GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Fakultät für Physik $[q,p]=i\hbar$

Master's Thesis

Classification of Top-Antitop Events in Association with Bosons Using Graph Neural Networks in Atlas

prepared by

Clinton Gonsalves

from Vasai, India.

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-MSc-2025-07

Thesis period: 1st April 2025 until 30th September 2025

First referee: Prof. Dr. Arnulf Quadt

Second referee: Prof. Dr. Ariane Frey

Abstract

At the ATLAS experiment at CERN, proton-proton (pp) collisions of an energy of 13.6 TeV produce complex final states containing a wide range of overlapping physical processes and share similar final state particles. This complexity creates major challenges to clearly distinguish rare signal events accurately. Additionally, limitations pertaining to the detector and uncertainties in particles' reconstruction lead us to the applications of sophisticated event classification techniques. This thesis presents a method applying a graph neural network (GNN) to the classification of final states containing a $t\bar{t}$ paired with additional particles collectively referred to as $t\bar{t}+X$, wherein X can represent a Higgs boson (H), a Z or W boson, or a photon (γ), using Monte Carlo simulated events. In contrast to standard techniques like graph convolutional neural networks, including those requiring information to be in grid-like structures, GNNs give options to inherently encode the physically relational and spatial nature of a particle on a per-event basis, representing particles and their interactions as nodes and edges of graphs. Utilizing an application of message passing amongst nodes of graphs, GNNs give a richer description of event topologies, thus aiding in a richer event identification. A series of model architectures are studied, which includes Graph Attention Network and Graph Transformer, to find the best version to achieve an optimization of multi-class classification across a spectrum of event types.

Contents

1 Introduction

The Standard Model (SM) [1–4] of particle physics is a highly successful and tested model. It represents our most complete theoretical framework for understanding fundamental particles and their interactions. The interactions in the Standard Model of particle physics include three of the four fundamental forces, the fourth being gravity. These forces are the electromagnetic (EM), weak nuclear, and strong nuclear forces. These interactions in nature are described using symmetry groups in the framework of gauge theories, which classify the elementary particles into quarks, leptons, gauge bosons and the Higgs boson.

The quarks and leptons form matter, gauge bosons mediate forces that govern the interactions between these particles, and the Higgs boson gives mass to these elementary particles. This study focuses on the top quark, which is observed to have an exceptionally large mass of 172.53 ± 0.33 GeV [5–9], which makes it the heaviest known elementary particle. This gives the top quark an important role in testing the Standard Model predictions, exploring the electroweak symmetry breaking mechanism, and providing sensitivity to potentially new physics phenomena beyond the Standard Model.

Experiments such as Atlas [10] at Cern record huge amounts of data from proton-proton (pp) collisions at an energy of 13.6 TeV to probe these fundamental interactions to test the Standard Model (SM) and search for physics phenomena beyond. In these processes, events that involve top quark pair production could also be accompanied by an additional particle. These events are collectively called $t\bar{t} + X$ events, with X representing additional particles such as a photon γ, Z^0 boson, $W^{\pm}$ boson, or Higgs boson, and are of high interest [11–13].

However, the accurate characterisation of such $t\bar{t} + X$ processes is greatly computationally intensive as well as analytically challenging due to the multiplicity as well as the complexity of the particles at the final state. Each of the collision processes

creates numerous jets, leptons, and photons, as well as missing transverse energy (E_T^{miss}), which create ambiguities of the respective observed particles towards the processes that they originated from. This augments the probable particle-to-origin assignments, which make the characterisation hard. Classical machine learning and cut-based methods are likely to struggle with this complexity. Graph Neural Networks (GNNs), introduced in recent years, are one of the alternatives presented, which can be used for graph-structured data that contain means to naturally express the relational and spatial structures of particle interactions event by event, with particles and their interactions as graph nodes and edges. By their design, GNNs are able to effectively propagate and build up relational information, enhancing the ability to capture the relationally intrinsic structure of particle collision data. The attention mechanism in Graph Attention Networks (GAT) makes them beneficial. In contrast to the classical graph convolutional network method, GAT calculates attention coefficients between connected nodes which assign an adaptive coefficient to edge features like invariant mass, ΔR (a measure of the angular separation between two particles in the detector), and node features like particle kinematics. This increases the accuracy of event classification by allowing GAT to highlight or downplay particular particle interactions. Graph transformers, with the use of query, key, and value transformations, allow each particle node to attend to every other particle in the event, enabling more expressive attention relationships. Incorporating GAT or transformer-like attention mechanisms into GNN frameworks thus represents a step for advancing the analysis of complex particle physics events. This thesis focuses on the classification of $t\bar{t}+X/t+X$ processes using graph neural networks, in which graph attention networks and graph transformers are explored. The study demonstrates performance in classification tasks on simulated events. Through this approach, the work aims to contribute more towards Standard Model measurements and improved sensitivity to potential new physics.

2 Theoretical Background

The Standard Model (SM) of particle physics is a highly successful theory. It describes fundamental particles and their interactions, excluding gravity. It incorporates three of the four known fundamental interactions including electromagnetic, weak, and strong. The fundamental interactions of nature are described using symmetry groups in the framework of gauge theories and classifies the elementary particles into quarks, leptons, gauge bosons and the Higgs Boson. This section discusses the theoretical framework of the Standard model of particle physics.

2.1 Particle Content

The Standard Model particles are shown in Figure 2.1 and are classified into two main categories which are fermions and bosons.

2.1.1 Fermions

The standard model of particle physics today categorises quarks and leptons as fermions, which make up matter content. Every fermion has an associated anti-fermion, they posses the same mass but carry opposite quantum numbers [14]. There are six quark flavours which are arranged as up type and down type. They are named up, down, charm, strange, top and bottom (also called truth and beauty) quarks. With up, charm, and top quarks categorised as up type and down, strange, and bottom quarks as down types. They participate in all three interactions which are described in the Standard Model. These interactions include strong, weak and electromagnetic (EM) forces. The leptons which are categorised as down type are electron, muon, and tau, and their corresponding down-type partners are ν_e, ν_μ and, ν_τ, which are known as neutrinos. Leptons only interact via electroweak

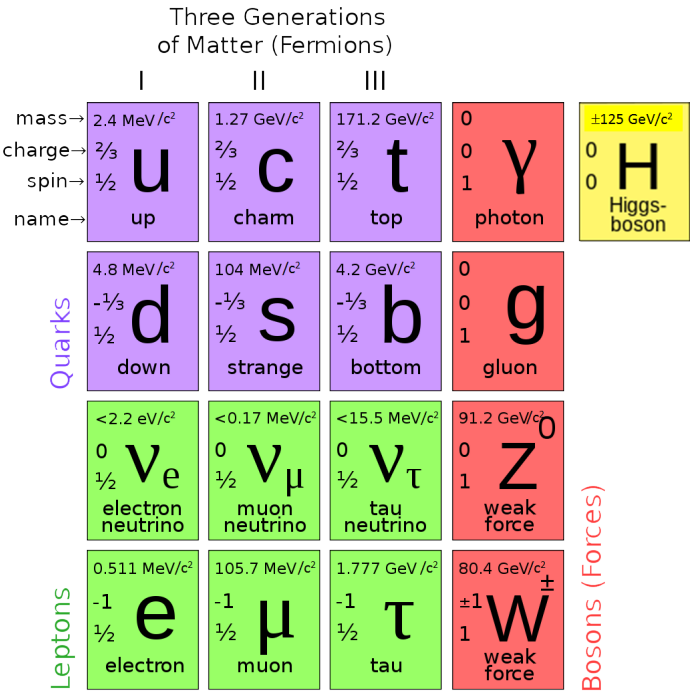


Figure 2.1: The Standard Model of Elementary Particles, illustrating the three generations of fermions (quarks and leptons), along with the fundamental bosons responsible for mediating forces. Each particle is labelled with its mass, electric charge, and spin. The left section categorises matter particles, while the right section displays the interaction carriers, which are gauge bosons (gluon, photon, W and Z bosons) and the Higgs boson.

force and do not have colour charge.

2.1.2 Bosons

The gauge bosons in the Standard Model include the photon (γ), the W boson ($W^{\pm}$), the Z boson (Z^0), and the gluon (g). They mediate the fundamental interactions between particles. The Standard Model includes a single scalar boson known as the Higgs boson H, which is responsible for the mechanism of mass generation through spontaneous symmetry breaking.

2.2 Gauge Structure of the Standard Model

The Standard Model is a gauge quantum field theory based on the symmetry group given by,

$$SU(3)_C \times SU(2)_L \times U(1)_Y \tag{2.1}$$

where $SU(3)_C$ corresponds to Quantum Chromodynamics (QCD), describing the strong interaction between quarks mediated by eight gluons. $SU(2)_L \times U(1)_Y$ governs the electroweak interaction, unifying electromagnetic and weak interactions at high energies [15–21].

2.2.1 Quantum Electrodynamics (QED)

The electron was first hypothesised as an indivisible quantity of charge in 1838 by Richard Laming, which was then discovered in 1897 by J.J. Thomson [22–24]. In 1926, the word 'photon' was brought into light by Gilbert Lewis to name Einstein's quantised light [25, 26]. Bosons were named after Satyandra Nath Bose by Paul Dirac for treating photons as indistinguishable particles by deriving Planck's radiation law [27–31]. After significant developments in quantum mechanics, quantum electrodynamics was introduced by Paul Dirac in 1927 [21, 32]. By 1949, the method of renormalisation was introduced [33–36]. Quantum electrodynamics (QED) is part of relativistic quantum field theory based on electrodynamics. It describes the interaction of electrically charged particles with the electromagnetic field and is mediated by exchanging massless photons. The mathematical framework is described using Abelian gauge theory. It is based on the $U(1)$ gauge symmetry group. It is the most precisely tested theory in the Standard Model. The Lagrangian density for QED is given by

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^{\mu} D\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{2.2}$$

where the field ψ denotes the Dirac spinor and $\bar{\psi}$ adjoint spinor representing initial state and final state charged fermions such as electrons and muons, while m

corresponds to the mass of the fermion. The term D_μ is given by,

$$D_\mu = \partial_\mu + ieA_\mu \tag{2.3}$$

defines the covariant derivative. The covariant derivative maintains local $U(1)$ gauge invariance by introducing the gauge field A_μ, which corresponds to the electromagnetic (photon) field. The constant e is the electromagnetic coupling strength, commonly known as the elementary charge. The field strength tensor $F_{\mu\nu}$ is,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.4}$$

which takes the electric and magnetic components of the electromagnetic field into account [21, 32–35, 37].

2.2.2 Weak Interaction

The weak interaction was first proposed by Enrico Fermi [38] to explain processes such as beta decay and neutrino scattering. Unlike the massless photon and gluons that mediate the electromagnetic and strong interactions, the weak interaction is unique because it violates parity (P) and charge-parity (CP) symmetries [39, 40]. This was followed by the discovery of W and Z bosons (carriers of weak interaction) [41–44], which were proposed by the Glashow-Weinberg-Salam electroweak theory (also known as the GWS model) [15, 16]. Since the weak interaction only couples to left-handed particles, the fermion fields are split into left-handed and right-handed fermion fields, which can be written as $\Psi_{L,R} = \frac{1}{2}(1 \mp \gamma_5)\Psi$. These fields are arranged in weak isospin doublets $I_3 = \frac{1}{2}$ and in weak isospin singlets $I_3 = 0$ [45–47]. The electroweak theory unifies the electromagnetic and weak interactions under a gauge group. The gauge invariance demands that all fundamental gauge bosons are massless. But the W and Z bosons were observed to be massive, and simply adding mass terms breaks the gauge symmetry, which makes the theory non-renormalisable. The Higgs mechanism was proposed by P. Higgs and F. Englert. In the Higgs mechanism, vector bosons can acquire mass without breaking gauge invariance due to spontaneous symmetry breaking of a complex scalar field [17, 48, 49]. The gauge symmetry $SU(2)_L \times U(1)_Y$ is spontaneously broken down to $U(1)_{\mathrm{EM}}$ via the

Higgs mechanism. This gives mass to W and Z bosons and therefore answers the question of why electromagnetism is a long-range interaction and explains why the weak force has a short range.
The Z boson has a mass of approximately 91.19 GeV and is charge neutral. It couples to fermions through the electroweak neutral current interaction given by

$$\mathcal{L}_{NC} = \frac{g}{\cos\theta_W} \sum_f \bar{f}\gamma^\mu (g_V - g_A\gamma^5) f Z_\mu, \tag{2.5}$$

where g is the weak coupling constant, θ_W is the Weinberg angle, and g_V and g_A are the vector and axial-vector couplings.

$$g_V = T_{3f} - 2Q_f \sin^2\theta_W,\ g_A = T_{3f}, \tag{2.6}$$

with T_{3f} being the third component of the weak isospin and Q_f the electric charge of the fermion f, and $\bar{f}$. The W boson has a mass of 80.385 GeV. It has an electric charge ± 1 and the charged current interaction is described by the Lagrangian,

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}}(\bar{u}\gamma^\mu(1-\gamma^5)dW_\mu^+ + \bar{d}\gamma^\mu(1-\gamma^5)uW_\mu^-), \tag{2.7}$$

the parameter g represents the weak coupling constant, which is associated with the $SU(2)_L$ gauge symmetry. The fields $\bar{u}$ and d are the Dirac spinors corresponding to up-type and down-type quarks. The term γ^μ represents the gamma matrices used in the formulation of relativistic currents, while γ^5 is employed to isolate the left-handed chiral components of fermions. The operator $(1-\gamma^5)/2$ acts as a projector, ensuring that only left-handed fermions participate in the interaction. The physical charged weak bosons $W^\pm$ arise from the linear combinations of the $SU(2)_L$ gauge fields and is given by,

$$W\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \tag{2.8}$$

where W_μ^1 and W_μ^2 are two of the three weak isospin gauge fields associated with the $SU(2)_L$ group. These charged bosons mediate transitions between up-type and down-type fermions within the same generation, such as $u \leftrightarrow d, c \leftrightarrow s,$ and $t \leftrightarrow b$

[15–17, 21, 50–52].

2.2.3 Quantum Chromodynamics (QCD)

In the Standard Model of particle physics, the theory of quantum chromodynamics (QCD) describes the strong force, or strong nuclear force. A major development was the discovery of the proton [53, 54], which gave us the understanding of the positively charged nucleus, and the discovery of the neutron by James Chadwick [55]. Before the 1970s we did not know what made the atomic nucleus bind together. A fundamental strong nuclear force was hypothesised to explain this [29]. Murray Gell-Mann and George Zweig proposed that protons and neutrons (known as baryons) could be made up of the elementary particles called quarks. This was shortly after Gell-Mann's 1961 formulation of the particle classification system known as the Eightfold Way or SU(3) flavour symmetry [56, 57]. Then the reasoning became clear that they are simply just permutations of the subparticles and their spins. The theory of QCD describes that this interaction is done via coupling with three different colour charges (red, green, and blue), which is carried by the quarks and is mediated by the eight massless gauge bosons, which are gluons. QCD is based on the gauge group $SU(3)_{QCD}$. The QCD Lagrangian is given by

$$\mathcal{L}_{\mathrm{QCD}} = \bar{q}(i\gamma^{\mu}D_{\mu} - m_q)q - \frac{1}{4}G^{a}_{\mu\nu}G^{a\mu\nu}, \tag{2.9}$$

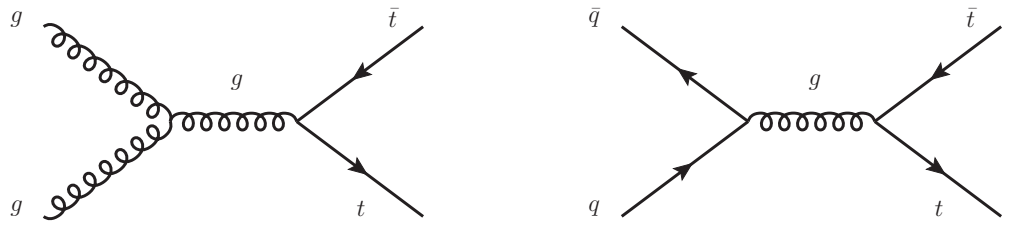
where q and $\bar{q}$ are the quark fields and represent Dirac spinor fields, carrying colour charge and flavour indices. D_{μ} is the covariant derivative with g_s as the strong coupling constant and $G^{a}_{\mu\nu}$ is the gluon field strength tensor which is given by,

$$D_{\mu} = \partial_{\mu} - ig_s T^{a} G^{a}_{\mu} \tag{2.10}$$

in which G^{a}_{μ} are the gluon fields, which are vector gauge bosons mediating the strong interaction.

$$G^{a}_{\mu\nu} = \partial_{\mu}G^{a}_{\nu} - \partial_{\nu}G^{a}_{\mu} + g_s f^{abc} G^{b}_{\mu} G^{c}_{\nu}, \tag{2.11}$$

where T^{a} represents the generators of the $SU(3)_C$ gauge group, these generators correspond to the Gell-Mann matrices λ^{a}. These matrices form the fundamental



(a) Gluon-gluon fusion. **(b)** Quark-antiquark annihilation.

Figure 2.2: Leading-order Feynman diagrams for top-antitop ($t\bar{t}$) production in proton-proton collisions.

representation of the color $SU(3)_C$ group, satisfying the commutation relation given by,

$$[T^a, T^b] = if^{abc}T^c \tag{2.12}$$

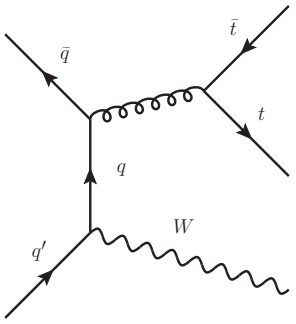
where f^{abc} are the structure constants of the $SU(3)_C$ algebra [18–20, 58–60].

2.3 $t\bar{t} + X$ processes

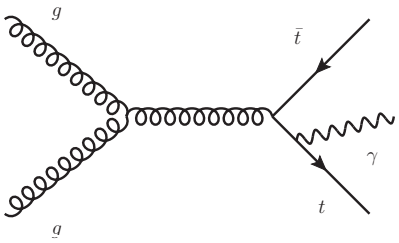
In high-energy proton-proton collision, top-antitop quark pair ($t\bar{t}$) production is a QCD process. At leading order, the Feynman diagrams are gluon-gluon fusion and quark-antiquark annihilation shown in Figure 2.2. Owing to the high mass, the top quark has a very short lifetime ($\sim 5 \times 10^{-25}$ s), below the timescale of QCD hadronisation. It then almost exclusively decays through $t \to Wb$ before hadronising, making experimental access to the properties through the kinematics of the decay product feasible.

