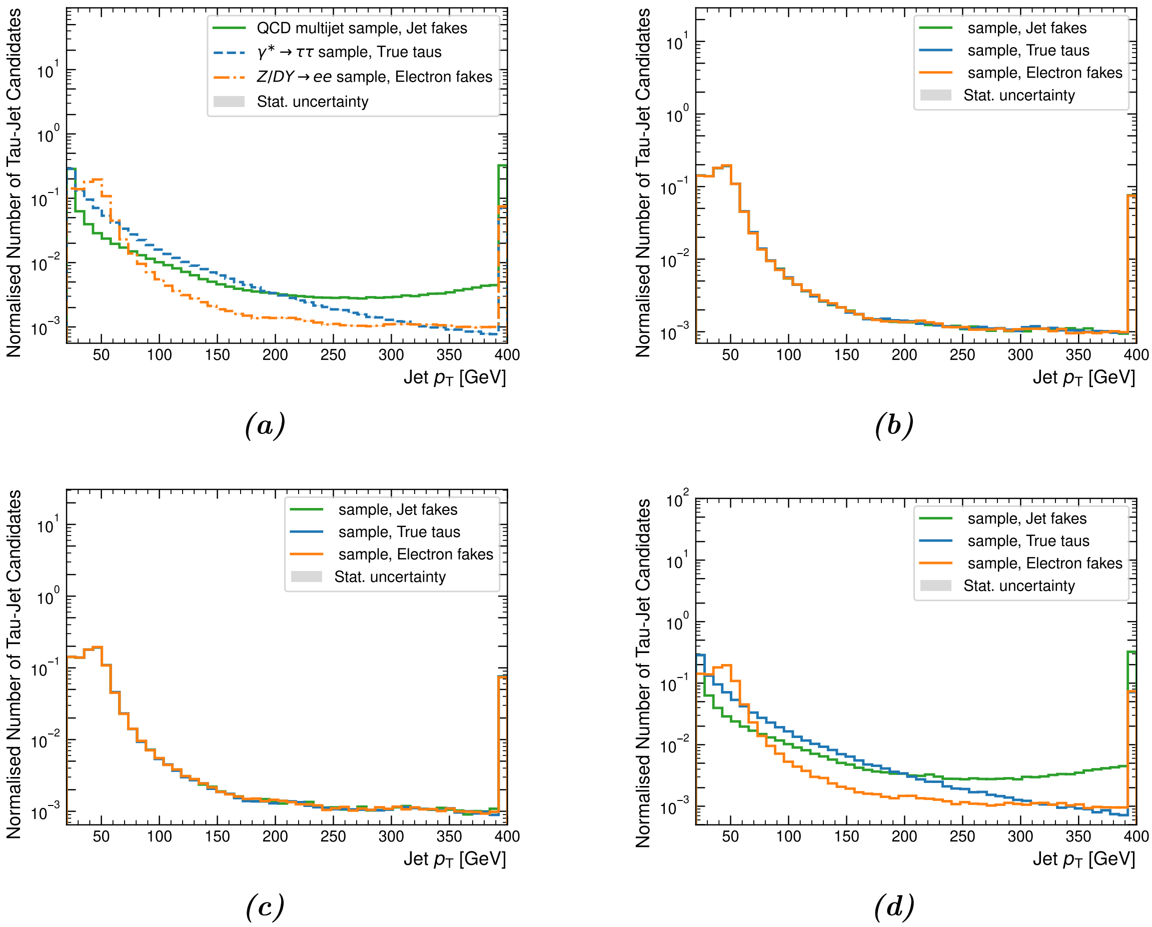


Figure 5.3.: Transverse momentum (p_T) distribution as in Figure 5.2, restricted to the low-p_T region where the resampling effect is most pronounced.

5.3.1. Umami Bookkeeping Variables

Besides the machine learning input features, the preprocessed datasets contain additional variables that are not presented to the neural network during training. These include truth labels, event identifiers, reconstruction flags, class assignments, and auxiliary quantities used during preprocessing and performance evaluation.

The truth information obtained from the Monte Carlo simulation allows reconstructed tau candidates to be matched to their generated particle origin and forms the basis for the supervised training procedure. Furthermore, these variables are required for the calculation of performance metrics such as signal efficiencies and background rejection (Section 6.1). Since they contain generator-level information that is unavailable in collision data, they are excluded from the network input to prevent information leakage.

5.3.2. Graph Neural Network Input Features

The Graph Neural Network receives information from several detector object collections, each describing a different aspect of the reconstructed tau candidate (see Section 3.3 and Section 3.2.3 for how these objects are reconstructed). Rather than relying solely on global jet properties, the network processes low-level detector information, allowing it to learn correlations between individual detector objects.

The input features are divided into five groups:

- **Jet variables**, describing the reconstructed tau candidate as a whole.
- **Track variables**, characterizing the charged particles reconstructed in the Inner Detector.
- **Cluster variables**, providing information on the calorimeter energy deposits associated with the tau candidate.
- **Neutral particle flow objects (Neutral PFOs)**, describing reconstructed neutral particles, primarily originating from neutral pion decays.
- **Shot particle flow objects (Shot PFOs)**, representing localized electromagnetic energy deposits used in the reconstruction of neutral pions.

Together, these feature groups provide complementary information about the topology, tracking properties, and calorimeter signatures of the reconstructed tau candidate, and form the five node types processed by the GNTau_eVeto architecture described in Section 4.5. The full list of low-level input variables used for jets, tracks, calorimeter clusters, and particle-flow objects (neutral PFOs and photon "shots") is given in Tables A.1–A.5. All variables are taken directly from the derivation-level tau, track, cluster, and PFO containers.

6. Performance Studies

This chapter quantifies the tau-identification and electron-rejection performance of the graph-neural-network tagger (GNTau [3], and its extended version with an integrated electron veto, GNTau_eVeto, developed in this thesis) relative to the existing baseline recurrent neural network (RNN [2]). All numbers are evaluated on a common, statistically independent test sample, so that all models are compared on exactly the same events.

6.1. Metrics and Conventions

Two complementary figures of merit are used throughout this chapter:

- **Background rejection** at a fixed signal (tau) efficiency, defined as the inverse of the background efficiency, $R = 1/\varepsilon_{\mathrm{bkg}}$, evaluated separately against jet fakes and against electron fakes.
- **Area under the ROC curve (AUC)**, summarising the full efficiency–rejection trade-off in a single number, again computed separately per background type.

Working points are defined *per prong multiplicity*, following the convention used for the reference tau-identification and electron-veto selections [26] and can be found in Table 6.1

Table 6.1.: Definition of the signal efficiency working points for the jet-fake rejection and electron-veto selections, split by reconstructed prong multiplicity.

Selection Type	1-prong	3-prong
Jet-fake rejection	60 %, 75 %, 85 %, 95 %	45 %, 60 %, 75 %, 95 %
Electron veto	85 %, 90 %, 95 %	90 %, 95 %, 97.5 %

Because the electron-fake background is limited in size, particularly for 3-prong taus, where only 628 fiducial electron-fake candidates are available, some working points fall into a regime where fewer than 10 background events are expected to pass the selection.

In this regime a single measured rejection value is no longer statistically meaningful, and the quoted number is instead reported as a one-sided 68 % confidence-level *lower limit* on the true rejection (Clopper–Pearson interval [45]), marked with a “>” sign in both the tables and the figures below. In the figures, the corresponding curve segments are drawn dashed with upward-pointing markers to distinguish them visually from an actual measurement.

6.2. Samples and Fiducial Selection

All numbers in this chapter are computed after applying a common fiducial selection: reconstructed jet transverse momentum $p_T > 15\,\mathrm{GeV}$, $|\eta| < 2.5$, with the calorimeter crack region $1.37 < |\eta| < 1.57$ excluded. For genuine taus, the reconstructed prong multiplicity is further required to match the truth prong multiplicity, so that 1-prong and 3-prong categories are not contaminated by prong mis-reconstruction; this requirement is not applied to the jet and electron backgrounds, which are selected purely on the *reconstructed* prong multiplicity. Table 6.2 summarises the resulting sample sizes.

Table 6.2.: Fiducial event counts used in the performance studies of this chapter, before (top row) and after (bottom two rows) splitting by reconstructed prong multiplicity.

Selection	Tau	QCD jet	Electron
Fiducial (inclusive)	460,743	469,349	460,305
1-prong (reconstructed)	321,122	24,669	422,444
3-prong (reconstructed)	82,900	37,794	628

Note the very different background sample sizes between the jet-fake and electron-fake studies, and especially between the 1-prong and 3-prong electron-fake samples (422,444 vs. 628 fiducial electrons): this directly explains why the 3-prong electron-veto working points are statistics-limited while the 1-prong ones are not (see Section 6.5). This is not an artefact of the particular Monte Carlo samples used, but reflects the underlying physics: since a genuine electron produces only a single reconstructed track, it can only fake the 1-prong tau topology, and a reconstructed 3-prong electron candidate can only arise from rare combinatorial effects such as an accidental overlap with unrelated tracks or a photon conversion (Section 2.3.1 of Chapter 2). The small 3-prong electron-fake sample is therefore an expected consequence of this physics, not merely a limitation of the available statistics.

6.3. ROC Curves: Jet-Fake and Electron-Fake Rejection

Figures 6.1 and 6.2 show the background rejection as a function of tau efficiency for 1-prong and 3-prong taus, respectively. In each figure, the left panel shows rejection against QCD-jet fakes for RNN, GNTau and GNTau_eVeto, and the right panel shows rejection against electron fakes (eVeto) for RNN and GNTau_eVeto only, since the baseline GNTau model does not produce an electron-veto discriminant. The lower sub-panel in each case shows the ratio of each curve to the RNN reference. Dotted segments and upward-pointing markers denote statistics-limited regions, as defined in Section 6.1.

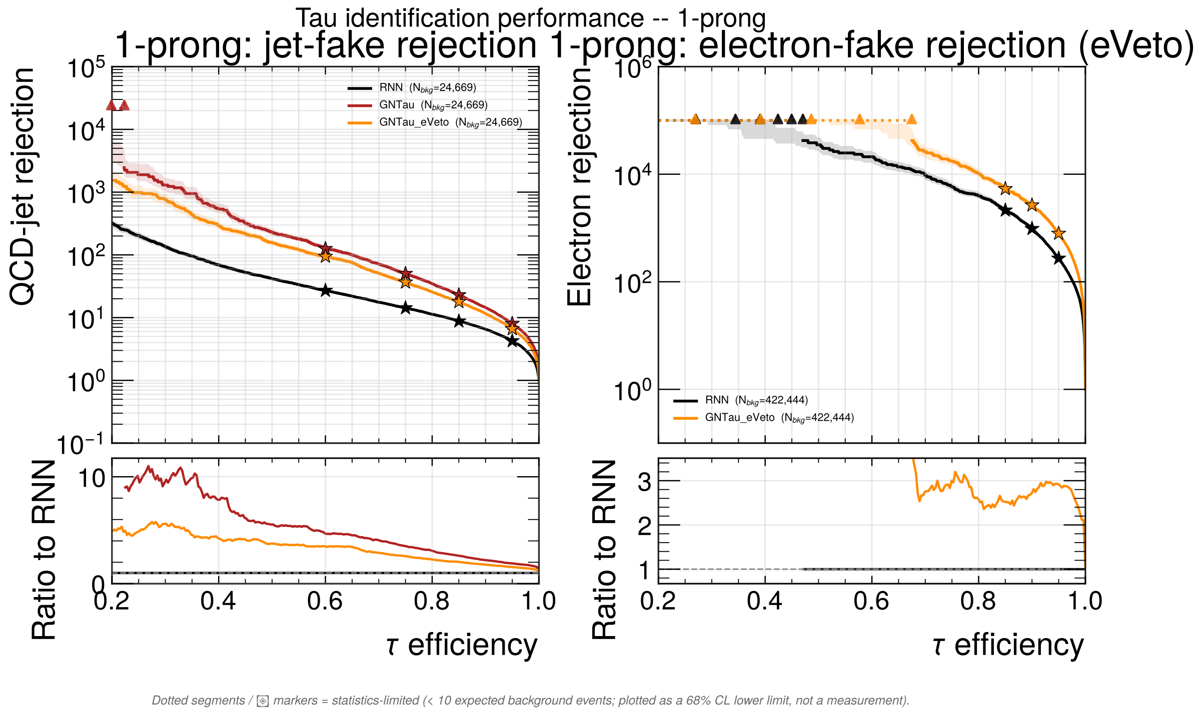


Figure 6.1.: 1-prong tau identification performance: jet-fake rejection (left) and electron-fake rejection (right) as a function of tau efficiency, for RNN, GNTau and GNTau_eVeto. Background sample sizes: $N_{\mathrm{bkg}} = 24{,}669$ (jet fakes), $N_{\mathrm{bkg}} = 422{,}444$ (electron fakes).

Across both prong categories, GNTau improves the jet-fake rejection by roughly a factor of 3–7 relative to RNN at fixed tau efficiency (see Table 6.3 and Section 6.5 for the exact figures). Adding the electron-veto capability (GNTau_eVeto) recovers most, though not all, of this improvement in jet-fake rejection, while simultaneously providing an electron-fake rejection that matches or exceeds the RNN electron veto.

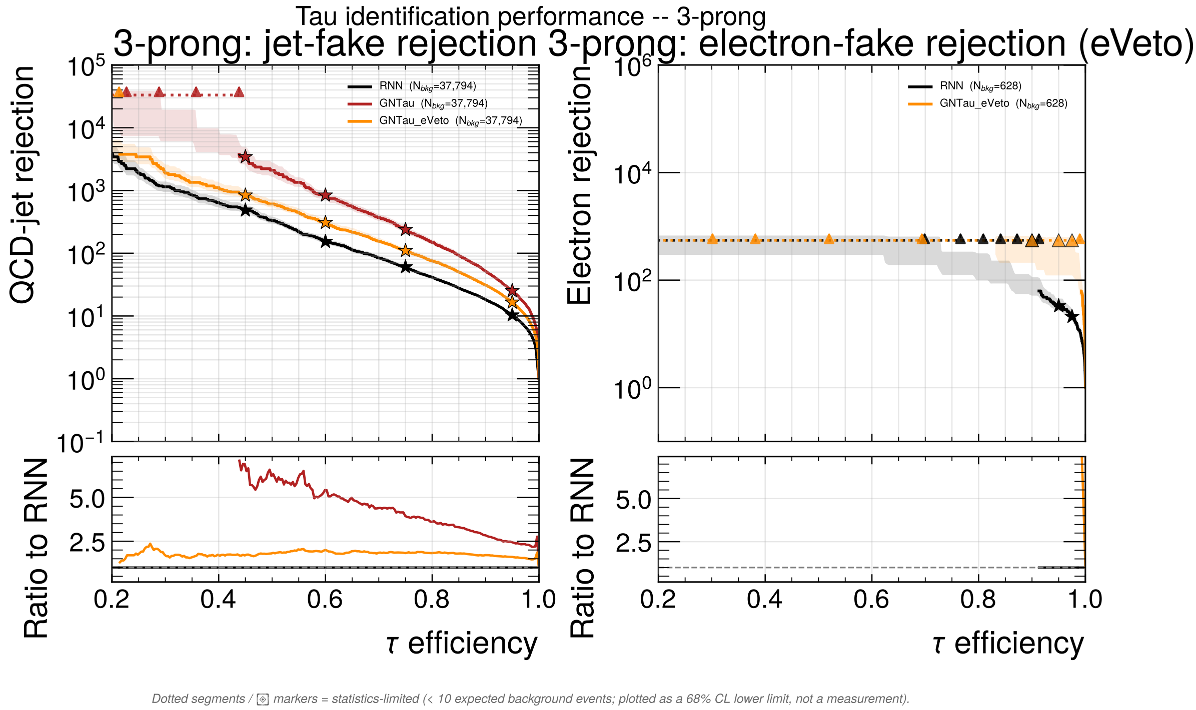


Figure 6.2.: 3-prong tau identification performance: jet-fake rejection (left) and electron-fake rejection (right) as a function of tau efficiency, for RNN, GNTau and GNTau_eVeto. Background sample sizes: $N_{\mathrm{bkg}} = 37{,}794$ (jet fakes), $N_{\mathrm{bkg}} = 628$ (electron fakes).

6.4. AUC Summary

Table 6.3 reports the area under the ROC curve for each model, prong category and fake type, together with the corresponding number of background events entering the calculation.

GNTau and GNTau_eVeto improve the AUC over RNN in every category. The margin is largest for jet-fake rejection (e.g. 0.9416 → 0.9754 for 1-prong, 0.9789 → 0.9922 for 3-prong), while the electron-veto AUCs were already close to unity for RNN and are improved further by GNTau_eVeto. Because AUC is dominated by the bulk of the ROC curve, it should be read together with the working-point rejections in Section 6.5, which are more directly relevant for an analysis operating at a specific efficiency point, and which show larger relative gains than the AUC numbers alone might suggest.

Table 6.3.: Area under the ROC curve (AUC) for each model, split by prong multiplicity and fake type. N_{bkg} is the number of fiducial background events (QCD jets or electrons) entering the calculation.

Model	Prong	Fake type	AUC	N_{bkg}
RNN	1	jet	0.9416	24,669
GNTau	1	jet	0.9754	24,669
GNTau_eVeto	1	jet	0.9690	24,669
RNN	3	jet	0.9789	37,794
GNTau	3	jet	0.9922	37,794
GNTau_eVeto	3	jet	0.9873	37,794
RNN	1	electron	0.9987	422,444
GNTau_eVeto	1	electron	0.9993	422,444
RNN	3	electron	0.9925	628
GNTau_eVeto	3	electron	0.9990	628

6.5. Working-Point Rejection Summary

Tables 6.4–6.7 list the background rejection at each nominal working point, for jet fakes and electron fakes, split by prong multiplicity. Entries marked with ">" are statistics-limited one-sided 68 % CL lower limits (Section 6.1), not point measurements.

Table 6.4.: Jet-fake rejection at each 1-prong tau-ID working point.

Model	60%	75%	85%	95%
RNN	27.0	14.3	8.8	4.3
GNTau	125.2	50.2	22.9	8.0
GNTau_eVeto	94.5	36.6	17.9	6.6

Table 6.5.: Jet-fake rejection at each 3-prong tau-ID working point.

Model	45%	60%	75%	95%
RNN	484.5	153.6	60.4	10.3
GNTau	3435.8	839.9	237.7	25.1
GNTau_eVeto	839.9	307.3	109.9	16.5

At fixed tau efficiency, GNTau improves the jet-fake rejection by a factor of ∼3–7 over RNN (e.g. a factor 4.6 at the 60 % 1-prong working point, and a factor 7.1 at the 45 % 3-prong working point), with the improvement shrinking towards the tightest, highest-efficiency working points as expected. Integrating the electron veto into the same network

Table 6.6.: Electron-fake (eVeto) rejection at each 1-prong working point.

Model	85%	90%	95%
RNN	2133.6	971.1	272.5
GNTau_eVeto	5347.4	2656.9	791.1

Table 6.7.: Electron-fake (eVeto) rejection at each 3-prong working point. All entries at these working points are statistics-limited given the small 3-prong electron-fake sample ($N_{\text{bkg}} = 628$); values are 68 % CL lower limits.

Model	90%	95%	97.5%
RNN	>551.7	33.1	20.9
GNTau_eVeto	>551.7	>551.7	>551.7

(GNTau_eVeto) reduces the jet-fake rejection somewhat relative to the jet-only GNTau baseline (e.g. 94.5 vs. 125.2 at the 1-prong 60 % working point), consistent with the network now sharing capacity between the tau/jet and tau/electron classification tasks, while still comfortably outperforming RNN, and while providing an electron rejection that is at least as good as, and in several working points markedly better than, the dedicated RNN electron veto.

For the 3-prong electron veto specifically, the entire working-point range tested is statistics-limited for GNTau_eVeto (Table 6.7): the measured rejection is consistent with being at least as large as 551.7 everywhere, without the current test sample being large enough to place a tighter upper bound. Note that the corresponding RNN entries at 95% and 97.5% *are* unambiguous measurements (fewer than 10 events survive, but the RNN rejection at these points is low enough that this is not a limit), which shows that the apparent GNTau_eVeto lower limits reflect an already very strong rejection rather than merely a wide statistical uncertainty band across the board.

6.6. Training Convergence

GNTau_eVeto is trained using the Salt training framework [43] (see Chapter 5, Section 4.5), which also produces the loss curves shown in this section directly from its training logs. Figure 6.3 shows the GNTau_eVeto training and validation loss as a function of epoch. The loss decreases sharply over the first three epochs and then flattens, with training and validation loss tracking each other closely throughout, indicating no significant overfitting up to the final epoch (30). The vertical dashed line marks a run boundary, i.e. where training was resumed from a checkpoint (e.g. after a restart or a change in job configuration) rather than a discontinuity in the optimisation itself; the loss curve is continuous across it. The final validation loss (≈ 0.098 at epoch 30) is used to select the checkpoint reported throughout this chapter.

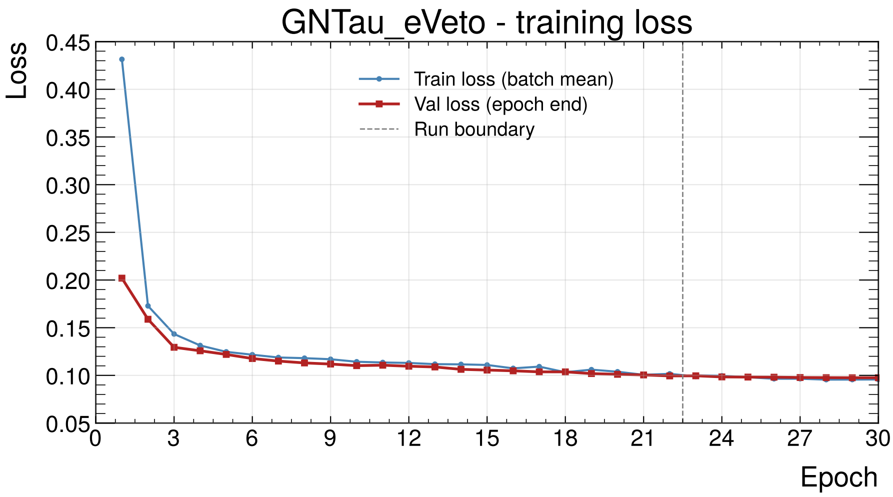


Figure 6.3.: Training (blue) and validation (red) loss of GNTau_eVeto as a function of training epoch. The dashed vertical line marks a run boundary (training resumed from a checkpoint), not a change in the loss definition.

6.7. Discriminant Distributions

Figures 6.4 and 6.5 show the tau-ID (tau-vs-jet) and eVeto discriminant distributions for signal and background, on the same log-odds scale as the ROC curves of Section 6.3, for 1-prong and 3-prong taus respectively. These distributions are the underlying quantities from which every rejection and AUC number in this chapter is derived, and are shown here mainly as a cross-check and for qualitative interpretation.

For both prong categories, the GN taggers' tau and background distributions are visibly better separated than the RNN distributions, consistent with the higher AUC and rejection values reported above; the effect is most pronounced for the 3-prong tau-ID discriminant, where the RNN tau and QCD distributions overlap over a much wider range than either GN model's.

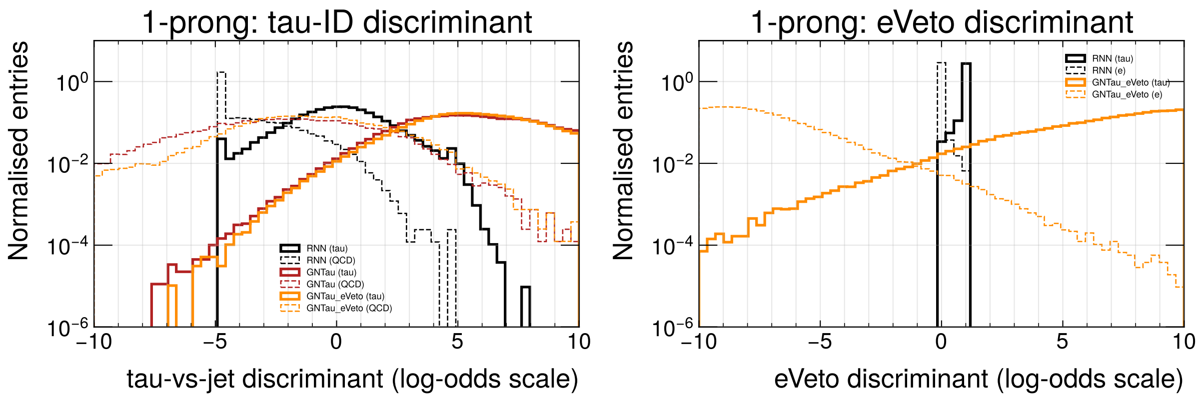


Figure 6.4.: 1-prong discriminant distributions on a log-odds scale: tau-ID (tau-vs-jet) discriminant for RNN, GNTau and GNTau_eVeto (left), and eVeto discriminant for RNN and GNTau_eVeto (right).

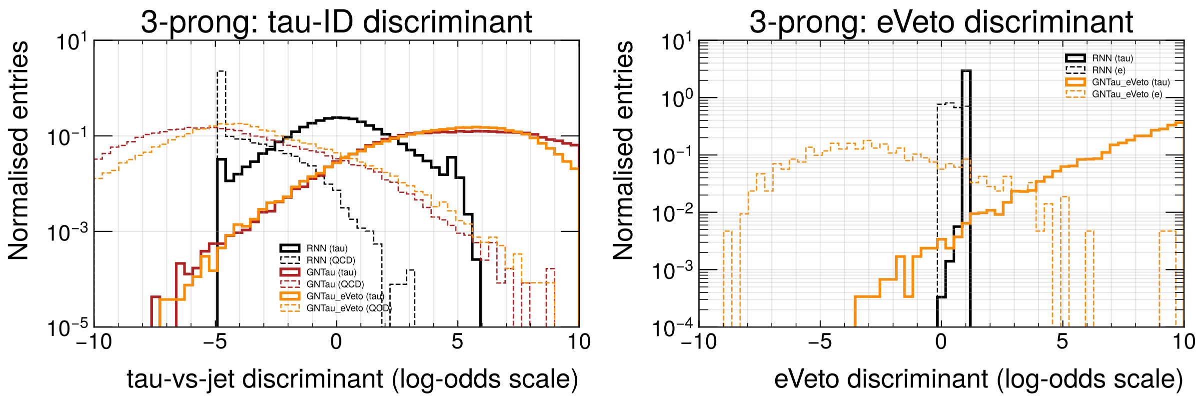


Figure 6.5.: 3-prong discriminant distributions on a log-odds scale: tau-ID (tau-vs-jet) discriminant for RNN, GNTau and GNTau_eVeto (left), and eVeto discriminant for RNN and GNTau_eVeto (right).

6.8. Correlation Between the Tau-ID and eVeto Discriminants

Since GNTau_eVeto produces both the tau-ID and eVeto discriminants from a single shared network (a 3-class softmax over tau/jet/electron), it is important to know how correlated the two outputs are: if they were highly correlated, selecting on both in sequence would add little on top of either individual selection; if they are largely independent, the two selections are closer to genuinely complementary information, and can be treated as approximately factorisable when combining working points. Figures 6.6 and 6.7 show the two discriminants against each other, split by truth class, together with the Pearson correlation coefficient ρ in each panel.

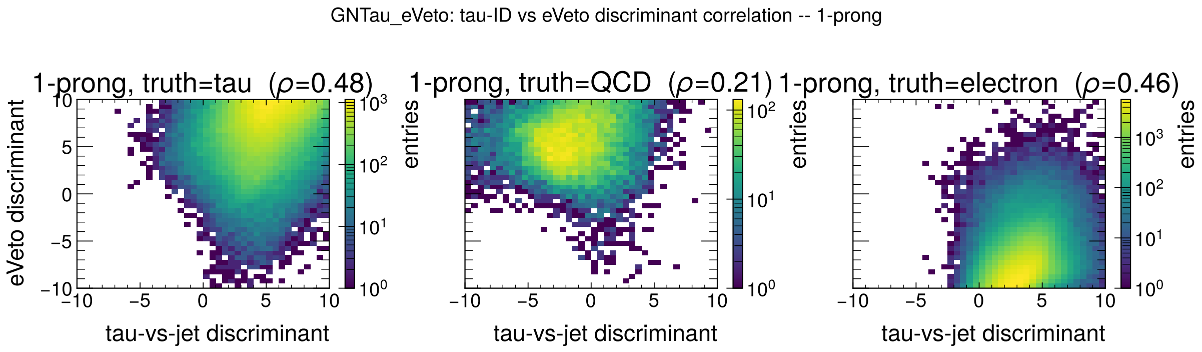


Figure 6.6.: 1-prong correlation between the GNTau_eVeto tau-ID and eVeto discriminants, split by truth class (tau, QCD jet, electron), with the Pearson correlation coefficient ρ given in each panel title.