In the Standard Model, the top quark decays through the charged weak current interaction into a W boson together with a bottom quark almost 100% of the time. Along with $t\bar{t}$ production, there are other processes where the top-antitop quark pair is accompanied by an additional boson. These processes are together called $t\bar{t} + X$ processes, where X is an additional boson like W, Z, Higgs boson (H), or a

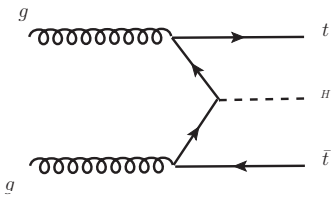
photon (γ). The $t\bar{t}+X$ processes are important because they play an important role in probing the electroweak couplings of the top quark [11–13, 61].
In the $t\bar{t}W$ process an additional W boson is radiated from an incoming quark line as shown in Figure 2.3a. In this process, the final state is defined by the decays of the three W bosons in the event. Two of them are from the decays of the top quarks, and the third one from associated production. Both W bosons can decay leptonically to a charged lepton and a neutrino, or hadronically to a quark pair. When one or two W bosons decay leptonically, the event contains one or two leptons in the final state. When all three W bosons decay leptonically, the event contains a final state with three charged leptons, such an event is called a tri-leptonic event. They typically include electrons or muons, missing transverse momentum (p_T^{miss}) from the undetected neutrinos, and several jets, including b-quark jets. A completely hadronic final state occurs when all three W bosons decay hadronically. A semi-leptonic final state occurs when one W boson decays leptonically and the other two hadronically. Di-leptonic final states are composed of two leptonic decays and one hadronic decay, and tri-leptonic final states happen when all three W bosons decay leptonically.
The production of the Z^0 boson in association with a top-antitop pair, denoted $t\bar{t}Z$, shown in the Figure 2.3d process, serves as a direct probe of the neutral-current interactions of the top quark. It enables the measurement of the top-Z coupling and can test for deviations arising from effective operators or new physics scenarios such as top compositeness or vector-like quarks. In the $t\bar{t}Z$ process, the decay mode of the Z boson determines the structure of the final state. The Z boson may decay leptonically into an electron or muon pair, hadronically into a pair of quarks, or invisibly into a pair of neutrinos. Consequently, the full event may contain multiple isolated leptons, hadronic jets, or significant missing transverse energy.
The production of a Higgs boson in association with a top-antitop pair, denoted $t\bar{t}H$ and shown in figure Figure 2.3c, provides a direct measurement of the top-Higgs Yukawa coupling, which is the strongest fermion-Higgs interaction in the Standard Model. The top-Higgs coupling is critical to understanding electroweak symmetry breaking and vacuum stability. The final states of the $t\bar{t}H$ process depend on the decay class of the Higgs boson. The dominant decay mode is $H \to b\bar{b}$, which results in a final state containing multiple b-jets. While this mode has the highest



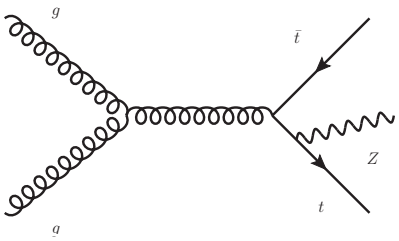
(a) $t\bar{t}W$ production



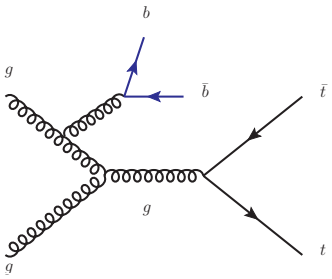
(b) $t\bar{t}\gamma$ production



(c) $t\bar{t}H$ production



(d) $t\bar{t}Z$ production



(e) $t\bar{t}b\bar{b}$ production

Figure 2.3: Feynman diagrams for representative $t\bar{t} + X$ processes considered in this analysis.

branching ratio, it also overlaps significantly with QCD backgrounds. The $H \to \gamma\gamma$ class leads to a final state with two isolated photons and is characterised by a clean and distinctive signature, albeit with a much lower branching fraction. Another important decay mode is $H \to WW^*$, which can contribute to final states with additional leptons or jets depending on whether the W bosons decay leptonically or hadronically. Similarly, the decay $H \to ZZ^*$ can produce four-lepton final states or combinations involving jets and missing transverse energy (E_T^{miss}), depending on the Z decay. Although rare, the $H \to \tau^+\tau^-$ class may also be included, leading to leptonic or hadronic tau signatures. These diverse final states from the Higgs boson, when combined with the decay products of the top quarks, offer a wide range of topologies.

The $t\bar{t}\gamma$ process shown in Figure 2.3b, wherein a photon is radiated either in the production or decay stage of the top quark, provides insight into the electromagnetic properties of the top quark, such as its electric charge and magnetic moment. This class is also sensitive to possible anomalous dipole moments or new charged particles that may modify the photon emission pattern. The presence of a high-p_T photon in the final state makes this class experimentally attractive, though care must be taken to isolate prompt photons from background sources such as final-state radiation or photon conversions.

The $t\bar{t}b\bar{b}$ process corresponds to the production of a top-antitop pair plus an extra bottom-antibottom quark pair that usually occurs by gluon-gluon fusion supplemented by gluon splitting, depicted in Figure 2.3e. The importance of the process in this context comes from the fact that it forms the leading irreducible QCD background of processes for which a boson decays into bottom quarks, namely $t\bar{t}H$ ($H \to b\bar{b}$) and $t\bar{t}Z$ ($Z \to b\bar{b}$). Therefore, studying the final state of $t\bar{t}b\bar{b}$ becomes important.

Production of a leading top quark pair in association with additional particles, known collectively as $t\bar{t} + X$, is important for testing the Standard Model (SM) at hadron colliders. The study of these processes provides tests of perturbative quantum field theory, also measuring the interactions of the top quark with both gauge and scalar bosons and physics beyond the SM. In practical cases, however, many classes produce exceedingly similar final states at the detector. Classification of such final states then becomes essential, as it reveals which processes are

most likely misidentified and provides deeper insight into the underlying physical properties that create such overlap.

2.4 Graph Neural Networks

Graph Neural Networks (GNNs) present a framework for extracting information from data organized as graphs, consisting of nodes and edges that embody pairwise associations. GNNs operate by iteratively aggregating information from within the local neighbourhood of a given node and updating the node's representation by the use of layers from the neural network. This framework finds its best use in areas such as particle physics data analysis, wherein the particle interactions are easily presented as a graph. In this project work, two of the most distinguishing architectures used are those of Graph Attention Networks (GATs) and Graph Transformers (GTs), whereby each of them utilizes an attention mechanism to selectively highlight the most important neighbouring nodes at the time of feature aggregation [62–65].

2.4.1 Graph Attention Network

Graph Attention Networks (GATs) [63, 66] use learnt attention coefficients at training time in order to fine-tune the traditional message passing framework. In doing that, the model is able to pay differential attention towards all of the neighbours, instead of paying attention to each of its neighbours equally. Considering a graph provided as $G = (V, E)$, by V denoting the set of nodes and by E denoting the set of edges. At the node-level of representation, each individual node $i \in V$ gets represented by an individual feature vector $h_i \in \mathbb{R}^F$, here F denotes the dimension of input features applicable towards each individual node.
A GAT layer initially applies a shared linear transformation $W \in \mathbb{R}^{F' \times F}$ on each of the features corresponding to each of the nodes, here F' is the size of the projected/hidden feature space. Subsequently, for each edge $(i, j) \in E$, an unnormalised attention coefficient α_{ij} is calculated as follows,

$$e_{ij} = \text{LeakyReLU}\left(\vec{a}^\top [Wh_i \,||\, Wh_j]\right), \tag{2.13}$$

where $\vec{a} \in \mathbb{R}^{2F'}$ is a learnable vector used to score the concatenated features of the nodes, and $||$ denotes concatenation. The coefficient e_{ij} quantifies the importance of node j's features to node i. These coefficients are then normalised using a softmax function over the neighbourhood $\mathcal{N}(i)$ of node i given by,

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})}. \tag{2.14}$$

The updated feature representation of node i is computed as a weighted sum of its neighbours' features as shown in,

$$h'_i = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W h_j \right), \tag{2.15}$$

where σ is a non-linear activation function such as ReLU. In this project, the attention mechanism also includes edge features. This variant, often referred to as edge-aware attention, computes attention coefficients with,

$$e_{ij} = \text{LeakyReLU} \left(\vec{a}^\top [W h_i \,||\, W h_j \,||\, W_e e_{ij}] \right), \tag{2.16}$$

where $e_{ij} \in \mathbb{R}^D$ represents the edge feature vector and $W_e \in \mathbb{R}^{F' \times D}$ is a learnable matrix that transforms the edge features into the same space as the node features. Here, D is the dimension of edge. This enables the model to condition its attention not only on node features but also on the characteristics of the interaction between them. To further enhance the expressiveness and stability of learning, GATs often employ multi-head attention. This involves running multiple independent attention mechanisms in parallel and then concatenating their outputs,

$$h'_i = \Big\|_{k=1}^{K} \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k)} W^{(k)} h_j \right), \tag{2.17}$$

where K is the number of attention heads, each with its own parameters $W^{(k)}$ and attention coefficients $\alpha_{ij}^{(k)}$. This approach helps in capturing diverse interactions between nodes [63, 66].

2.4.2 Graph Transformer

Graph Transformers (GTs) [64, 66] generalise the Transformer architecture, originally developed for natural language processing, to graph-structured data. The self-attention mechanism in Graph Transformers computes attention weights using query, key, and value projections. For node i, its query vector is $Q_i = h_i W_Q$; for node j, its key and value vectors are $K_j = h_j W_K$ and $V_j = h_j W_V$, respectively. Here, $W_Q, W_K, W_V \in \mathbb{R}^{F \times d}$ are learnable projection matrices, d is the attention dimension and F is the dimension of node features. The edge feature $e_{ij} \in \mathbb{R}^{d_e}$ is linearly projected using $W_e \in \mathbb{R}^{1 \times d_e}$ and added to the raw attention score as shown,

$$\alpha_{ij} = \frac{\exp\left(\frac{Q_i \cdot K_j^\top + W_e e_{ij}}{\sqrt{d}}\right)}{\sum_k \exp\left(\frac{Q_i \cdot K_k^\top + W_e e_{ik}}{\sqrt{d}}\right)}. \tag{2.18}$$

These attention coefficients are used to compute the aggregated message passed to node i, weighted by the value vectors given by,

$$m_i = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} V_j. \tag{2.19}$$

The resulting vector is then processed by a Feed-Forward Network (FFN), which typically consists of two linear layers with a non-linearity in between. The updated node representation is computed using a residual connection followed by layer normalisation,

$$h_i' = \text{LayerNorm}\left(h_i + \text{FFN}(m_i)\right). \tag{2.20}$$

In addition to self-attention, Graph Transformers may incorporate structural encodings that preserve the underlying graph topology. These encodings include node degrees, shortest path distances, Laplacian eigenvectors, or learned embeddings for edge types. They can be added to node features or used to bias the attention mechanism. By leveraging both node and edge-level information, Graph Transformers provide a highly expressive framework for capturing intricate dependencies in graph-structured data [67]. In summary, Graph Attention Networks and Graph Transformers offer approaches for learning from graph data. GATs provide efficient and adaptive neighbourhood aggregation, especially when local interactions domi-

nate, while GTs offer a more flexible architecture capable of modelling complex dependencies. In this project, both architectures were explored and evaluated for the classification of $t\bar{t} + X$ processes in simulated events.
For graph-level classification tasks, the information contained in the individual node embeddings must be aggregated into a fixed-size representation of the entire graph. One common approach is mean pooling, which computes the average of the node embeddings across the graph. For a graph which has N nodes and node representations $h_1, h_2, \ldots, h_N$, the pooled graph representation h_G is defined as

$$h_G = \frac{1}{N} \sum_{i=1}^{N} h_i. \tag{2.21}$$

This operation produces a permutation-invariant representation that preserves global information about the graph while ensuring a consistent input size for subsequent classification layers.
The loss function used is the weighted cross-entropy loss, which is suitable for multi-class classification problems with class imbalance. In this classification task, the model is trained using the weighted cross-entropy loss function, which is suitable for multi-class problems with imbalanced class distributions. For a classification problem with C classes, the standard cross-entropy loss is defined as

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log(p_c), \tag{2.22}$$

where y_c denotes the true label (equal to 1 for the correct class and 0 otherwise), and p_c is the predicted probability for class c. In the weighted formulation, a class-specific weight w_c is introduced to scale the contribution of each class in the loss calculation:

$$\mathcal{L}_{\text{weighted}} = -\sum_{c=1}^{C} w_c \, y_c \log(p_c). \tag{2.23}$$

3 Experimental Setup

The ATLAS Detector [10] is known as the largest particle detector in the world and one of the four major experiments at the Large Hadron Collider (LHC) at CERN. It is designed to study a broad range of physics phenomena by taking advantage of the high-energy proton-proton collisions. Today studies from the Higgs boson and top quark to searches for supersymmetry and dark matter take place at this detector. As a general-purpose particle detector, ATLAS plays a big role in studying the fundamental constituents of matter and the forces governing their interactions.

The ATLAS Detector, being the largest volume detector ever constructed at a particle collider, stands 46 metres in length and 25 metres in height and weighs around 7,000 tonnes, as shown in Figure 3.1. Being a symmetric detector, it is built around the collision point of proton-proton collisions and has high-end detector technologies designed to detect and identify the different kinds of particles and measure their momentum, energies, and trajectories, which helps reconstruct complex final states arising from high-energy collisions.
The detector is organised into several concentric layers, each optimised for a specific function. These layers include the Inner Detector (ID), which is used for tracking charged particles; the Electromagnetic and Hadronic Calorimeters, used for measuring energy deposits; and the Muon Spectrometer, which identifies and measures muons. These subsystems are embedded within a system of powerful magnets that bend the paths of charged particles, allowing momentum measurements via the curvature of their tracks.
In the following sections, each of the subsystems of the ATLAS detector are described, focusing on their design, purpose, and contribution to particle reconstruction and identification.

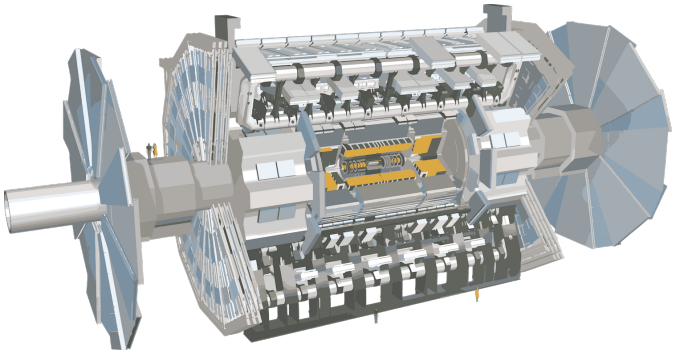


Figure 3.1: ATLAS Detector ©CERN.

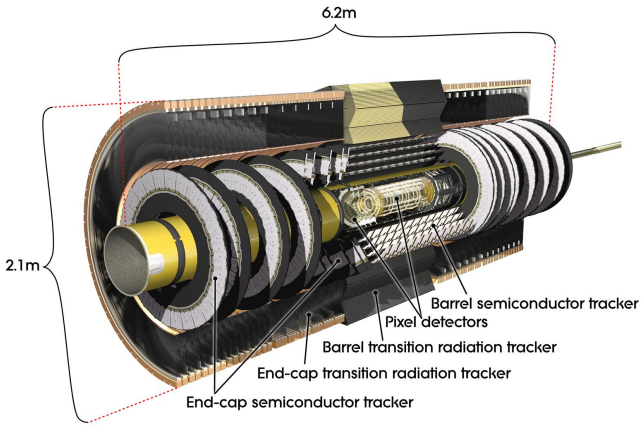


Figure 3.2: Diagram of Inner Detector in ATLAS Detector ©CERN.

3.1 The Inner Detector

The inner detector is the tracking system situated in the innermost part of the ATLAS Detector, directly closest to the beam pipe. It is used for reconstructing the trajectories of the charged particles. It is surrounded by superconducting solenoid magnets which generate a 2 Tesla axial magnetic field which is responsible for bending the trajectories of the charged particles, allowing precise determination of the particle momenta and charge by measuring the particle curvature.
The Inner Detector covers a pseudorapidity range of $|\eta| < 2.5$ and is made of three sub-detectors [10].

3.1.1 Pixel Detector

The Pixel Detector is the innermost detector of the ATLAS Inner Detector. It is located at the closest interaction point of the proton-proton collisions. It provides the highest granularity and spatial resolution, enabling precise reconstruction of charged-particle tracks and vertices, including those from heavy-flavour decays. The system consists of four cylindrical barrel layers, including the Insertable B-Layer (IBL), and three endcap discs on each side, covering $|\eta| < 2.5$ with radii from 33 mm (IBL) to 122 mm. Each pixel sensor has a typical size of 50 μm $\times$ 250 μm, delivering an impact parameter resolution of about 10 μm in the transverse plane. With over 100 million readout channels, the Pixel Detector is the most granular ATLAS tracking subsystem and is essential for vertexing, pile-up mitigation, and b-tagging [10].

3.1.2 Semiconductor Tracker (SCT)

The Semiconductor Tracker (SCT), located between the Pixel Detector and the Transition Radiation Tracker, provides high-precision tracking in both the barrel and endcap regions. It employs silicon microstrip sensors to measure charged-particle trajectories in the r-ϕ plane, essential for determining transverse momentum. The system consists of four cylindrical barrel layers and nine endcap disks on each side, amounting to over six million readout classs. Each module comprises pairs of single-sided silicon strips mounted back-to-back at a stereo angle of 40 mrad, enabling accurate two-dimensional hit reconstruction. With a typical strip pitch of 80 μm, the SCT achieves a spatial resolution of about 17 μm in the transverse plane. Covering radii from 299 to 560 mm and operating in the 2 T solenoidal magnetic field, the SCT provides precise momentum measurements while remaining radiation-hard up to fluences of 2×10^{14} $\mathrm{n_{eq}/cm^2}$. It complements the high-granularity Pixel Detector at small radii and the TRT at larger radii, forming a central element of the ATLAS tracking system [10].

3.1.3 Transition Radiation Tracker (TRT)

The Transition Radiation Tracker (TRT) is the outermost component of the ATLAS Inner Detector and provides both precise charged particle tracking and electron identification through transition radiation. Positioned outside the Semiconductor Tracker (SCT), it complements the silicon detectors by adding a large number of hits per track, improving momentum resolution and track reconstruction in forward and high-multiplicity regions. The TRT consists of about 300,000 thin-walled drift tubes, or "straws", each 4 mm in diameter and filled with a Xenon-based gas mixture. These straws act as drift chambers, giving spatial measurements with a resolution of about 130 μm in the r-ϕ plane, and are arranged in three barrel layers and 14 endcap wheels, covering radii from 563 to 1066 mm and pseudorapidities up to $|\eta| < 2.0$. A distinctive feature of the TRT is its ability to detect transition radiation, emitted when relativistic particles cross boundaries of different dielectric constants. Radiators of low-density fibres or foils enhance this effect, and the resulting photons are absorbed in the Xenon gas, producing higher ionisation signals that allow discrimination between electrons and hadrons [10].
The region outside the Inner Detector in ATLAS is occupied by the Outer Detector system, which are responsible for measuring the particle energies and identifying muons. These systems complement the high-resolution tracking of the inner detector, thereby providing near complete coverage of final states produced in proton-proton collisions. The Outer Detector consists primarily of the calorimeter system and muon spectrometer, both of which play important part in reconstructing the energy flow, particle types, and global event kinematics.