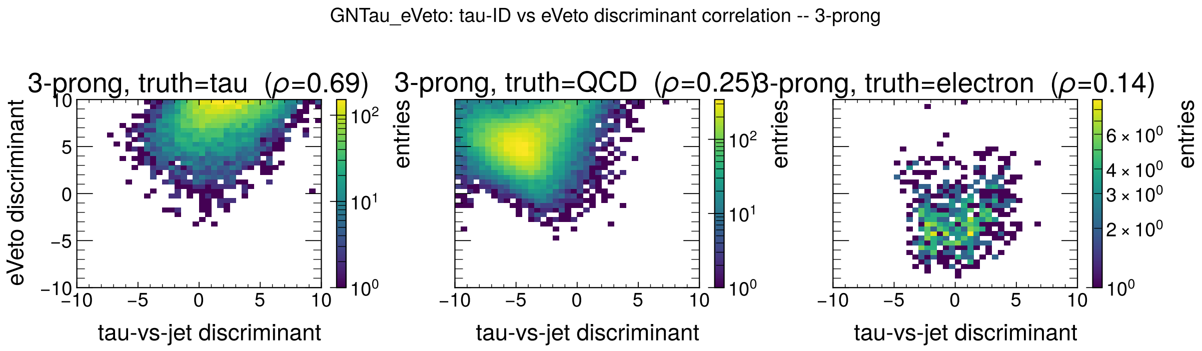


Figure 6.7.: 3-prong correlation between the GNTau_eVeto tau-ID and eVeto discriminants, split by truth class (tau, QCD jet, electron), with the Pearson correlation coefficient ρ given in each panel title.

The two discriminants are moderately, but not strongly, correlated for every truth class and both prong categories: $\rho = 0.48$ (tau), 0.21 (QCD) and 0.46 (electron) for 1-prong, and $\rho = 0.69$ (tau), 0.25 (QCD) and 0.14 (electron) for 3-prong. None of these values are close to ± 1, so the tau-ID and eVeto outputs of GNTau_eVeto carry meaningfully complementary information rather than being a relabelling of the same underlying score, supporting the choice of a single shared network for both tasks. At the same time, the correlation for genuine taus is not negligible, and is largest for 3-prong taus ($\rho = 0.69$). **This has a direct practical consequence: the combined efficiency of applying a tau-ID working point and an eVeto working point simultaneously cannot be obtained by simply multiplying the two one-dimensional signal efficiencies**, as that implicitly assumes independence. Any analysis that needs a combined operating point should instead read the joint (tau-ID, eVeto) signal efficiency directly off the 2-D distribution (or an equivalent lookup table) at the two chosen thresholds.

6.9. Kinematic Flatness of the Working Points

The ROC curves and working-point tables above are all evaluated inclusively in p_T and $|\eta|$. For a working point to be usable in an analysis without a dedicated correction, the corresponding signal efficiency (and, ideally, the background rejection) should remain reasonably flat as a function of the kinematics of the tau candidate. Here, a single discriminant threshold is calibrated once, on the inclusive sample, to give the nominal working-point efficiency quoted in each figure, and is then re-applied without modification in bins of p_T or $|\eta|$: any drift in the resulting efficiency reflects the tagger's kinematic dependence rather than a change in the threshold itself. Figures 6.8 –6.14 show the results for the tau-ID (jet-fake) and eVeto (electron-fake) working points, respectively.

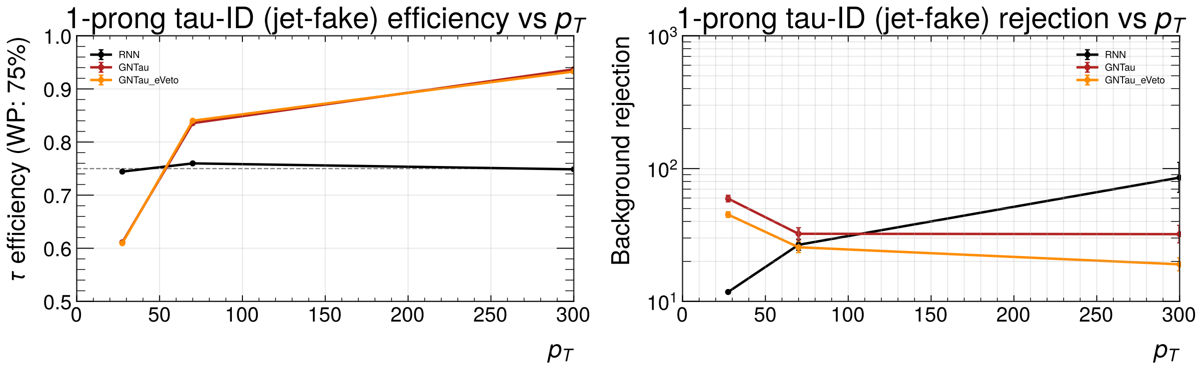


Figure 6.8.: 1-prong tau-ID working point (nominal efficiency 75%): signal efficiency (left) and jet-fake rejection (right) as a function of p_T, at a threshold fixed from the inclusive sample.

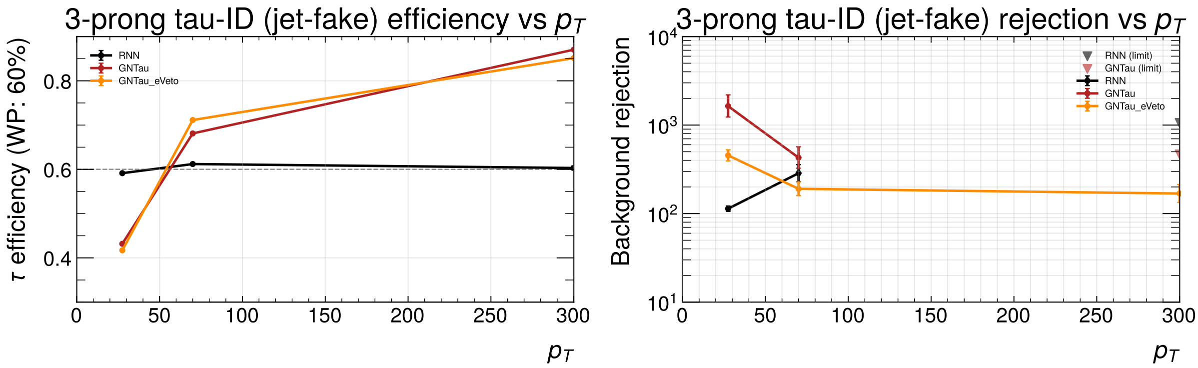


Figure 6.9.: 3-prong tau-ID working point (nominal efficiency 60%): signal efficiency (left) and jet-fake rejection (right) as a function of p_T, at a threshold fixed from the inclusive sample.

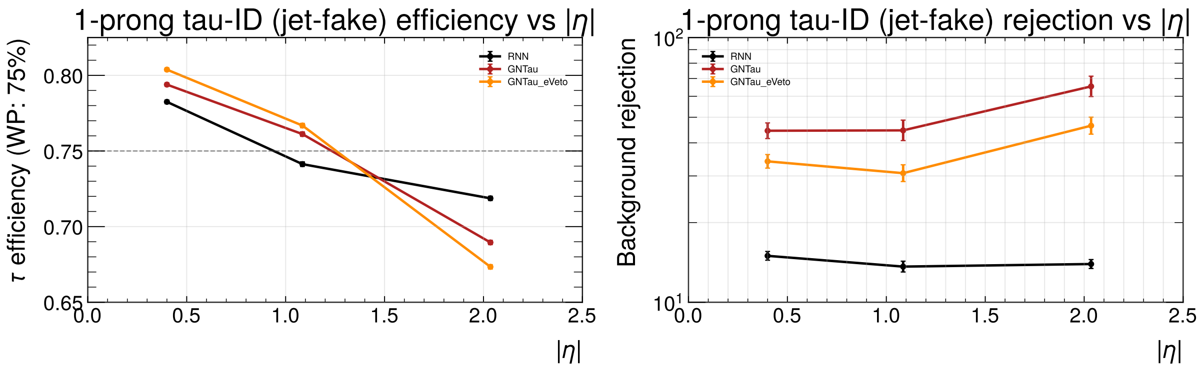


Figure 6.10.: 1-prong tau-ID working point (nominal efficiency 75%): signal efficiency (left) and jet-fake rejection (right) as a function of $|\eta|$, at a threshold fixed from the inclusive sample.

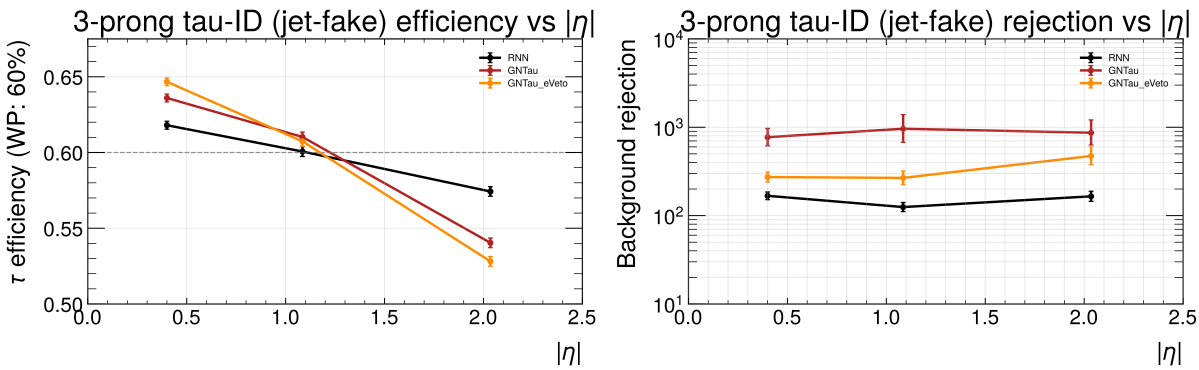


Figure 6.11.: 3-prong tau-ID working point (nominal efficiency 60%): signal efficiency (left) and jet-fake rejection (right) as a function of $|\eta|$, at a threshold fixed from the inclusive sample.

Low-p_T efficiency undershoot for the GN taggers. The single most notable feature in these flatness studies is at the lowest p_T bin ($p_T \approx$ 25–30 GeV, Figures 6.8 and 6.9): both GNTau and GNTau_eVeto fall well *below* their nominal working-point efficiency (1-prong: $\approx$ 61% measured vs. the 75% nominal target; 3-prong: $\approx$ 42–43% measured vs. the 60% nominal target), whereas RNN remains close to its nominal efficiency in the same bin. Because the threshold is calibrated once on the inclusive sample, some kinematic dependence is expected by construction; the point here is that the *size* of the shortfall is substantially larger for the GN taggers than for RNN at low p_T. In practice, an analysis applying this inclusive working point to a low-p_T-enriched selection would obtain a lower tau efficiency than the nominal working-point label suggests, unless a p_T-dependent threshold is used instead.

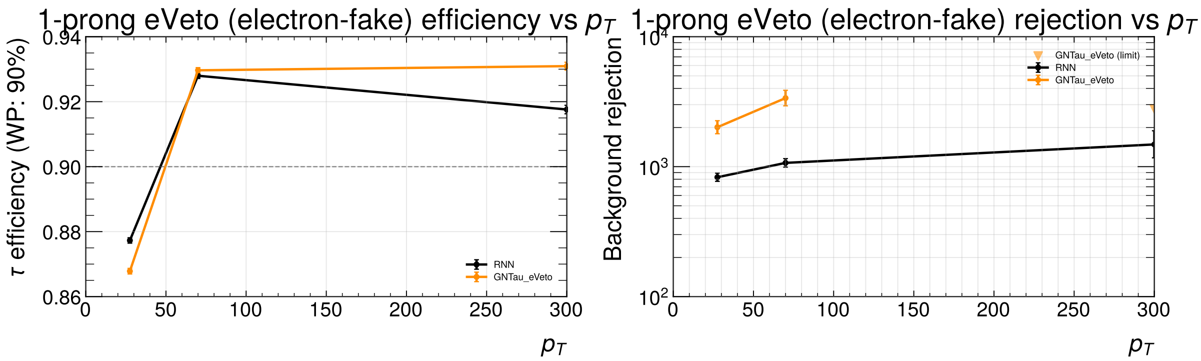


Figure 6.12.: 1-prong eVeto working point (nominal efficiency 90%): signal efficiency (left) and electron-fake rejection (right) as a function of p_T, at a threshold fixed from the inclusive sample.

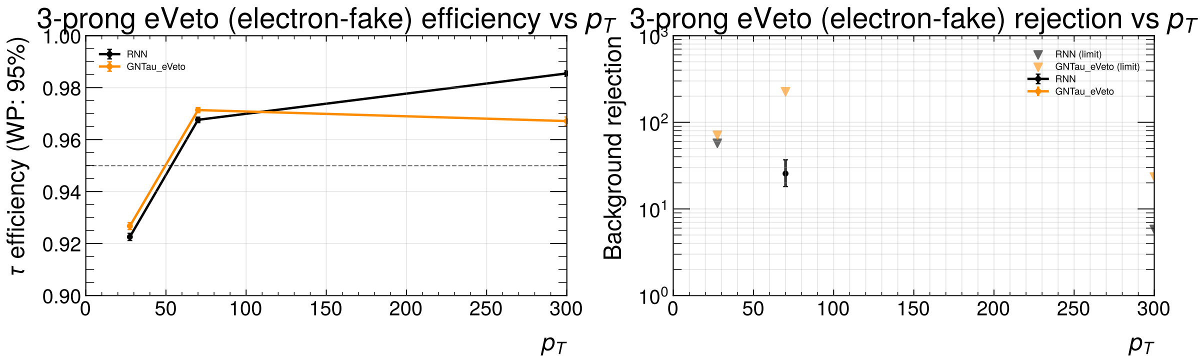


Figure 6.13.: 3-prong eVeto working point (nominal efficiency 95%): signal efficiency (left) and electron-fake rejection (right) as a function of p_T, at a threshold fixed from the inclusive sample. Most points are statistics-limited given the small 3-prong electron-fake sample (Section 6.2).