3.2 Calorimeter System

The calorimeter system surrounds the Inner Detector and provides hermetic coverage up to pseudorapidities of $|\eta| < 4.9$. It is designed to measure the energies of electrons, photons, and hadrons, and it is segmented finely in both the longitudinal and transverse directions to allow shower-shape analysis and particle identification. A fundamental purpose of the calorimeter system is to stop incident particles by forcing them to undergo a cascade of interactions with

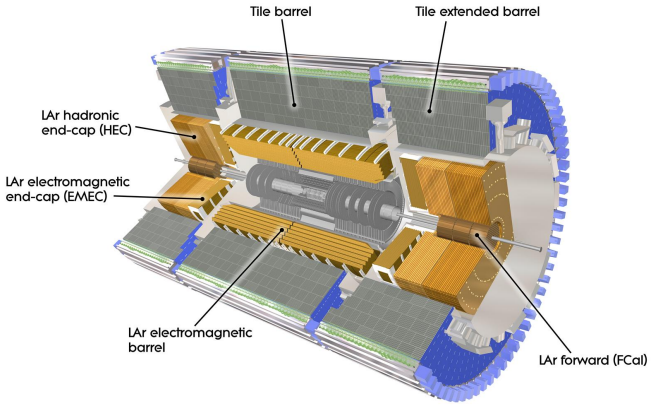


Figure 3.3: Diagram of Calorimeter system in ATLAS Detector ©CERN.

the dense absorber material. In this way, the particles deposit, ideally, their entire energy inside the calorimeter volume. The active medium then records this deposited energy, allowing for a measurement of the particle's initial energy. For electromagnetic showers, produced by electrons and photons, the calorimeter ensures near-complete energy deposition within the electromagnetic calorimeter, while hadronic showers, being more extended, are measured using both the electromagnetic and hadronic calorimeters.

In the liquid Argon (LAr) calorimeters, energy deposition by charged particles produces ionisation electrons in the Argon, which drift under a strong electric field and are collected by readout electrodes, where the signal is amplified and digitised. This technology provides high stability, fine granularity, and excellent linearity of response. In the Tile Calorimeter, which employs plastic scintillator tiles, the passage of charged particles excites the scintillator molecules, and as they de-excite they emit visible photons of light. These photons are collected via wavelength-shifting fibres and guided to photomultiplier tubes (PMTs), which convert the light signal into an electronic pulse proportional to the deposited energy.

The electromagnetic calorimeter (ECAL), based on lead absorbers with liquid Argon as the active medium, employs an accordion-shaped geometry that ensures complete azimuthal coverage without gaps. It measures the energies of electrons and photons with high precision and provides essential information for

reconstructing their directions and identifying electromagnetic showers. Located around the ECAL, the hadronic calorimeter (HCAL) measures the energy of strongly interacting particles such as pions, kaons, and jets. In the central barrel region ($|\eta| < 1.7$), the HCAL consists of steel absorbers interleaved with plastic scintillator tiles, forming the Tile Calorimeter. In the endcap and forward regions, where particle fluxes are higher, Copper or Tungsten absorbers combined with liquid Argon are employed. Together, the electromagnetic and hadronic calorimeters form a system capable of measuring the total transverse energy flow in an event. They are essential for the reconstruction of jets, the identification of isolated photons and electrons, and the determination of missing transverse momentum (E_T^{miss}), which provides sensitivity to neutrinos and potentially new physics [10].

3.3 Muon Spectrometer

The muon spectrometer is the outermost component of the ATLAS detector and is dedicated to the identification and precise momentum measurement of muons, which typically traverse the calorimeters with minimal energy loss. Covering the pseudorapidity range $|\eta| < 2.7$ for tracking and up to $|\eta| < 2.4$ for triggering, the system is embedded in large superconducting air-core toroidal magnets. The magnetic field bends muon trajectories, enabling an independent momentum measurement that dominates at high transverse momenta, while complementing the Inner Detector at lower momenta.

Muon detection relies on gaseous detectors that convert the passage of muons into electronic signals through ionisation of the gas medium. Precision tracking is achieved mainly with Monitored Drift Tubes (MDTs), while Cathode Strip Chambers (CSCs) are employed in the high-rate forward regions. Triggering requires fast detectors, provided by Resistive Plate Chambers (RPCs) in the barrel and Thin Gap Chambers (TGCs) in the endcaps, which supply timing and coarse spatial information to identify the correct bunch crossing.

Together, these technologies provide both accurate track reconstruction and efficient triggering. In combination with the Inner Detector, the muon spectrometer ensures excellent muon identification and momentum resolution, which is crucial for a wide

range of ATLAS physics analyses, from Standard Model measurements to searches for new phenomena.

4 Monte Carlo Samples and Machine Learning Methodology

In this study we use Monte Carlo (MC) simulated event samples generated for the ATLAS detector Run-2 at the Large Hadron Collider. The processes are $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}\gamma$, $t\bar{t}b\bar{b}$, and $t\bar{t}H$.

4.0.1 Monte Carlo Event Samples and Pre-processing

The MC events come in ROOT [68] format, which was developed by CERN originally for data analysis. The samples in root format consist of reconstructed physics objects such as electrons, muons, jets, photons, and missing transverse energy (E_T^{miss}). A custom processing pipeline is created through which each collision event is transformed into graph-based representations suitable for Graph Neural Network (GNN) architectures. The complete processing and training workflow is summarised in Figure 4.1, which illustrates the flow of information from Monte Carlo ROOT files to the final evaluation and plotting of results.

Object Definitions and Event Selections

The reconstructed physics objects in the Monte Carlo sample are used as inputs for the event classification. These objects include electrons, muons, jets, photons, and missing transverse energy (E_T^{miss}). Each reconstructed object (electrons, muons, jets, etc.) has to satisfy an identification and isolation criterion which guarantees that objects are genuine and have well-measured kinematic properties. Electrons interact with the detector material and leave a track in the detector; they are required to have some kinematic requirements and identification selections on the basis of likelihood-based (LH) identification with LooseLH and TightLH conditions,

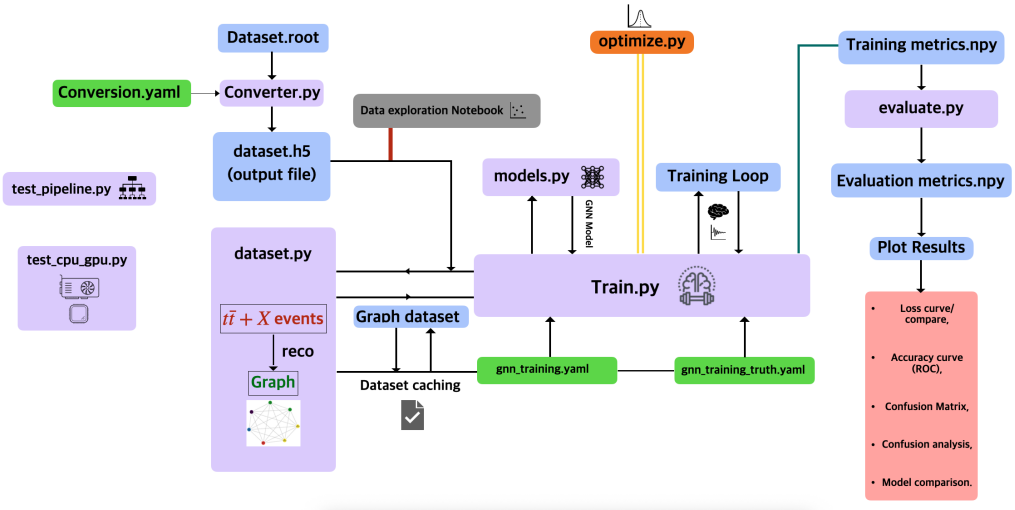


Figure 4.1: Diagram visualising the pipeline of the $t\bar{t} + X$ graph-neural-network classification, showing dataset conversion, graph construction, model training, and evaluation.

which are working points that present a certain threshold that reduces contributions from nearby particles. This is calculated using

$$L_{S(B)}(x) = \prod_{i=1}^{n} P_{S(B),i}\Big(x_i\Big). \tag{4.1}$$

Here x is the measured discriminating variable for a candidate object, and $P_{S(B),i}$ is the probability density function. Crack veto is used to remove electrons in the calorimeter transition region, which is between the barrel and endcap, where the detector response is poorly modelled. An isolation criterion has to be put in place to separate particle tracks and is done with an isolation cone with variable radius (VarRad) around the particle track, with 'loose' and 'tight' as working point options.

Muons are also required to have some kinematic requirements, identification and isolation requirements and are reconstructed by combining tracks in the Inner Detector (ID) with those in the Muon Spectrometer (MS). Along with kinematic requirements, jets are clustered using particle-flow (EMPFlow) objects using the anti-kt algorithm with radius parameter R=0.4, a standard jet definition in ATLAS. To separate jets from additional jets coming from the proton-proton interactions,

Object	Kinematics	Identification	Isolation / PU
Electrons	$p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 2.47$	Likelihood-based ID (LooseLH, TightLH), crack veto	Track and calorimeter isolation (Loose/Tight VarRad)
Muons	$p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 2.5$	Combined ID+MS tracks, Medium quality	Track and calorimeter isolation (Loose/Tight VarRad)
Jets	$p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 2.5$	anti-k_t ($R = 0.4$), EMPFlow inputs	Jet Vertex Tagger (JVT) and NNJvt for pile-up suppression
b-jets	$p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 2.5$	GN2v01 b-tagger (e.g. FixedCutBEff 77)	JVT for pile-up suppression
Photons	$p_{\mathrm{T}} > 25$ GeV, $\lvert\eta\rvert < 2.5$	EM shower-shape ID (Loose/Tight)	Calorimeter isolation (Loose/Tight)

Table 4.1: Definitions of reconstructed physics objects and their selection criteria used in this analysis.

Jet Vertex Tagger (JVT) with its neural network counterpart (NNJvt). The b-jets are identified using b-tagging, which identifies jets originating from b-quarks using a neural network (GN2) algorithm with a working point called FixedCutBEff 77, which means the working point is tuned to 77 % efficiency for true b-jets. Photons in the ATLAS detector are identified by their electromagnetic calorimeter shower shape ID and Loose/Tight again means different strictness [69–76]. The Table 4.1 and Table 4.2 summarise the object definitions and list the pre-selections.

4.0.2 Graph-based Event Representation

The pipeline begins with converting the n-tuples from ROOT format to HDF5 format. To do that, we take the samples having different Dataset IDs (DSIDs) that are extracted from file paths with key components. ROOT files contain variable-length arrays which are flattened, and offsets are stored as HDF5 files. The graphs

Table 4.2: Event pre-selection requirements for the $t\bar{t} + X$ dataset, including the common $t\bar{t}$ baseline and lepton-multiplicity selections.

Region	Leptons	Jets	b-jets
Baseline	≥ 1 b-tag	≥ 1 jet ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$)	≥ 1
1ℓ	$= 1$ e/μ ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$)	≥ 4	≥ 1
2ℓ	$= 2$ e/μ ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$)	≥ 3	≥ 1
3ℓ	≥ 3 e/μ ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$)	≥ 2	≥ 1

are constructed with each physical object selected as a node which has a set of features. In this study we used the transverse momentum (p_T), pseudorapidity (η), azimuthal angle (ϕ), and object-specific features such as b-tagging discriminants for jets and particle identification information along with its charge for leptons and photons. In the case of MET, the node features used were the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ and its azimuthal direction ϕ_{MET}. The pseudorapidity η and the fourth feature, reserved for quantities such as the particle identification information or a b-tagging score, were set to zero.

To make the best of the kinematics present, we used fully connected graphs for this study. Which means that edges are constructed between all possible pairs of nodes that are present in an event. This makes sure that every particle (node) in the event can exchange information using the message passing. Now considering an edge connecting two nodes i and j, a set of edge features are computed which gives extra information to the GNN about the relationship between the two particles. In case of edges that connect two visible particles like jets, leptons, or photons the features used are the angular distance ΔR and the invariant mass m_{ij}. The angular distance in the η-ϕ plane is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{4.2}$$

where $\Delta\eta = \eta_i - \eta_j$ and $\Delta\phi = \phi_i - \phi_j$. The invariant mass of the two-particle system,

$$m_{ij}^2 = (E_i + E_j)^2 - (\vec{p}_i + \vec{p}_j)^2, \tag{4.3}$$

is Lorentz-invariant and encodes the combined energy-momentum structure of the pair. In case of edges between visible particles and the MET node, the features used are the angular distance to MET, denoted as ΔR_{MET}, and the transverse mass m_T. The angular distance to MET is defined as

$$\Delta R_{\text{MET}} = \sqrt{(\eta_{\text{obj}} - \eta_{\text{MET}})^2 + (\phi_{\text{obj}} - \phi_{\text{MET}})^2}, \tag{4.4}$$

where the MET vector is treated as an object characterised by its azimuthal angle ϕ_{MET}, and η_{MET} is set to zero. MET gives the information about the alignment of visible objects with the missing transverse momentum, which ise usually used to identify neutrinos. The transverse mass is given by

$$m_T = \sqrt{2p_T^{\text{obj}} E_T^{\text{miss}} \Big(1 - \cos(\Delta\phi)\Big)}, \tag{4.5}$$

where p_T^{obj} is the transverse momentum of the visible object, E_T^{miss} is the MET magnitude, and $\Delta\phi$ is their azimuthal separation. The transverse mass is used for identifying events like leptonic $W \to \ell\nu$ decays. The diagram of the graph structure is shown in Figure 4.2.

To build the final event classes for training, $t\bar{t}H$ and $t\bar{t}Z(Z \to q\bar{q})$ events were separated using parton-level (truth) information. The parton history of each generated event is inspected to identify the underlying hard process and to distinguish Higgs decays from hadronic Z decays, which can otherwise produce very similar final states. The 14 classes chosen for classification were $t\bar{t}(H \to b\bar{b})$, $t\bar{t}(H \to WW_{\text{semilep}})$, $t\bar{t}(H \to WW_{\text{dilep}})$, $t\bar{t}(H \to WW_{\text{had}})$, $t\bar{t}(H \to \tau^+\tau^-)$, $t\bar{t}(Z \to$ light quarks (Ljs)), $t\bar{t}(Z \to b\bar{b})$, $t\bar{t}(Z \to \mu^+\mu^-)$, $t\bar{t}(Z \to e^+e^-)$, $t\bar{t}(Z \to \tau^+\tau^-)$, $t\bar{t}(Z \to \nu\bar{\nu})$, $t\bar{t}\gamma$, $t\bar{t}W$, and $t\bar{t}b\bar{b}$.

4.0.3 Graph Neural Network Training and Optimisation

In this study we used Graph Attention Network (GATNet) and Graph Transformer (GT). Models are defined by a set of hyperparameters like the number of input classes, the hidden layer dimensions, the number of output classes, the dimensionality of edge features, the pooling method, and the dropout probability applied during training.

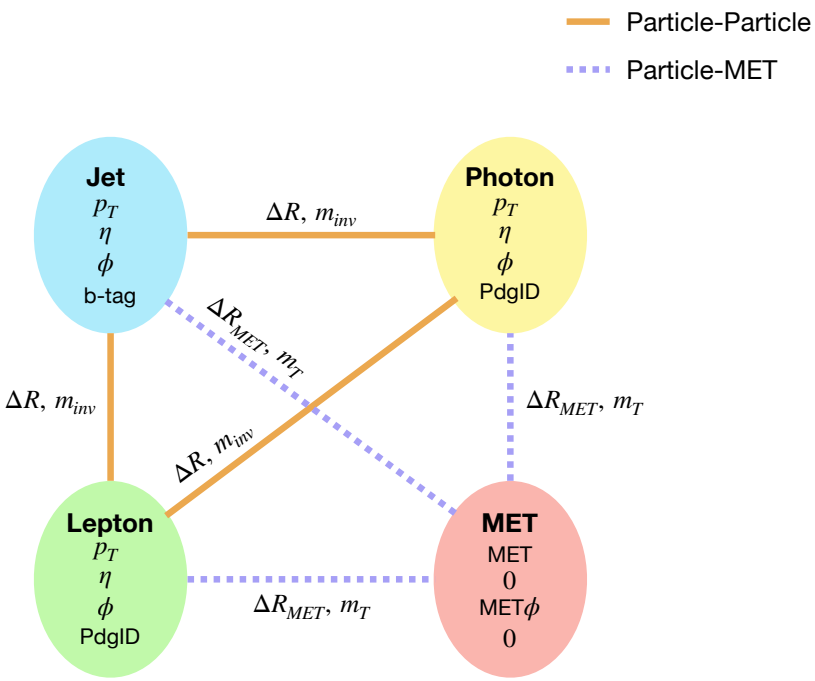


Figure 4.2: Event graph used for GNN input. Nodes represent Jet (blue), Photon (yellow), Lepton (green), and MET (red) with key kinematic features. Orange edges denote particle-particle links ($\Delta R, m_{\text{inv}}$); lavender edges denote particle-MET links ($\Delta R_{\text{MET}}, m_T$).

Graph Attention Network (GATNet)

The GATNet architecture (GATv2Conv) used in this study was determined through hyperparameter optimisation using the Optuna framework. This optimised model takes four input features per node. The training parameters are shown in Table 4.3. Every GAT layer uses ten attention heads. Dropout is used which is a regularising technique. In this technique, some fraction of neurons are randomly deactivated in a given layer during each forward pass mechanism. This prevents the network from relying heavily on certain nodes, which helps the model to learn the overall patterns. Also, mean pooling is used as the global aggregation function.

Table 4.3: Summary of key hyperparameters and training settings for the optimised GATNet and GTNet (GTNet) models.

Parameter	**GATNet**	**GTNet**
Hidden units	64, 16	48, 48, 48
Attention heads	10 per layer	12 per layer
Dropout	0.374	0.142
Pooling	mean	mean
Learning rate	1.03×10^{-4}	1.41×10^{-3}
Weight decay	3.08×10^{-6}	1.68×10^{-3}
Batch size	128	128
Epochs (max)	150	150
Loss	weighted cross-entropy	weighted cross-entropy
Events per class	1.0×10^{5}	1.0×10^{5}
Early stopping:		
parience	20	20
LR scheduler:		
patience	ReduceLROnPlateau	ReduceLROnPlateau
factor	0.8	0.8

Graph Transformer Network (GTNet)

The GTNet architecture used in this study was determined using hyperparameter optimisation with the Optuna framework. The training parameters are shown in Table 4.3. Each GT layer uses 12 attention heads. In addition, mean pooling is used as the global aggregation function.