High-p_T/high-$|\eta|$ reversal in 1-prong jet-fake rejection. In the highest kinematic bins for 1-prong taus, $p_T \approx 300\,\text{GeV}$ (Figure 6.8) and $|\eta| \approx 2.0$ (Figure 6.10) — RNN's jet-fake rejection overtakes both GN taggers: RNN reaches a rejection of $\approx$ 85–90 while GNTau plateaus around $\approx$ 33 and GNTau_eVeto *decreases* to $\approx$ 20. This is the opposite ordering to the inclusive result of Section 6.5 and is, to our knowledge, the only region in this entire study where RNN outperforms both GN taggers. Before this is reported as a genuine physics conclusion (e.g. a limitation of the graph-based features at high $p_T/|\eta|$), the per-bin background counts in this region should be checked explicitly, since a small residual background count could produce a similar effect statistically; this is flagged here as an open item for further investigation rather than an established result.

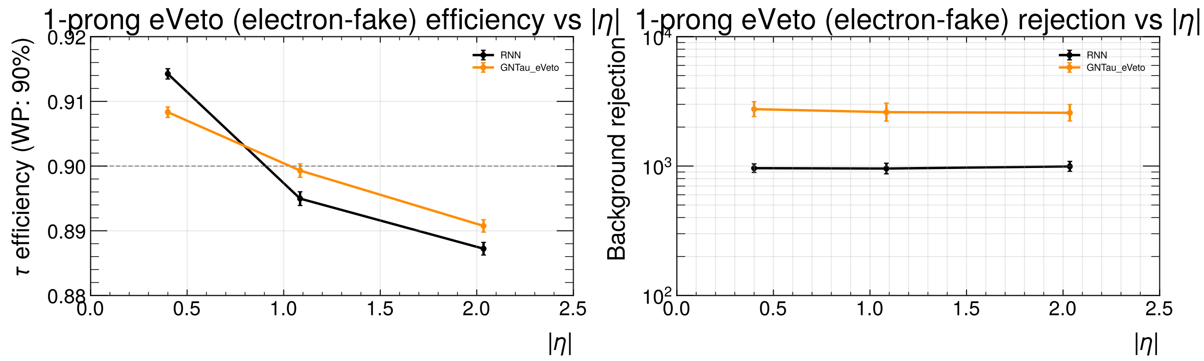


Figure 6.14.: 1-prong eVeto working point (nominal efficiency 90%): signal efficiency (left) and electron-fake rejection (right) as a function of $|\eta|$, at a threshold fixed from the inclusive sample.

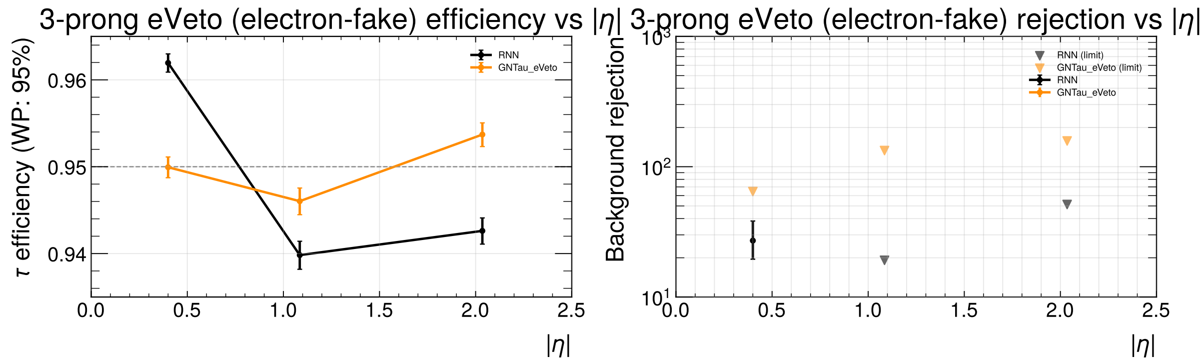


Figure 6.15.: 3-prong eVeto working point (nominal efficiency 95%): signal efficiency (left) and electron-fake rejection (right) as a function of $|\eta|$, at a threshold fixed from the inclusive sample. Most points are statistics-limited given the small 3-prong electron-fake sample (Section 6.2).

eVeto flatness. The 1-prong eVeto working point (Figures 6.12 and 6.14) is comparatively well behaved: GNTau_eVeto tracks RNN's kinematic dependence closely and maintains an approximately constant factor $\sim$ 2–3$\times$ higher rejection across the full p_T and $|\eta|$ range probed, which is the desired outcome for a working point intended for direct use in an analysis. The corresponding 3-prong eVeto slices (Figures 6.13 and 6.15) are, as anticipated from the small 3-prong electron-fake sample, dominated by statistics-limited points in most kinematic bins; they are shown here as a qualitative consistency check (not inconsistent with a flat response) rather than a quantitative flatness measurement.

7. Conclusion

This thesis presents the development and evaluation of GNTau_eVeto, a Graph Neural Network capable of simultaneously distinguishing hadronically decaying tau leptons from both QCD jets and electrons. The motivation for this work was to replace the previously separate jet- and electron-rejection networks with a single unified model, while retaining competitive classification performance. Summary of results. Both the jet-only GNTau model and the combined GNTau_eVeto model outperform the existing RNN baseline across all prong multiplicities and fake-background types studied, in terms of both AUC (Table 6.3) and background rejection at the nominal working points (Tables 6.4–6.7). Combining the tau-identification and electron-veto tasks into a single network comes with a modest reduction in jet-fake rejection relative to the jet-only GNTau baseline; however, this is offset by removing the need for a separate electron-veto model, and by an electron-fake rejection that is competitive with, or better than, the dedicated RNN electron veto. The 3-prong electron-veto working points remain limited by the size of the available electron-fake test sample rather than by the discriminating power of the model itself. Overall, these results show that a unified Graph Neural Network architecture can successfully replace the previous two-network approach while maintaining, and in most respects improving on, its performance. Two findings from the kinematic-flatness studies (Section 6.9) qualify this overall picture and should be kept in mind for any downstream use of these working points: the GN taggers' efficiency falls short of the nominal working-point target at low p_T by a larger margin than RNN's, and RNN unexpectedly overtakes both GN taggers in jet-fake rejection at the highest p_T and $|\eta|$ for 1-prong taus, a region that warrants further investigation. In addition, the moderate correlation between the tau-ID and eVeto discriminants for genuine taus (Section 6.8) means that combined working-point efficiencies must be read off the joint distribution rather than approximated as a product of the two individual efficiencies.

8. Outlook

Although GNTau_eVeto already demonstrates competitive performance, several directions could further improve its classification capabilities. More training data. The current model was trained using approximately 3.5 million jets per class. Substantially larger datasets are already available, roughly 55 million hadronically decaying tau candidates, 770 million QCD dijet samples, and 12 million electron candidates. Since increasing the training set from one million to 3.5 million samples per class already produced a noticeable performance gain in this work, scaling up further is expected to improve generalization performance even more. Finer treatment of the electron background. Rather than representing all electrons as a single class, the electron category could be split according to production mechanism, for instance, separating electrons from electroweak boson decays from those produced in tau decays. Although both categories produce similar calorimeter signatures, electrons from tau decays originate from a displaced secondary vertex due to the tau lepton's finite lifetime and therefore typically have larger impact parameters. Exploiting this distinction explicitly could allow the network to make better use of these differences and improve overall classification performance. Input feature optimization. A systematic study of the currently used input variables could identify redundant or weakly informative features and simplify the input representation. At the same time, additional detector-level variables that better capture the differences between hadronic tau decays, QCD jets, and electrons could be explored to further improve discrimination power. Such studies would also offer insight into which detector information is most relevant for Graph Neural Network-based tau identification. Further investigation of the high-p_T high-$|\eta|$ region. The unexpected reversal in relative jet-rejection performance between RNN and the GN taggers at the kinematic extremes for 1-prong taus deserves dedicated study, both to understand its origin and to assess whether it affects the reliability of the derived working points in that regime. Taken together, this work demonstrates the feasibility and promise of Graph Neural Networks for simultaneous multi-class tau identification and provides a solid foundation on which these further improvements can build.

A. Appendix

A.1. Tables of input variables for Training GNTau_eVeto

The jet-level variables (Table A.1) describe global kinematic and shower-shape properties of the reconstructed tau candidate as a whole.

The track-level variables (Table A.2) characterize the individual charged-particle tracks associated with the tau candidate, including their kinematics, impact-parameter significances relative to the tau vertex, and hit-pattern information from the inner-detector subsystems.

The calorimeter-cluster variables (Table A.3) describe the energy deposits and shower shapes of the topological clusters associated with the tau candidate, and are particularly relevant for separating hadronic and electromagnetic energy depositions.

The neutral particle-flow object (PFO) variables (Table A.4) describe reconstructed neutral hadronic energy deposits, combining calorimeter-cluster kinematics with cell-based shower-shape moments.

The photon "shot" PFO variables (Table A.5) describe reconstructed sub-cluster photon candidates within the tau shower, which are used to resolve individual photons from π^0 decays.

Table A.1.: Jet-level input variables.

Variable	Unit
TauJets_eta	–
TauJets_pt	GeV
TauJets_centFrac	–
TauJets_sumEMCellEtOverLeadTrkPt	–
TauJets_dRmax	–
TauJets_trFlightPathSig	–
TauJets_etOverPtLeadTrk	–
TauJets_ptRatioEflowApprox	–
TauJets_mEflowApprox	GeV
TauJets_massTrkSys	GeV
TauJets_SumPtTrkFrac	–
TauJets_EMPOverTrkSysP	–
TauJets_isolFrac	–

Table A.2.: Track-level input variables.

Variable	Unit
TauTracks_d0TJVA	mm
TauTracks_d0SigTJVA	–
TauTracks_z0sinthetaTJVA	mm
TauTracks_z0sinthetaSigTJVA	–
TauTracks_z0TJVA	mm
TauTracks_dz0_TV_PV0	mm
TauTracks_d0_old	mm
TauTracks_dEta	–
TauTracks_dPhi	rad
TauTracks_dRJetSeedAxis	–
TauTracks_dRIntermediateAxis	–
TauTracks_dRDetectorAxis	–
TauTracks_dEtaJetSeedAxis	–
TauTracks_dPhiJetSeedAxis	rad
TauTracks_trackPt	GeV
TauTracks_qOverP	1/GeV
TauTracks_theta	rad
TauTracks_charge	e
TauTracks_rConv	mm
TauTracks_rConvII	mm
TauTracks_numberOfInnermostPixelLayerHits	–
TauTracks_numberOfPixelHits	–
TauTracks_numberOfPixelSharedHits	–
TauTracks_numberOfPixelDeadSensors	–
TauTracks_numberOfPixelHoles	–
TauTracks_numberOfContribPixelLayers	–
TauTracks_nPixHits	–
TauTracks_nIBLHitsAndExp	–
TauTracks_expectInnermostPixelLayerHit	–
TauTracks_expectNextToInnermostPixelLayerHit	–
TauTracks_numberOfSCTHits	–
TauTracks_numberOfSCTSharedHits	–
TauTracks_numberOfSCTDeadSensors	–
TauTracks_numberOfSCTHoles	–
TauTracks_nSCTHits	–
TauTracks_nSiHits	–
TauTracks_numberOfTRTHighThresholdHits	–
TauTracks_numberOfTRTHits	–
TauTracks_eProbabilityHT_offl	–
TauTracks_eProbabilityNN_offl	–
TauTracks_eProbabilityNNorHT_offl	–

Table A.3.: Calorimeter-cluster input variables.

Variable	Unit
TauClusters_cls_et	GeV
TauClusters_cls_dPhi	rad
TauClusters_cls_dEta	–
TauClusters_cls_SECOND_R	mm^2
TauClusters_cls_SECOND_LAMBDA	mm^2
TauClusters_cls_CENTER_LAMBDA	mm
TauClusters_cls_FIRST_ENG_DENS	GeV/mm^3
TauClusters_cls_EM_PROBABILITY	–
TauClusters_cls_CENTER_MAG	mm

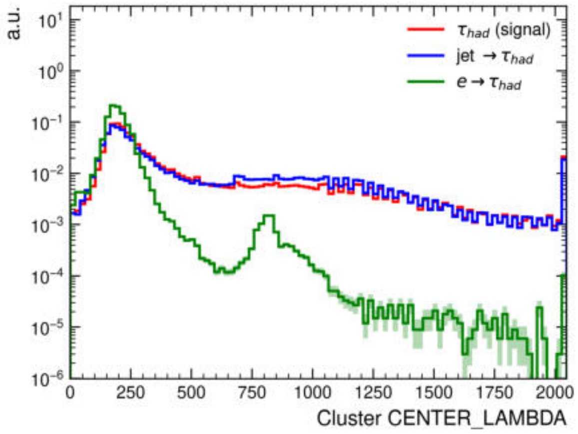
Table A.4.: Neutral particle-flow object (PFO) input variables.