For both of the models, we use Adam optimizer [77], which updates the network parameters during the gradient descent. Both models also use a ReduceLROnPlateau scheduler. This is designed to decrease the learning rate by a factor. During training, the model parameters are periodically saved (checkpointed). After training is completed, the checkpoint with the lowest validation loss is used for the final evaluation on the test set. To take care of the imbalance between classes, a maximum number of events per class can be specified. For this study, graph

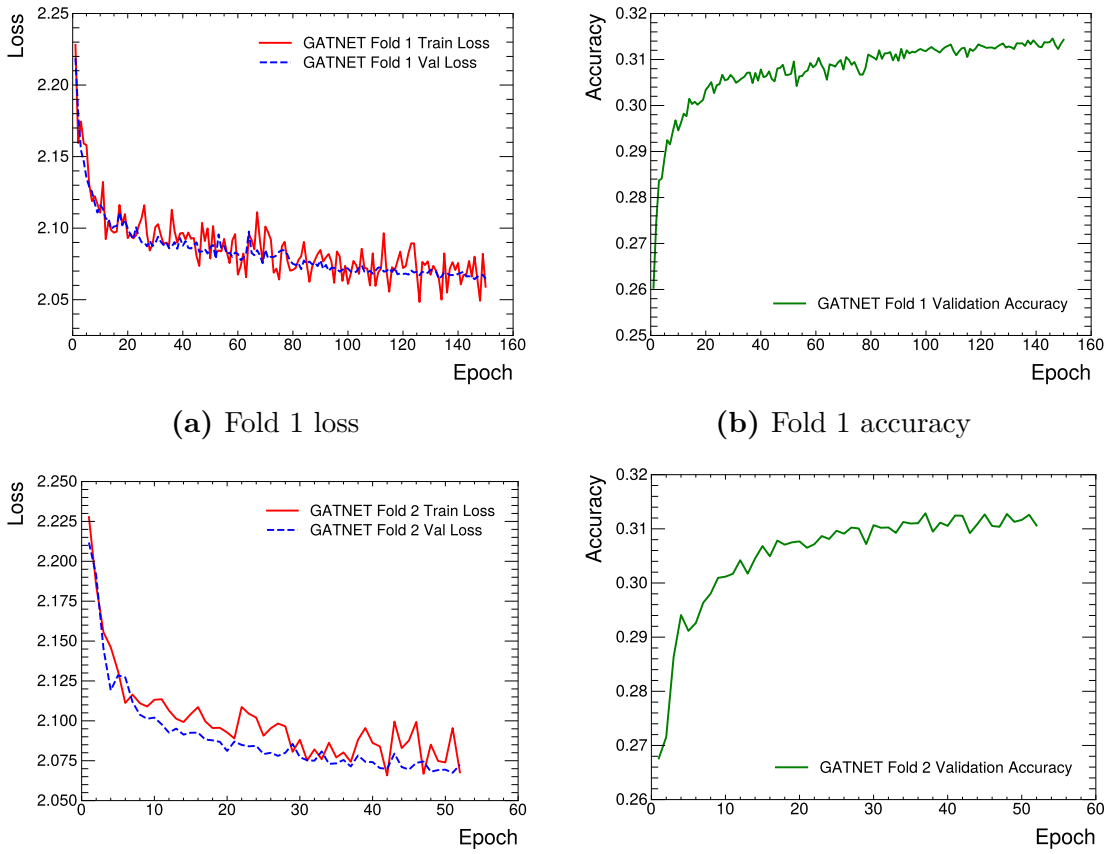
dataset with 100,000 events was constructed from each $t\bar{t}+X$ process, making sure that all classes are represented with equal amount of statistics. The storing method of the graph datasets was done by designing a hash based system to guarantee reproducibility. The classes are defined from dedicated samples for each process, with the option of applying truth-level selections for separating processes further. The training uses the weighted cross-entropy loss, this assigns class weights inversely proportional to their frequencies. This dual strategy makes sure that the model gets to learn from a balanced dataset. The training is done using the stratified K-fold cross validation technique [78] with 2 folds. This means that the graph dataset is split into 2 folds with equal amount of events (graphs). Each fold contains 70% of the events for training set, 20% for validation set and 10% for test set.

4.0.4 Performance Evaluation

The performance, prediction ability and reliability of the trained models was performed through an assessment. After training with the Monte Carlo samples of $t\bar{t}+X$ events, the final models are subjected to checks that test both their internal convergence and how well they can classify. Model evaluation is done using the test set.

Training and Validation Monitoring

The convergence behaviour of the GATNet and GTNet models was accomplished by observing how training losses, validation losses and validation accuracy evolved, over the full training period with early stopping [79, 80]. The results for the GATNet are shown in Figure 4.3a, Figure 4.3b, Figure 4.3c and Figure 4.3d.
The training and validation loss curves show a quick decrease during the initial epochs, followed by a slow plateauing as the models approach convergence. After 40 to 50 epochs the training losses show stabilisation around a value of 2.06 for both the models with no significant separation between the training and validation loss curves. This shows us that the networks can learn effectively from the available data without overfitting.
The validation accuracy shows a similar behaviour. It rises steeply during the early stages and gradually approaches a plateau. The final accuracy remains stable



(a) Fold 1 loss

(b) Fold 1 accuracy

(c) Fold 2 loss

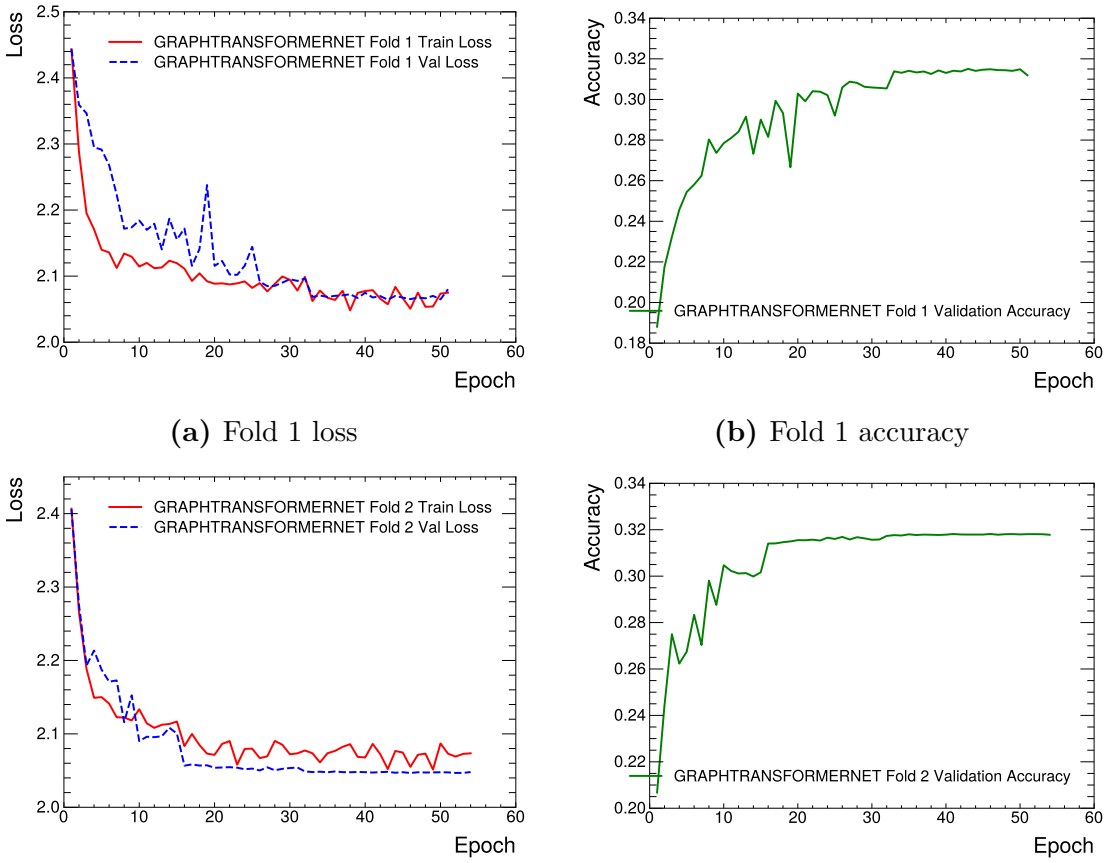
(d) Fold 2 accuracy

Figure 4.3: Training and validation monitoring for the GTNet model. Subfigures (a,b) show the evolution of loss and accuracy for Fold 1, and (c,d) for Fold 2.

during the later epochs around the value of 0.31. This confirms that the optimised learning rate scheduler and early stopping criteria successfully guide the training to a well-converged minimum with both architectures. The results for the GTNet model are shown in Figure 4.4a, Figure 4.4b, Figure 4.4c and Figure 4.4d.

Classification Quality on Test Data

The confusion matrices in Figure 4.5 (GATNet) and Figure 4.6 (GTNet) show the true and predicted class labels for all $t\bar{t}+X$ classes. Each row shows a true physics process, and each column shows the predicted class. The diagonal elements in the



(a) Fold 1 loss

(b) Fold 1 accuracy

(c) Fold 2 loss

(d) Fold 2 accuracy

Figure 4.4: Training and validation monitoring for the GTNet model. Subfigures (a,b) show the evolution of loss and accuracy for Fold 1, and (c,d) for Fold 2.

confusion matrix show the fraction of correctly classified events per class, and the off-diagonal elements show misclassified events. The hard confusion matrices are obtained by applying the argmax function directly to the model's raw logits output so that each event is assigned to the class with the highest raw score and then compared to its true label. The diagonal elements dominate in both the models, showing that the networks classify correctly by capturing the features. For some classes, high diagonal values are seen, showing that they have clean signatures, for example $t\bar{t}\gamma$, in which up to about 70 % of the events are correctly identified. Processes with more complex or overlapping final states, such as $t\bar{t}H(\tau^+\tau^-)$ or

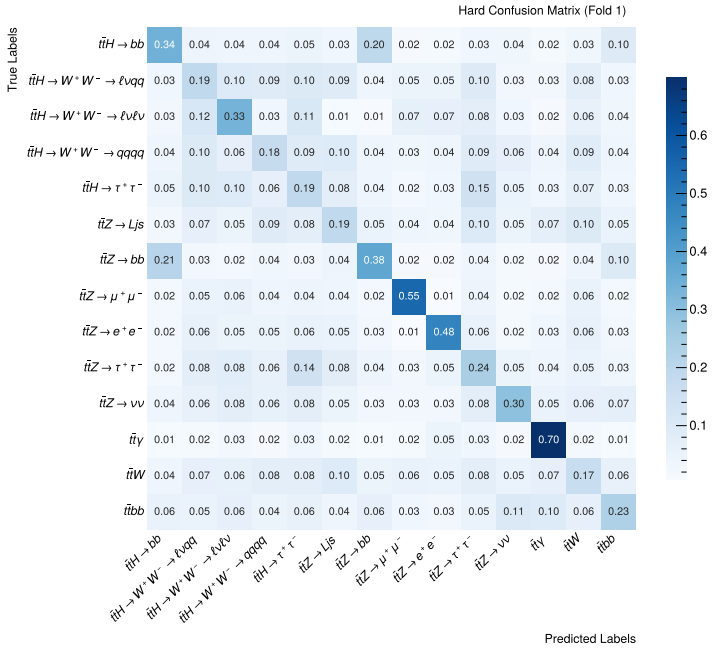
$t\bar{t}W$, show lower diagonal fractions and larger off-diagonal entries, which shows confusion between these classes.
The $t\bar{t}H$ process names are given based on the decay products of the associated boson. For example, $t\bar{t}H \to WW^*$, $lvqq$ shows a semi-leptonic final state with one W boson decaying leptonically ($\ell\nu$) and the other hadronically (qq).
The classification performance of the trained GATNet and GTNet models can be further examined by looking at the receiver-operating-characteristic (ROC) curves and their area-under-curve (AUC) values. For every physics process the multi-class output of the network is treated in a one-versus-rest fashion. So for a given class C, all events of C are considered as signal, and all other events are considered as background. The soft probability assigned to C is then scanned over all possible thresholds to construct the curve of true-positive rate versus false-positive rate. The AUC is the integral of this curve and quantifies how often the network assigns a higher probability to a chosen true class C event than to a randomly chosen background event.
The ROC plots in Figure 4.7 (GATNet) and Figure 4.8 (GTNet) show steep rises toward the upper left corner, stating a discriminating power. For both folds of each model, almost all classes reach AUC values above 0.95, approaching unity (AUC ≈ 0.99). Classes with more complicated and similar looking final states still show better AUCs in the range of $0.95 - 0.98$.
The precision-recall (PR) curves in Figure 4.9 (GATNet) and Figure 4.10 (GTNet) show for each of the fourteen $t\bar{t} + X$ processes the change of classification purity called precision as the selection is loosened to higher signal efficiencies called recall. On the left of the plot, which shows very low recall, the most confidently classified events are accepted. In this region precision is close to unity for all classes, because all of the accepted events truly belong to the target class. The curves then slowly show lower for some classes as the threshold is loosened, allowing the less certain events, which shows low precision. In the high-precision, low-recall region for some classes we see a very high amount of fluctuations. This is a statistical effect and is because at these tight thresholds only a handful of events pass the selection. Which means removal of a single event can change the precision in a sharp way and cause fluctuations. As we go to the higher recall region, more events contribute to the estimate, and the curves quickly smooth out. The overall position and steepness of

the curves show differences in the ability of the model to classify successfully. So the classes like leptonic $t\bar{t}Z$ final states ($t\bar{t}(Z \to \mu^+\mu^-, e^+e^-)$) and $t\bar{t}\gamma$ show almost flat curves with average precision (AP) value close to one. This means that these classes can be identified with high precision even when the threshold is loosened and a large fraction of their events is retained. And compared to them, classes like $t\bar{t}W$ and $t\bar{t}Z \to \tau^+\tau^+$ reach only AP $\simeq$ 0.4-0.5, which is substantial contamination once the threshold is relaxed.
With these results and comparing both the folds, one can conclude that both the network architectures show similar patterns. The differences in both of the folds are small.

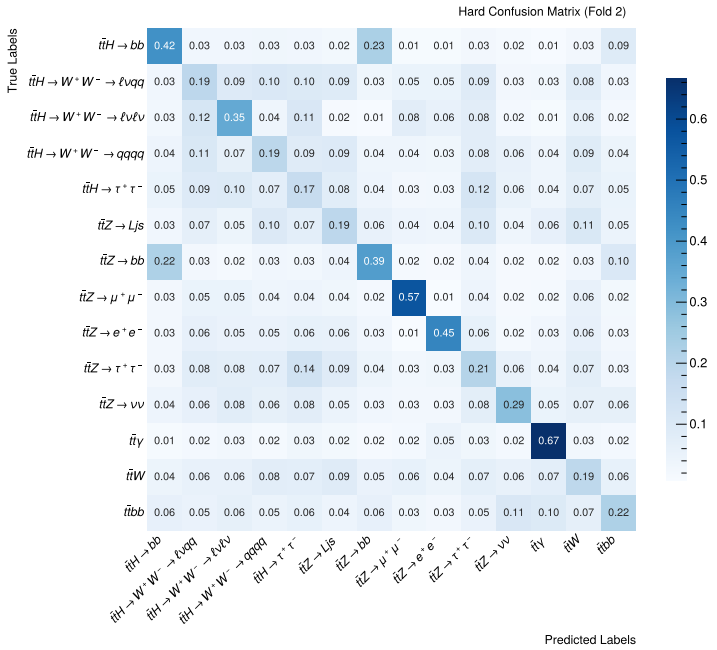
Probability Distributions

In this section we will look at the softmax probability distributions which display a view of how confidently the models assign each event to its true physics category. Figure 4.11 (GATNet) and Figure 4.12 (GTNet) show, for several representative final states. This is calculated by using a softmax function on the model outputs, which are in the form of logits. On the x-axis we have the predicted probability which is assigned by the trained network, while the y-axis shows the normalised count of events for each of the six physics processes considered in the classification task. Every line is from the output of the different true class and is shown in a different colour. Here we will discuss five interesting ones; the others can be seen in the Appendix section. The figures show that the model assigned for every true class event probability is assigned to be larger than 0.1, which shows that the model is able to recognise patterns and show good separation capability. The processes like $t\bar{t}(H \to b\bar{b})$, $t\bar{t}(Z \to b\bar{b})$ and $t\bar{t}b\bar{b}$ share a common $b\bar{b}$ final state, which makes their separation challenging. The signal peaks are broad and extend higher than $p \simeq 0.3$, showing good discrimination, but the overlap or confusion between $t\bar{t}(H \to b\bar{b})$ and $t\bar{t}(Z \to b\bar{b})$ is highly visible. The clean leptonic and photonic channels $t\bar{t}(Z \to \mu^+\mu^-)$ and $t\bar{t}\gamma$ show peaks near unity, which is a good indication and suggests that the networks assign higher probabilities to most of these events. The $t\bar{t}W$ class seems to be the most difficult to separate; $t\bar{t}W$ events in the histogram are between 0.1 and 0.2, suggesting high uncertainty. and only

rarely extends to larger values.
The probability distributions from the GTNet show similar behaviour. This confirms that the conclusions are consistent between architectures which highlight that they are able to learn the underlying physical patterns. GTNet shows good separation capability. Like GATNet, GTNet shows consistent behaviour for processes like $t\bar{t}(H \rightarrow b\bar{b})$, $t\bar{t}(Z \rightarrow b\bar{b})$ and $t\bar{t}b\bar{b}$ which share a common $b\bar{b}$. And the clean leptonic and photonic channels of $t\bar{t}(Z \rightarrow \mu^+\mu^-)$ and $t\bar{t}\gamma$ also show peaks near unity. Similarly, the classes such as $t\bar{t}W$ show that they are difficult to separate, highlighting their similar final state.

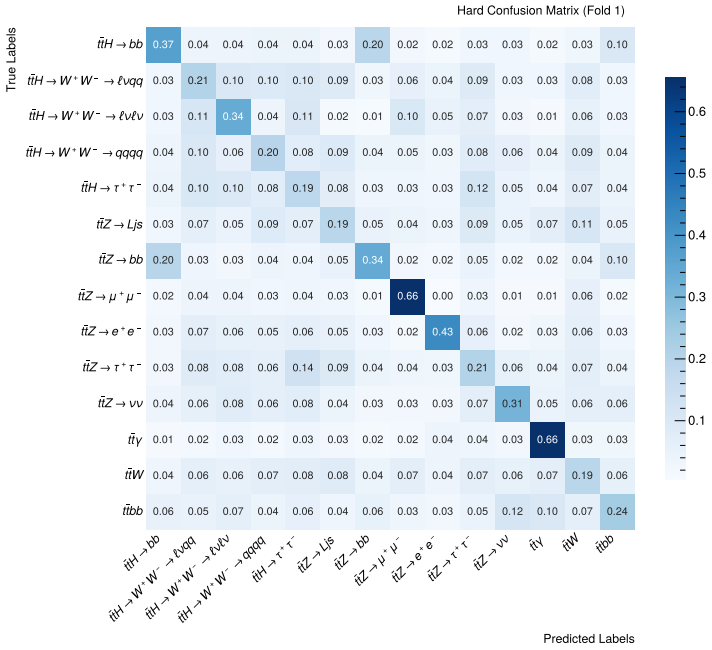


(a) Fold 1

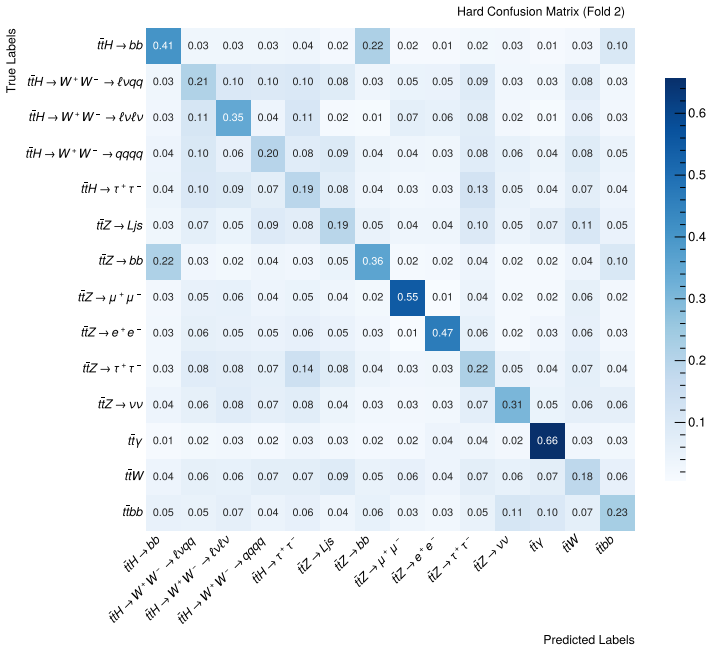


(b) Fold 2

Figure 4.5: Confusion matrices for the GATNet model. Subfigures (a) and (b) show the normalised classification performance on independent test data for Folds 1 and 2, respectively.

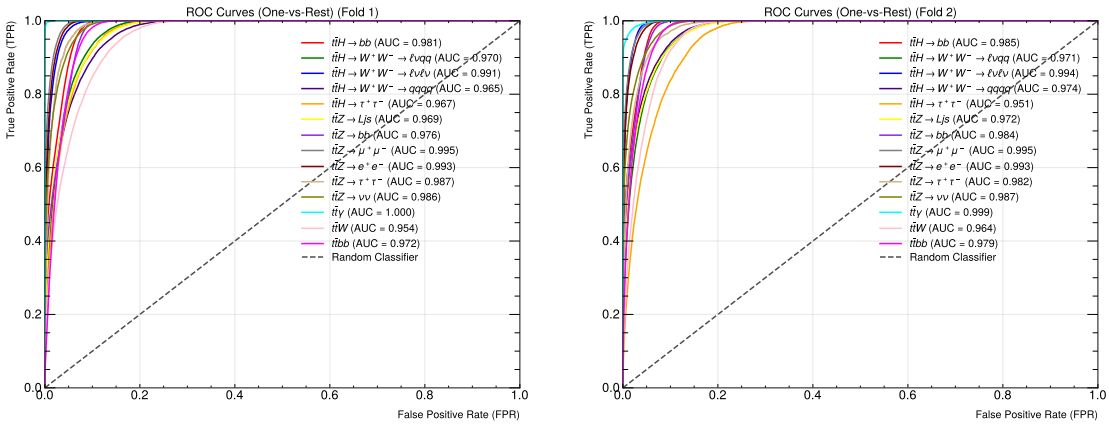


(a) Fold 1



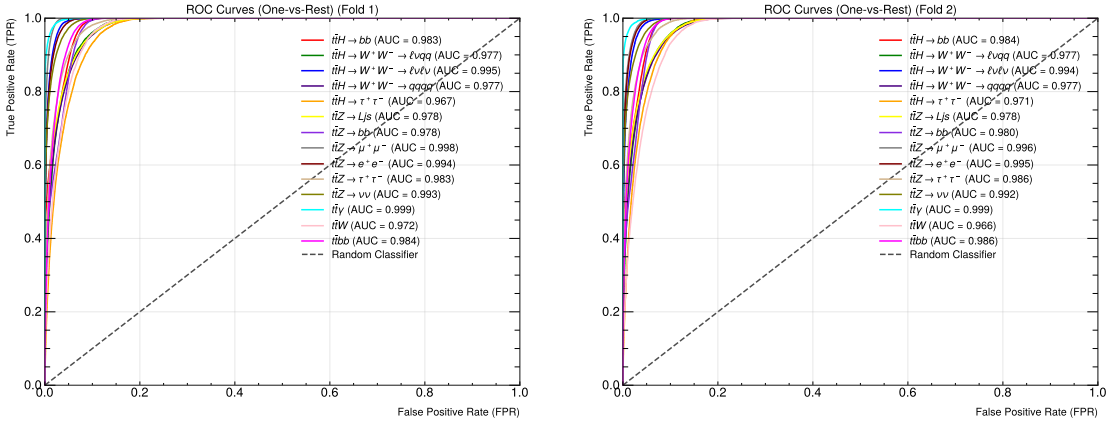
(b) Fold 2

Figure 4.6: Confusion matrices for the GTNet model. Subfigures (a) and (b) show the normalised classification performance on independent test data for Folds 1 and 2, respectively.



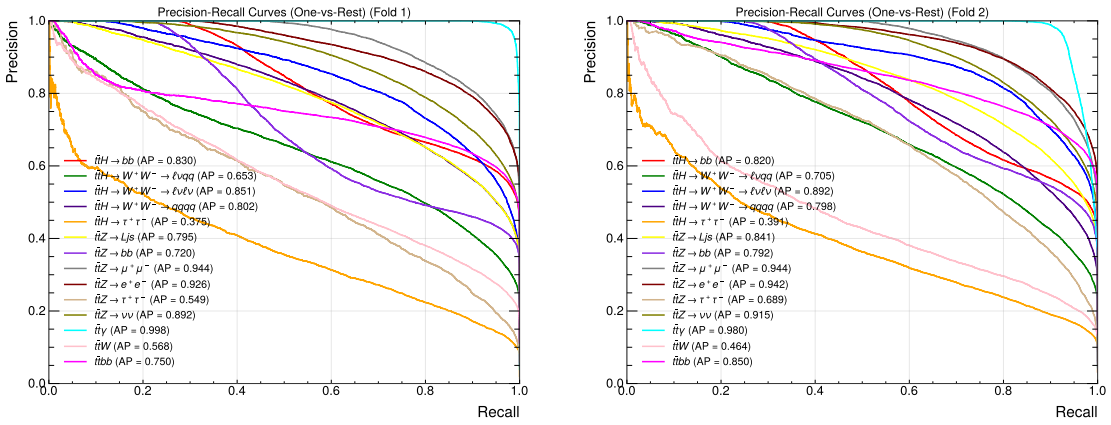
(a) Fold 1 **(b)** Fold 2

Figure 4.7: Receiver operating characteristic (ROC) curves for all signal classes in the $t\bar{t} + X$ classification. The true positive rate is shown versus the false positive rate for (a) GATNet fold 1, (b) GTNet fold 2.



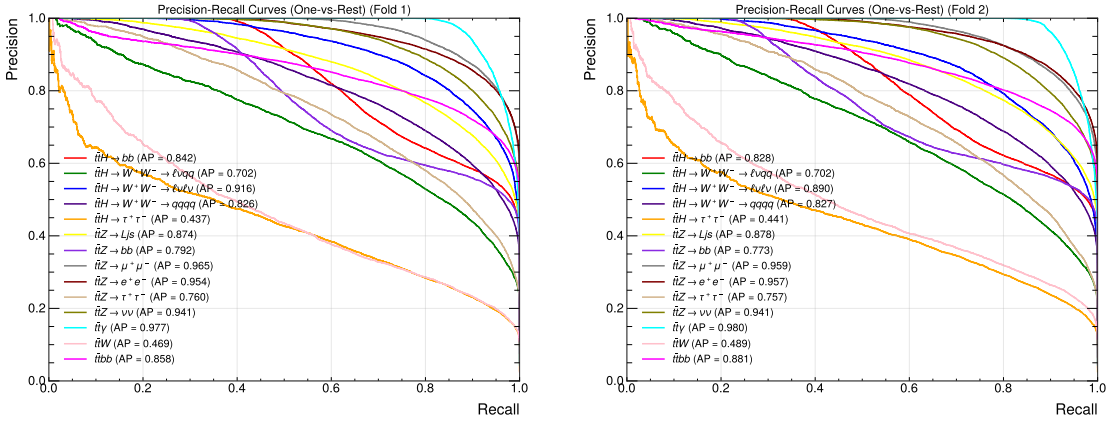
(a) Fold 1 **(b)** Fold 2

Figure 4.8: Receiver operating characteristic (ROC) curves for all signal classes in the $t\bar{t} + X$ classification. The true positive rate is shown versus the false positive rate for (a) GTNet fold 1, (b) GTNet fold 2.



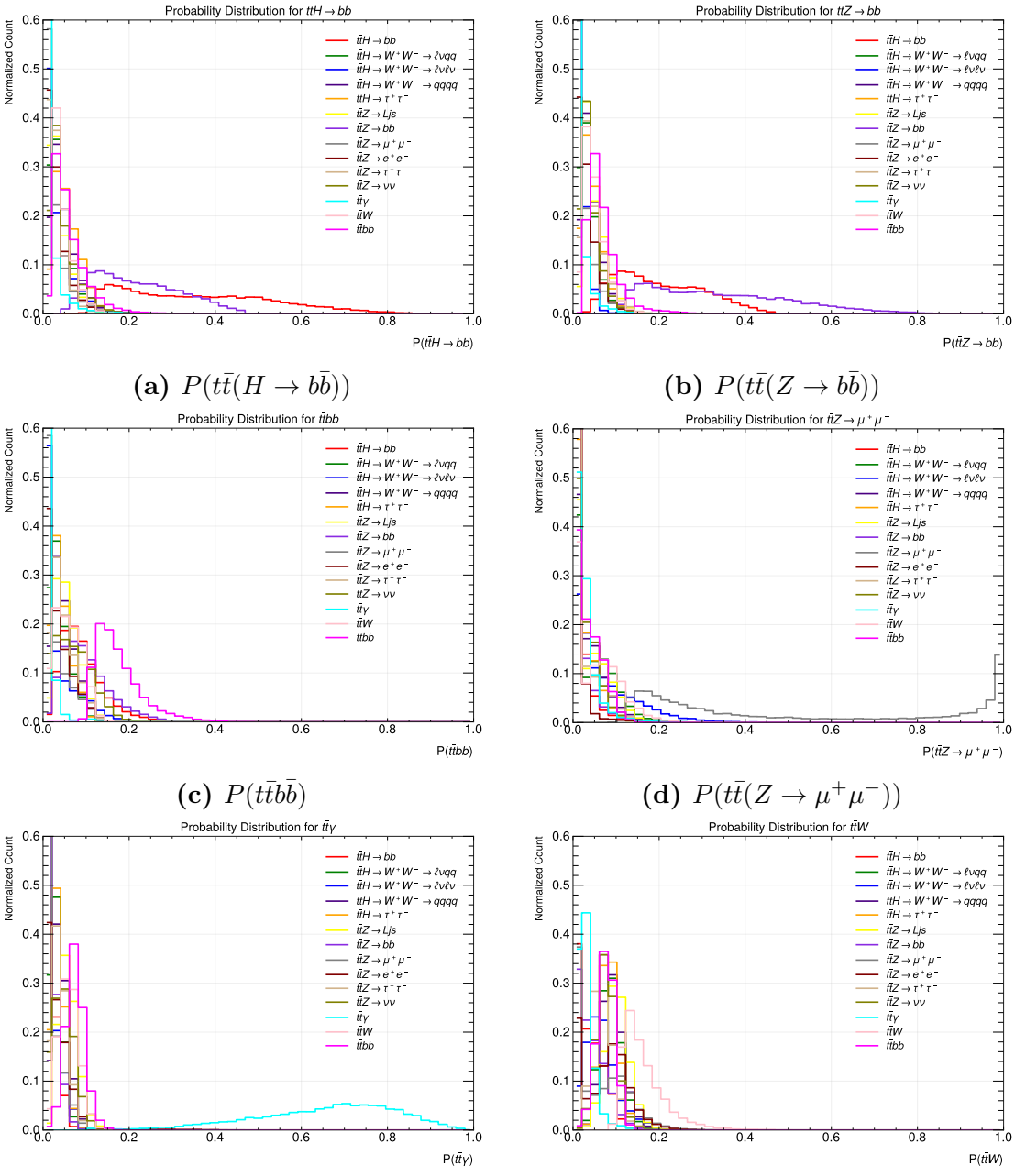
(a) Fold 1 **(b)** Fold 2

Figure 4.9: Precision recall (PR) curves for all signal classes in the $t\bar{t} + X$ classification. Precision is shown versus recall for (a) GATNet fold 1, (b) GATNet fold 2.



(a) Fold 1 **(b)** Fold 2

Figure 4.10: Precision recall (PR) curves for all signal classes in the $t\bar{t} + X$ classification. Precision is shown versus recall for (a) GTNet fold 1, (b) GTNet fold 2.

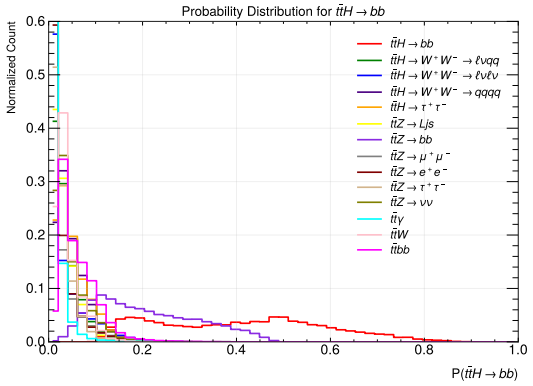


(a) $P(t\bar{t}(H \to b\bar{b}))$ **(b)** $P(t\bar{t}(Z \to b\bar{b}))$

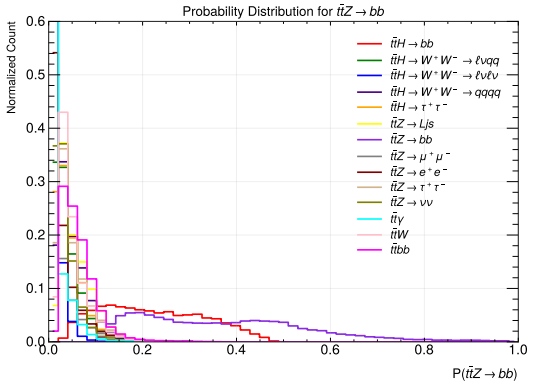
(c) $P(t\bar{t}b\bar{b})$ **(d)** $P(t\bar{t}(Z \to \mu^+\mu^-))$

(e) $P(t\bar{t}\gamma)$ **(f)** $P(t\bar{t}W)$

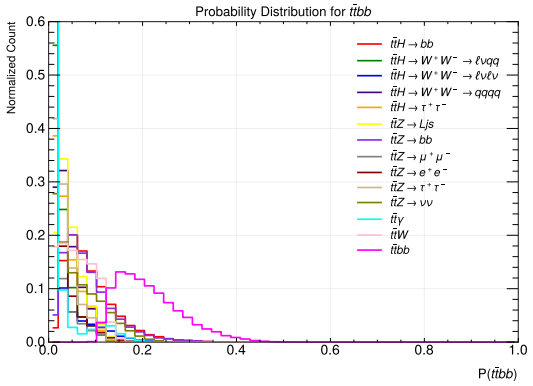
Figure 4.11: Softmax distribution of predicted class probability $P(\text{class})$ from the GATNet for events with the indicated true label.



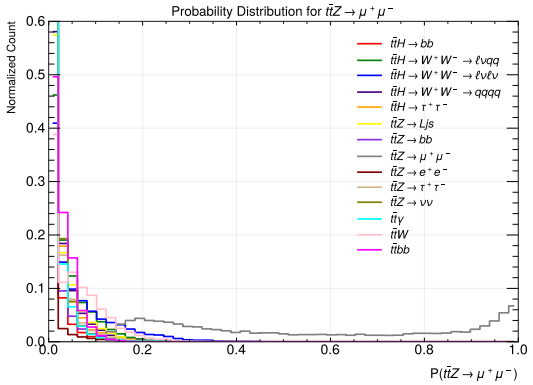
(a) $P(t\bar{t}(H \to b\bar{b}))$



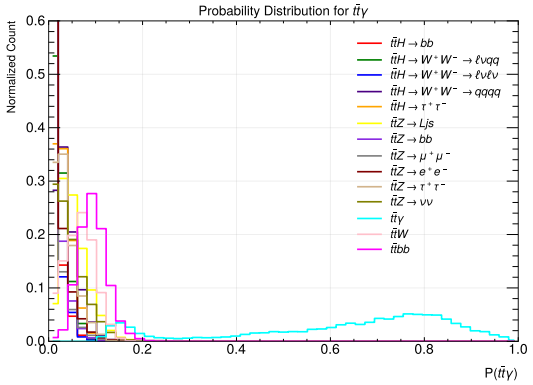
(b) $P(t\bar{t}(Z \to b\bar{b}))$



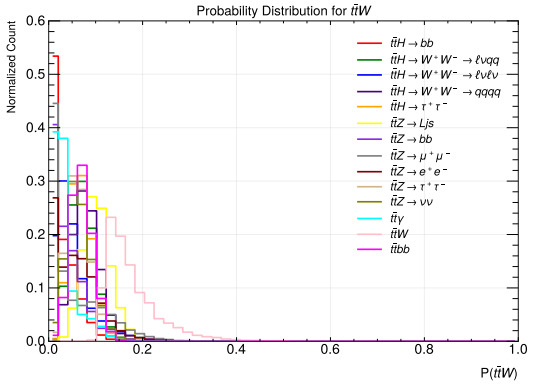
(c) $P(t\bar{t}b\bar{b})$



(d) $P(t\bar{t}(Z \to \mu^+\mu^-))$



(e) $P(t\bar{t}\gamma)$



(f) $P(t\bar{t}W)$

Figure 4.12: Softmax distribution of predicted class probability P(class) from the GTNet for events with the indicated true label.

5 Results and Conclusions

In the performance evaluation, the model demonstrated good separation capability, and the model was able to learn the underlying physical patterns of different classes from their final state. From these separation capabilities, one can study which observables can be used to discriminate the final states and act as effective indicators for distinguishing these final states from each other.

The $t\bar{t}\gamma$ class is the best-performing class. This class involves a top-antitop pair production in association with a photon, which leads to final states involving two W bosons and two b-jets along with this associated photon (γ). The defining experimental feature of this class is a single isolated energetic photon reconstructed in the electromagnetic calorimeter, typically with large transverse momentum due to the production of the photon from the hard scattering process. These characteristics act as an indicator for the model to separate it from the other classes and explain why the models achieve relatively high classification accuracies for this class, as shown in Figure 4.5 and Figure 4.6. The softmax probability distribution in Figure 4.11e for GATNet and Figure 4.12e for GTNet shows $P(t\bar{t}\gamma)$ across all the classes. The model assigned $t\bar{t}\gamma$ softmax probability extends from the range of 0.1 to 0.9, with the softmax probabilities of the other classes ranging from 0 to 0.2. This is reflected in the ROC curves from GAT and GT in Figure 4.7 and Figure 4.8 with AUC score close to unity. The PR curve shown in fig. 4.9 for GATNet and fig. 4.10 for GTNet is consistently smooth with high precision showing flat behaviour ranging from low recall to high recall. This suggests that relaxing thresholds do not impact classification quality due to the clean features separating it from the rest of the classes. Hence, the defining experimental feature of this class is therefore a single isolated energetic photon reconstructed in the electromagnetic calorimeter, typically with large transverse momentum.