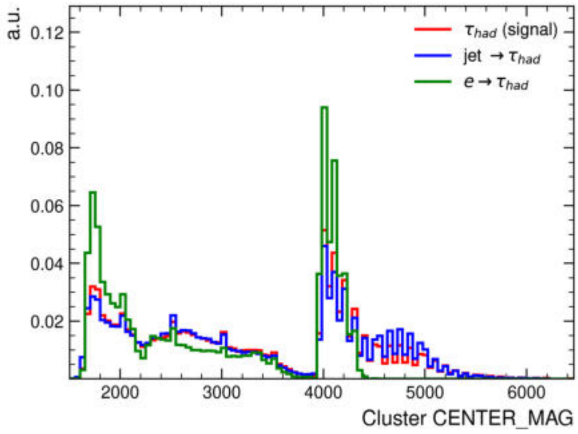
Variable	Unit
pfo_neutral_e	GeV
pfo_neutral_eta	–
pfo_neutral_phi	rad
pfo_neutral_pT	GeV
pfo_neutral_cellBased_CENTER_LAMBDA	mm
pfo_neutral_cellBased_SECOND_R	mm^2
pfo_neutral_cellBased_SECOND_LAMBDA	mm^2
pfo_neutral_cellBased_ENG_FRAC_CORE	–
pfo_neutral_cellBased_ENG_FRAC_EM	–
pfo_neutral_cellBased_LATERAL	–
pfo_neutral_cellBased_LONGITUDINAL	–

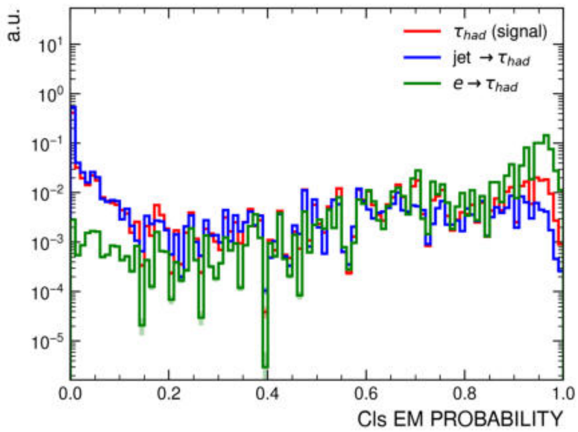
Table A.5.: Photon "shot" PFO input variables.

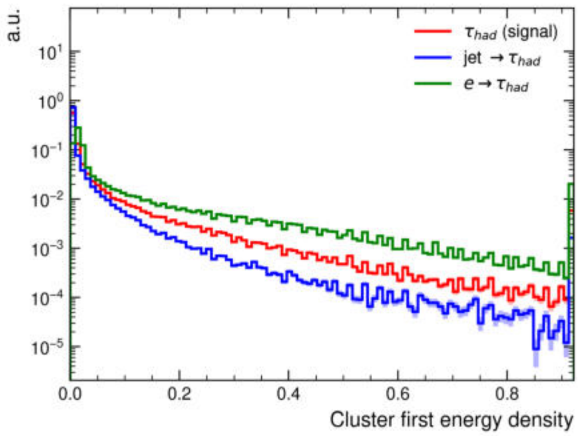
Variable	Unit
pfo_shots_e	GeV
pfo_shots_eta	–
pfo_shots_phi	rad
pfo_shots_nPhotons	–

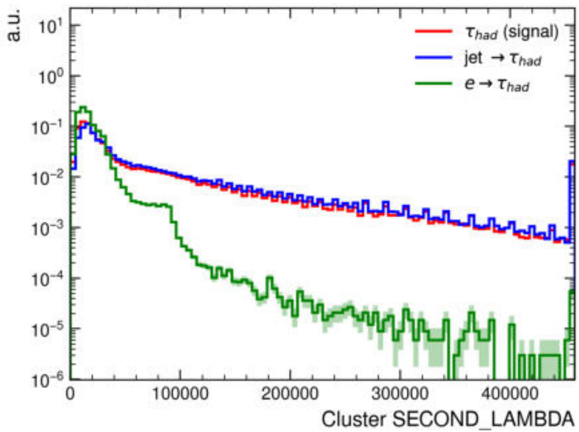
A.2. Normalised Input Variables after umami resampling