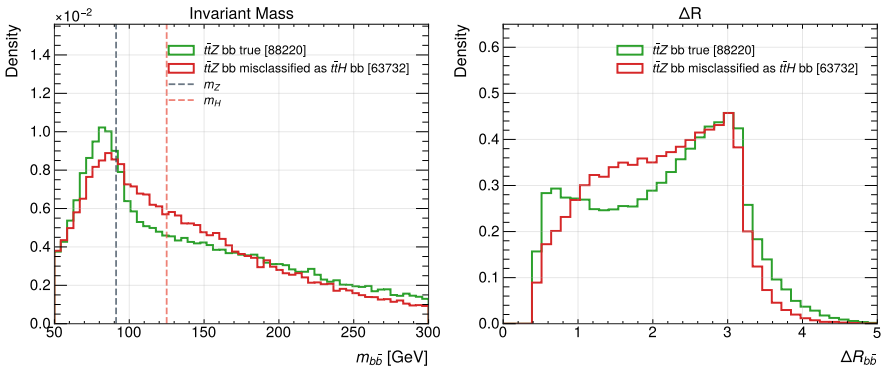
Next in line is the $t\bar{t}(Z \rightarrow \mu^+\mu^-)$ class. In this class, the Z boson produces

a pair of two oppositely charged muons along with a $t\bar{t}$ pair. The presence of $Z \to \mu^+\mu^-$ results in a clean isolated signal of two oppositely charged muons with high transverse momentum. The combined invariant mass of these muons is close to the Z pole mass. The muons identified based on the muon quality requirements with reconstruction using the Muon Spectrometer and Inner Detector sets it apart from the $t\bar{t}(Z \to e^+e^-)$ class. The confusion matrix in Figure 4.5 and Figure 4.6 show 50% to 70% of the events correctly classified as $t\bar{t}(Z \to \mu^+\mu^-)$. The presence of muons in other classes contributes to the misclassification of this class. The softmax probability distribution in Figure 4.11d and Figure 4.12d shows $P(t\bar{t}(Z \to \mu^+\mu^-))$ across all the classes. The model assigned $P(t\bar{t}(Z \to \mu^+\mu^-))$ softmax probability extends from the range of 0.1 to 1, with the softmax probabilities of the other classes ranging from 0 to 0.3, with $t\bar{t}(H \to WW^* \to l\nu l\nu)$ (ranging from 0 to 0.3) showing a minor overlap in the lower end ranging from 0.1 to 0.3. The $t\bar{t}(Z \to \mu^+\mu^-)$ process shows a peak at 1, showing model confidence for separating this class. This is reflected in the ROC curves from GAT and GT in Figure 4.7 and Figure 4.8 with AUC scores close to unity. The PR curve is consistently smooth with high precision, showing flat behaviour ranging from low recall to high recall. The curve starts dipping at a recall of 0.6. This suggests that thresholds that correspond to recall values below 0.6 do not impact classification quality due to the clean physical features separating it from the rest of the classes.

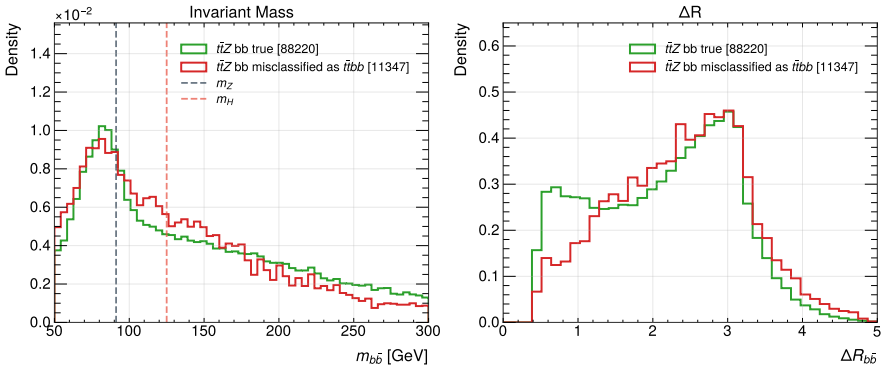
The $t\bar{t}(Z \to e^+e^-)$ class shows similar performance to $t\bar{t}(Z \to \mu^+\mu^-)$ due to the similarity in physical properties. In this class the Z boson produces an electron-positron pair along with a $t\bar{t}$ pair. The presence of $Z \to e^+e^-$ results in a clean isolated signal of an electron-positron pair with high transverse momentum from the hard scattering. The electrons and positrons identified based on their tracks in the inner detectors and energy depositions in the electromagnetic calorimeter set this class apart from the $t\bar{t}(Z \to \mu^+\mu^-)$ class. The invariant mass of these electron-positron pair lies close to the Z pole mass, like the previous class. The confusion matrix in Figure 4.5 and Figure 4.6 shows 40% to 50% of the events to be correctly classified as $t\bar{t}(Z \to e^+e^-)$. The presence of an electron-positron pair in other classes leads to misclassification of this channel. The softmax probability distribution in Figure 4.11e and Figure 4.12e shows $P(t\bar{t}(Z \to e^+e^-))$ across all the classes. The model assigned $P(t\bar{t}(Z \to e^+e^-))$ softmax probability extends from

the range of 0.1 to 1, with the softmax probabilities of the other classes ranging from 0 to 0.3, with $t\bar{t}(H \to WW^* \to l\nu l\nu)$ (ranging from 0 to 0.3) showing a minor overlap in the lower end ranging from 0.1 to 0.3. The $t\bar{t}(Z \to e^+e^-)$ process shows a peak at 1, showing model confidence for separating this class. This is reflected in the ROC curves from GAT and GT in Figure 4.7 and Figure 4.8 with AUC score close to unity. The PR curve is consistently smooth with high precision, showing flat behaviour ranging from low recall to high recall. The curve starts dipping at a recall of 0.4. This suggests that thresholds corresponding to the recall values below 0.4 do not impact classification quality due to the clean physical features separating it from the rest of the classes. But it does have minor overlap with other classes having similar final states.

In the $t\bar{t}(Z \to b\bar{b})$ class, the Z boson decays into a pair of bottom quarks, adding two additional b-jets to the two already present from the top decays. The presence of a high b-tag score and a combined $b\bar{b}$ invariant mass close to the Z pole mass and $\Delta R_{b\bar{b}}$ helps it to be separated from the other classes. The network can exploit correlations in b-tag scores and reconstructed invariant masses near the Z-boson mass to enhance discrimination. The confusion matrix shows that around 30% of these events were classified correctly as $t\bar{t}(Z \to b\bar{b})$. But this class shows a high amount of confusion of this class with $t\bar{t}(H \to b\bar{b})$ and $t\bar{t}b\bar{b}$. This is due to the physical similarities of these processes. These processes have additional b-jets along with a $t\bar{t}$ pair. The distributions of softmax probabilities are shown in Figure 4.11b and Figure 4.12b shows broadly distributed $t\bar{t}Z \to b\bar{b}$ with a clear overlapping distribution of $t\bar{t}(H \to b\bar{b})$. The $t\bar{t}(Z \to b\bar{b})$ distribution ranges from 0.1 to 1 and shows clear separation capability from other classes distributed from 0 to 0.5 (considering $t\bar{t}(H \to b\bar{b})$). The ROC curves show good separation with high AUC score. The PR curve shows no fluctuations in the high-precision and low-recall region. The curve takes a steep drop at 0.3 with a medium AP score. This suggests that the separation power of this class drops highly at thresholds corresponding to the recall values above 0.3. The misclassification of the $t\bar{t}(Z \to b\bar{b})$ process can be studied by examining whether the model classifies it based on the combined invariant mass and ΔR of the b-jets. This can be seen in Figure 5.1 which shows the combined $m_{b\bar{b}}$ and $\Delta R_{b\bar{b}}$ distributions from Z boson decay, in which the b-jets are ΔR matched with the parton-history (truth). In Figure 5.1a the invariant mass



(a) $t\bar{t}(Z \to b\bar{b})$ misclassified as $t\bar{t}H \to b\bar{b}$

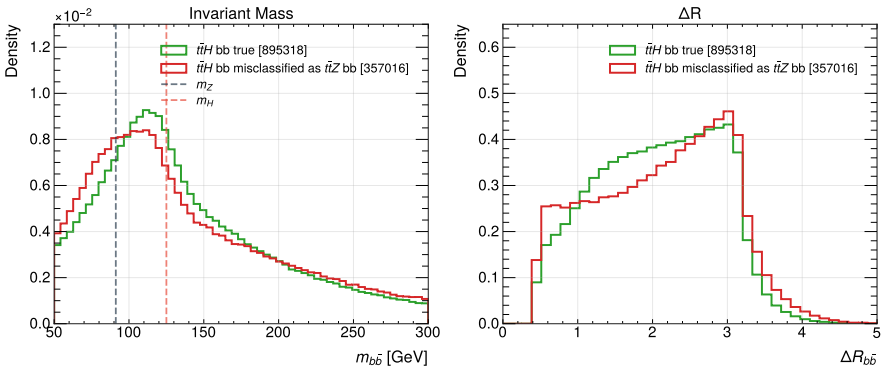


(b) $t\bar{t}(Z \to b\bar{b})$ misclassified as $t\bar{t}b\bar{b}$

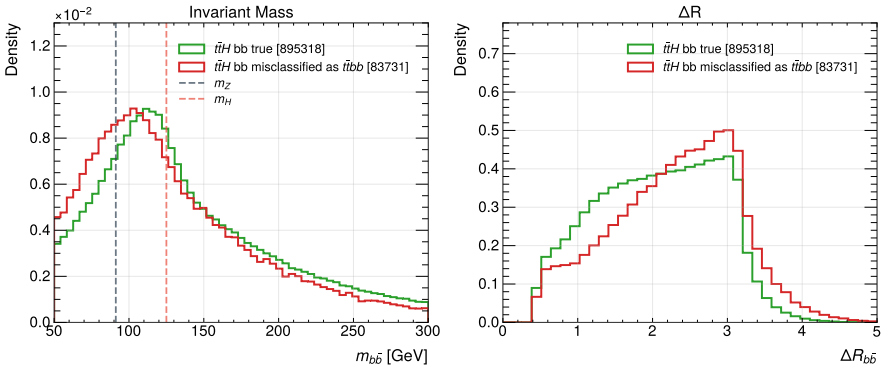
Figure 5.1: Distributions of the invariant mass $m_{b\bar{b}}$ and angular separation $\Delta R_{b\bar{b}}$ of the two truth-matched b-jets from Z boson decays in correctly classified $t\bar{t}(Z \to b\bar{b})$ events and those misclassified as (a) $t\bar{t}(H \to b\bar{b})$ or (b) $t\bar{t}b\bar{b}$.

of the $b\bar{b}$ ($m_{b\bar{b}}$) from $t\bar{t}(Z \to b\bar{b})$ peaks around m_Z, and the $t\bar{t}(Z \to b\bar{b})$ events that are misclassified as $t\bar{t}(H \to b\bar{b})$ events show a broader peak. The ΔR_{bb} distributions show different characteristics in the misclassified events. Similarly, the Figure 5.1b shows $t\bar{t}(Z \to b\bar{b})$ misclassified as $t\bar{t}b\bar{b}$.

The $t\bar{t}(H \to b\bar{b})$ class shows similar behaviour in the confusion matrix, ROC and PR curve. Its final state contains two b-quarks from the Higgs boson and two additional b-quarks from the top decays, with light jets and leptons originating from the W bosons. The softmax probability distribution in Figure 4.11a and Figure 4.11b shows similar behaviour. The invariant mass of b-jet pairs peaks near the Higgs boson mass shown in Figure 5.2. The $t\bar{t}(H \to b\bar{b})$ events that are misclassified as



(a) $t\bar{t}H \to b\bar{b}$ misclassified as $t\bar{t}Z \to b\bar{b}$

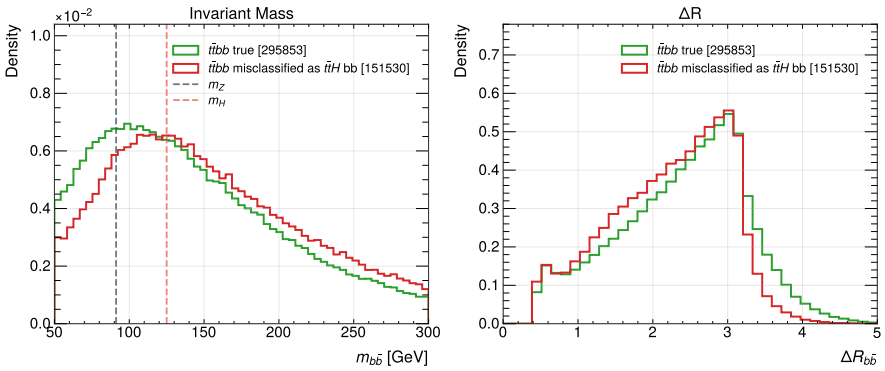


(b) $t\bar{t}H \to b\bar{b}$ misclassified as $t\bar{t}b\bar{b}$

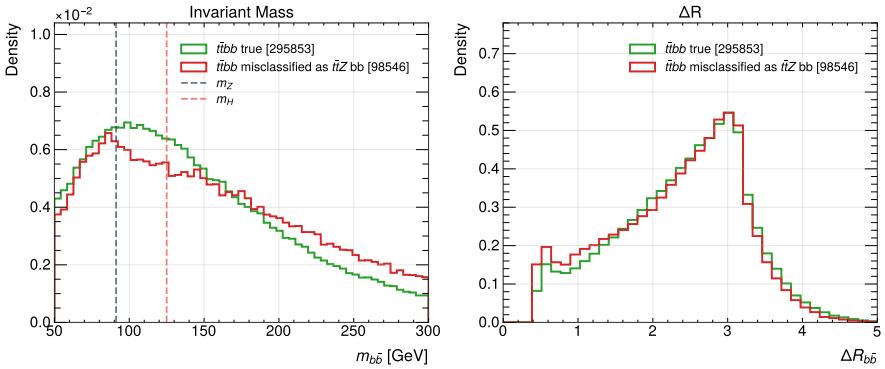
Figure 5.2: Distributions of the invariant mass $m_{b\bar{b}}$ and angular separation $\Delta R_{b\bar{b}}$ of the two truth-matched b-jets from Z boson decays in correctly classified $t\bar{t}(H \to b\bar{b})$ events and those misclassified as (a) $t\bar{t}(Z \to b\bar{b})$ or (b) $t\bar{t}b\bar{b}$.

$t\bar{t}(Z \to b\bar{b})$ events show a shift in the invariant mass distribution near the Z mass and show a ΔR_{bb} distribution similar to ΔR_{bb} from true $t\bar{t}Z \to b\bar{b}$ in Figure 5.1. The $t\bar{t}(H \to b\bar{b})$ events that are misclassified as $t\bar{t}b\bar{b}$ show a continuous invariant mass distribution but show a ΔR_{bb} distribution similar to the ΔR_{bb} distribution from true $t\bar{t}b\bar{b}$ in Figure 5.3.

The $t\bar{t}b\bar{b}$ class shows a different behaviour since it has no resonance in invariant mass. The $t\bar{t}b\bar{b}$ process produces a top-antitop pair accompanied by an additional $b\bar{b}$ pair from QCD radiation. The $t\bar{t}b\bar{b}$ class shows around 22% as correctly classified as $t\bar{t}b\bar{b}$ in the confusion matrix. But it shows better behaviour in the ROC curve with high AUC and PR curves with high AP. In the softmax probability distribution



(a) $t\bar{t}b\bar{b}$ misclassified as $t\bar{t}(H \to b\bar{b})$



(b) $t\bar{t}b\bar{b}$ misclassified as $t\bar{t}(Z \to b\bar{b})$

Figure 5.3: Distributions of the invariant mass $m_{b\bar{b}}$ and angular separation $\Delta R_{b\bar{b}}$ of the two truth-matched b-jets from Z boson decays in correctly classified $t\bar{t}b\bar{b}$ events and those misclassified as (a) $t\bar{t}(H \to b\bar{b})$ or (b) $t\bar{t}(Z \to b\bar{b})$.

$t\bar{t}b\bar{b}$ shown in Figure 4.11c and Figure 4.12c has a relatively narrow peak from 0.1 to around 0.4, with rest of the classes ranging from 0 to 0.2. The Figure 5.3 shows similar behaviour.

In the semi-leptonic $t\bar{t}(H \to W^+W^- \to \ell\nu qq)$ class, the Higgs boson decays into a W^+W^- pair, and one W of the Higgs boson decays leptonically and the other one decays hadronically. This makes this class very complex. This means along with the $t\bar{t}$ decay products, there is a Higgs boson decaying into semi-leptonic WW^* out of which at least one W must be off-shell since the Higgs boson mass is smaller than twice the W mass. This final state hence has one lepton with high transverse momentum, $E_{\mathrm{T}}^{\mathrm{miss}}$ and two light quark jets. In the confusion matrix, this class

shows around 20% of the correctly classified events as $t\bar{t}(H \to W^+W^- \to \ell\nu qq)$. The softmax distribution in the Appendix Figure 6.1a and Figure 6.3a shows a relatively narrow peak between 0.1 and 0.4. The rest of the distributions from other classes overlap significantly and are situated further than 0. These classes include $t\bar{t}(H \to W^+W^- \to \ell\nu\ell\nu)$, $t\bar{t}(H \to W^+W^- \to qqqq)$, $t\bar{t}(H \to \tau\tau)$, $t\bar{t}(Z \to Ljs)$ and $t\bar{t}(Z \to \tau\tau)$. This is due to combinatorial ambiguities in which W bosons from top quarks decay and τ lepton decays are confused with W bosons decayed from Higgs bosons. This can be seen in the PR curve, which shows a poor AP score and fluctuations at low recall values with an instant drop in the curve as recall increases. The relaxing threshold introduces misclassified events. But at high thresholds there is good separation against the rest, which explains the ROC curve with high AUC score. Similarly, $t\bar{t}(H \to W^+W^- \to \ell\nu\ell\nu)$, $t\bar{t}(H \to W^+W^- \to qqqq)$, $t\bar{t}(H \to \tau\tau)$, $t\bar{t}(Z \to ljs)$ and $t\bar{t}(Z \to \tau\tau)$ have major overlaps due to combinatorial ambiguities, with $t\bar{t}(H \to W^+W^- \to \ell\nu\ell\nu)$ having better separation and $t\bar{t}(H \to \tau\tau)$ having the worst separation among all classes shown in the Appendix 6.
The $t\bar{t}(Z \to \nu\nu)$ channel has two additional neutrinos originating from Z boson decay. This results in a large $E_{\mathrm{T}}^{\mathrm{miss}}$. The confusion matrix shows around 30% correctly classified as $t\bar{t}(Z \to \nu\nu)$. The $t\bar{t}W$ is the most confused class in the confusion matrix. The PR curve shows similar behaviour to $t\bar{t}(H \to \tau\tau)$. The associated W boson can decay leptonically and hadronically. This hadronically decaying W boson results in two jets. The combined invariant mass is close to the W boson mass. The W boson mass is close to Z boson, which also decays to two jets. This causes $t\bar{t}W$ to be misclassified as $t\bar{t}Z \to Ljs$ (light quarks). This is seen in the softmax probability distribution in Figure 4.11f and Figure 4.12f.
This work has demonstrated the application of Graph neural networks to the multi-class classification of $t\bar{t} + X$ processes in ATLAS. A complete end-to-end pipeline was built. Both Graph Attention Network and Graph Transformer show capabilities of separating $t\bar{t} + X$ final states. The model's performance shows which observables help separate final states successfully. This shows where the future of experimental analysis will gain most from such neural networks.

Bibliography

[1] S. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22**, 579 (1961)

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

[3] A. Salam, *Weak and Electromagnetic Interactions*, ed. Nobel Symposium No. 8 (Almqvist & Wiksell, Stockholm, 1968) (1968)

[4] A. Salam, *Weak and Electromagnetic Interactions*, Almqvist & Wiksell, Stockholm, nobel symposium 8 edition (1968)

[5] M. Kobayashi, T. Maskawa, CP Violation in the Renormalizable Theory of Weak Interaction, *Prog. Theor. Phys.* **49**, 652 (1973)

[6] CDF Collaboration, Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab, *Phys. Rev. Lett.* **74**, 2626 (1995)

[7] D0 Collaboration, Observation of the Top Quark, *Phys. Rev. Lett.* **74**, 2632 (1995)

[8] M. B. et al., Top-Quark Physics, *Phys. Rep.* **803**, 1 (2019)

[9] ATLAS and CMS Collaborations, Combination of Measurements of the Top Quark Mass from Data Collected by the ATLAS and CMS Experiments at $\sqrt{s} = 7$ and 8 TeV, *Phys. Rev. Lett.* **132**, 261902 (2024)

[10] ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3, *JINST* **19**, P05063 (2024)

[11] ATLAS Collaboration, Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ Cross Sections in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector, *Phys. Rev. D* **99**, 072009 (2019)

[12] CMS Collaboration, Observation of Top Quark Pairs Produced in Association with a Vector Boson in pp Collisions at $\sqrt{s} = 8$ TeV, *Phys. Rev. Lett.* **115**, 032002 (2015)

[13] ATLAS Collaboration, Inclusive and Differential Cross-Section Measurements of $t\bar{t}Z$ Production in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector, Including EFT and Spin-Correlation Interpretations, *JHEP* **07**, 163 (2024)

[14] S. N. et al., P. D. Group, Review of Particle Physics, *Phys. Rev. D* **110**, 030001 (2024)

[15] S. L. Glashow, Partial Symmetries of Weak Interactions, *Nucl. Phys.* **22**, 579 (1961)

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

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

[18] C.-N. Yang, R. L. Mills, Conservation of Isotopic Spin and Isotopic Gauge Invariance, *Phys. Rev.* **96**, 191 (1954)

[19] D. J. Gross, F. Wilczek, Ultraviolet Behavior of Non-Abelian Gauge Theories, *Phys. Rev. Lett.* **30**, 1343 (1973)

[20] H. D. Politzer, Reliable Perturbative Results for Strong Interactions?, *Phys. Rev. Lett.* **30**, 1346 (1973)

[21] E. Noether, Invariante Variationsprobleme, *Nachr. Königl. Ges. Wiss. Göttingen, Math.-Phys. Kl.* pages 235–257 (1918)

[22] R. Laming, On the Electrical Phenomena Accompanying the Disruption of Water and Other Liquids by Voltaic Action, *Philos. Mag.* **13**, 322 (1838)

[23] R. Laming, On the General Theory of Electricity, *Philos. Mag.* **14**, 318 (1839)

[24] J. J. Thomson, Cathode Rays, *Philos. Mag.* **44**, 293 (1897)

[25] A. Einstein, Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt, *Ann. Phys.* **17**, 132 (1905)

[26] G. N. Lewis, The Conservation of Photons, *Nature* **118**, 874 (1926)

[27] S. N. Bose, Plancks Gesetz und Lichtquantenhypothese, *Z. Phys.* **26**, 178 (1924)

[28] A. Einstein, Quantentheorie des Einatomigen Idealen Gases, *Sitzungsber. Preuss. Akad. Wiss.* pages 261–267 (1924)

[29] H. Yukawa, On the Interaction of Elementary Particles. I, *Proc. Phys.-Math. Soc. Jpn.* **17**, 48 (1935)

[30] M. Planck, Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum, *Verh. Dtsch. Phys. Ges.* **2**, 237 (1900)

[31] M. Planck, Über das Gesetz der Energieverteilung im Normalspectrum, *Ann. Phys.* **4**, 553 (1901)

[32] P. A. M. Dirac, The Quantum Theory of the Emission and Absorption of Radiation, *Proc. R. Soc. Lond. A* **114**, 243 (1927)

[33] J. Schwinger, On Quantum Electrodynamics and the Magnetic Moment of the Electron, *Phys. Rev.* **73**, 416 (1948)

[34] R. P. Feynman, Space–Time Approach to Quantum Electrodynamics, *Phys. Rev.* **76**, 769 (1949)

[35] F. J. Dyson, The S-Matrix in Quantum Electrodynamics, *Phys. Rev.* **75**, 1736 (1949)

[36] W. E. L. Jr., R. C. Retherford, Fine Structure of the Hydrogen Atom by a Microwave Method, *Phys. Rev.* **72**, 241 (1947)

[37] S. Tomonaga, On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields, *Prog. Theor. Phys.* **1**, 27 (1946)

[38] E. Fermi, Tentativo di una Teoria dell'Emissione dei Raggi β, *Nuovo Cim.* **11**, 1 (1934)

[39] T. D. Lee, C. N. Yang, Question of Parity Conservation in Weak Interactions, *Phys. Rev.* **104**, 254 (1956)

[40] E. A. C. S. Wu, R. W. H. et al., Experimental Test of Parity Conservation in Beta Decay, *Phys. Rev.* **105**, 1413 (1957)

[41] UA1 Collaboration, Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV, *Phys. Lett. B* **122**, 103 (1983)

[42] UA1 Collaboration, Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider, *Phys. Lett. B* **126**, 398 (1983)

[43] UA2 Collaboration, Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $\bar{p}p$ Collider, *Phys. Lett. B* **122**, 476 (1983)

[44] UA2 Collaboration, Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider, *Phys. Lett. B* **129**, 130 (1983)

[45] E. Fermi, Tentativo di una teoria dell'emissione dei raggi β, *Nuovo Cim.* **11**, 1 (1934), english translation: *Z. Phys.* **88**, 161 (1934).

[46] E. C. G. Sudarshan, R. E. Marshak, Chirality Invariance and the Universal Fermi Interaction, *Phys. Rev.* **109**, 1860 (1958)

[47] R. P. Feynman, M. Gell-Mann, Theory of the Fermi Interaction, *Phys. Rev.* **109**, 193 (1958)

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

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

[50] Gargamelle Neutrino Collaboration, Observation of Neutrino-Like Interactions Without Muon or Electron in the Gargamelle Neutrino Experiment, *Phys. Lett. B* **46**, 138 (1973)

[51] UA1 Collaboration, Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV, *Phys. Lett. B* **122**, 103 (1983)

[52] UA2 Collaboration, Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN pp Collider, *Phys. Lett. B* **122**, 476 (1983)

[53] E. Rutherford, Collision of α Particles with Light Atoms. IV. An Anomalous Effect in Nitrogen, *Philos. Mag.* **37**, 581 (1919)

[54] E. Rutherford, Bakerian Lecture: Nuclear Constitution of Atoms, *Proc. R. Soc. Lond. A* **97**, 374 (1920)

[55] J. Chadwick, The Existence of a Neutron, *Proc. R. Soc. Lond. A* **136**, 692 (1932)

[56] M. Gell-Mann, A Schematic Model of Baryons and Mesons, *Phys. Lett.* **8**, 214 (1964)

[57] G. Zweig, An SU(3) Model for Strong Interaction Symmetry and Its Breaking, Technical Report 8182/TH.401, CERN (1964)

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

[59] J. D. Bjorken, E. A. Paschos, Inelastic Electron–Proton and Gamma–Proton Scattering, and the Structure of the Nucleon, *Phys. Rev.* **185**, 1975 (1969)

[60] UA2 Collaboration, Evidence for Gluon Bremsstrahlung in e^+e^- Annihilations, *Phys. Lett. B* **118**, 167 (1982)

[61] ATLAS Collaboration, Measurements of $t\bar{t} + X$ Production: Summary of Associated Production with Jets, Bosons, and Heavy Flavour, ATLAS Public Results (2025)

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

[63] S. P. Z. Wu, F. C. et al., A Comprehensive Survey on Graph Neural Networks, *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 4 (2021)

[64] V. P. Dwivedi, X. Bresson, A Generalization of Transformer Networks to Graphs, *Patterns* **2**, 100377 (2021)

[65] M. J. S. Yun, R. Kim, Graph Transformer Networks, *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 2706 (2020)

[66] A. V. et al., Attention Is All You Need, in Adv. Neural Inf. Process. Syst., volume **30**, pages 5998–6008 (2017)

[67] Y. S. et al., Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification, in Proc. 30th Int. Joint Conf. Artif. Intell. (IJCAI), pages 1548–1554 (2021)

[68] R. Brun, F. Rademakers, ROOT: An Object-Oriented Data Analysis Framework, *Nucl. Instrum. Meth. A* **389**, 81 (1997)

[69] ATLAS Collaboration, Electron Reconstruction and Identification in the ATLAS Experiment Using the 2015 and 2016 LHC Proton–Proton Collision Data at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **79**, 639 (2019)

[70] ATLAS Collaboration, Muon Reconstruction and Identification Efficiency in ATLAS Using the Full Run 2 *pp* Collision Data Set at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **81**, 578 (2021)

[71] G. P. Salam, Towards Jetography, *Eur. Phys. J. C* **67**, 637 (2010)

[72] ATLAS Collaboration, Measurement of the Photon Identification Efficiencies with the ATLAS Detector Using LHC Run 2 Data Collected in 2015 and 2016, *Eur. Phys. J. C* **79**, 205 (2019)

[73] ATLAS Collaboration, Electron and Photon Performance Measurements with the ATLAS Detector Using the 2015–2017 LHC Proton–Proton Collision Data, *JINST* **14**, P12006 (2019)

[74] ATLAS Collaboration, Jet Energy Scale and Resolution Measured in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector, *Eur. Phys. J. C* **80**, 1104 (2020)

[75] ATLAS Collaboration, Jet Energy Measurement with the ATLAS Detector in Proton–Proton Collisions at $\sqrt{s} = 7$ TeV, *Eur. Phys. J. C* **73**, 2304 (2013)

[76] ATLAS Collaboration, Measurement of Inclusive Jet and Dijet Cross Sections in Proton–Proton Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector, *Phys. Rev. D* **86**, 014022 (2012)

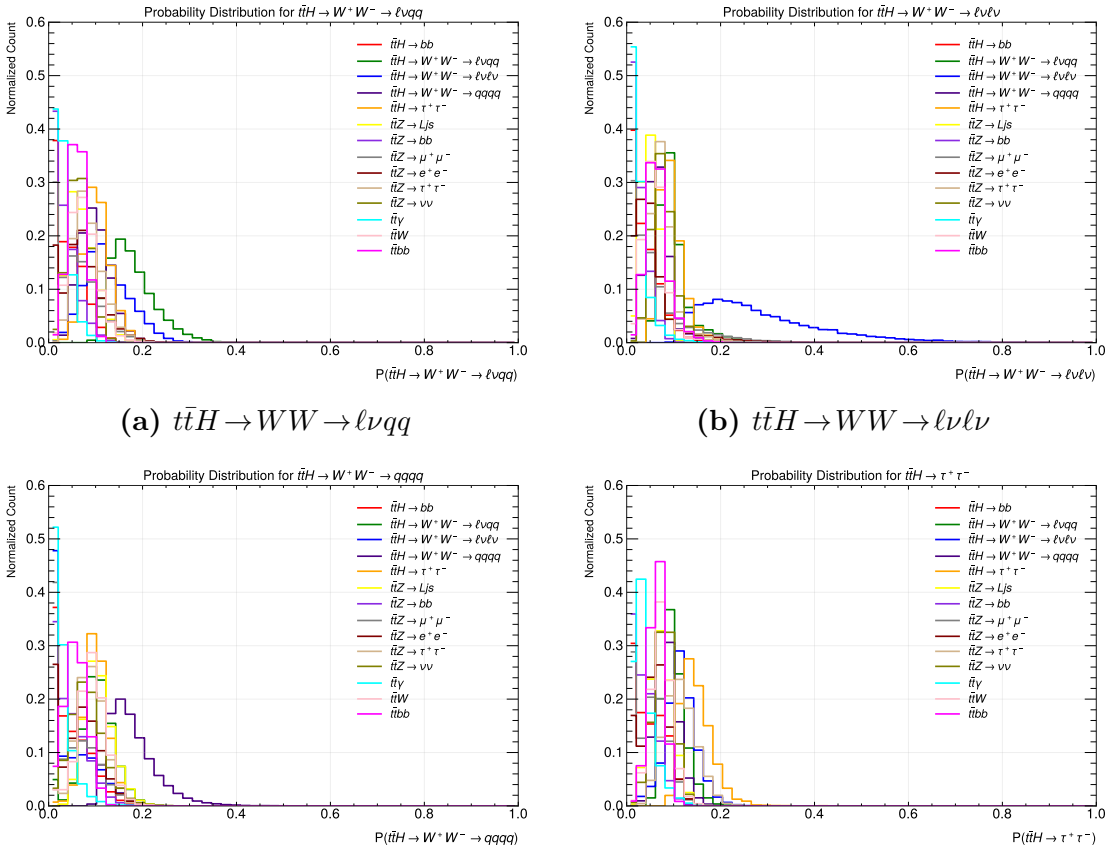
[77] D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, *arXiv preprint* (2014)

[78] R. Kohavi, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, in Proc. 14th Int. Joint Conf. Artif. Intell. (IJCAI), volume **2**, pages 1137–1143 (1995)

[79] Y. B. I. Goodfellow, A. Courville, Deep Learning, MIT Press (2016)

[80] L. Prechelt, Automatic Early Stopping Using Cross Validation: Quantifying the Criteria, *Neural Networks* **11(4)**, 761 (1998)

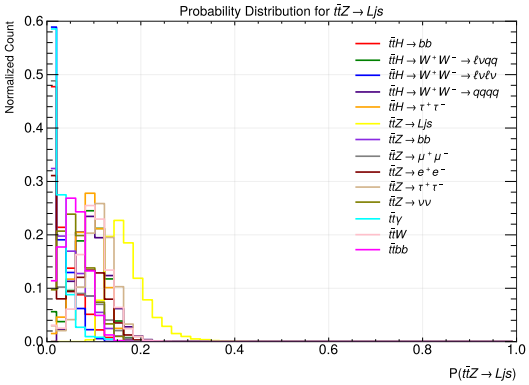
6 Appendix



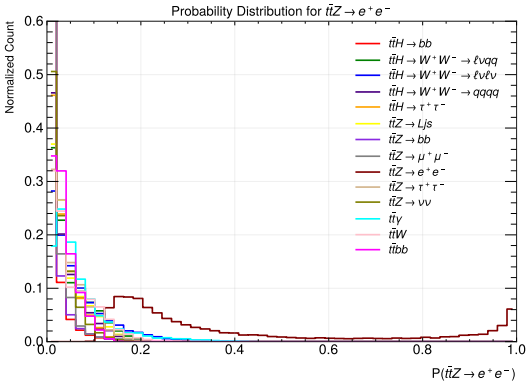
(a) $t\bar{t}H \to WW \to \ell\nu qq$ **(b)** $t\bar{t}H \to WW \to \ell\nu\ell\nu$

(c) $t\bar{t}(H \to WW \to qqqq)$ **(d)** $t\bar{t}(H \to \tau^+\tau^-)$

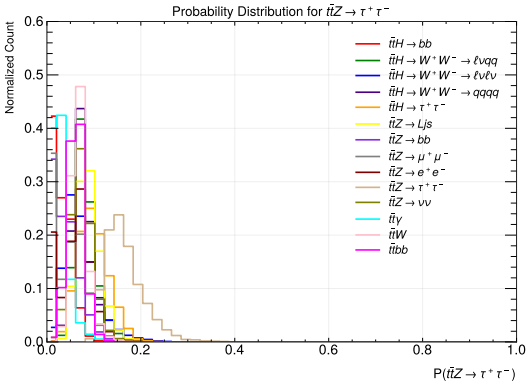
Figure 6.1: Appendix: normalised distributions of GATNet Fold-1 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.