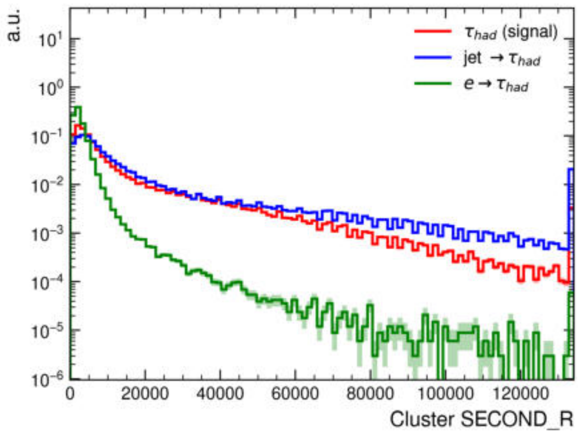


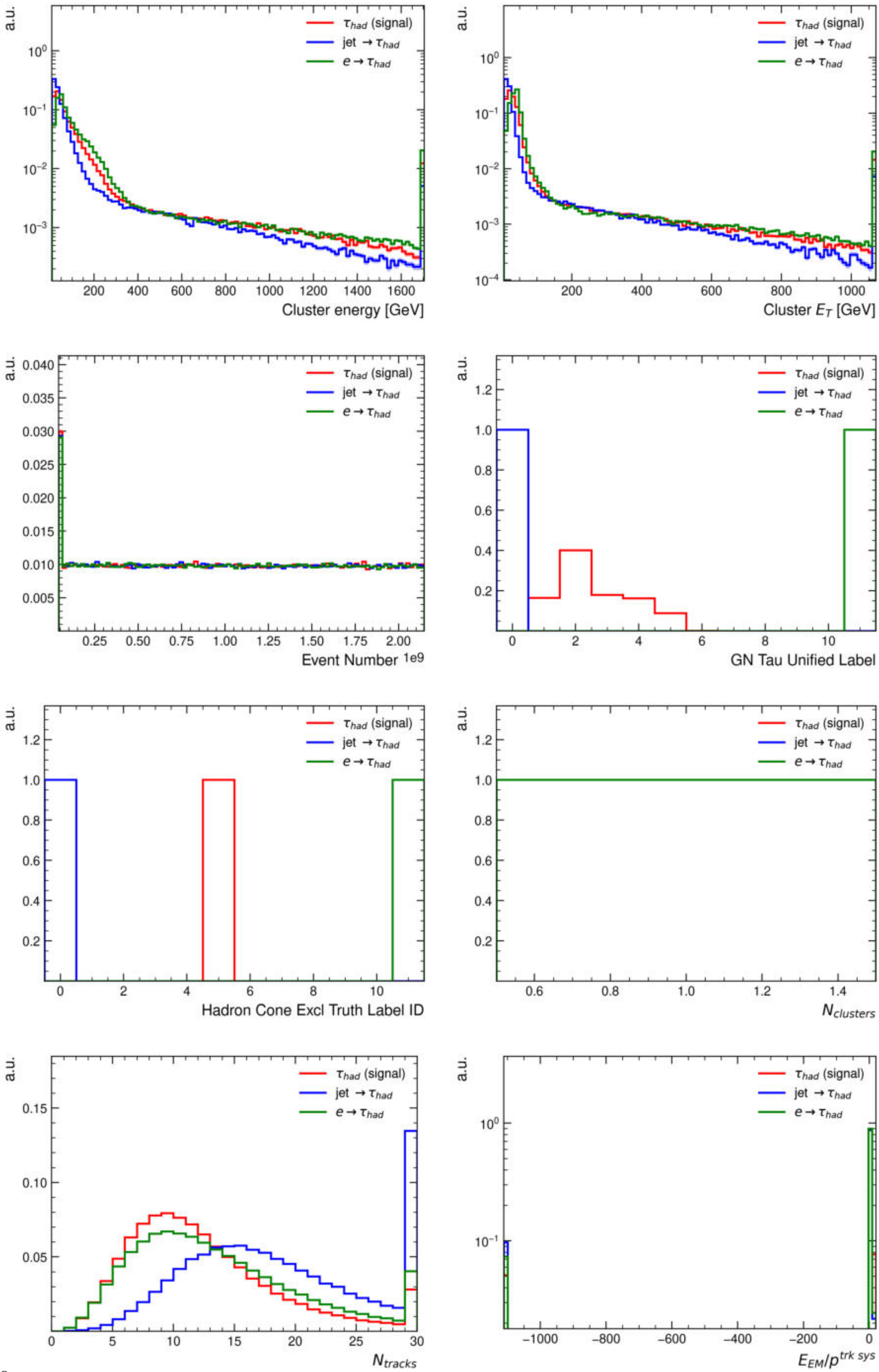


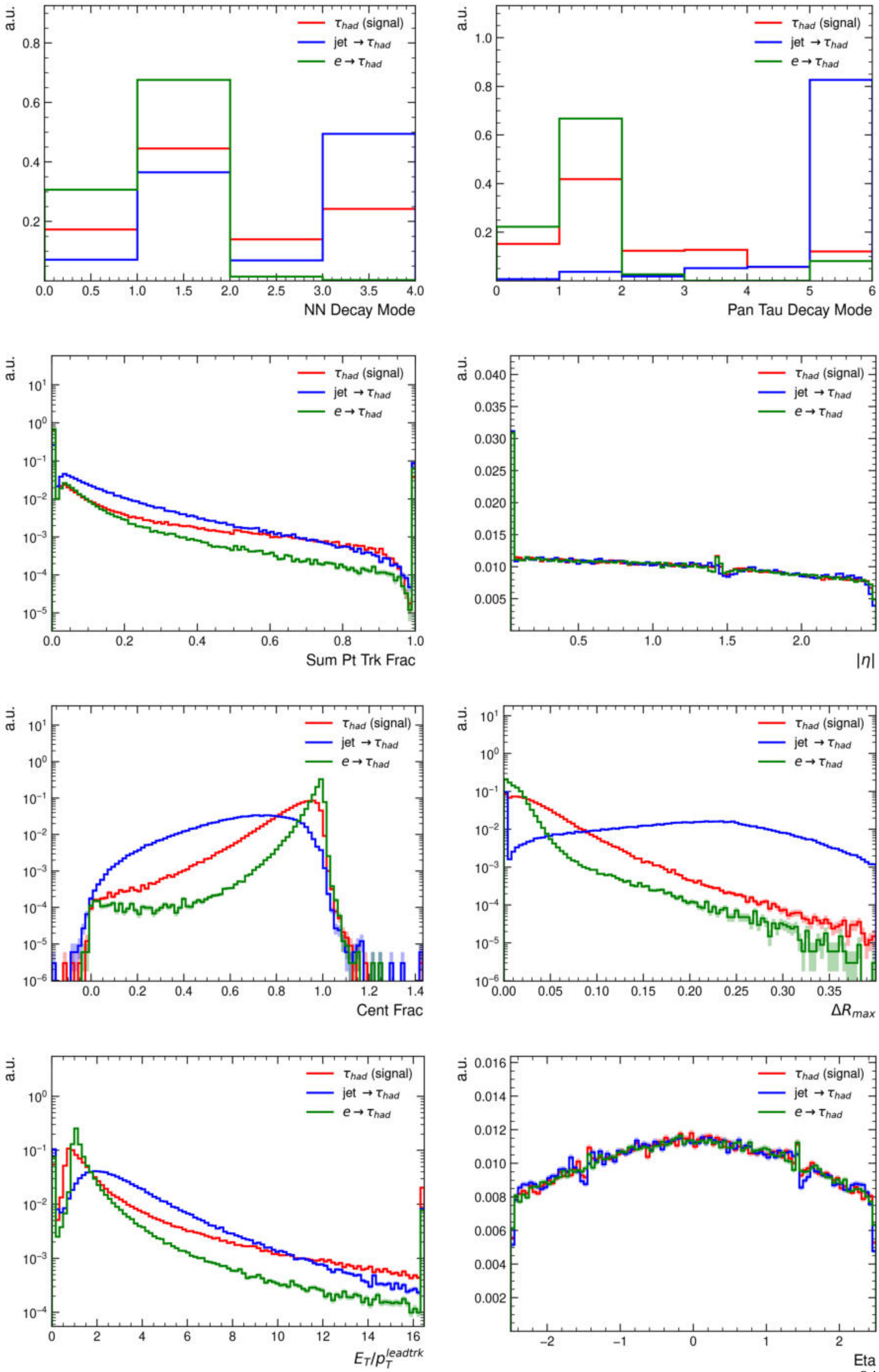


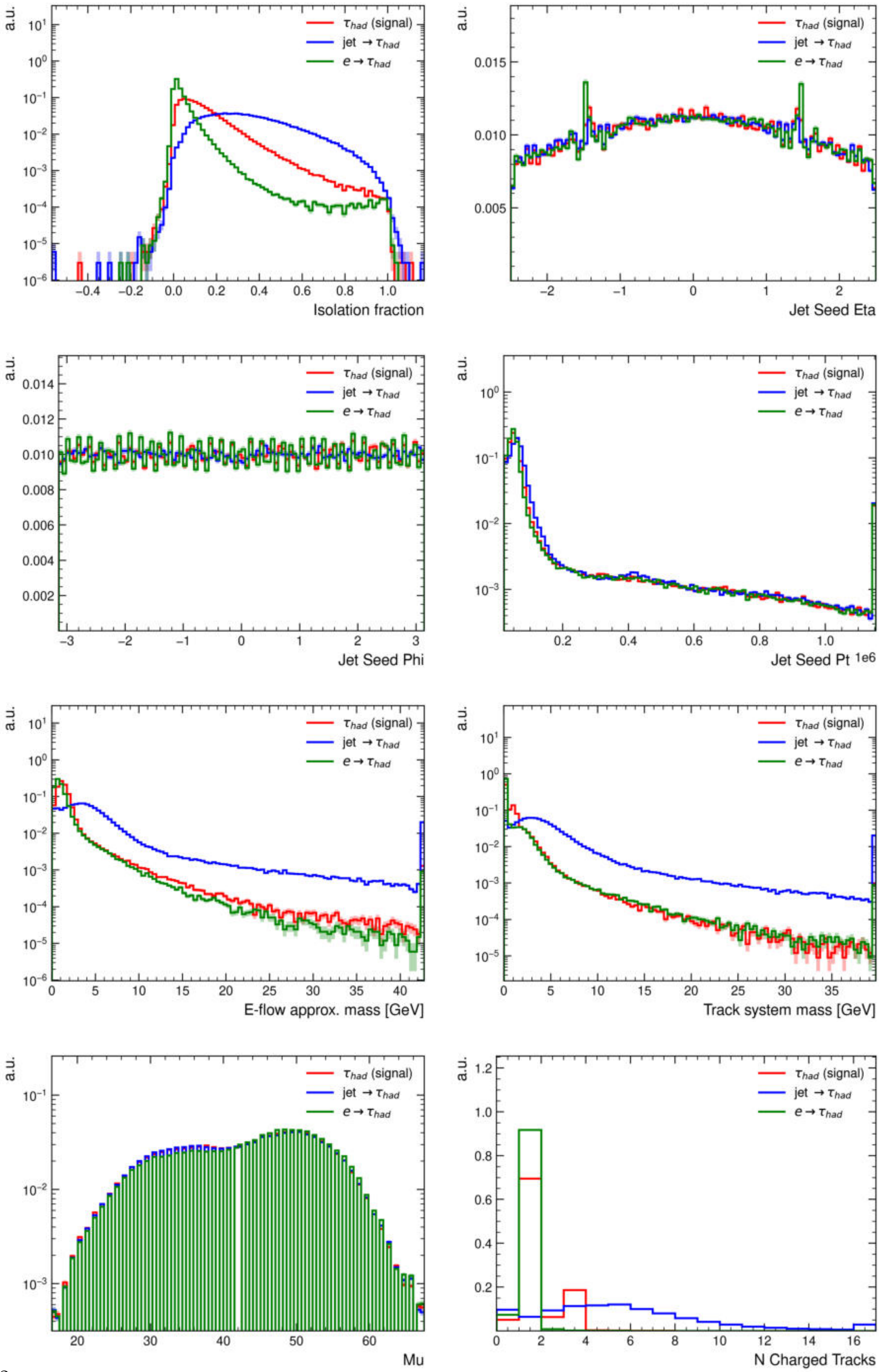


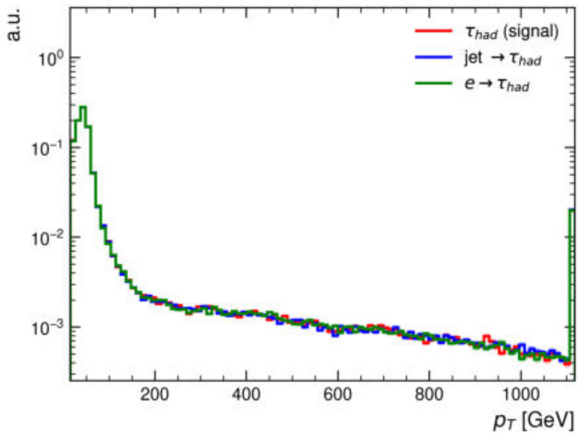


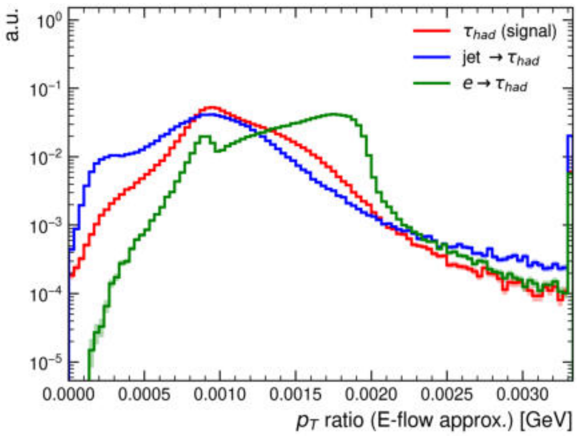


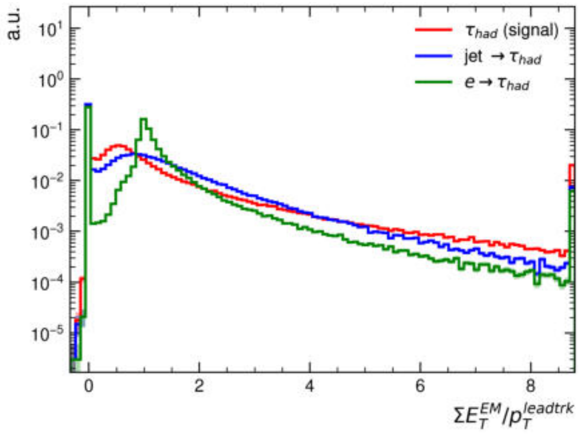

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Cluster energy [GeV]
Cluster E_T [GeV]
Event Number 1e9
GN Tau Unified Label
Hadron Cone Excl Truth Label ID
$N_{clusters}$
N_{tracks}
$E_{EM}/p^{trk\ sys}$

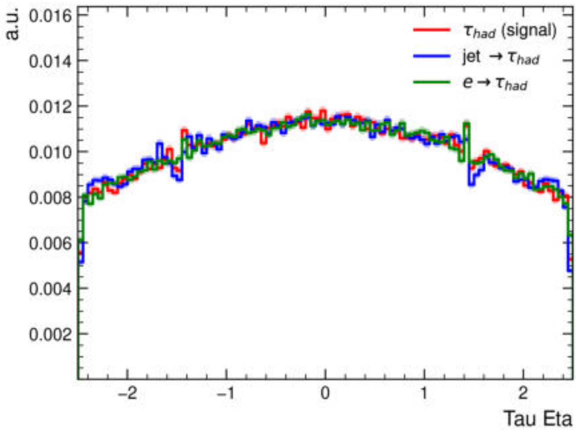

a.u.
τ_{had} (signal)
jet $\to \tau_{had}$
$e \to \tau_{had}$
NN Decay Mode
Pan Tau Decay Mode
Sum Pt Trk Frac
$|\eta|$
Cent Frac
ΔR_{max}
$E_T/p_T^{leadtrk}$
Eta

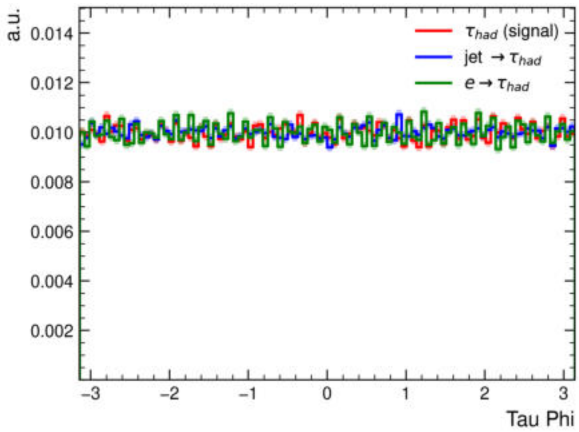

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Isolation fraction
Jet Seed Eta
Jet Seed Phi
Jet Seed Pt 1e6
E-flow approx. mass [GeV]
Track system mass [GeV]
Mu
N Charged Tracks

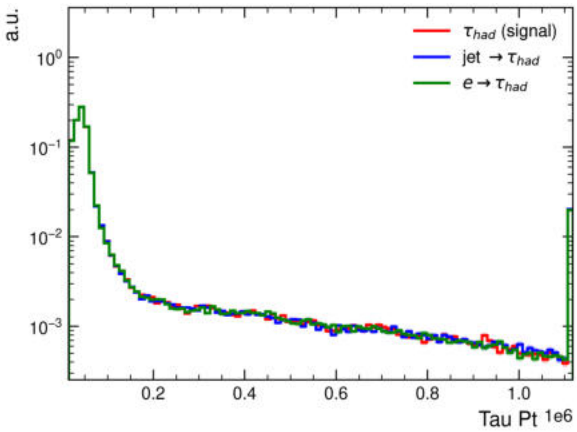

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
p_T [GeV]

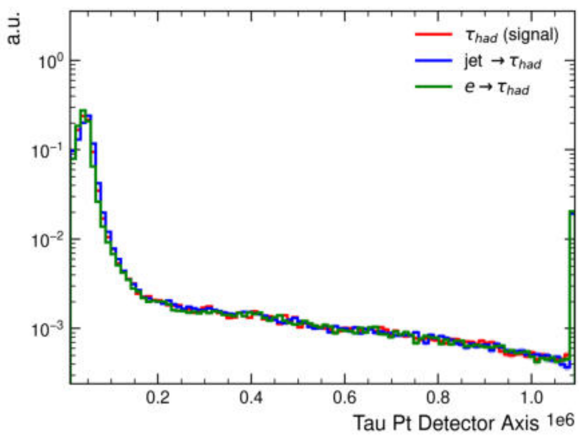

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
p_T ratio (E-flow approx.) [GeV]

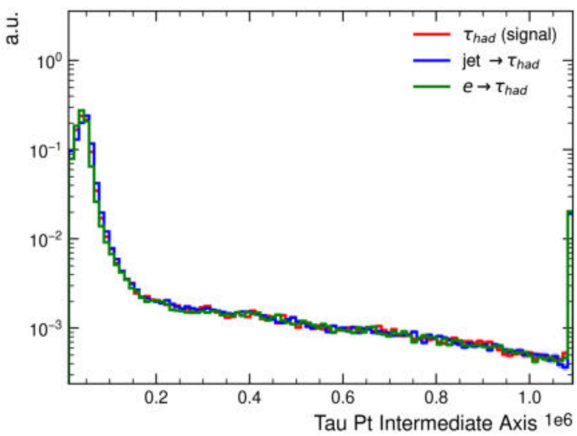

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
$\Sigma E_T^{EM}/p_T^{leadtrk}$

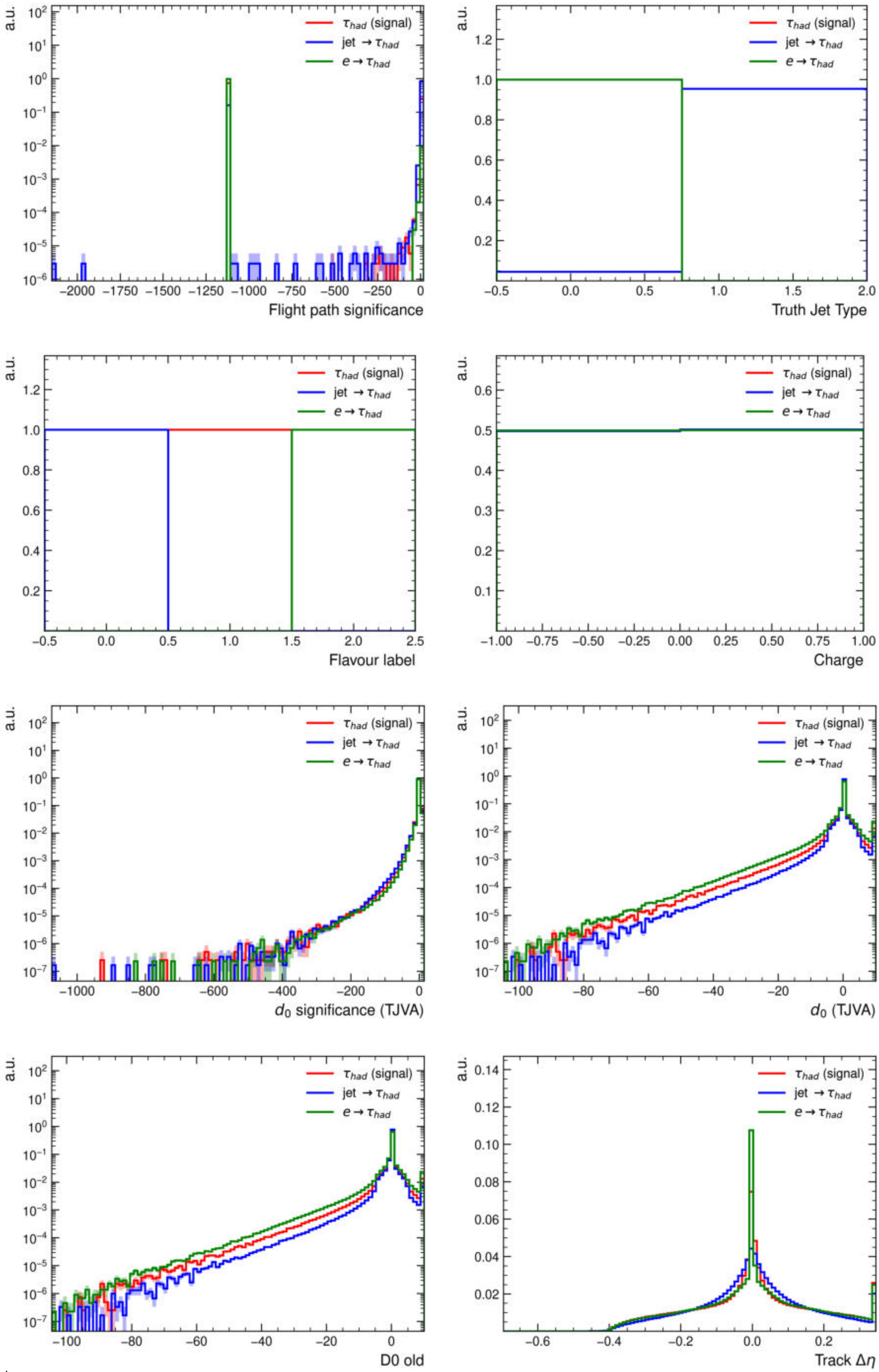

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Tau Eta


a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Tau Phi

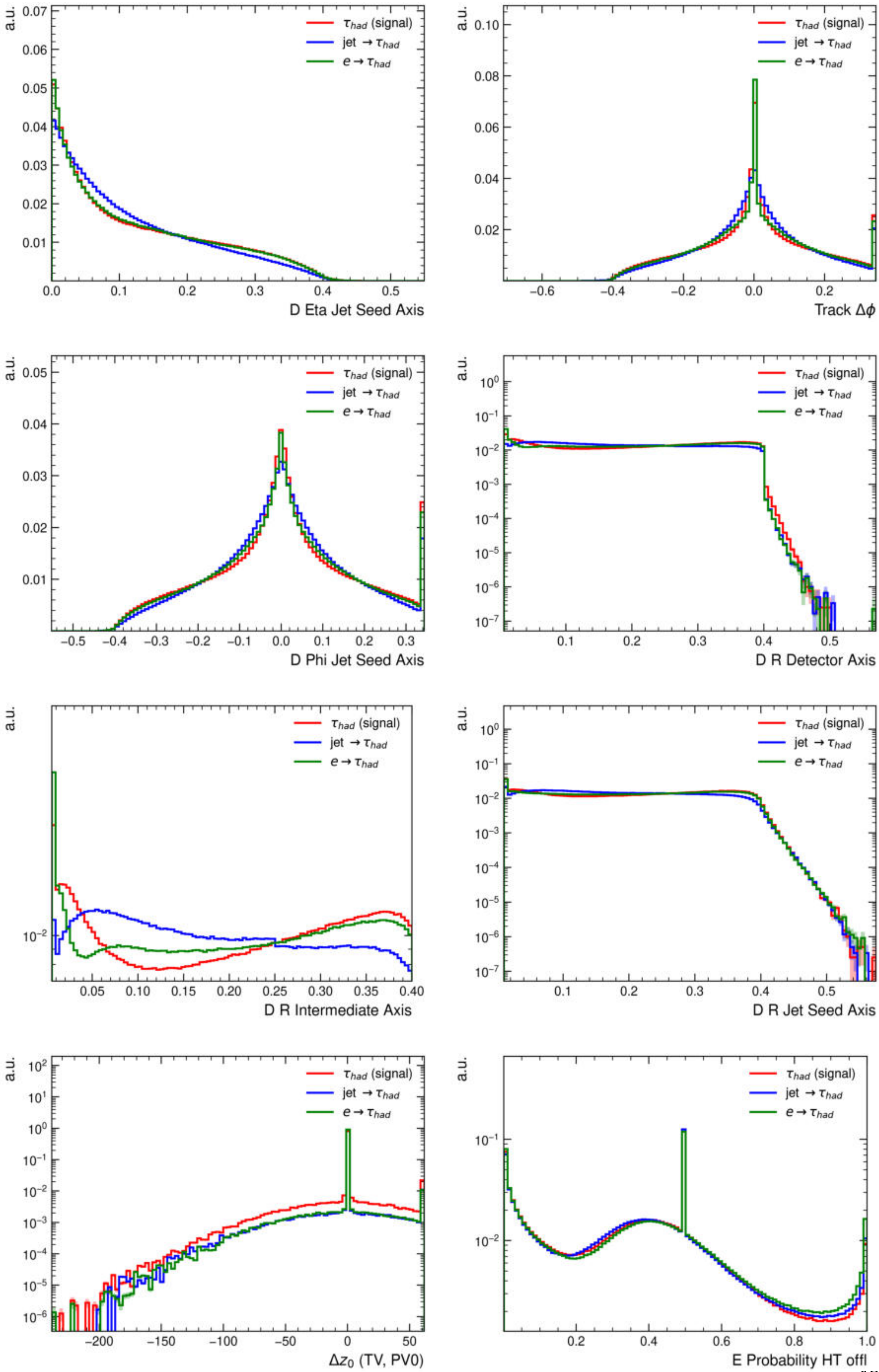

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Tau Pt
1e6

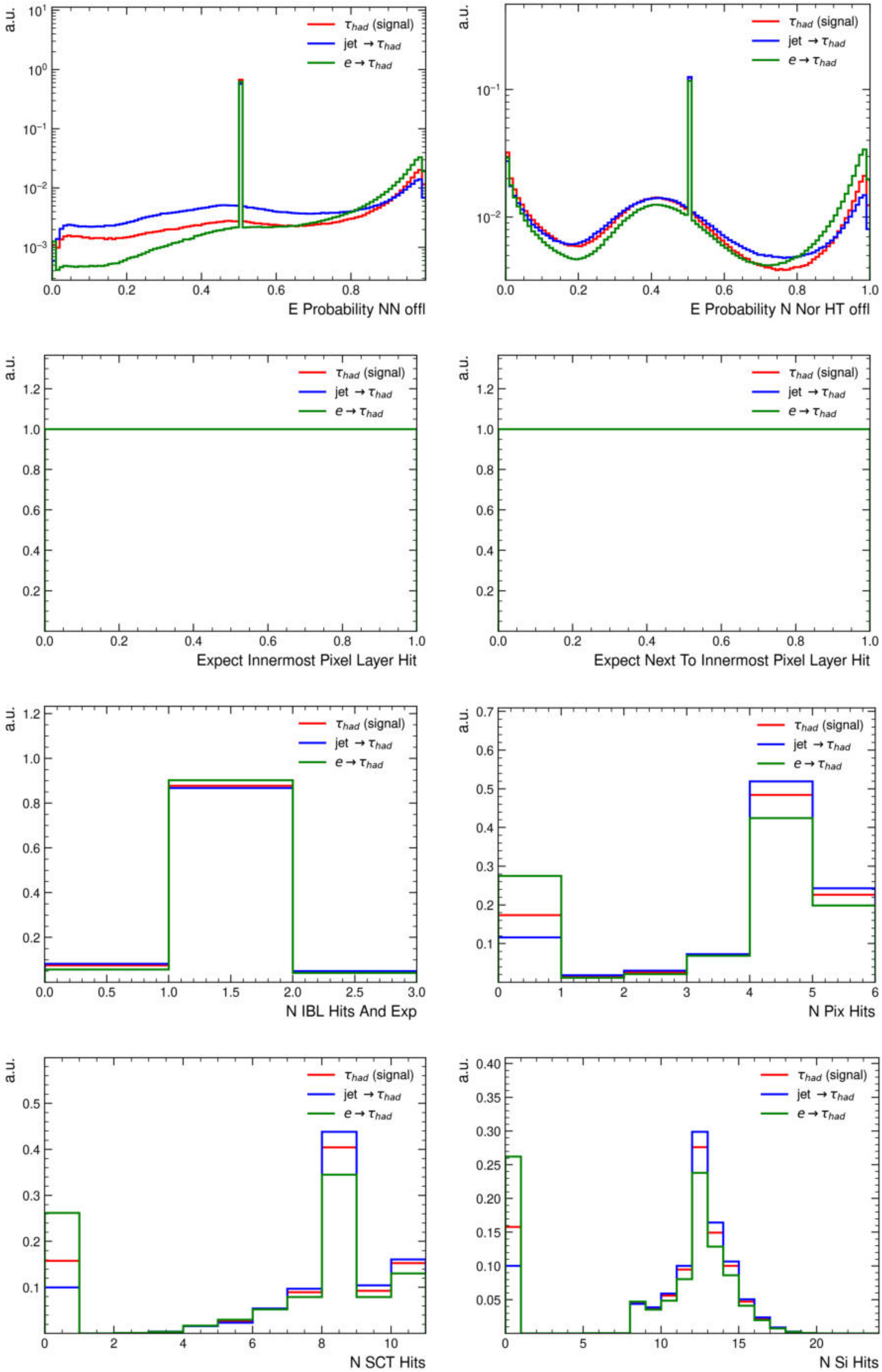

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Tau Pt Detector Axis
1e6

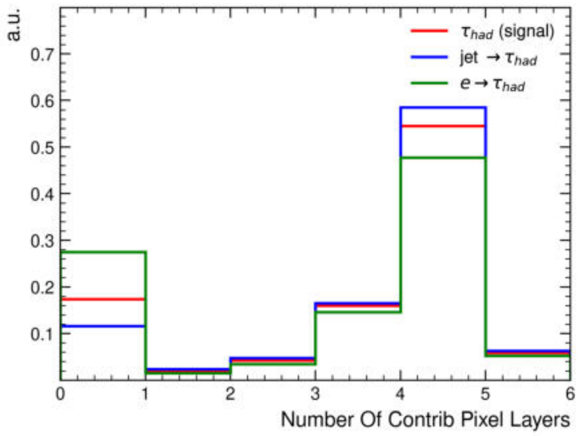

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Tau Pt Intermediate Axis
1e6

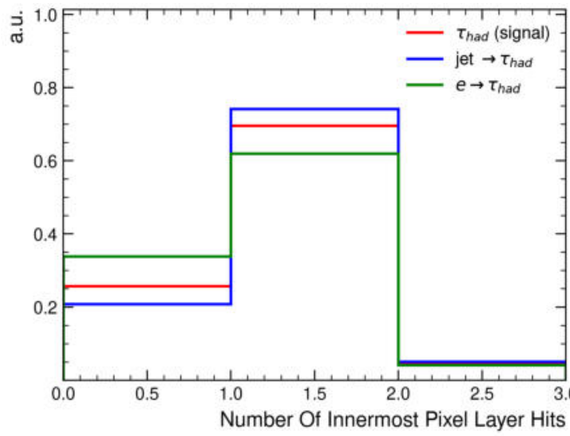


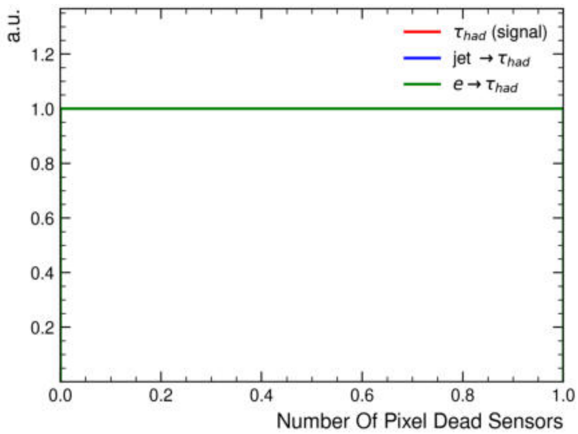
a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Flight path significance
Truth Jet Type
Flavour label
Charge
d_0 significance (TJVA)
d_0 (TJVA)
D0 old
Track $\Delta\eta$

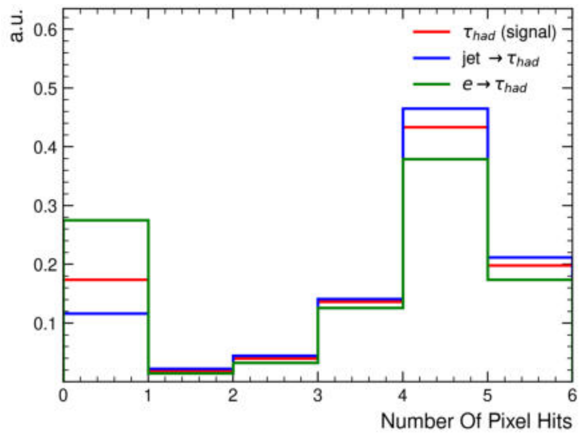

τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
a.u.
D Eta Jet Seed Axis
Track $\Delta\phi$
D Phi Jet Seed Axis
D R Detector Axis
D R Intermediate Axis
D R Jet Seed Axis
Δz_0 (TV, PV0)
E Probability HT offl

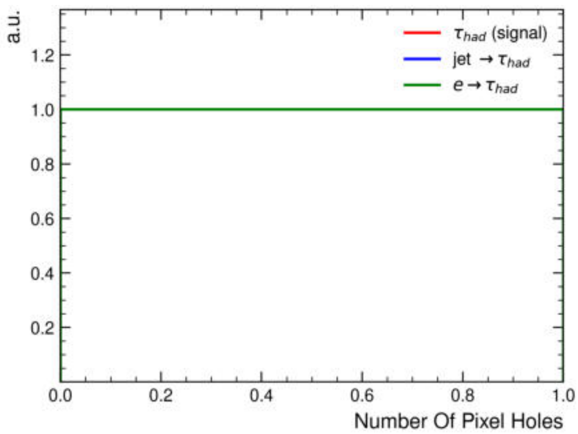

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
E Probability NN offl
E Probability N Nor HT offl
Expect Innermost Pixel Layer Hit
Expect Next To Innermost Pixel Layer Hit
N IBL Hits And Exp
N Pix Hits
N SCT Hits
N Si Hits

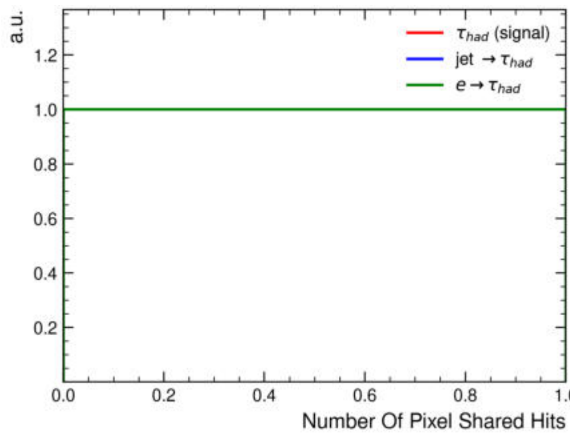

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of Contrib Pixel Layers

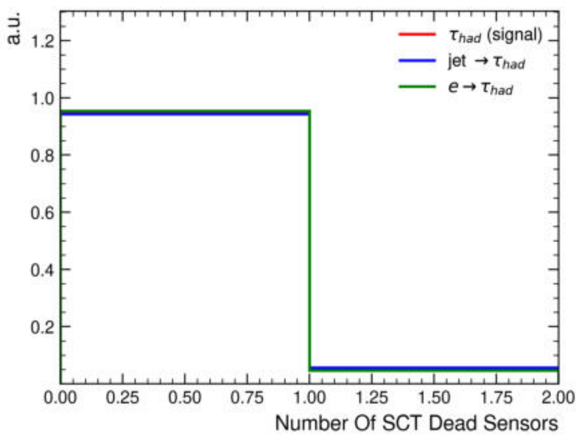

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of Innermost Pixel Layer Hits