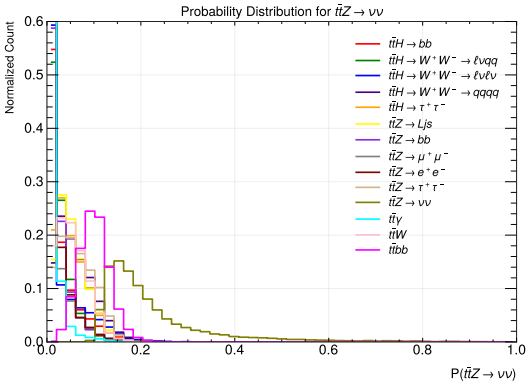
(e) $t\bar{t}(Z \to \mathrm{ljs})$



(f) $t\bar{t}(Z \to e^+e^-)$

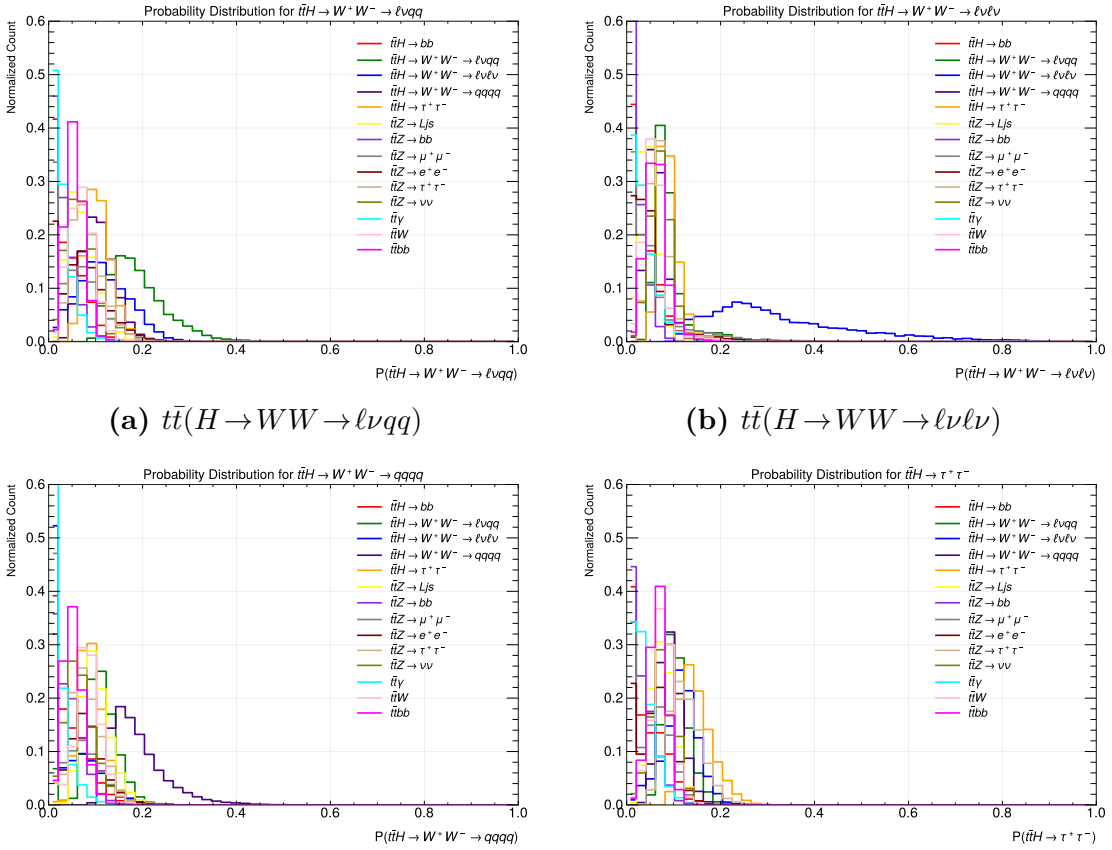


(g) $t\bar{t}(Z \to \tau^+\tau^-)$



(h) $t\bar{t}(Z \to \nu\bar{\nu})$

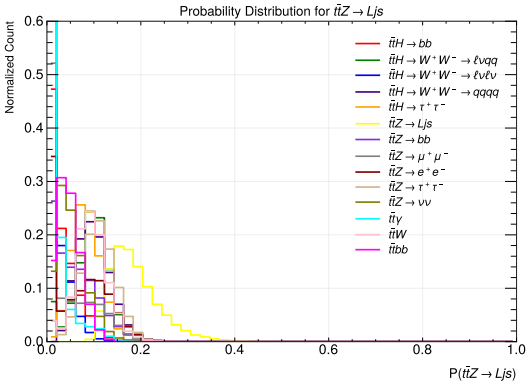
Figure 6.1: Appendix (continued): normalised distributions of GATNet Fold-1 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.



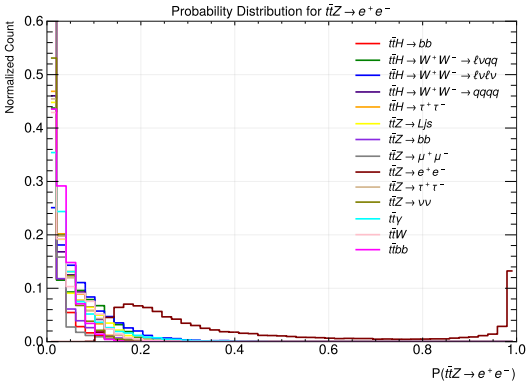
(a) $t\bar{t}(H \to WW \to \ell\nu qq)$ **(b)** $t\bar{t}(H \to WW \to \ell\nu\ell\nu)$

(c) $t\bar{t}(H \to WW \to qqqq)$ **(d)** $t\bar{t}(H \to \tau^+\tau^-)$

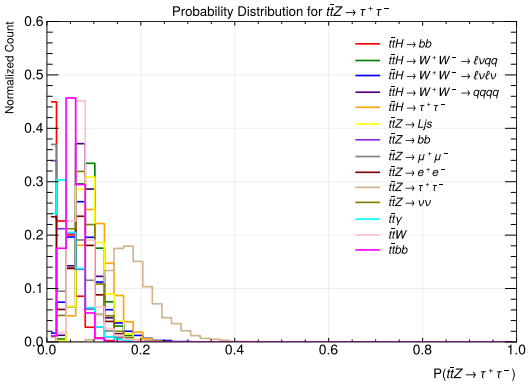
Figure 6.2: Appendix: normalised distributions of GATNet Fold-2 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.



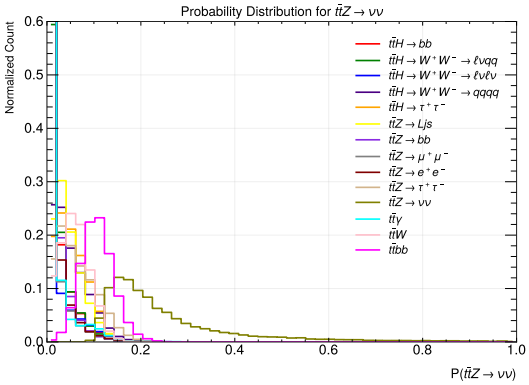
(e) $t\bar{t}(Z \to \mathrm{ljs})$



(f) $t\bar{t}(Z \to e^+e^-)$

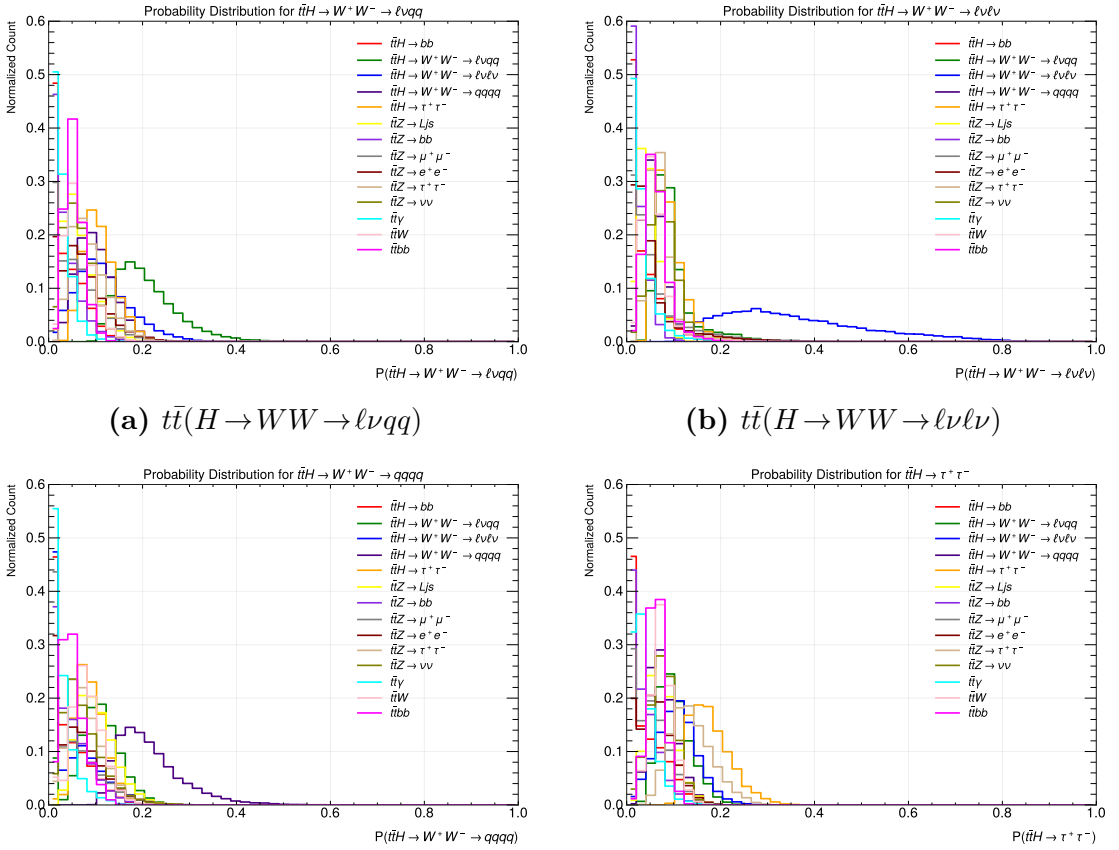


(g) $t\bar{t}(Z \to \tau^+\tau^-)$



(h) $t\bar{t}(Z \to \nu\bar{\nu})$

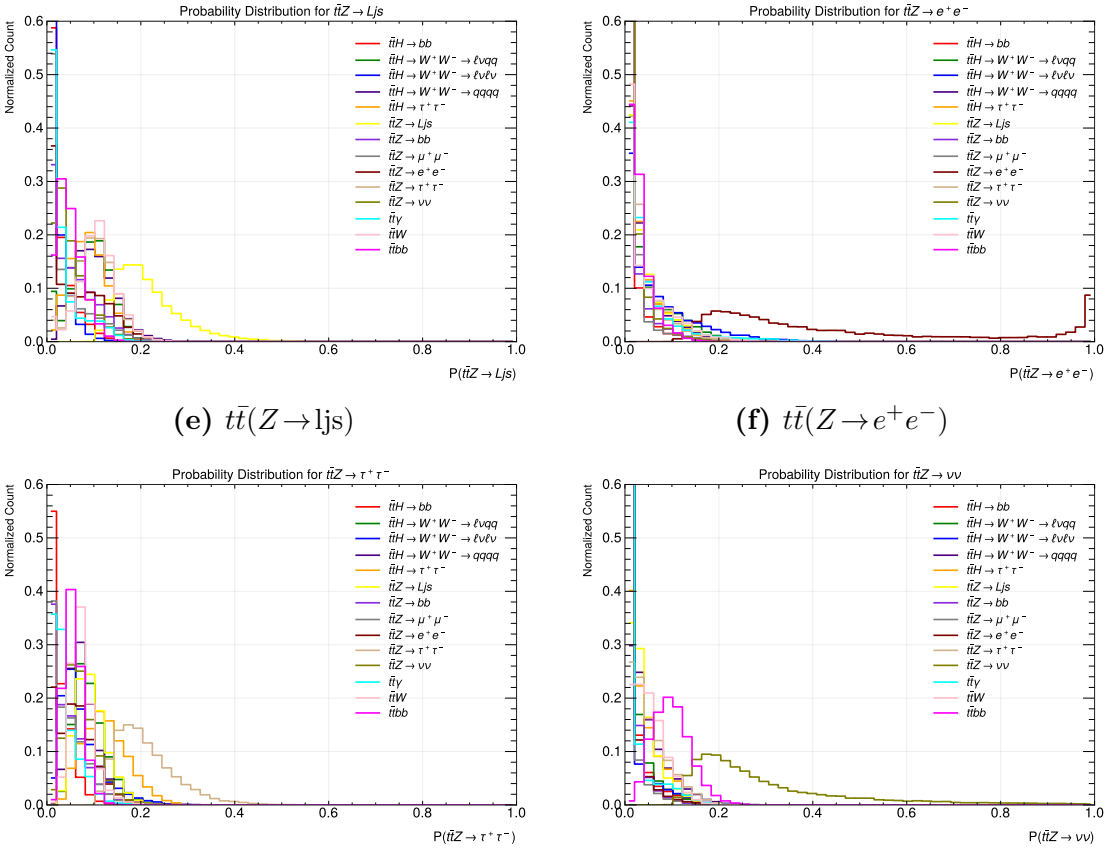
Figure 6.2: Appendix (continued): normalised distributions of GATNet Fold-2 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.



(a) $t\bar{t}(H \to WW \to \ell\nu qq)$ **(b)** $t\bar{t}(H \to WW \to \ell\nu\ell\nu)$

(c) $t\bar{t}(H \to WW \to qqqq)$ **(d)** $t\bar{t}(H \to \tau^+\tau^-)$

Figure 6.3: Appendix: normalised distributions of GTNet Fold-1 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.



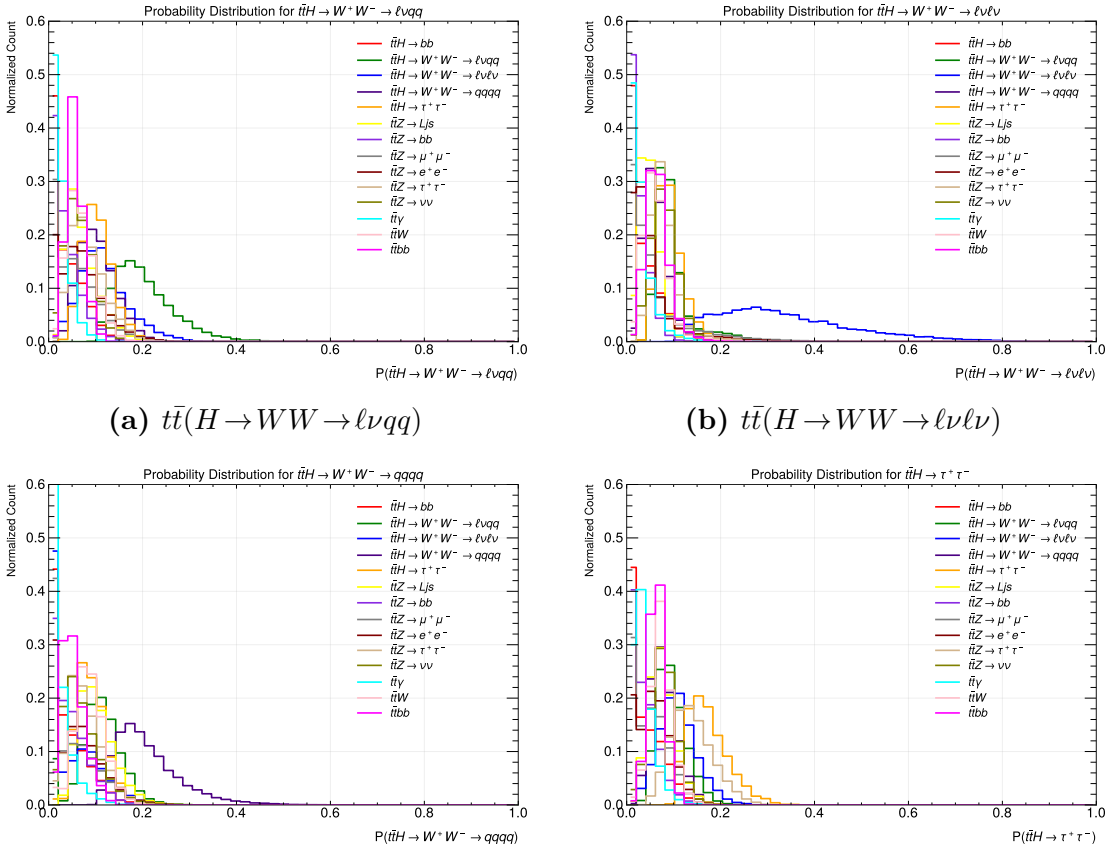
(e) $t\bar{t}(Z \to \mathrm{ljs})$

(f) $t\bar{t}(Z \to e^+e^-)$

(g) $t\bar{t}(Z \to \tau^+\tau^-)$

(h) $t\bar{t}(Z \to \nu\bar{\nu})$

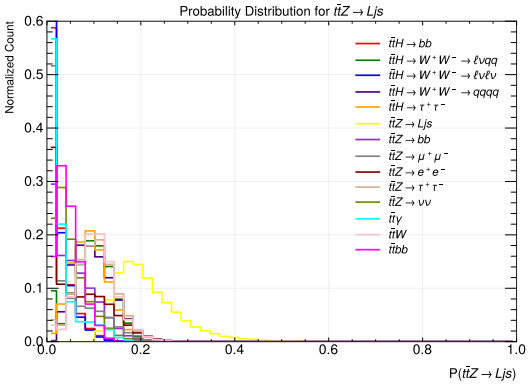
Figure 6.3: Appendix (continued): normalised distributions of GTNet Fold-1 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.



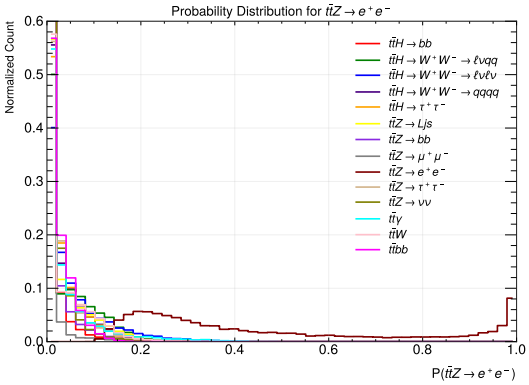
(a) $t\bar{t}(H \to WW \to \ell\nu qq)$ **(b)** $t\bar{t}(H \to WW \to \ell\nu\ell\nu)$

(c) $t\bar{t}(H \to WW \to qqqq)$ **(d)** $t\bar{t}(H \to \tau^+\tau^-)$

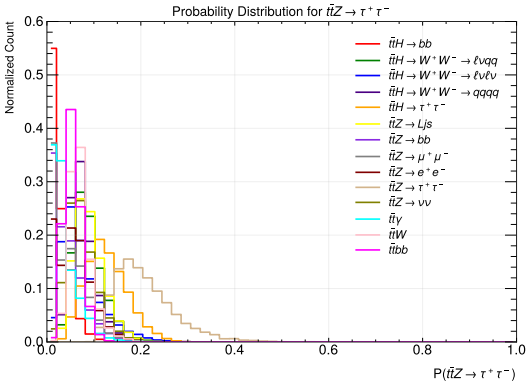
Figure 6.4: Appendix: normalised distributions of GTNet Fold-2 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.



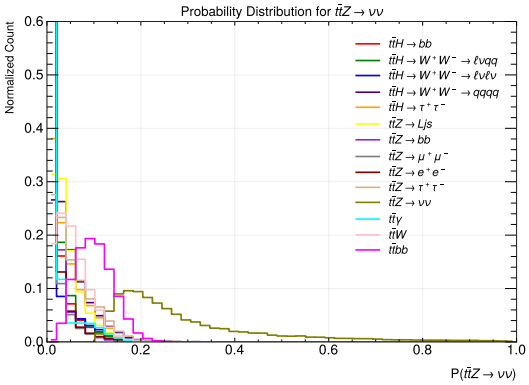
(e) $t\bar{t}(Z \to \mathrm{ljs})$



(f) $t\bar{t}(Z \to e^+e^-)$



(g) $t\bar{t}(Z \to \tau^+\tau^-)$



(h) $t\bar{t}(Z \to \nu\bar{\nu})$

Figure 6.4: Appendix (continued): normalised distributions of GTNet Fold-2 predicted class probability $P(\text{class})$ for the remaining classes. Each panel shows $P(c)$ for events whose true label is c.

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

In this paper, I have used ChatGPT or another AI as follows.:

☐ not at all

☐ during brainstorming

☐ when creating the outline

☐ to write individual passages, altogether to the extent of _____% of the entire text

☐ for the development of software source texts

☑ for optimising or restructuring software source texts

☐ for proofreading or optimising

☐ further, namely: _____

I hereby declare that I have stated all uses completely. Missing or incorrect information will be considered as an attempt to cheat.

Place/Date: Göttingen, Germany 30/09/25

Signature:

Erklärung nach §17(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 7. November 2025

(Clinton Gonsalves)