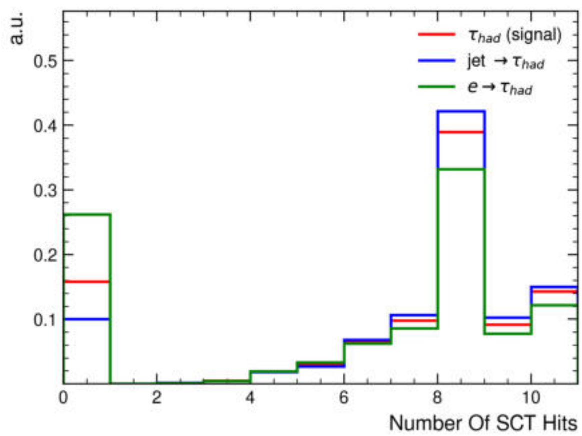

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of Pixel Dead Sensors

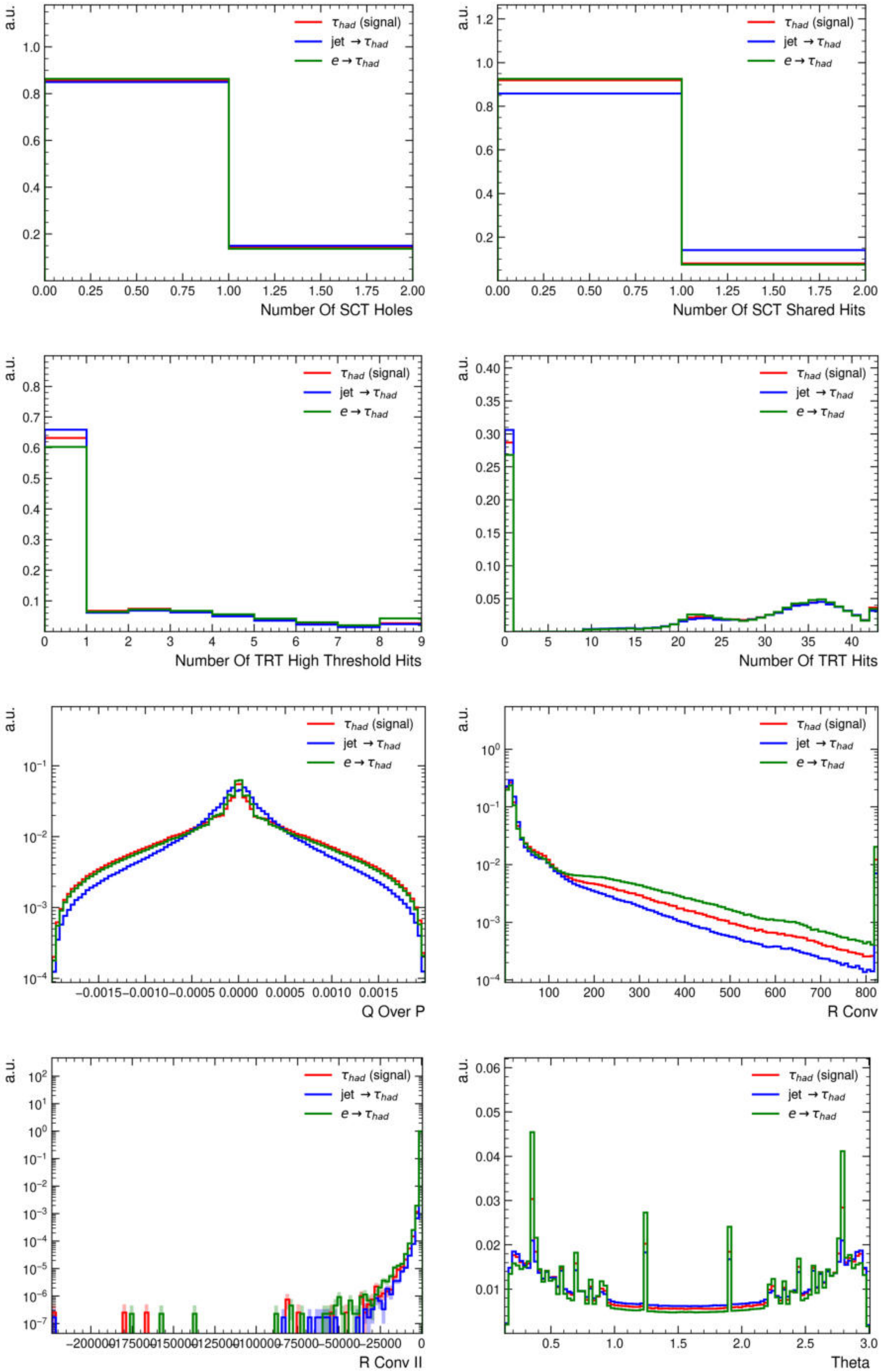

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of Pixel Hits


a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of Pixel Holes

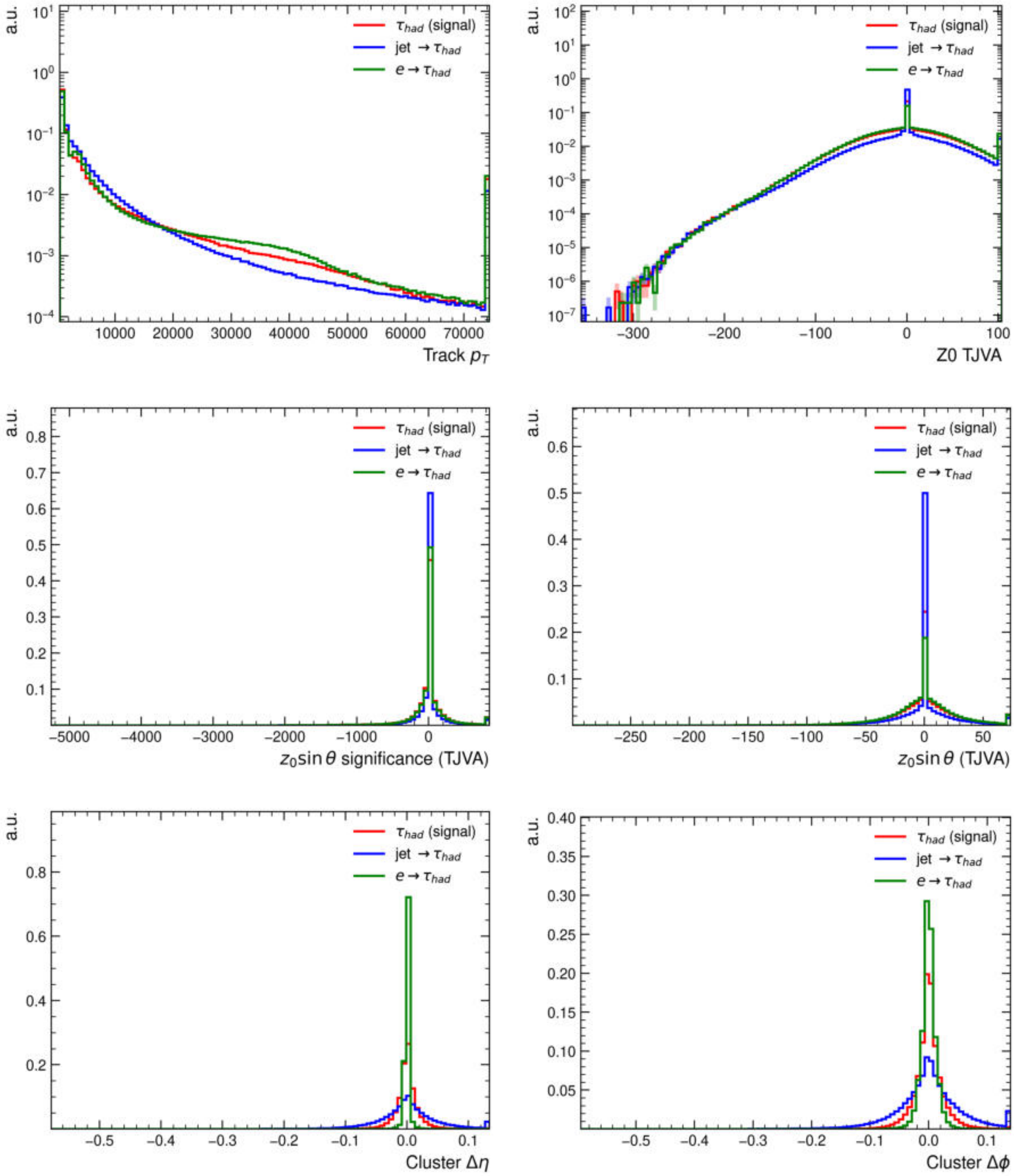

a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of Pixel Shared Hits


a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of SCT Dead Sensors


a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of SCT Hits



a.u.
τ_{had} (signal)
jet $\rightarrow \tau_{had}$
$e \rightarrow \tau_{had}$
Number Of SCT Holes
Number Of SCT Shared Hits
Number Of TRT High Threshold Hits
Number Of TRT Hits
Q Over P
R Conv
R Conv II
Theta


a.u.
τhad (signal)
jet → τhad
e → τhad
Track pT
Z0 TJVA
z0 sin θ significance (TJVA)
z0 sin θ (TJVA)
Cluster Δη
Cluster Δφ

Bibliography

[1] S. Navas, et al. (Particle Data Group), *Review of Particle Physics*, PTEP **2024(8)**, 083C01 (2024)

[2] *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, Technical Report ATL-PHYS-PUB-2019-033, CERN, Geneva (2019), all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-033, URL `https://cds.cern.ch/record/2688062`

[3] A. Huang, et al., *Heterogeneous Graph Neural Network for identifying hadronically decayed tau leptons at the High Luminosity LHC*, JINST **18(07)**, P07001 (2023), URL `http://dx.doi.org/10.1088/1748-0221/18/07/P07001`

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

[5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19(21)**, 1264 (1967)

[6] A. Salam, J. C. Ward, *Weak and electromagnetic interactions*, Nuovo Cim. **11(4)**, 568 (1959)

[7] M. Thomson, *Modern Particle Physics*, Cambridge University Press (2013)

[8] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716(1)**, 1 (2012)

[9] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, 321 (1964)

[10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, 508 (1964)

[11] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13**, 585 (1964)

[12] H. Fritzsch, M. Gell-Mann, H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. B **47**, 365 (1973)

[13] *Tikz Tau Decay Visualization*, available at: `https://tikz.net/files/tau_decay-001.png` (accessed 2026-07-15)

[14] E. Fermi, *High Energy Nuclear Events*, Prog. Theor. Phys. **5(4)**, 570 (1950)

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

[16] E. Mobs, *The CERN accelerator complex in 2019* (2019), general photo. Available at: `https://cds.cern.ch/record/2684277` (accessed 2026-07-15)

[17] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3(08)**, S08003 (2008)

[18] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3**, S08004 (2008)

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

[20] LHCb Collaboration, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30(07)**, 1530022 (2015)

[21] ATLAS Collaboration, *ATLAS experiment schematic or layout illustration* (2022), photo by R. M. Bianchi. Available at: `https://cds.cern.ch/record/2837191` (accessed 2026-07-15)

[22] J. Pequenao, *Computer generated image of the ATLAS inner detector* (2008), available at: `https://cds.cern.ch/record/1095926` (accessed 2026-07-15)

[23] J. Pequenao, *Computer Generated image of the ATLAS calorimeter* (2008), available at: `https://cds.cern.ch/record/1095927` (accessed 2026-07-15)

[24] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77(7)** (2017)

[25] J. Pequenao, *Computer generated image of the ATLAS Muons subsystem* (2008), available at: `https://cds.cern.ch/record/1095929` (accessed 2026-07-16)

[26] *Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data*, Technical report, CERN, Geneva (2022), all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-044, URL `https://cds.cern.ch/record/2827111`

[27] *Tikz Tau Decay Cone Visualization for 1-prong*, available at: `https://tikz.net/wp-content/uploads/2021/11/jet_tau-001.png` (accessed 2026-07-16)

[28] *Tikz Tau Decay Cone Visualization for 3-prong*, available at: `https://tikz.net/wp-content/uploads/2021/11/jet_tau-002.png` (accessed 2026-07-16)

[29] *Tikz Tau Decay Cone Visualization for an electron*, available at: `https://tikz.net/wp-content/uploads/2021/11/jet_tau-003.png` (accessed 2026-07-16)

[30] *Tikz Tau Decay Cone Visualization for a Jet*, available at: `https://tikz.net/wp-content/uploads/2021/11/jet_tau-005.png` (accessed 2026-07-16)

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

[32] A. Parmezan, V. Alves de Souza, G. Batista, *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*, Inf. Sci. (2019)

[33] J. Heaton, I. Goodfellow, Y. Bengio, *Deep learning: The MIT Press*, Genet. Program. Evolvable Mach. **19** (2017)

[34] A. G. Ivakhnenko, *Polynomial theory of complex systems*, IEEE Trans. Syst. Man Cybern. **SMC-1(4)**, 364 (1971)

[35] *Tikz Neural Network Architecture*, available at: `https://tikz.net/wp-content/uploads/2021/12/neural_networks-002.png` (accessed 2026-07-16)

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

[37] V. Nair, G. E. Hinton, *Rectified Linear Units Improve Restricted Boltzmann Machines*, in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, volume 27, pages 807–814 (2010)

[38] N. Srivastava, G. Hinton, A. Krizhevsky, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, J. Mach. Learn. Res. **15(56)**, 1929 (2014)

[39] *Tikz Neural Network Dropout*, available at: `https://tikz.net/janosh/dropout.png` (accessed 2026-07-16)

[40] A. Vaswani, N. Shazeer, N. Parmar, *Attention Is All You Need*, in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (2017)

[41] F. Scarselli, M. Gori, A. C. Tsoi, *The Graph Neural Network Model*, IEEE Trans. Neural Netw. **20(1)**, 61 (2009)

[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)

[43] J. Barr, D. Biswas, M. Draguet, *Salt: Multimodal Multitask Machine Learning for High Energy Physics*, JOSS **10(112)** (2025)

[44] J. Barr, J. Birk, M. Draguet, *Umami: A Python toolkit for jet flavour tagging*, JOSS **9(102)**, 5833 (2024)

[45] C. J. Clopper, E. S. Pearson, *The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial*, Biometrika **26(4)**, 404 (1934)

Danksagung

I would like to begin by thanking my parents. Studying here in Göttingen, and having the chance to pursue an education like this at all, is not something I take for granted — it was made possible by their constant support, both financial and emotional. I am deeply grateful for everything they have given me along the way.

To my girlfriend, thank you for standing by me throughout this thesis. Your patience during the stressful stretches and your steady encouragement kept me going more than you probably realize.

I would also like to thank my supervisor, Arnulf, along with everyone in AG Quadt. Working alongside this group has given me a new appreciation for just how much skill and intuition goes into particle physics, the ability to see structure in what looks, at first, like pure noise. It's been genuinely humbling to watch.

Finally, my thanks go to Karsten, who advised me closely throughout this project. His willingness to take the time to explain things, answer my questions, and guide me through the rough patches made this thesis possible.

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 sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 25. August 2026

(Jerrit Rouven Hamester)